Question A1.
Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

where $\mathbf{y}$ is a $T \times 1$ vector of observations on a dependent variable, $\mathbf{X}$ is a $T \times k$ full-rank matrix of observations on exogenous variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters, and $\mathbf{u}$ is a $T \times 1$ vector of unobserved normally distributed disturbances, with $E(\mathbf{u}) = 0$,and with $E(\mathbf{u}\ \mathbf{u}') = \sigma^2\mathbf{I}$.

1. (a) [20%] Derive the ordinary least squares estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, and show it is unbiased.
   (b) [20%] Derive the variance-covariance matrix for $\hat{\boldsymbol{\beta}}$.
   (c) [20%] Let $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, show $\hat{\mathbf{u}} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{u}$, where $\mathbf{M} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$
   (d) [20%] What is $E(\hat{\mathbf{u}}'\ \hat{\mathbf{u}})$? Using this derive an unbiased estimator of σ^2.
   (e) [20%] How would you estimate the standard error of the ith element of $\hat{\boldsymbol{\beta}}$, $\hat{\beta}_i$?

Answer.
a. Minimise

$$\begin{aligned}
S &= \hat{\mathbf{u}}'\hat{\mathbf{u}} = \left(y'y + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{2}\hat{\boldsymbol{\beta}}\mathbf{X}'y\right) \\
\frac{\partial S}{\partial \hat{\boldsymbol{\beta}}} &= 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{2X}'y = 0 \\
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}
\end{aligned}$$

So given exogeneity implies independence and $E(\mathbf{u}) = 0$,

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + \mathbf{E}\ \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right\} \\
&= \boldsymbol{\beta} + \mathbf{E}\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}E(\mathbf{u})
\end{aligned}$$

$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, it is unbiased.
b. From above $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$, so

$$V(\hat{\boldsymbol{\beta}}) = \mathbf{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'$$

$$\begin{aligned}
V(\hat{\boldsymbol{\beta}}) &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right)') \\
&= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\ \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\ \mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

c. $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{My}$

$\hat{\mathbf{u}} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{X}\boldsymbol{\beta}+\mathbf{u}) = (\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'u = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'u$

$\hat{\mathbf{u}} = \mathbf{u} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{u} = \mathbf{Mu}$

d. $E(\hat{\mathbf{u}}'\ \hat{\mathbf{u}}) = \mathbf{E}(\mathbf{u}'\mathbf{M}'\mathbf{Mu}) = \mathbf{E}(\mathbf{u}'\mathbf{Mu}) = \mathbf{E}(\mathbf{tr}(\mathbf{u}'\mathbf{u})\mathbf{trM}) = \boldsymbol{\sigma}^2(T-k)$, so $E(\hat{\mathbf{u}}'\ \hat{\mathbf{u}}/(\mathbf{T}-\mathbf{K})) = \boldsymbol{\sigma}^2$, estimate $\boldsymbol{\sigma}^2$ by $s^2 = \hat{\mathbf{u}}'\ \hat{\mathbf{u}}/(\mathbf{T}-\mathbf{K})$.

e estimate $V(\hat{\boldsymbol{\beta}})$ by $s^2\,(\mathbf{X}'\mathbf{X})^{-1}$ then the square root of the ith diagonal element is the standard error of $\hat{\beta}_i$.

Question A2

The structural form of a simultaneous system for two variables

$$\mathbf{By}_t = \boldsymbol{\gamma} + \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \mathbf{u}_t;\ E(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Sigma}.$$

$$\begin{bmatrix} 1 & 0 \\ \beta_{21} & 1 \end{bmatrix}\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

has reduced form

$$\mathbf{y}_t = \mathbf{a} + \mathbf{Ay}_{t-1} + \boldsymbol{\varepsilon}_t;\ E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \Omega.$$

where $E(u_{1t}) = E(u_{2t}) = 0$; $E(u_{1t}^2) = \sigma_{11}$; $E(u_{2t}^2) = \sigma_{22}$; $E(u_{1t}u_{2t}) = \sigma_{12} = 0$; $E(u_{it}u_{j,t-h}) = 0$; $i,j = 1,2$, $h \neq 0$, for $t = 1, 2, ..., T$.

1. (a) [25%] Express $\mathbf{a}, \mathbf{A}, \boldsymbol{\varepsilon}_t, \Omega.$ in terms of $\mathbf{B}, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \mathbf{u}_t, \boldsymbol{\Sigma}.$
   (b) [25%] How would you estimate $\mathbf{a}, \mathbf{A}, \boldsymbol{\varepsilon}_t, \Omega$?
   (c) [25%] Discuss the identification of the structural form. How would you estimate $\mathbf{B}, \boldsymbol{\gamma}, \boldsymbol{\Gamma}, \mathbf{u}_t, \boldsymbol{\Sigma}$?
   (d) [25%] Explain how the structural model relates to the Choleski decomposition used to calculate impulse response functions in a VAR.

Answer

(a)

$$\begin{aligned} \mathbf{y}_t &= \mathbf{a} + \mathbf{Ay}_{t-1} + \boldsymbol{\varepsilon}_t;\ E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \Omega. \\ \mathbf{y}_t &= \mathbf{B}^{-1}\boldsymbol{\gamma} + \mathbf{B}^{-1}\boldsymbol{\Gamma}\mathbf{y}_{t-1} + \mathbf{B}^{-1}\mathbf{u}_t;\ E(\mathbf{B}^{-1}\mathbf{u}_t\mathbf{u}_t'\mathbf{B}^{-1'}) = \mathbf{B}^{-1}\boldsymbol{\Sigma}\mathbf{B}^{-1'}. \end{aligned}$$

(b) In $\mathbf{y}_t = \mathbf{a} + \mathbf{Ay}_{t-1} + \boldsymbol{\varepsilon}_t$; $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \Omega$

$\hat{\mathbf{a}}$ and $\hat{\mathbf{A}}$ are estimated by ordinary least squares on each equation $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{a}} - \hat{\mathbf{A}}\mathbf{y}_{t-1}$, $\hat{\Omega} = \hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'/T$.

$$\begin{aligned} y_{1t} &= a_{10} + a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + \varepsilon_{1t} \\ y_{2t} &= a_{20} + a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + \varepsilon_{2t} \end{aligned}$$

$\widehat{\omega}_{11} = \sum\widehat{\varepsilon}_{1t}^2/T$, $\widehat{\omega}_{22} = \sum\widehat{\varepsilon}_{2t}^2/T$, $\widehat{\omega}_{12} = \sum\widehat{\varepsilon}_{1t}\widehat{\varepsilon}_{2t}/T$,

(c) Since $\mathbf{B}$ is triangular and $\boldsymbol{\Sigma}$ is diagonal, the system is recursive and each structural equation can be estimated by OLS.

(d) For the same ordering the Choleski decomposition is the same as a recursive system.

Question A3

In the stochastic processes below ε_t is an independent normal white noise process, that is $\varepsilon_t\tilde{}IN(0,\sigma^2)$, and there is a sample of data for y_t, $t = 1, 2, ..., T$.

1. (a) [20%] The trend stationary process is

$$y_t = \alpha + \rho y_{t-1} + \delta t + \varepsilon_t.$$

How would you estimate the coefficients in this equation and using those estimates how would you forecast y_{T+1} and y_{T+2}?

(b) [20%] The ARIMA(1,1,1) process is

$$\Delta y_t = \alpha + \rho \Delta y_{t-1} + \varepsilon_t + \mu\varepsilon_{t-1.},$$

How would you estimate the coefficients in this equation and and using those estimates how would you forecast y_{T+1} and y_{T+2}?

(c) [20%] The random walk with drift is

$$y_t = \alpha + y_{t-1} + \varepsilon_t.$$

Show that y_t can be written as the sum of a deterministic trend, a stochastic trend and an initial condition, y_0.

(d) [20%] How would you estimate the random walk with drift equation and using those estimates how would you forecast y_{T+1} and y_{T+2}. What are the variances of your forecasts?

(e) [20%] In the ARIMA(1,1,0) process:

$$\Delta y_t = \alpha + \rho \Delta y_{t-1} + u_t,$$

where u_t may be serially correlated, how would you implement an augmented Dickey-Fuller test to test the null hypothesis that y_t was $I(2)$ against the alternative that it was $I(1)$.

Answer

a. Estimate the parameters by OLS and the forecasts would be

$$\begin{aligned} y^f_{T+1} &= \hat{\alpha} + \hat{\rho} y_T + \hat{\delta}(T+1). \\ y^f_{T+2} &= \hat{\alpha} + \hat{\rho} y^f_{T+1} + \hat{\delta}(T+2) \end{aligned}$$

b.Estimate the parameters by maximum likelihood using an interative non-linear procedure. The forecasts are

$$\begin{aligned} \Delta y^f_{T+1} &= \hat{\alpha} + \hat{\rho}\Delta y_T + \hat{\mu}\hat{\varepsilon}_{T.}, \\ \Delta y^f_{T+2} &= \hat{\alpha} + \hat{\rho}\Delta y^f_{T+1}; \\ y^f_{T+1} &= y_T + \Delta y^f_{T+1} \\ y^f_{T+2} &= y^f_{T+1} + \Delta y^f_{T+2}. \end{aligned}$$

c. Substituting back

$$
\begin{aligned}
y_t &= \alpha + (\alpha + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
y_t &= 2\alpha + y_{t-2} + \varepsilon_t + \varepsilon_{t-1} \\
y_t &= \alpha t + y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i}.
\end{aligned}
$$

d. $\hat{\alpha}$ is just the mean of Δy_t

$$
\begin{aligned}
y^f_{T+1} &= \hat{\alpha} + y_T;\ \text{var}{=}\sigma^2 \\
y^f_{T+2} &= 2\hat{\alpha} + y_T,\ \text{var}{=}2\sigma^2
\end{aligned}
$$

e. write it as

$$
\Delta^2 y_t = \alpha + (\rho - 1)\Delta y_{t-1} + \sum_{i=1}^{p} \Delta^2 y_{t-i} + \varepsilon_{tt},
$$

where p is chosen to remove serial correlation and test $b = (\rho - 1) = 0$ using the non standard Dickey-Fuller critical values.

Question B1

1. An unrestricted VAR was estimated on US annual data, 1875-2011 for short interest rates, R_t, long interest rates, RL_t, and inflation, INF_t, in % at annual rates. The Akaike Information Criterion, AIC, suggested 2 lags and the Schwarz Bayesian Information Criterion, BIC, 1 lag. Using a VAR(2) trace tests for r, the number of cointegrating vectors, and 5% critical values assuming unrestricted intercept no trend are:

H_0	H_1	$trace$	CV
$r = 0$	$r \geq 0$	75.64	29.8
$r = 1$	$r \geq 2$	22.95	14.3
$r = 2$	$r = 3$	2.54	3.8

While assuming unrestricted intercept restricted trend ,they are

H_0	H_1	$trace$	CV
$r = 0$	$r \geq 0$	92.34	42.91
$r = 1$	$r \geq 2$	38.31	25.87
$r = 2$	$r = 3$	3.6	12.52

Assuming $r = 2$, the estimated cointegrating relations (standard errors) [t statistics] are

Cointegrating Eq:	CointEq1	CointEq2
R(-1)	1.000000	0.000000
RL(-1)	0.000000	1.000000
INF(-1)	-1.110979 (0.20301) [-5.47262]	-1.225494 (0.18568) [-6.60000]
C	-2.425593	-2.130975

The coefficients (standard errors) [t statistics] of the ECM, cointegrating equation terms, in the VECM are given below (constants and lagged changes are not reported)

Error Correction:	D(R)	D(RL)	D(INF)
CointEq1	-0.350208	0.042025	-1.048366
	(0.09837)	(0.04471)	(0.35338)
	[-3.55998]	[0.94002]	[-2.96666]
CointEq2	0.317551	-0.052172	1.499932
	(0.09422)	(0.04282)	(0.33846)
	[3.37028]	[-1.21845]	[4.43158]

(a) [20%] Explain how the AIC and BIC are calculated and why the BIC suggests a shorter lag than the AIC.

(b) [20%] Explain what unrestricted intercept, restricted trend means in the context of an estimated VECM. Explain how to interpet the test statistics and why they suggest $r = 2$ in both cases.

(c) [20%] For $r = 2$, how many just identifying restrictions are required on the cointegrating vectors? Explain the just identifying restrictions used here.

(d) [20%] Test the individual hypotheses that the coefficients on inflation are equal to minus one: $\beta_{13} = -1; \beta_{23} = -1$. A likelihood ratio test of the joint hypothesis that both coefficients are equal to minus one gave a test statistic of 1.74. What is the distribution of this statistic? Is the joint hypothesis rejected?

(e) [20%] Comment on the economic interpretation of the system and the pattern of adjustment. Which variable appears exogenous?

Answer

a. AIC is -2(MLL-K) BIC is -2(MLL-0.5KlnT) where T is the sample size and K the number of parameters. BIC penalises extra parameters more than AIC so suggests a shorter lag length.

b in the VECM the intercept μ lies outside the cointegrating space the trend inside it, restricted to r rather than 3 coefficients.

$$\Delta y_t = \mu + \alpha(\beta y_{t-1} + \gamma t)$$

The test statistic is greater than the critical value for the null of $r = 0$, and $r = 1$, but not for r=2, so we do not reject the hypothesis r=2.

c. 2 on each CV, there they are $\beta_{11} = 1, \beta_{12} = 0, \beta_{21} = 0, \beta_{22} = 1$, These just identifying restrictions have each interest rate determined by the rate of inflation.

d. (-1.11+1)/0.2=-0.55, (-1.22+1)/0.18=-1.22, not rejected in either case. LR test statistic is Chi-squared 2, critical value 5.99, so not rejected.

e. The two cointegrating vectors can be interpreted as long-run real interest rates, both have significant feedbacks on the short rate and inflation, but no feedback on the long-rate which may suggest it is exogenous. Since the coefficients on the ECM terms are roughly equal and opposite, it may be interpred as a feedback from the spread on interest rates and inflation.

Question B2

1. Data for 189 countries were used to explain life expectancy in each country, LE, which ranged from 48.3 to 83.4 years, by the logarithm of per capita income, LPCI. The data were ordered by LPCI starting with the lowest and going to the highest. The EViews output is given below.

Dependent Variable: LE
Method: Least Squares
Date: 01/31/18 Time: 11:26
Sample: 1 189
Included observations: 189

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.27829	2.568390	9.452729	0.0000
LPCI	5.172650	0.279985	18.47472	0.0000

R-squared	0.646045	Mean dependent var	71.28771
Adjusted R-squared	0.644152	S.D. dependent var	8.050099
S.E. of regression	4.802125	Akaike info criterion	5.986520
Sum squared resid	4312.296	Schwarz criterion	6.020824
Log likelihood	-563.7261	Hannan-Quinn criter.	6.000417
F-statistic	341.3155	Durbin-Watson stat	2.176566
Prob(F-statistic)	0.000000		

Analysis of the residuals showed a skewness coefficient of -1.32 and a kurtosis coefficient of 5.76 and some large outliers. The Jarque-Bera, JB, test statistic, which is $\chi^2(2)$ under the null of normality was 115. The BPG heteroskedasticity test had a p value of 0.6043 and the RESET test for non-linearity, using the squared fitted value had a p value of 0.1977.

The equation was re-estimated adding dummy variables for the observations corresponding to the four largest residuals at observations 53, 60, 97 and 162: Lesotho, Swaziland, Gabon and Equatorial Guinea. The results are shown on the next page.

Dependent Variable: LE
Method: Least Squares
Date: 01/31/18 Time: 11:48
Sample: 1 189
Included observations: 189

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.26248	2.187408	11.09188	0.0000
LPCI	5.216575	0.238511	21.87140	0.0000
D53	-20.43737	4.085622	-5.002266	0.0000
D60	-16.32016	4.074528	-4.005411	0.0001
D97	-16.63978	4.082461	-4.075918	0.0001
D162	-19.06107	4.071816	-4.681222	0.0000

R-squared	0.752412	Mean dependent var	71.28771
Adjusted R-squared	0.745648	S.D. dependent var	8.050099
S.E. of regression	4.059935	Akaike info criterion	5.671442
Sum squared resid	3016.402	Schwarz criterion	5.774355
Log likelihood	-529.9513	Hannan-Quinn criter.	5.713135
F-statistic	111.2264	Durbin-Watson stat	2.085193
Prob(F-statistic)	0.000000		

This regression had skewness of -0.64 and a kurtosis of 3.3 and a JB statistic of 13.6. The BPG p value was 0.4431 and the RESET p value was 0.0647.

(a) [20%] What values would you expect for the coefficients of skewness and kurtosis if the distribution was normal? Comment on the skewness and kurtosis in the two regressions. Did including the dummy variables make the distribution of the residuals more normal? Explain.

(b) [20%] What effect does non-normality of the errors have on the properties of the least squares estimate of the coefficient of LPCI? Are the results consistent with this theoretical result? Explain.

(c) [20%] Explain how the dummy variables for the outlier observations are constructed and explain how you would interpret the coefficients of the dummy variables.

(d) [20%] Explain how the Durbin-Watson statistic is calculated and what it tells you in this case where the data are not time series but were ordered by the independent variable.

(e) [20%] Explain how the heteroskedasticity and RESET tests are calculated and what they tell you in the case of these two regressions.

Answer

a. Skewness 0, kurtosis 3. Normality is massively rejected in the first regression by the JB stat, shows negative skew, towards the lower tail and excess kurtosis, fatter tails than a normal.. The dummies reduced skewness and kurtosis though normality is still rejected, the rejection is much less strong a much smaller value of the JB.

b. Normality is not one of the Gauss Markov assumptions, so if the other assumptions hold, it remains BLUE and fixing the non-normality with the dummies did not change the estimates of interecept and slope much.

c. D53 is zero except for observation 53,and one for that observation. The coefficient measures the prediction error for observation 53, life expectancy is 20 years less than would be expected at that level of income.

d. $DW = (\sum \Delta\widehat{u}_i^2)/\sum \widehat{u}_i^2$. Since it uses the change it depends on the ordering and unlike time series there is no natural ordering for cross-section. If they are ordered by independent variable it may pick up non-linearity.

e. BPG regresses squared residuals on explanatory variables, reset test regresses residuals on squared fitted values. No evidence of problems in either case at 5% level, but dummies may have increased power of the reset test.

Question B3

US data, 1872-2014, on log dividends, d_t, and log earnings, e_t, for firms in the S&P500 were used to estimate the following ARDL and 2 ECM models:

$$\begin{aligned}
ARDL &: \quad d_t = \alpha_0 + \alpha_1 d_{t-1} + \beta_0 e_t + \beta_1 e_{t-1} + u_t; \\
ECM1 &: \quad \Delta d_t = a_0 + a_1 d_{t-1} + b_0 \Delta e_t + b_1 e_{t-1} + u_t; \\
ECM2 &: \quad \Delta d_t = \lambda_1 \Delta d_t^* + \lambda_2 (d_{t-1}^* - d_{t-1}) + u_t \\
d_t^* &= \quad \theta_0 + \theta e_t.
\end{aligned}$$

The estimated coefficients (standard errors) are

$$\begin{array}{lllllllll}
d_t = & -0.13 & +0.65 & d_{t-1} & +0.20 & e_t & +0.12 & e_{t-1} & +\hat{u}_t \quad R^2 = 0.997, \\
 & (0.02) & (0.03) & & (0.03) & & (0.03) & & \quad SER = 0.086
\end{array}$$

$$\begin{array}{lllllllll}
\Delta d_t = & -0.13 & -0.35 & d_{t-1} & +0.20 & \Delta e_t & +0.32 & e_{t-1} & +\hat{u}_t \quad R^2 = 0.526, \\
 & (0.02) & (0.03) & & (0.03) & & (0.03) & & \quad SER = 0.086
\end{array}$$

$$\begin{array}{lllllllll}
\lambda_1 = & 0.22 & \lambda_2 = & 0.35 & \theta_0 = & -0.38 & \theta = & 0.91 & R^2 = 0.526, \\
 & (0.03) & & (0.03) & & (0.02) & & (0.01) & SER = 0.086
\end{array}$$

1. (a) [20%] Explain the relationship between the parameters in the ARDL and ECM equations.

   (b) [20%] Explain ECM2 and interpret its four parameters. What is the long-run elasticity of earnings to dividends? Is it significantly different from one? How do you interpret $\theta_0 < 0$.

   (c) [20%] Wald tests on ECM1 for (i) $H_0 : a_1 + b_1 = 0$ and (ii) $H_0 : a_1/b_1 + 1 = 0$, give χ^2 test statistics of 27.1 and 48.6 respectively. Interpret the hypotheses. Give the 5% critical values using the table at the beginning of the exam.. Are the hypotheses rejected? What does this example reveal about Wald tests?

(d) [20%] Diagnostic tests on ECM1 and p values are: heteroskedasticity (regression of $\hat{u}_t^2$ on the regressors) $p = 0.0035$; normality (skewness-kurtosis): $p = 0.0000$; up to second order serial correlation: $p = 0.0672$. In which cases do the tests indicate failure of the assumptions at the 5% level.? What are the consequences of the failures?

(e) [20%] ECM2 was estimated by non-linear least squares, using starting values $\lambda_1 = 0.3$, $\lambda_2 = 0.3$, $\theta_0 = 1$, $\theta = 1$,and converged after 5 iterations. Explain the procedure for estimating non-linear models, including the role of starting values and convergence criteria. Comment on the choice of starting values.

Answer

a. Reparameterised

$$d_t - d_{t-1} = \alpha_0 + (\alpha_1 - 1)d_{t-1} + \beta_0(e_t - e_{t-1}) + (\beta_0 + \beta_1)e_{t-1} + u_t$$

b. There is a long run relationship determining target dividends and a short run adjustment to the change in target, λ_1, and the previous error, λ_2. θ is the long run elasticity, it is significantly different from one, $(0.91 - 1)/0.01 = 9$, the underlying long run equation is

$$D_t = KE_t^\theta$$

so $\theta_0 = \ln K$, can be negative if $K < 1$.

c. In both cases the hypothesis is that $\theta = 1$, the test statistic is chi-squared(1), CV=3.84, so both are rejected. Shows Wald tests are not invariant to how the non-linear restriction is written.,

d. Evidence of heteroskedasticity, non-normality, at 5% level. With heteroskedasticity estimates remain unbiased but not minimum variance and the standard errors are wrong. Normality is not one of the Gauss-Markov assumptions but test statistics may be wrong in small samples.

e. Use initial values θ^0 to evaluate the log-likelihood LL^0 (or SSR), and first and second derivates, use these to determine the direction up (down) and the size of the next step to θ^0 Continue until it converges in that $LL^n - LL^{n-1} < \epsilon$, $\partial LL/\partial\theta < \epsilon$, where ϵ is the convergence criteria. The λ are adjustment coefficients so starting values between zero and one are sensible, you expect θ close to 1, θ_0 small negative number might have been better.