QF604 Practice 1 Test Questions with Suggested Answers

Q1. Show one stochastic process of stock price P_t whereby $e^{-rT}P_T$ is a martingale for any T>t. Write out the details of this price stochastic process with $\ln(P_{t+1})$ on the LHS.

Ans:

$\ln(P_{t+1}) = \ln(P_t) + (r - ½\sigma^2) + \sigma\varepsilon_{t+1}$ where ε_{t+1} is distributed as N(0,1). You can also write $\ln(P_{t+\Delta}) = \ln(P_t) + (r - ½\sigma^2)\Delta + \sigma\varepsilon_{t+\Delta}$ where $\varepsilon_{t+\Delta}$ is distributed as $N(0,\Delta)$. $\varepsilon_{t+\Delta}$ can also be written as W_Δ

Q2. Suppose R_t is excess stock return of an asset at week t. M_t is excess market return at week t. We run an OLS regression as follows.

$$R_t = a_0 + a_1M_t + \varepsilon_t \text{ for } t = 1,2,\ldots,60$$

Assume R_t and M_t are normally distributed. Given $\sum_{t=1}^{60} R_tM_t$=0.008, $\sum_{t=1}^{60} M_t^2 = 0.006$, $\bar{M} = 0.004$, $\bar{R} = 0.0035$, SSR (or RSS) = 6.0×10^{-5}, find OLS estimates $\hat{a}_0$ and $\hat{a}_1$.

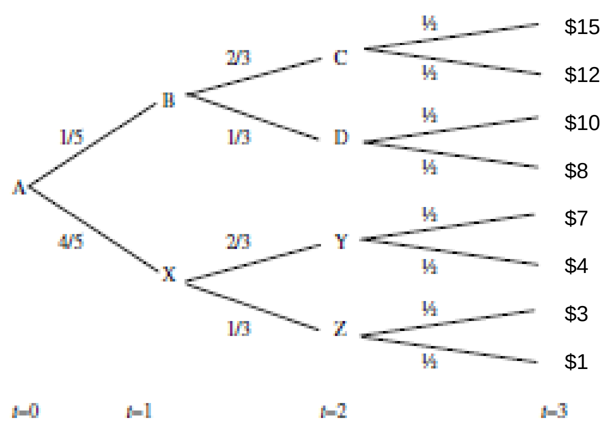
Ans:

$\hat{a}_1$= [0.008 – 0.0035 x (60 x 0.004)]/[0.006 – 0.004 x (60 x 0.004)] = <u>1.4206</u>

$\hat{a}_0$ = 0.0035 – 1.4206 x 0.004 = <u>–0.00218</u>

Q3. The tree below shows the resolution of uncertainties over time. Each branch represents a state of nature occurring at that time period (t,t+1]. The number on the branch indicates the probability of occurrence of the state. The alphabet at each node denotes the state at that time t. By t=3, the security prices P_3 are revealed as shown on the nodes.

Find conditional expectations $E(P_3|B)$ and $E(P_3|X)$ at t=1. Hence, find the expectation of P_3 at t=1 when there is no information. Suppose the per period risk-adjusted discount rate of the security is 10%, independent of the information. What would be the different prices of the security at t=1 given information B, given information X, and given no information?



Ans:

Conditional expectation of P_3 at t=1 given B is \$4.333, given X is \$12, and with no info is \$5.867 .

Price at t=1 given info B is \$3.581, given X is \$9.917, and given no info is \$4.848.

Q4. In an event study consisting of 100 cases of Mergers and Acquisitions at announcement dates by each acquiring firm, assume each firm stock's $AR_{jt} \sim N(0,\sigma_j^2)$ for each t. Moreover, $\sum_{j=1}^{100} \sigma_j^2 = 0.5$. Assume the AR_{jt}'s across j=1,2,…,100 are all independent of each other. Find the exact distribution of CAAR (-10,τ) in terms of τ within the event window. If CAAR (-10,-1) = 0.35, CAAR (-10,0) = 0.5, and CAAR (-10,1) = 0.45, what can you infer about the event, using 2-tail 95% level of significance?

Ans:

$CAR_j(-10,\tau) \sim N(0, [\tau+11]\, \sigma_j^2)$. Hence CAAR (-10,τ) ~ N(0, 0.0001 × [τ+11] × 0.5).

Now CAAR (-10,-1) ~ N(0, 0.0005). CAAR (-10, 0) ~ N(0, 0.00055). CAAR (-10, 1) ~ N(0, 0.0006).

$0.35/\sqrt{0.0005} = 15.7$, $0.5/\sqrt{0.00055} = 21.3$, $0.45/\sqrt{0.0006} = 18.4$.

Critical region to reject null of no abnormality is when |z-statistic| ≥ 1.96, ie at 2-tail 5% significance level. Hence CAAR is significantly > zero the different periods. There appears to be a strong leakage up to 10 days earlier till event day +1.

Q5. The following represent quarterly real GDP per capita data (first column, Y) and full-time median weekly per head real wage (second column, X) in USA. The data are from 2015 first quarter to 2019 last quarter. Data are all in USD and collected from Federal Reserve Bank of St Louis. (You need Excel or a computer to work on this.)

53983	341
54295	339
54368	340
54280	345
54464	346
54633	345
54827	346
55005	349
55240	352
55458	353
55806	352
56210	345
56503	351
56927	351
57257	354
57336	355
57719	355
57946	357
58167	360
58392	362

Run an OLS regression $\ln(Y_t) = A + B \ln(X_t) + U_t$ where it is assumed that X_t, U_t are independent. $U_t \sim N(0,\sigma^2)$. Find OLS estimate $\hat{A}$ and $\hat{B}$. (More sophisticated versions of this kind of regression are used in economics verifying the capital-labor substitution theories.)

Ans:
$\hat{A}$ = 3.454 and $\hat{B}$ = 1.277

SUMMARY OUTPUT								
Regression Statistics								
Multiple F	0.889977							
R Square	0.792059							
Adjusted I	0.780506							
Standard I	0.012368							
Observati	20							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regressio	1	0.010489	0.010489	68.56285	1.5E-07			
Residual	18	0.002754	0.000153					
Total	19	0.013242						
	Coefficient	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	3.454093	0.90307	3.824836	0.001241	1.556815	5.351372	1.556815	5.351372
X Variable	1.276593	0.154173	8.280269	1.5E-07	0.952687	1.600498	0.952687	1.600498

Q6. Using summary output information in Q5, test the null hypothesis H_0: B = 0. Test also H_0: B = 1. Also find the 95% confidence interval of the estimate of B. (You need Excel or a t-distribution table.)

Ans:
Test H_0 : B = 0. Std err. of $\hat{B}$ = 0.154173. t-stats = 1.276593/0.154173=8.28. p-value (2-tail t-distribution, d.f. 18) is 1.5 x 10^{-7}. Reject H_0 at 2-tail 0.1% significance level (p-value < 0.001)
Test H_0 : B = 1. Std err. of $\hat{B}$ = 0.154173. t-stats = (1.276593-1)/0.154173=1.7940. (2-tail t-distribution, d.f. 18) p-value is 0.0448. Reject H_0 at 2-tail 10% significance, but cannot reject H_0 at 2-tail 5% significance level.
95% confidence interval means both the upper and lower end tails have 2.5% areas. For the t-statistic with 18 df, 95% confidence interval about zero is +/- 2.1, i.e. Prob(t_{18} < 2.1)=0.975. Hence the c.i. here is 1.276593 +/- 0.154173 * 2.1 = (0.95283 , 1.60036)

Q7. In testing the Sharpe-Lintner two-parameter CAPM, the following cross-sectional regression is run for each time period t, t=1,2,…..,T.

$$R_i = \delta_0 + \delta_1 \hat{\beta}_i + e_i \text{ for assets } i = 1,2,\ldots,N$$

where R_i is return of asset i, and $\hat{\beta}_i$ is the beta of asset i estimated from another separate procedure, and e_i is the residual error of return to i. The OLS regression will produce estimates of δ_0 and δ_1. What are the financial meanings of the estimates of δ_0 and δ_1? If δ_1 is constant in every time period, can you find more efficient estimates of δ_0 and δ_1 by using only one cross-sectional regression?

Ans:
Estimates of δ_0 and δ_1 are estimates of the riskfree rate and of the market risk premium (excess expected market return). Running only in one period assumes the riskfree rate and the market risk premium are time-invariant or constant over the sample time periods. We can use the time series average of returns R_i for each i as the dependent variable in the one cross-sectional regression. In this case, the estimators for

δ_0 and δ_1 should be more efficient with smaller estimator standard errors since the sample variance of time series average of each e_i, $\bar{e}_i$, should be smaller than for each var(e_i).

Q8. Suppose return rate R_t follows the following GARCH (1,1) process:

$$R_t = 0.1 + 0.2X_t + \sqrt{h_t}\, e_t \text{ , and } h_t = 0.16\, h_{t-1} + 0.1e_{t-1}^2 \text{ ,}$$

where e_t is N(0,1) and independent of X_t. var(X_0) = 1, var(R_0)=0.05. Find the value of h_2 if X_0 = 0.15, X_1 = 0.10, R_1 =0.15, and you assume e_0 = 0.

Ans:

R_0 = 0.1 + 0.2 (0.15) = 0.13. var(R_0) = 0.04(1) + h_0 , so h_0 = 0.05 – 0.04 = 0.01. h_1 = 0.0016.

$\sqrt{h_1}\, e_1$ = 0.15 – 0.1 – 0.2(0.10) = 0.03. So e_1 = 0.03/0.04 = 0.75.

So $h_2 = 0.16(0.0016)+0.1(0.75^2) = 0.056506$.

Q9. Suppose returns $R_t = (\mu - ½\sigma_t^2) + \sigma_t Z_t$ where Z_t is distributed as i.i.d. N(0,1).

And $\sigma_t^2 = a_0 + a_1\sigma_{t-1}^2 + a_2\sigma_{t-1}^2 Z_{t-1}^2 + a_3(Z_{t-1} - \lambda)^2$ where μ, a_0, a_1, a_2, a_3, and λ are constants. Find the conditional variance $E_t(\sigma_{t+2}^2)$ in terms of a_0, a_1, a_2, a_3, λ, σ_t^2 and Z_t.

Ans:

$\sigma_{t+1}^2 = a_0 + a_1\sigma_t^2 + a_2\sigma_t^2 Z_t^2 + a_3(Z_t - \lambda)^2$, so $E_t[\sigma_{t+1}^2] = a_0 + a_1\sigma_t^2 + a_2\sigma_t^2 Z_t^2 + a_3(Z_t - \lambda)^2$

$\sigma_{t+2}^2 = a_0 + a_1\sigma_{t+1}^2 + a_2\sigma_{t+1}^2 Z_{t+1}^2 + a_3(Z_{t+1} - \lambda)^2$, so $E_t[\sigma_{t+2}^2] = a_0 + a_1 E_t[\sigma_{t+1}^2] + a_2 E_t[\sigma_{t+1}^2 Z_{t+1}^2] + a_3 E_t (Z_{t+1} - \lambda)^2$

Thus $E_t[\sigma_{t+2}^2] = a_0 + a_1 E_t[\sigma_{t+1}^2] + a_2 E_t[\sigma_{t+1}^2] + a_3 [1+\lambda^2]$

$= (a_0 + a_3 [1+\lambda^2]) + (a_1+a_2) [a_0 + a_1\sigma_t^2 + a_2\sigma_t^2 Z_t^2 + a_3(Z_t - \lambda)^2]$

QF604 Practice 2 Test Questions with Suggested Answers

Q1. The following linear regression output shows a cross-sectional regression of the increase in return rate of the stock of a food chain company as dependent variable, and its chain business volume as explanatory variable.

Dependent Variable: IRET / Method: Least Squares / Sample: 1, 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0900	0.0300	3.000	0.0034
VOLUME	0.527	**????**	**????**	
R-squared	**????**	Mean dependent var		0.1130
Adjusted R-squared	**????**	S.D. dependent var		0.133
S.E. of regression	0.0610	$F_{1,98}$ -statistic		**????**
Sum squared resid	0.3650	Prob(F-statistic)		0.000000

(a) Find R^2 and adjusted R^2

(b) Find the F-statistic testing H_0: VOLUME Coefficient = 0

(c) Find the t-statistic and standard error of the VOLUME Coefficient

Ans:

(a) SST=.133^2*99=1.7512; R^2 = 1 – SSE/SST = 1 - .3650/1.7512=0.7916

Adj R^2 = (1-k)/(N-k) + (N-1)/(N-k) x R^2 = -1/98 + 99/98 x .7916 = 0.7895

(b) $F_{1,98}$ = $[R^2/1]/[(1-R^2)/98]$ = 372.18

(c) t_{98} = √(372.18) = 19.30 . Standard error of VOLUME = 0.527/19.30= 0.0273

Q2. Let R_i be the average return over a particular month of portfolio i comprising stocks with betas that are largely similar. Let β_i be the beta of portfolio i's. In a cross-sectional linear regression of $R_i = a + b\,\beta_i + e_i$ where there are 100 observations per variable, and e_i is assumed to be homoskedastic and normally distributed with mean zero. OLS estimates $\hat{a} = 0.005, \hat{b} = 0.015,\ \sum_{i=1}^{100} \hat{e}_i^2 = 0.15$,

$X^TX = \begin{bmatrix} 100 & 80 \\ 80 & 120 \end{bmatrix}$ and $X^T = \begin{bmatrix} 1 & 1 & 1 & \cdots & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & \cdots & x_{100} \end{bmatrix}$

(a) Find the t-statistics of the OLS estimators $\hat{a}$, $\hat{b}$ under the null H_0: a=0 , H_A: a ≠ 0; and H_0: b=0 , H_A: b≠0.

(b) What is the average return of all the portfolio average returns R_i's?

(c) How do you interpret the economic meanings of estimates of a and b?

Ans:

(a) \hat u = sqrt(0.15/98) = 0.0391. $(X^TX)^{-1} = \begin{bmatrix} 0.0214 & -0.0143 \\ -0.0143 & 0.0179 \end{bmatrix}$

Std dev (\hat a) = 0.0391*√0.0214= 0.0057, std dev (\hat b) = 0.0391*√0.0179 = 0.0052;

t_a = 0.877, t_b = 2.885

(b) \bar Y = \hat a + \hat b * \bar X = 0.005 + 0.015*80 /100= 0.017 or 1.7%

(c) \hat a is estimate of riskfree rate; \hat b is estimate of expected excess market return or market risk premium for that particular month

Q3. Suppose Y_t is the excess stock return at month t, and X_t is the excess market return at time t. Suppose we run an OLS regression

$Y_t = a + b X_t + e_t$

on a sample size 60. Assume normally distributed returns.

(a) Find the alpha and the beta estimate of the stock using the following sample data.

$$\sum_{t=1}^{60} X_t Y_t = 5.0\times10^{-4} \quad \sum_{t=1}^{60} X_t^2 = 2.0\times10^{-4} \quad \overline{Y} = 0.05\times10^{-2} \quad \overline{X} = 0.02\times10^{-2}$$

(b) Further, given SSR = 6.0×10^{-5}, provide a test of H_0: a = 0, H_0: b = 1. (Show the computed t-statistics.)

(c) Suppose you had instead run an OLS regression

$Y_t = c + d Z_t + n_t$

where Z_t is another factor variable e.g. market trading volume. If OLS estimates $\hat{c}$ and $\hat{d}$ are significantly different from zero, and R^2 is high, does it imply that the market premium is not important in this case? Explain why or why not.

Ans:

(a) \hat b = (sum XY – 60 x bar X x bar Y)/(sum X^2 – 60 x bar X^2) = 2.50

\hat a = bar Y - \hat b \bar X = 0

(b) $\hat{\sigma}_e^2$ = 0.00006/58 = 1.03448 E-6

var(\hat b) = 1.03448 E-6 x (1/[sum X^2 – 60* bar X^2]) =0.005235237

var(\hat a) = 1.03448 E-6 x (1/60 + bar X^2/[sum X^2 – 60* bar X^2]) = 1.74508 x 10^{-8}

Under H_0: a =0, t_{58} = 0/√(1.74508 x 10^{-8}) = 0

Under H_0: b=1, t_{58} = (2.50-1)/√ 0.005235237= 20.73

Hence, H_0: b=1 hypothesis is rejected at very low significance levels. H_0: a =0 is not rejected.

(c) No, market premium could still be important. One possibility to explain both the results is that X and Z are highly correlated. For example, if $X_t = p + qZ_t + u_t$ where u_t is i.i.d. noise, then $Y_t = (a+bp) + bq Z_t + (bu_t+e_t)$ which produces the regression results of the second regression.

Q4. An OLS regression $Y_t = A + BX_t + e_t$ (Eq. 1) is run, where A, B are constants, and $e_t \sim N(0,\sigma_e^2)$ is independent of X_t. The OLS estimates are $\hat{A}$ and $\hat{B}$. Suppose instead run the regression as follows: $X_t = C + DY_t + \eta_t$ (Eq. 2). Is the OLS estimate of D in Eq. (2) equal to 1/OLS estimate B in Eq. (1)? Explain.

Ans:

In Eq. 1, X=-A/B + 1/B Y –e/B. So by (2), C=-A/B, D=1/B, and $\eta_t = -e_t/B$. The problem of course is that in Eq. 2, η_t and Y_t are perfectly correlated from Eq. 1.

No. \hat B from (Eq. 1) = sum(Y-barY)(X-barX)/sum(X-barX)^2

But \hat D from (Eq. 2) = sum(X-barX)(Y-barY)/sum(Y-barY)^2. \hat D is not inverse of \hat B.

Q5. The following Y, Z data are given. (Use a PC for computations)

Y	4	3	2	-1	3	8	2	0	-3	3

Z	3	4	2	5	1	0	6	3	4	2

(a) Find the OLS estimates of a, b in regression Y = a + bZ + i.i.d. residual error.
(b) Find the t-statistics of OLS estimates a, b under assumption of homoscedastic residual errors.
(c) Suppose the residual errors are heteroskedastic but have a diagonal covariance matrix. Find the HCCME t-statistics of the OLS estimates.

Ans:
(a) 5.2, -1.0333 (b) 3.327, -2.290
(c) the variance matrix is found as $(X'X)^{-1}(X'\Omega X)(X'X)^{-1}$ where Ω is a diagonal matrix with diagonal elements equal to fitted squared residuals. t-values are 3.793, -2.293
Adjustment using N/N-k or 10/8 times $(X'X)^{-1}(X'\Omega X)(X'X)^{-1}$ acceptable. So, $\sqrt{0.8}$ x (3.793, -2.293) or 3.393, -2.051

Y	Z
4	3
3	4
2	2
-1	5
3	1
8	0
2	6
0	3
-3	4
3	2

SUMMARY OUTPUT

Regression Statistics	
Multiple F	0.629255
R Square	0.395962
Adjusted I	0.320457
Standard I	2.471504
Observati	10

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressio	1	32.03333	32.03333	5.244202	0.051263
Residual	8	48.86667	6.108333		
Total	9	80.9			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	5.2	1.563117	3.326687	0.010436	1.595447	8.804553	1.595447	8.804553
X Variable	-1.03333	0.451233	-2.29002	0.051263	-2.07388	0.007212	-2.07388	0.007212

X matrix

1	3
1	4
1	2
1	5
1	1
1	0
1	6
1	3
1	4
1	2

X'

1	1	1	1	1	1	1	1	1	1
3	4	2	5	1	0	6	3	4	2

X'(Omega)X

48.86667	168.4667
168.4667	753.8533

inv(X'X)

0.4	-0.1
-0.1	0.033333

HCCME estimate of variance of \hat B is inv(X'X) (X'OmegaX)inv(X'X)

1.879867	-0.53662
-0.53662	0.20317

t-stats	a	b
	3.792624	-2.2925

fitted U	fitted U^2
1.9	3.61
1.933333	3.737778
-1.13333	1.284444
-1.03333	1.067778
-1.16667	1.361111
2.8	7.84
3	9
-2.1	4.41
-4.06667	16.53778
-0.13333	0.017778

Omega =diag(fitted U^2)

3.61	0	0	0	0	0	0	0	0	0
0	3.737778	0	0	0	0	0	0	0	0
0	0	1.284444	0	0	0	0	0	0	0
0	0	0	1.067778	0	0	0	0	0	0
0	0	0	0	1.361111	0	0	0	0	0
0	0	0	0	0	7.84	0	0	0	0
0	0	0	0	0	0	9	0	0	0
0	0	0	0	0	0	0	4.41	0	0
0	0	0	0	0	0	0	0	16.53778	0
0	0	0	0	0	0	0	0	0	0.017778

Q6. $Y_i = a + bX_i + e_i$ where e_i are i.i.d. disturbances distributed as $N(0,\sigma^2)$. Find the log likelihood function of the sample N of disturbances. Thus, solve to obtain the maximum likelihood estimates of a and b. Are they similar to the OLS estimates. If so, why?

Ans:
log likelihood = $-N/2 \log(2\pi) - N/2 \log(\sigma^2) - 1/(2\sigma^2) \times \sum_{i=1}^{N}(Y_i - a - bX_i)^2$
Max likelihood estimates of \hat a, \hat b are \hat b = $\sum_{i=1}^{N} X_i(Y_i - \bar{Y}) / \sum_{i=1}^{N} X_i(X_i - \bar{X})$
and \hat a = $\bar{Y} - \hat{b}\bar{X}$.
Why? -- max log likelihood is equivalent to min $\sum_{i=1}^{N}(Y_i - a - bX_i)^2$ or OLS.

Q7. In the generalized method of moments, suppose the theory says $E_t[\, f(P_{t+1}; a,b,c)\,] = 0$, where a, b, c are parameters to be estimated, $f(P_{t+1})$ is a nonlinear function of price P_{t+1}, and W_t, U_t are available instruments. Formulate three moment conditions to estimate the parameters. State what are other necessary conditions in order to attain the consistent GMM estimators?

Ans:

$E_t[f(P_{t+1})] = 0$, $E_t[f(P_{t+1} W_t)] = 0$, $E_t[f(P_{t+1} U_t)] = 0$

P, U, W should be stationary-ergodic, so sample moments converge to $E[f(P_{t+1})] = 0$, $E[f(P_{t+1} \times W_t)] = 0$, $E[f(P_{t+1} \times U_t)] = 0$. Second moments and derivatives need exist for getting the optimal weighting matrix.

Q8. Suppose an ARMA (1,1) process $Y_t = \sigma_u^2 + ½ Y_{t-1} + u_t – ½ u_{t-1}$ is covariance-stationary and $var(u_t) = \sigma_u^2$. The error u_t is identically distributed over time and $cov(u_t,u_{t-k}) = 0$ for any $k > 0$. Find the unconditional mean and variance of Y. Also find the first order autocovariance and autocorrelation of Y. You may assume $Eu_t = 0$. (Express your answers in terms of σ_u^2.)

Ans:

Suppose unconditional mean of Y is μ, then $E Y_t = \sigma_u^2 + ½ E Y_{t-1} + 0 – 0$ or $\mu = \sigma_u^2 + ½ \mu$, so $\mu = \underline{\mathbf{2\sigma_u^2}}$.

(Suppose you had assumed $Eu_t = a$, then $E Y_t = \sigma_u^2 + ½ E Y_{t-1} + a – ½ a$ or $\mu = \sigma_u^2 + ½ \mu + ½ a$, so $\mu = 2\sigma_u^2+a$).

Let unconditional variance of Y be V. Then $V = ¼ V + \sigma_u^2 + ¼ \sigma_u^2 + 2 Cov(½Y_{t-1}, – ½u_{t-1})$ since $cov(u_t, u_{t-1}) = 0$ and $cov(Y_{t-1}, u_t) =0$. The latter arises from expansion of the ARIMA whereby Y_{t-1} comprises linear combinations of lagged u_{t-1}, u_{t-2}, u_{t-3}, etc. that are not correlated with u_t.

Now $V = ¼ V + \sigma_u^2 + ¼ \sigma_u^2 – ½ Cov(Y_{t-1},u_{t-1}) = ¼ V + 5/4 \sigma_u^2 – ½ Cov(½ Y_{t-2} + u_{t-1} – ½ u_{t-2} , u_{t-1}) =$ $¼ V + 5/4 \sigma_u^2 – ½ Cov(u_{t-1} , u_{t-1}) = ¼ V + 5/4 \sigma_u^2 – ½ \sigma_u^2$. Therefore $¾ V = ¾ \sigma_u^2$. Hence $\underline{\mathbf{V = \sigma_u^2}}$. (This holds whether a = 0 or not.)

$Cov(Y_t, Y_{t-1}) = cov(\sigma_u^2 + ½ Y_{t-1} + u_t – ½ u_{t-1}, Y_{t-1}) = ½ var(Y_t) – ½ cov(u_{t-1}, Y_{t-1}) =$ $½ \sigma_u^2 – ½ cov(u_{t-1}, ½ Y_{t-2} + u_{t-1} – ½ u_{t-2}) = ½ \sigma_u^2 – ½ cov(u_{t-1},u_{t-1})$ $\underline{\mathbf{= 0}}$. First autocovariance is zero. This can also be obtained from $Y_t = \sigma_u^2(1+1/2+1/4+1/8+....) + u_t$, $Y_{t-1} = \sigma_u^2(1+1/2+1/4+1/8+....) + u_{t-1}$, and taking their cov. Hence also first autocorrelation = first autocovariance divided by unconditional variance $\underline{\mathbf{= 0}}$.

Q9. A theoretical model is postulated as $Y_t = \dfrac{e^{a+bX_t+cZ_t+\eta_t}}{1+e^{a+bX_t+cZ_t+\eta_t}}$, where a, b, and c are constants, and η_t is a zero mean i.i.d. noise. In addition, it is thought that the restriction b+c = 1 holds. If time series data of $\{X_t, Y_t, Z_t\}$ are available, explain how you would perform a linear regression to estimate a, b, and c. Perform a manual computation of the OLS estimates of a, b, and c using the following small sample, and test H_0: a=1 and b=0. Show your workings. (Use a PC for computations.)

t	1	2	3	4
X_t	10	15	5	20
Y_t	0.6	0.5	0.7	0.8
Z_t	3	6	2	5

Ans:

Run $\ln(Y_t/[1-Y_t]) = Y^*= a+bX_t+(1-b)Z_t +\eta_t$. $Y^*-Z = a + b(X-Z) +\eta$. Then find OLS est of a, b first. <u>\hat a = -1.50, \hat b = -0.217. Therefore \hat c = 1-\hat b = 1.217.</u>

Y*	0.405465	0	0.847298	1.386294										
						Regression Statistics								
Y*-Z	-2.59453	-6	-1.1527	-3.61371		Multiple R	0.531167							
X-Z	7	9	3	15		R Square	0.282138							
						Adjusted R	-0.07679							
vector Y		matrix X				Standard Er	2.117363							
-2.594535		1	7			Observatior	4							
-6		1	9											
-1.152702		1	3			ANOVA								
-3.613706		1	15				*df*	*SS*	*MS*	*F*	*gnificance F*			
						Regression	1	3.524051	3.524051	0.786052195	0.468833			
						Residual	2	8.966454	4.483227					
						Total	3	12.4905						
							Coefficient:	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
						Intercept	-1.49773	2.332305	-0.64217	0.586548263	-11.5328	8.537373	-11.5328	8.537373
						X Variable 1	-0.21677	0.244492	-0.8866	0.468833021	-1.26873	0.835199	-1.26873	0.835199
fitted u	0.420585	-2.55134	0.995338	1.135574										
u'u/2	4.483227													
matrix X'X	4	34		inv (X'X)		1.2133333	-0.11333							
	34	364				-0.1133333	0.013333							
F(2,2)=((\hat a-1), \hat b) (X'X) (\hat a -1, \hat b)'/2 divided by u'u/2														
(\hat a -1, \hat b)		-2.49773	-0.21677			(\hat a -1, \hat b)'		-2.49773						
								-0.21677						
F(2,2)	8.796636													

Test with restrictions

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and $F_{2,2} = \dfrac{(R\hat{B}-r)^T\left[R\left(X^TX\right)^{-1}R^T\right]^{-1}(R\hat{B}-r)/2}{\hat{u}^T\hat{u}\,/2}$ = 8.7966, where $\hat{B} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$.

p-value = 0.1021, hence cannot reject H_0 at 1-tail 10% significance level.

QF604 Practice 3 Test Questions with Suggested Answers

Q1. The following estimation output table shows a linear regression of number of oil rigs in USA against the $ price per barrel of crude oil. The dependent variable of number of oil rigs is called COUNT. The $ price per barrel of oil is called PRICE. A constant is added for the regression. (Assume there is no simultaneous equation bias.)

Dependent Variable: COUNT Method: Least Squares Sample: 1 751

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-678.2789	59.66590	-11.36795	0.0000
PRICE	84.29917	????	????	????
R-squared	????	Mean dependent var		1131.280
Adjusted R-squared	????	S.D. dependent var		584.6037
S.E. of regression	385.7005	Akaike info criterion		14.75066
Sum squared resid	1.114246E+08	Schwarz criterion		14.76297
Log likelihood	-5536.873	$F_{1,749}$ -statistic		????
Durbin-Watson stat	0.041786	Prob(F-statistic)		????

(a) Find the F-statistic, R^2, and adjusted R^2 statistics. (All the information required are available in the table).
(b) Find the standard error of the PRICE coefficient estimate and its t-statistic.
(c) What would be a likely source of problem in the estimates? Name the method you would use to provide more efficient estimates.

Ans:
(a) SST=584.6^2*750=256,317,870; R^2 = 1 – SSE/SST = 1 – 111424892/256,317,870=0.5653
Adj R^2 = (1-k)/(N-k) + (N-1)/(N-k) x R^2 = 0.5647. $F_{1,749}$ = [R^2/1]/[(1-R^2)/749] = 974.
(b) t_{749} = sqrt(974) = 31.2. Standard error of PRICE coefficient = 84.3/31.2= 2.702
(c) DW too low – positive autocorrelation. Use feasible GLS.

Q2. The following estimation output table shows a linear regression of an international pollution index of CO_2 emissions against the perceived corruption index produced by Transparency International. The dependent variable CO2 is in million tons per year. The explanatory variable is the perceived corruption index (PCI). Higher PCI denotes lower corruption. A constant is added for the regression. The regression is cross-sectional across 180 countries in 2018. (Major producing countries of PRC, USA, Japan, India, and Russia are left out of the sample.)

Dependent Variable: CO2 Method: Least Squares		Sample: 1 158		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	41.125	1.724	23.849	6.67E-54
PCI	0.0222	????	????	????
R-squared	????	Mean dependent var		79.34
Adjusted R-squared	????	S.D. dependent var		19.24
S.E. of regression	19.027	Akaike info criterion		13.2074
Sum squared resid	????	Schwarz criterion		13.6593
Log likelihood	-5978.873	$F_{1,156}$ -statistic		????

(a) Find the Sum of square residuals RSS, F-statistic, and R^2. (All the information required are available in the table).

(b) Find the standard error of the PCI coefficient estimate and its t-statistic.

(c) How would you conclude if countries with low PCI affect emissions amount?

Ans:

(a) SSR = SE residuals^2 x 156 = 19.027^2 x 156 = <u>56476.2</u>; $R^2 = 1 - SSR/TSS =$ 1 - 56476.2/(19.24^2 x 157) = <u>0.0283</u> $F_{1,156} = (R^2/1)/[(1-R^2)/598]$=<u>4.54</u>

(b) t_{156} = sqrt(4.54) = <u>2.13</u>. Standard error of PCI coefficient = 0.0222/2.13= <u>0.0104</u>

(c) Countries with low PCI are less developed and therefore less emissions.

Q3. Suppose a record of the last 50 years was taken, and 10 of the years had negative GDP growth while the other 40 years had positive GDP growth. In the 10 years with negative GDP growth or recession, 7 of these years saw inverted yield curves the year just prior to recession. In the other 40 years of positive GDP growth, 7 of those years saw inverted yield curves the year just prior to growth. If you see an inverted yield curve this year, what is the probability of a negative GDP growth next year given just the above information? If you could run an OLS linear regression using actual GDP growth numbers, Y, as dependent variable and actual yield curve slope, X, as explanatory variable, and suppose the OLS result is

$$Y_t = 0.01 + 1.2*X_{t-1}$$

where estimated residual variance $\hat{\sigma}_u^2 = 0.0004$. Assuming the residual is i.i.d. normally distributed with zero mean, and that this period $X_t = -0.02$, what is the estimated probability of a negative GDP growth next year using the linear model? If the two answers above do not match, explain why is it so? (A couple of sentences will do here.) If they match, you do not need to explain.

Ans:
R: event of recession or neg GPD growth
I: event of inverted YC the year earlier
$\text{Prob}(R|I) = \text{Prob}(R \cap I)/\text{Prob}(I) = [7/50]/[14/50]=$<u>50%</u>
$E(Y_{t+1})=0.01-1.2*0.02= -0.014$. $Y_{t+1} \mid X_t \sim N(-0.014, 0.0004)$.
So $\text{Prob}(Y_{t+1}<0) = \text{Prob}(\{Y_{t+1}-[-0.014]\}/0.02<0.014/0.02) = \text{Prob}(z<0.7) =$ <u>0.76</u>
The latter is parametric, and the distribution contains more information than the original non-parametric. It could be that this period negative yield slope is especially large, larger than average. The non-parametric answer does not have this additional piece of info about the yield slope quantity.

Q4. A portfolio's return rate process r_t is modeled as follows:
$r_t = a + bx_t + h_t^{1/2} e_t$ where a, b are constants. x_t is a normally distributed variable that is observed, and e_t is an unobserved i.i.d. disturbance and is distributed as N(0,1). h_t is an unobserved process that is uncorrelated with x_t and is modeled as follows:

$$h_t = 0.01 + 0.2\, h_{t-1} + 0.3\, h_{t-1}e_{t-1}^2$$

Suppose consistent estimates of a, b are 0.03 and 0.5 respectively. h_{t-1} is estimated consistently at 0.01. x_{t-1} has value 0.10 and current return r_{t-1} is 5%. Conditional on information at t-1, next period value of x_t is expected to be 0.12 and variance of x_t is 0.02. Show how you would compute the probability of the event $r_t < -10\%$. (You need a normal distribution table.)

Ans:
$h_{t-1}^{1/2}\, e_{t-1} = 0.05 - 0.03 - 0.5*0.10 = -0.03$
$0.1\, e_{t-1} = -0.03$, and so $e_{t-1} = -0.3$
$h_t = 0.01 + 0.2\, h_{t-1} + 0.3(0.01)\, e_{t-1}^2 = 0.01+0.2(0.01) + 0.3\,(0.01)\, 0.3^2 = 0.01227$
$\text{var}_{t-1}(r_t) = 0.5^2*0.02 + h_t = 0.005 + 0.01227 = 0.01727$
std dev $= \sqrt{(0.01727)} = 0.131415$
$E(r_t) = 0.03+0.5*0.12 = 0.09$
Therefore $r_t \sim N(0.09, 0.01727)$. $\Pr(r_t < -0.1) = \Phi([-0.1-0.09]/0.131415) = \Phi(-1.4458) =$ <u>0.074</u>.

	given values					
h(t-1)	0.01					
a	0.03					
b	0.5					
x(t-1)	0.1					
r(t-1)	0.05					
E_t-1(x(t))	0.12					
var_t-1(x(t))	0.02					
E(r(t)=a+bE(x(t))						
therefore E(r(t))=		0.09				
e(t-1)^2=(r_t-1-a-bx_t-1)^2/h(t-1)						0.09
var_t-1(r(t))=b^2var_t-1(x(t))+(0.01+0.2h(t-1)+0.3h(t-1)e(t-1)^2)						0.01727
std dev_t-1(r(t)) =						0.131415
value=	-0.1					
Z=	-1.445797376					
Prob(r(t)<-0.10)=	0.074117021					

Q5. A portfolio's return rate process r_t is modeled as follows:
$r_t = a + bx_t + h_t^{1/2} e_t$ where a, b are constants. x_t is a normally distributed variable that is observed, and e_t is an unobserved i.i.d. disturbance and is distributed as N(0,1). h_t is an unobserved process that is uncorrelated with x_t and is modeled as follows:

$$h_t = 0.005 + 0.5\, h_{t-1} + 0.2\, h_{t-1}e_{t-1}^2$$

Suppose consistent estimates of a, b are 0.12 and 0.07 respectively. h_{t-1} is estimated consistently at 0.024. x_{t-1} has value 0.1 and current return r_{t-1} is 4%. Conditional on information at t-1, next period value of x_t is expected to be 0.15 and variance of x_t is 0.05. Show how you would compute the probability of the event $r_t < -5\%$. (You need a normal distribution table.)

Ans:

	given values
h(t-1)	0.024
a	0.12
b	0.07
x(t-1)	0.1
r(t-1)	0.04
E_t-1(x(t))	0.15
var_t-1(x(t))	0.05

E_t-1(r(t)=a+bE_t-1(x(t))
therefore E_t-1(r(t))= 0.1305

e(t-1)^2=(r_t-1-a-bx_t-1)^2/h(t-1)	0.315375
var_t-1(r(t))=b^2var_t-1(x(t))+(0.005+0.5h(t-1)+0.2h(t-1)e(t-1)^2)	0.018759
std dev_t-1(r(t)) =	0.136963

value=	-0.05
Z=	-1.317876396
Prob(r(t)<-0.10)=	0.093772514

Q6. A researcher is attempting to model the duration time between trades on a stock. Suppose the time to wait, X, before the next stock trade happens on a trading system follows a distribution Probability ($X \leq x$ seconds) = $1 - \exp(-a^2x)$. The probability of the next trade arriving is higher when time elapsed is longer. Find the probability density function, and hence the log likelihood function of the sequential trades that happened in a day. The average time between each trade is 10 seconds on this trading day. Find the maximum likelihood estimate of a. What is the probability of the next trade arriving in 10 seconds or less?

Ans:

$\int_0^x f(x)dx = 1 - e^{-a^2x}$ where f(x) is the pdf.

The pdf of X is f(x) = $\frac{d}{dx}(1 - e^{-a^2x}) = a^2e^{-\lambda x}$. The likelihood function of observing the arrival times of N sequential trades is L = $\prod_{k=1}^{N} (a^2e^{-a^2x_k})$. Log L = N ln a^2 - $a^2 \sum_{k=1}^{N} x_k$. Max ln L yields $\widehat{a^2} = \left(\frac{\sum_{k=1}^{N} x_k}{N}\right)^{-1}$ so $\hat{a} = \left(\frac{\sum_{k=1}^{N} x_k}{N}\right)^{-1/2}$ where MLE of a is inverse of square root of sample average of trade arrival time. If trade arrivals are frequent, $\hat{a}$ is larger. If sample average of time is 10 seconds, then $\hat{a}$ is $\sqrt{0.1}$ = 0.3163 (Note prob (X<= 10 seconds) =$1-e^{-0.1 \times 10}$ = 0.63.

set2-6 use $a^{1/2}$, $a^{3/2}$, a^3, e^a , $2a^2$ instead.
\hat a = 0.01, 0.215447, 0.464, -2.303, 0.2236

Q7. All stocks in the market are formed into 20 large portfolios by sorting them according to their individual preliminary betas that were estimated using sample covariance of individual stock return rate with the market return rate divided by the sample market return rate variance. Suppose you take the two portfolios with the highest and the lowest betas. You calculate the equal-weighted average return rate in each time period of stocks within each of these two portfolios, and define these return rates as R_{1t} and R_{2t}, where subscripts 1, 2 denote the two different portfolios respectively, and t denotes the time period index. Their excess return rates are respectively r_{1t} and r_{2t}. These are excess over the risk-free rate prevailing at t. The excess market return rate is r_{mt}.

The Sharpe-Lintner CAPM implies that $r_{1t} = A_0 + A_1 r_{mt} + e_{1t}$ (1), and $r_{2t} = A_0 + A_2 r_{mt} + e_{2t}$ (2), where A_1, A_2 are the CAPM betas of the average returns of the two portfolios, and e_{1t}, e_{2t} are residual random disturbances that are uncorrelated with r_{mt}. The specifications in equations (1) and (2) hold for t = 1,2,3,….,T where T is the sample size. Let random vector $e_1=(e_{11}, e_{12}, e_{13}, \ldots, e_{1T})'$ and $e_2=(e_{21}, e_{22}, e_{23}, \ldots, e_{2T})'$. $A_0 = 0$ for a well-diversified portfolio with no consistent positive alphas. $cov(e_1) = \sigma_1^2 I_{TxT}$, $cov(e_2) = \sigma_2^2 I_{TxT}$, and $cov(e_1,e_2) = E(e_1,e_2') = \sigma_{12} I_{TxT}$ where I_{TxT} is identity matrix of dimension T.

Suppose you want to use equations (1) and (2) to test the restriction under null hypothesis that $A_0 = 0$. You combine multiple time series (1) and (2) into one single regression Y = XB + E where $Y_{2T \times 1} = (r_{11}, r_{12}, r_{13}, \ldots, r_{1T}, r_{21}, r_{22}, r_{23}, \ldots, r_{2T})'$, and $E = (e_1', e_2')'$.

(a) Show what the matrices X and B would look like, including their dimensions.
(b) Show the matrix cov(E) including its dimensions.
(c) Suppose now you are given information that $\sigma_2^2 = 2\sigma_1^2 = 2$ and $\sigma_{12} = \sigma_1^2$, T =100. Sample means of r_{mt} , r_{1t}, r_{2t}, are 0.01, 0.02, -0.01. Sample variance of r_{mt} is 0.02, and sample covariance of r_{1t} and r_{mt} is 0.003, and that of r_{2t} and r_{mt} is 0.001. (Use divisor T for all sample moments.) Find a best linear unbiased estimate of A_0.

(d) If you assume E is multivariate normal, find the t-statistic in the test of H_0: $A_0 = 0$, and interpret the result.

(Hint: If square matrix C = $\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$,

then $C^{-1} = \begin{pmatrix} [C_{11} - C_{12}C_{22}^{-1}C_{21}]^{-1} & -C_{11}^{-1}C_{12}[C_{22} - C_{21}C_{11}^{-1}C_{12}]^{-1} \\ -C_{22}^{-1}C_{21}[C_{11} - C_{12}C_{22}^{-1}C_{21}]^{-1} & [C_{22} - C_{21}C_{11}^{-1}C_{12}]^{-1} \end{pmatrix}$.

Please also use excel to find matrix inverses if that is faster. Your answers in parts (a) and (b) should be in terms of T, σ_1^2, σ_2^2, σ_{12}, and functions of r_{mt}, though not all these terms necessarily appear.)

Ans:

(a)

$$X_{2T\times 3} = \begin{pmatrix} 1 & r_{m1} & 0 \\ \vdots & \vdots & \vdots \\ 1 & r_{mT} & 0 \\ 1 & 0 & r_{m1} \\ \vdots & \vdots & \vdots \\ 1 & 0 & r_{mT} \end{pmatrix}, \qquad B_{3x1} = \begin{pmatrix} A_0 \\ A_1 \\ A_2 \end{pmatrix}$$

(b)

$$\text{cov}(E) = \Omega_{2Tx2T} = \begin{pmatrix} \sigma_1^2 & \dots & 0 & \sigma_{12} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_1^2 & 0 & \dots & \sigma_{12} \\ \sigma_{12} & \dots & 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{12} & 0 & \dots & \sigma_2^2 \end{pmatrix}, \qquad \sigma_{21} = \sigma_{12}$$

(c)

$$\text{cov}(E) = \Omega_{2Tx2T} = \begin{pmatrix} \sigma_1^2 & \dots & 0 & \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_1^2 & 0 & \dots & \sigma_1^2 \\ \sigma_1^2 & \dots & 0 & 2\sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_1^2 & 0 & \dots & 2\sigma_1^2 \end{pmatrix} = \sigma_1^2 \begin{pmatrix} 1\otimes \mathrm{I}_{T\times T} & 1\otimes \mathrm{I}_{T\times T} \\ 1\otimes \mathrm{I}_{T\times T} & 2\otimes \mathrm{I}_{T\times T} \end{pmatrix}$$

$$\Omega^{-1} = \frac{1}{\sigma_1^2}\begin{pmatrix} 2\otimes \mathrm{I}_{T\times T} & -1\otimes \mathrm{I}_{T\times T} \\ -1\otimes \mathrm{I}_{T\times T} & 1\otimes \mathrm{I}_{T\times T} \end{pmatrix}.$$

$$(X'\Omega^{-1}X) = \frac{1}{\sigma_1^2}\begin{pmatrix} T & \sum_{t=1}^T r_{mt} & 0 \\ \sum_{t=1}^T r_{mt} & 2\sum_{t=1}^T r_{mt}^2 & -\sum_{t=1}^T r_{mt}^2 \\ 0 & -\sum_{t=1}^T r_{mt}^2 & \sum_{t=1}^T r_{mt}^2 \end{pmatrix} = \begin{pmatrix} 100 & 1 & 0 \\ 1 & 4.02 & -2.01 \\ 0 & -2.01 & 2.01 \end{pmatrix}$$

$$(X'\Omega^{-1}X)^{-1} = \begin{pmatrix} 0.01005 & -0.005 & -0.005 \\ -0.005 & 0.5 & 0.5 \\ -0.005 & 0.5 & 0.997512 \end{pmatrix}$$

$$(X'\Omega^{-1}Y) = \frac{1}{\sigma_1^2}\begin{pmatrix} \sum_{t=1}^T r_{1t} \\ 2\ \sum_{t=1}^T r_{1t}r_{mt} - \sum_{t=1}^T r_{2t}r_{mt} \\ -\sum_{t=1}^T r_{1t}r_{mt} + \sum_{t=1}^T r_{2t}r_{mt} \end{pmatrix} = \begin{pmatrix} 2 \\ 2(0.3) - 0.1 \\ -0.3 + 0.1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0.5 \\ -0.2 \end{pmatrix}$$

$$(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) = \begin{pmatrix} 0.0186 \\ 0.14 \\ 0.0405 \end{pmatrix}$$

\hat $A_0 = 0.0186$

(d) t-stats for \hat $A_0 = 0.0184/\sqrt{0.01} = 0.184$. Hence cannot reject H_0: $A_0 = 0$. (workings for finding inverse below)

T	100	right is (X'Omega^-1X)	100	1	0
SAMPLE MEAN RM	0.01		1	4.02	-2.01
SAMPLE MEAN R1	0.02		0	-2.01	2.01
SAMPLE MEAN R2	-0.01				
SAMPLE VAR(RM)	0.02				
SAMPLE COV (R1,RM)	0.003	inv(X'Omega^-1X)	0.01005	-0.005	-0.005
SAMPLE COV(R2,RM)	0.001		-0.005	0.5	0.5
SIGMA_1^2	1		-0.005	0.5	0.997512
Computed		X'Omega^-1Y	2		
sum of RMt	1		0.5		
sum of (RMt)^2	2.01		-0.2		
sum of R1	2				
sum of R2	-1	(X'Omega^-1X)^-1(X'Omega^-1Y)			0.0186
sum of cov(R1,RM)	0.3				0.14
sum of cov(R2,RM)	0.1				0.040498
		t-stats for A0			0.185537
				cannot reject null	
				H0: A0=0	
				CAPM is OK	

Q8. In a regression, $Y_i^* = b_0 + b_1X_i + b_2Z_i + u_i$, where i = 1,2,3,…., N, denotes firms. Assume there is a large number of firms in the sample. b_0, b_1, b_2 are constants. u_i has zero mean and a unit constant variance. Moreover, $Z_i = c_1X_i + c_2W_i + v_i$, where c_1, c_2 are constants, and v_i has zero mean and a constant variance σ_v^2. u_i and v_i are both independent of X_i and W_i. However, $cov(u_i,v_i) \neq 0$. Furthermore, $Y_i = 1$ if $Y_i^* > 0$, but $= 0$ if $Y_i^* \leq 0$. $Y_i = 1$ indicates default. The regression model is about explaining default.

(a) If u_i and v_i are normally distributed, find $E(Y_i|X_i,W_i)$ in terms of the normal cumulative distribution function $\Phi(\cdot)$.
(b) Find the likelihood function.
(c) Alternatively, show how you can estimate the parameters b_0, b_1, b_2, c_1, c_2, σ_v^2 and σ_{uv} using the generalized method of moments with over-identification in the orthogonality conditions.

Ans:

(a) Firstly, note Z_i is endogenous, i.e. $cov(Z_i, u_i) \neq 0$.

$E(Y_i|X_i,W_i) = Prob(Y_i^*>0|X_i,W_i) = Prob(b_0 + b_1X_i + b_2Z_i + u_i >0|X_i,W_i)$

$= Prob(u_i > - b_0 - b_1X_i - b_2Z_i \ |X_i,W_i) = Prob(u_i < b_0 + b_1X_i + b_2Z_i \ |X_i,W_i)$

Since $cov(Z_i, u_i) \neq 0$, we cannot write the last quantity as $\Phi(b_0 + b_1X_i + b_2Z_i \ |X_i,W_i)$.

(This is like saying Prob(u<Z), When Z=1/2 u, then in fact the prob becomes Prob(1/2 u < 0)! since Z is not exogenous.) Thus $E(Y_i|X_i,W_i) \neq \Phi(b_0 + b_1X_i + b_2Z_i \ |X_i,W_i)$.

But $E(Y_i|X_i,W_i) = \text{Prob}(u_i < b_0 + b_1X_i + b_2(c_1X_i + c_2W_i + v_i) \,|X_i,W_i)$

$= \text{Prob}(u_i - b_2v_i < b_0 + b_1X_i + b_2(c_1X_i + c_2W_i) \,|X_i,W_i)$

But $u_i - b_2v_i \sim N(0, 1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv})$.

So $E(Y_i|X_i,W_i) = \text{Prob}(u_i - b_2v_i < b_0 + b_1X_i + b_2(c_1X_i + c_2W_i) \,|X_i,W_i)$

$= \text{Prob}([u_i - b_2v_i]/\sqrt{1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv}} < [b_0 + b_1X_i + b_2(c_1X_i + c_2W_i)]/\sqrt{1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv}} \,|X_i,W_i)$

$= \Phi([b_0 + b_1X_i + b_2(c_1X_i + c_2W_i)]/\sqrt{1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv}} \,|X_i,W_i)$.

(b) Likelihood function $L = \prod_{i=1}^{N} \Phi^{Y_i} (1-\Phi)^{1-Y_i}$

where $\Phi = \Phi([b_0 + b_1X_i + b_2(c_1X_i + c_2W_i)]/\sqrt{1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv}} \,|X_i,W_i)$.

(c) From (a), $E(Y_i - \Phi([b_0 + b_1X_i + b_2(c_1X_i + c_2W_i)]/\sqrt{1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv}}) \,|X_i,W_i) = 0$.

Similarly, $E\{X_i(Y_i - \Phi([b_0 + b_1X_i + b_2(c_1X_i + c_2W_i)]/\sqrt{1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv}}) \,|X_i,W_i)\} = 0$.

$E\{W_i(Y_i - \Phi([b_0 + b_1X_i + b_2(c_1X_i + c_2W_i)]/\sqrt{1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv}}) \,|X_i,W_i)\} = 0$.

Let $\varepsilon_I = Y_i - \Phi([b_0 + b_1X_i + b_2(c_1X_i + c_2W_i)]/\sqrt{1 + b_2^2\sigma_v^2 - 2b_2\sigma_{uv}} \,|X_i,W_i)$ where $E(\varepsilon_I|X_i,W_i)=0$.

Then $E(X_i\varepsilon_I \,|\, X_i,W_i) = 0$, $E(W_i\varepsilon_I \,|\, X_i,W_i) = 0$.

Moreover, $E(Z_i - c_1X_i - c_2W_i \,|\, X_i,W_i) = 0$, $E(X_i[Z_i - c_1X_i - c_2W_i] \,|\, X_i,W_i) = 0$, $E(W_i[Z_i - c_1X_i - c_2W_i] \,|\, X_i,W_i) = 0$,

$E([Z_i - c_1X_i - c_2W_i]^2 - \sigma_v^2 \,|\, X_i,W_i) = 0$.

And we can add more instruments such as X_i^2, W_i^2, etc.

Hence unconditional expectation $E[\varepsilon_I] = E_{X,W}(E[\varepsilon_I|\, X_i,W_i]) = 0$. Then the unconditional orthogonality conditions are (integrating over X_i and W_i)

$E(\varepsilon_I) = 0$, $E(X_i\varepsilon_I) = 0$, $E(W_i\varepsilon_I) = 0$, $E(Z_i - c_1X_i - c_2W_i) = 0$, $E(X_i[Z_i - c_1X_i - c_2W_i]) = 0$,

$E(W_i[Z_i - c_1X_i - c_2W_i]) = 0$, $E([Z_i - c_1X_i - c_2W_i]^2 - \sigma_v^2 \,|\, X_i,W_i) = 0$,

$E(X_i^2\varepsilon_I) = 0$, $E(W_i^2\varepsilon_I) = 0$,

to estimate 7 parameters: b_0, b_1, b_2, c_1, c_2, σ_v^2 and σ_{uv} with 9 moments restrictions and 2 over-identifying conditions. The overidentification must be "solved" by minimizing the covariance matrix of the sample moments.

Note: c_1, c_2, σ_v^2 can be estimated as a BLUE and consistent estimates using $Z_i = c_1X_i + c_2W_i + v_i$ (OLS or MLE). Call these $\hat{c}_1$, $\hat{c}_2$ $\hat{\sigma}_v^2$. These then reduces the need to use the following restrictions:

$E(Z_i - c_1X_i - c_2W_i) = 0$.

$E\{X_i(Z_i - c_1X_i - c_2W_i)\} = 0$.

$E\{W_i(Z_i - c_1X_i - c_2W_i)\} = 0$.