A Vector Error Correction Model (VECM) of Stockmarket Returns

By

Nagaratnam J Sreedharan

School of Economics
University of Tasmania
GPO Box 252-85
Hobart Tasmania 7001
E-mail: nj.sreedharan@utas.edu.au
Phone: +61 03 62267671
Fax: +61 03 62267587

April 2004

Abstract

Empirical observations on security prices and other financial time series usually not only include the closing prices (C_t), but also the opening, the highest and the lowest prices (O_t, H_t, L_t) for specific horizons such as days, weeks and months. A multivariate vector of prices (O_t, H_t, L_t, C_t) is obviously more informative than just the close prices (C_t) for modelling and forecasting. In this paper we attempt to capture the return generation process of security prices using all the quoted prices (O_t, H_t, L_t, C_t) via a vector error correction (VEC) process.

The results of the empirical models using US daily Dow Jones Industrial (DJI) index data from 1990 to 2000 (11 years) indicate some interesting stylised facts regarding security returns. We show, via the return generation process (RGP) proposed, that the "cointegrating" returns exhibit significant explanatory power. Some insights are also provided as to why ΔC_t logarithmic returns tend to be non-normally distributed.

Key words: vector autoregression (VAR), vector error correction (VEC), cointegration (CI), return generation process (RGP) and return distributions.

JEL classification:C32

1 Introduction

This paper is based on the premise that "it is possible that the unconditional distribution of asset returns may become normal, once the static and dynamic relationships are accounted for" Markellos (2002). In other words, normality may be falsely rejected due to the fact that the return generation process (RGP) was mis-specified.

The goal of this paper is to account for the static and dynamic relationships in asset returns using all the price and return vectors. A dynamic model of asset returns using the vector error correction model (VECM) representation of Engle and Granger (1987) is applied with the insight that even though open, high, low and close prices are non-stationary they might be cointegrated. In doing so, we are able to separate the non-marginal (information-based) and marginal (expectations-based) aspects of the return generation process. It turns out that the price generation process is the error correction process (ECP) of the VECM representation. Surprisingly, we still find that the close-to-close residuals, after accounting for static and dynamic relationships, are not only non-normally distributed but also abnormally distributed.

Section 2 describes the model adopted. Section 3 describes the dataset. Section 4 fits a VEC model. Section 5 fits a VECM model. In Section 6 we introduce the VECM-lead(CointEq1) Model. Section 7 attempts the VAR lead-lag(CointEq1,2,3) Model. The cointegrating vectors are highlighted in Section 8. In Section 9 we take stock of the ARCH process. Section 10 summarises the findings and suggests future directions for research.

2 The Model

Financial theory assumes that the behaviour of asset returns is the result of current and past information. In an informationally efficient market, "price changes must be unforecastable if they are properly anticipated, i.e., if they incorporate the expectations and information of all market participants" Lo and MacKinlay (1999). Since "expectations" is another form of "information", expectations are subsumed under information and we habitually ignore the effect of expectations by stating "prices reflect all available information" Samuelson (1965), rather than "prices reflect all expectations and available information". This could be partly due the subjectivity implied by expectations and objectivity implied by information.

The timeseries model proposed in this paper attempts to distinguish "expectations" from "information" in the price generation process. We assume "changes in current price is dependent on changes in past changes in price, current and immediate-past information and expectations". Thus we define a VECM-lead(CointEq1,2,3) model given as:

(1) $$\Delta P_t = \delta + \sum_{i=1}^{p} \alpha_i \Delta P_{t-i} + \left\{ \sum_{j=0}^{1} \beta_j \xi_{t-j} + \sum_{j=0}^{1} \gamma_j^{+} \xi_{t-j}^{+} + \sum_{j=0}^{1} \gamma_j^{-} \xi_{t-j}^{-} \right\} + \varepsilon_t$$

where ζ_t (zeta sub t) are current and immediate-past "normal" information and ξ_t (xi sub t) are the current and immediate-past positive and negative expectation(s) and ε_t (epsilon sub t) are the current abnormal disturbances. The terms in the within the square brackets are the error correction terms and δ is the long-run risk premium.

We determine the current informational and expectational "disturbances", ζ_t and ξ_t through a series of steps. We first analyse the multivariate data using a VAR representation. We then adopt a VEC representation to extract the cointegrating vectors. The cointegrating vectors have a lag of 1 by construction and proxy the lag 1 informational and expectational disturbances. Hence the "lead 1 cointegrating vectors" are assumed to capture the current informational and expectational disturbances. This subsequently enables us to extract the abnormal informational shocks (ε_t) by using a VAR representation with the leading cointegrating vectors and the exogenous variables.

3 The Dataset

The dataset is the daily DJI30 index prices from 1/1/1990 to 1/1/2001 covering a period of 11 years (2780 x 4 points). We use index data instead of a single asset because the index generally reflects the behaviour the market as a whole. We use an US index data as the US is one of the largest and most researched markets.

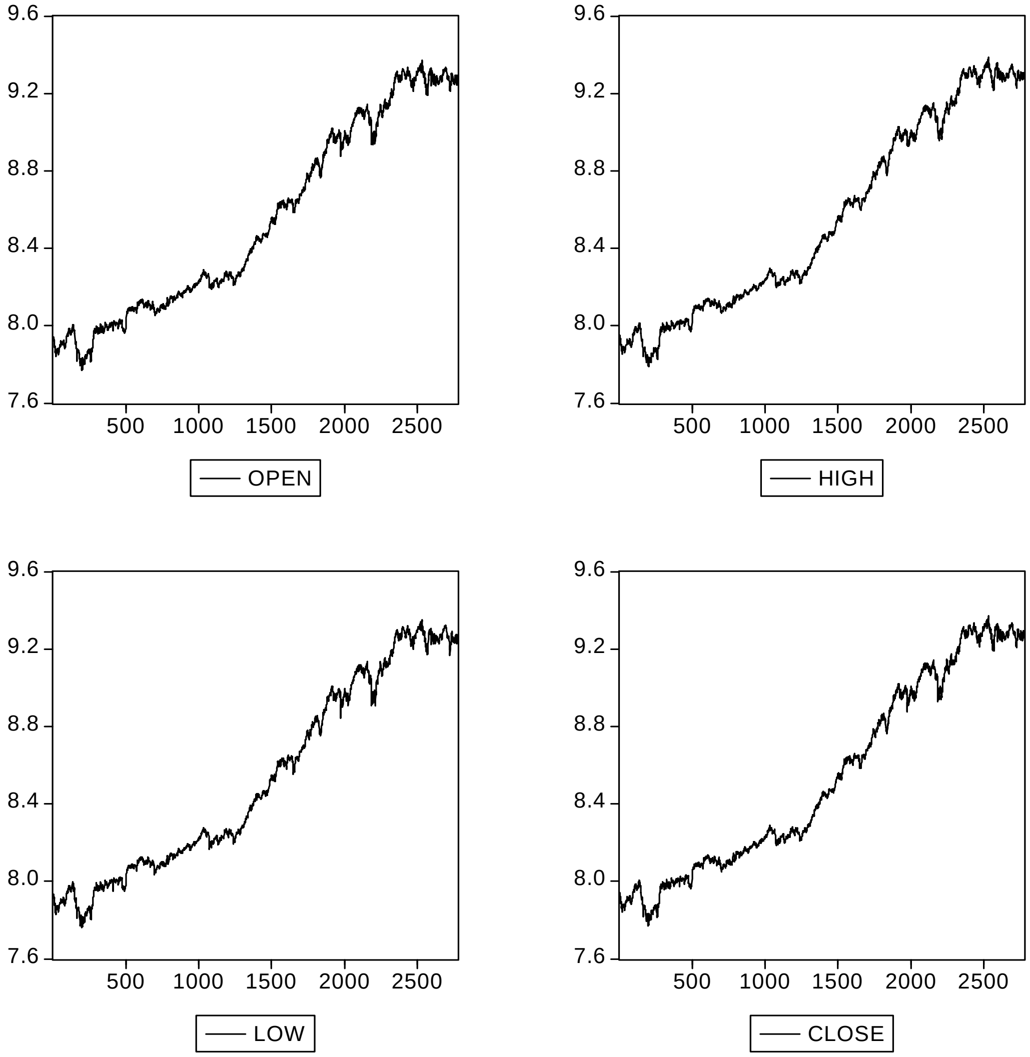


Figure 3-1 DJI30 OPEN LOW HIGH CLOSE Log-Prices (1/1/1990-1/1/2001)

Figure 3-1 shows the open, high, low and close logarithmic prices for chosen period.

Null Hypothesis: CLOSE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on Modified HQ, MAXLAG=30)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.094453	0.9483
Test critical values:	1% level	-3.432512	
	5% level	-2.862381	
	10% level	-2.567262	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CLOSE)
Method: Least Squares
Date: 05/26/04 Time: 09:28
Sample(adjusted): 2 2780
Included observations: 2779 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CLOSE(-1)	-3.40E-05	0.000360	-0.094453	0.9248
C	0.000774	0.003077	0.251589	0.8014
R-squared	0.000003	Mean dependent var		0.000484
Adjusted R-squared	-0.000357	S.D. dependent var		0.009372
S.E. of regression	0.009373	Akaike info criterion		-6.501182
Sum squared resid	0.243984	Schwarz criterion		-6.496915
Log likelihood	9035.393	F-statistic		0.008921
Durbin-Watson stat	1.944175	Prob(F-statistic)		0.924756

Table 3-1 Augmented Dickey-Fuller test statistic for CLOSE Log-Prices

Non-stationary of the log-close prices cannot be rejected using the augmented Dickey-Fuller test as shown in Table 3-1. Similar results were obtained for the other variables.

Figure 3-2 depicts the timeseries plots of the differenced logarithmic prices (logarithmic returns). The log-return series appear similar but are not identical, meaning $\Delta O_t \approx \Delta H_t \approx \Delta L_t \approx \Delta C_t$ and they seem to share similar shocks or disturbances.

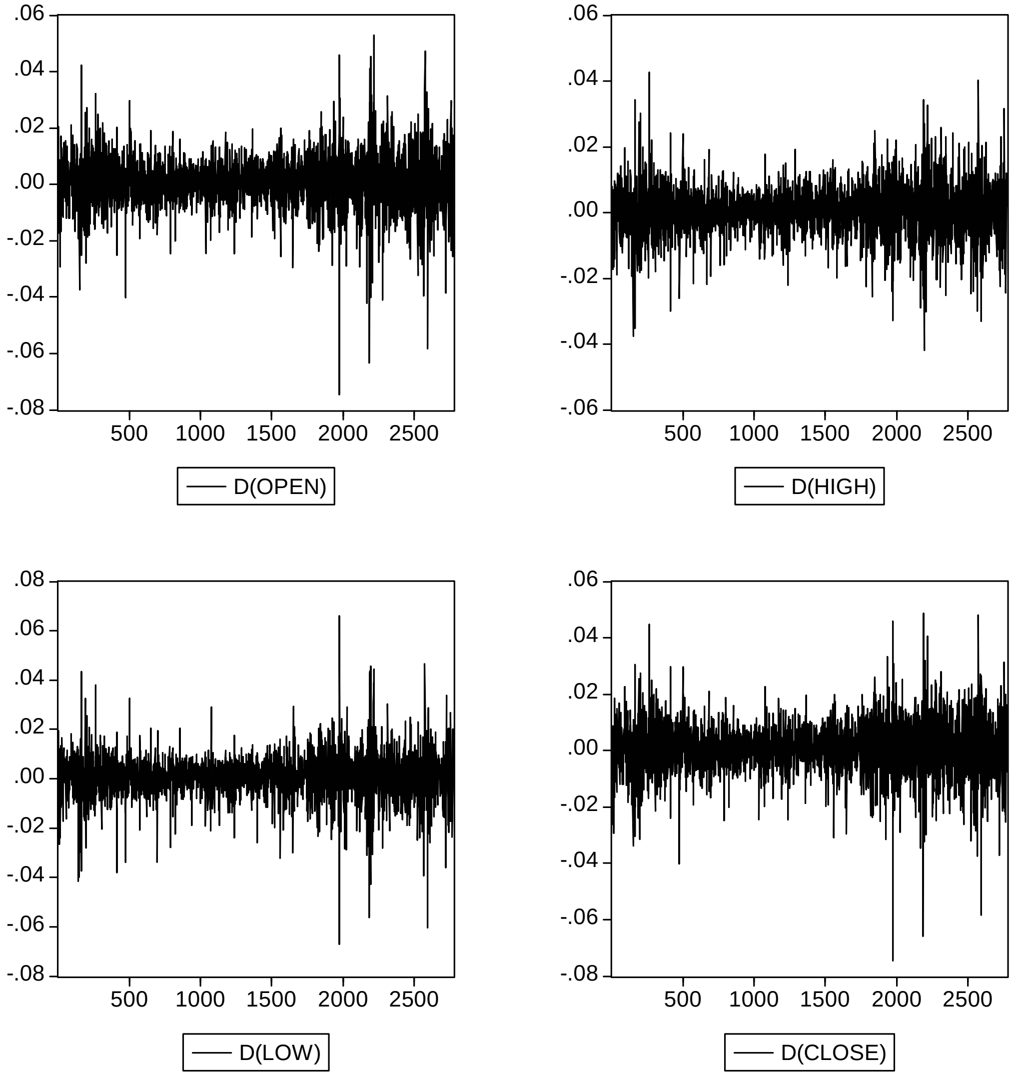


Figure 3-2 D(OPEN) D(LOW) D(HIGH) D(CLOSE) Log-Returns

	D(OPEN)	D(HIGH)	D(LOW)	D(CLOSE)
Mean	4.58E-18	5.86E-18	-4.43E-18	-2.05E-18
Median	8.36E-05	0.000275	0.000299	0.000117
Maximum	0.052373	0.042203	0.065505	0.048121
Minimum	-0.075043	-0.042524	-0.067535	-0.075033
Std. Dev.	0.009287	0.007751	0.009038	0.009372
Skewness	-0.406476	-0.177760	-0.366280	-0.412063
Kurtosis	8.123666	5.583676	9.167421	7.732747
Jarque-Bera	3116.283	787.5887	4466.515	2672.249
Probability	0.000000	0.000000	0.000000	0.000000
Sum	7.74E-15	-8.35E-15	-7.71E-15	-8.29E-15
Sum Sq. Dev.	0.239618	0.166885	0.226937	0.243985
Observations	2779	2779	2779	2779

Table 3-2 DJI30 Log-returns Summary Statistics

Table 3-2 summarises the summary statistics for the dataset used. The means do not differ significantly across the variables [see Table 3-3] but the variances differ significantly [see Table 3-4]. All the variables are negatively skewed with moderately high kurtosis. However, the Jarque-Bera statistics significantly rejects the normal distribution for all variables indicating a non-normality of their unconditional distributions.

Test for Equality of Means Between Series
Sample: 1 2780
Included observations: 2780

Method	df	Value	Probability
Anova F-statistic	(3, 11112)	8.82E-28	1.0000

Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	3	2.09E-31	6.96E-32
Within	11112	0.877424	7.90E-05
Total	11115	0.877424	7.89E-05

Category Statistics

Variable	Count	Mean	Std. Dev.	Std. Err. of Mean
D(OPEN)	2779	4.58E-18	0.009287	0.000176
D(HIGH)	2779	5.86E-18	0.007751	0.000147
D(LOW)	2779	-4.43E-18	0.009038	0.000171
D(CLOSE)	2779	-2.05E-18	0.009372	0.000178
All	11116	9.91E-19	0.008885	8.43E-05

Table 3-3 Test for Equality of Means Between Series

Test for Equality of Variances Between Series
Sample: 1 2780
Included observations: 2780

Method	df	Value	Probability
Bartlett	3	124.0436	0.0000
Levene	(3, 11112)	16.29509	0.0000
Brown-Forsythe	(3, 11112)	16.37086	0.0000

Category Statistics

Variable	Count	Std. Dev.	Mean Abs. Mean Diff.	Mean Abs. Median Diff.
D(OPEN)	2779	0.009287	0.006672	0.006672
D(HIGH)	2779	0.007751	0.005705	0.005701
D(LOW)	2779	0.009038	0.006304	0.006301
D(CLOSE)	2779	0.009372	0.006742	0.006741
All	11116	0.008885	0.006356	0.006354

Bartlett weighted standard deviation: 0.008886

Table 3-4 Test for Equality of Variances Between Series

Null Hypothesis: D(CLOSE) has a unit root			
Exogenous: Constant			
Lag Length: 0 (Automatic based on Modified HQ, MAXLAG=30)			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-51.23833	0.0001
Test critical values:	1% level	-3.432512	
	5% level	-2.862381	
	10% level	-2.567262	
*MacKinnon (1996) one-sided p-values.			

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CLOSE,2)
Method: Least Squares
Date: 05/26/04 Time: 09:37
Sample(adjusted): 3 2780
Included observations: 2778 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(CLOSE(-1))	-0.972240	0.018975	-51.23833	0.0000
C	0.000471	0.000178	2.644139	0.0082
R-squared	0.486056	Mean dependent var		-2.63E-06
Adjusted R-squared	0.485871	S.D. dependent var		0.013070
S.E. of regression	0.009371	Akaike info criterion		-6.501591
Sum squared resid	0.243796	Schwarz criterion		-6.497322
Log likelihood	9032.710	F-statistic		2625.366
Durbin-Watson stat	1.997477	Prob(F-statistic)		0.000000

Table 3-5 Augmented Dickey-Fuller test statistic for D(CLOSE) Log-Returns

Form Table 3-5 the log-difference series can be taken to be stationary. The null hypothesis that D(CLOSE) has a unit root can be rejected. Similar results were obtained for the other variables. Hence, for all our vectors, the logarithmic price series are non-stationary and the logarithmic returns series are stationary.

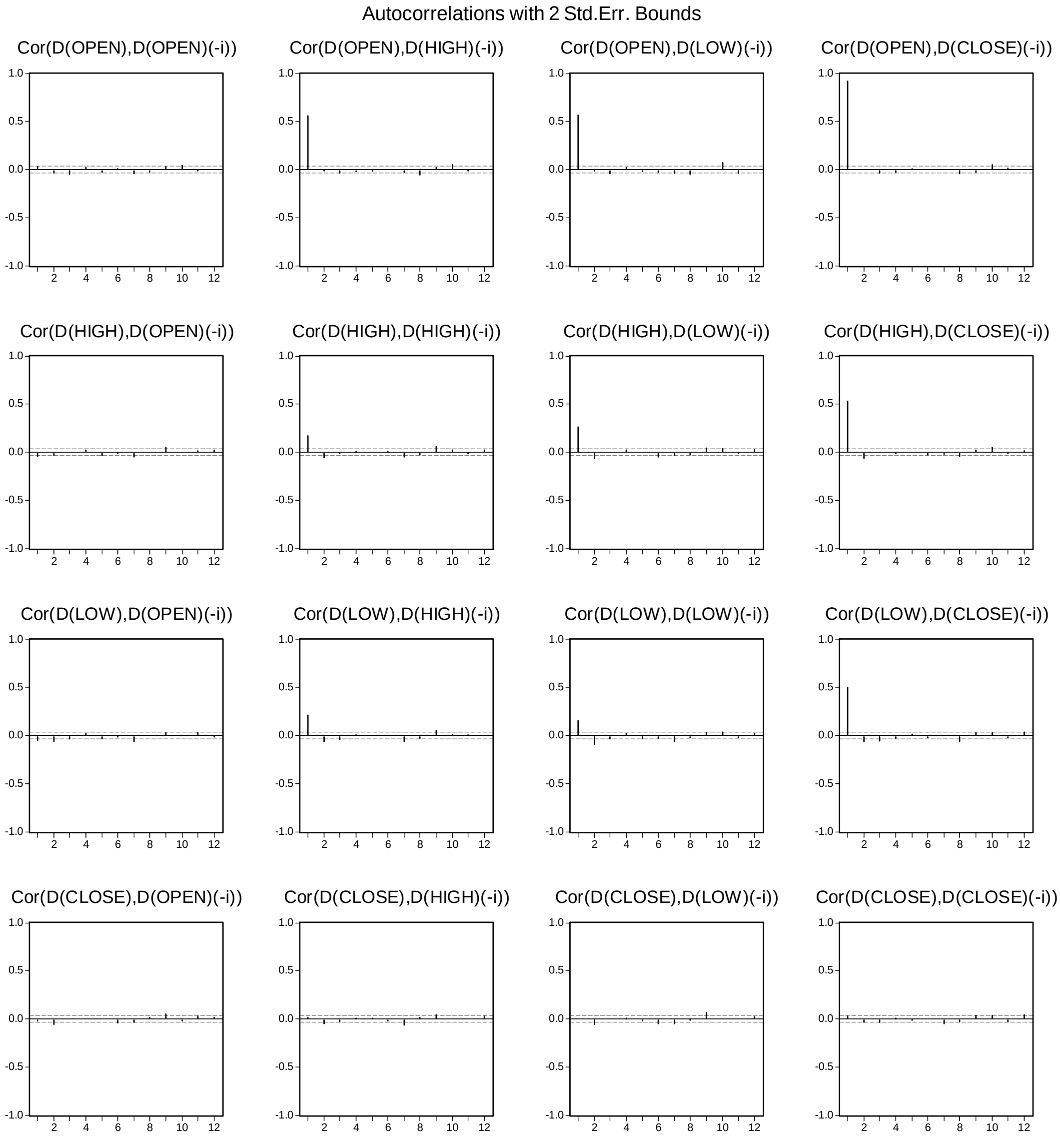


Figure 3-3 LS 0 0 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C

The auto- and cross-correlation plots in Figure 3-3 confirm the lag 1 autoregression for some of the variables. Clearly, the D(OPEN) and D(CLOSE) series are not auto-correlated. In fact, there is no significant autocorrelations for all combinations with D(OPEN). There is significant autocorrelations at lag 1 for the D(HIGH) and D(LOW) series.

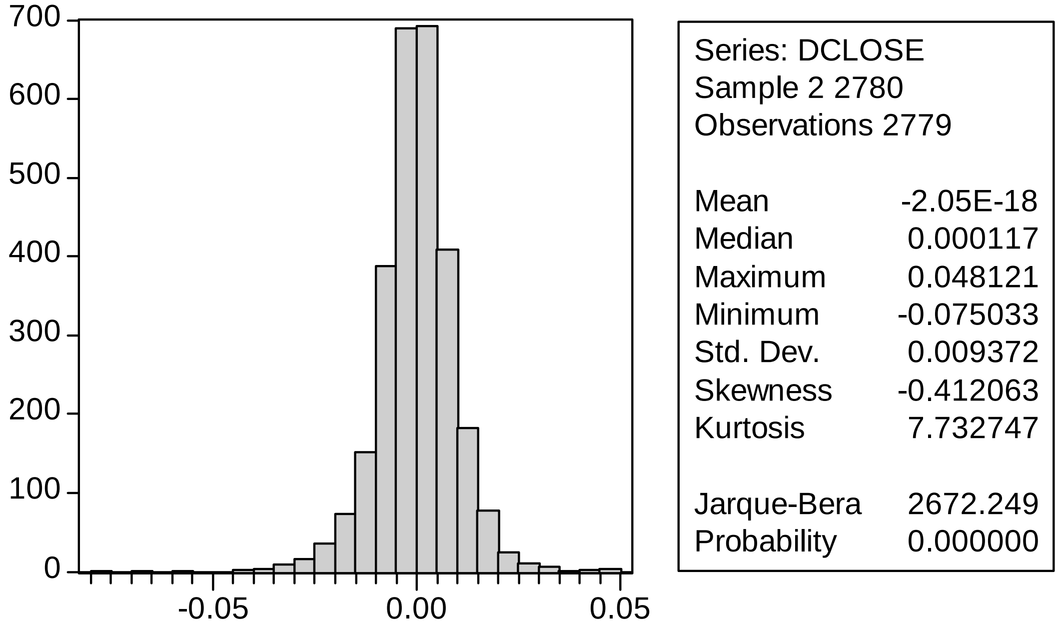


Figure 3-4 D(CLOSE) Histogram and Stats

The unconditional histogram of the close log-returns is highly peaked and moderately skewed. Note that the skewness and kurtosis are the same as for the "raw" logarithmic returns. The maximum and minimum exceed more than 3 standardised deviations and the Jarque-Bera test rejects the normal distribution.

Current financial theory attributes the non-normality to serial correlations and heteroskedasticity in the log-returns series. Thus, we attempt to remove the serial correlations from the dataset but instead of considering the univariate approach whereby an ARMA model is fitted to the close log-return series we use a multivariate approach where we consider all the log-price and log-return series i.e. we use the VECM formulation.

4 The VAR Model

First we undertake a VAR Lag Order selection process. The results for various selection criteria are listed in Table 4-1. The SC selects 10 lags, the HQ selects 11 lags and the rest select 29 lags, including the AIC. In this paper we adopt the HQ criteria and use 11 lags.

VAR Lag Order Selection Criteria
Endogenous variables: D(OPEN) D(HIGH) D(LOW) D(CLOSE)
Exogenous variables: C
Sample: 1 2780
Included observations: 2749

Lag	LogL	LR	FPE	AIC	SC	HQ
0	40148.66	NA	2.43E-18	-29.20673	-29.19812	-29.20362
1	43513.39	6717.226	2.13E-19	-31.64306	-31.60000	-31.62750
2	44111.81	1192.931	1.39E-19	-32.06680	-31.98928	-32.03879
3	44477.09	727.1019	1.08E-19	-32.32091	-32.20895	-32.28046
4	44756.52	555.4008	8.91E-20	-32.51256	-32.36615	-32.45966
5	44893.44	271.7591	8.16E-20	-32.60054	-32.41968	-32.53520
6	45006.29	223.6385	7.61E-20	-32.67100	-32.45569	-32.59321
7	45065.48	117.1264	7.37E-20	-32.70242	-32.45266	-32.61218
8	45126.66	120.8896	7.13E-20	-32.73529	-32.45108	-32.63260
9	45187.37	119.7984	6.91E-20	-32.76782	-32.44916	-32.65269
10	45273.46	169.6111	6.56E-20	-32.81882	-32.46570*	-32.69124
11	45321.49	94.48233	6.41E-20	-32.84212	-32.45455	-32.70209*
12	45352.15	60.22450	6.34E-20	-32.85278	-32.43077	-32.70031
13	45383.69	61.87449	6.27E-20	-32.86409	-32.40763	-32.69917
14	45409.31	50.17436	6.23E-20	-32.87109	-32.38017	-32.69372
15	45434.41	49.07500	6.19E-20	-32.87771	-32.35234	-32.68789
16	45453.84	37.95389	6.17E-20	-32.88021	-32.32039	-32.67794
17	45479.65	50.31227	6.13E-20	-32.88734	-32.29307	-32.67263
18	45505.22	49.78114	6.09E-20	-32.89430	-32.26558	-32.66715
19	45523.54	35.61895	6.08E-20	-32.89599	-32.23282	-32.65639
20	45566.74	83.84744	5.96E-20	-32.91578	-32.21816	-32.66373
21	45583.00	31.52592	5.96E-20	-32.91597	-32.18390	-32.65148
22	45611.09	54.36237	5.90E-20	-32.92477	-32.15825	-32.64782
23	45623.10	23.20563	5.92E-20	-32.92186	-32.12089	-32.63247
24	45664.84	80.53151	5.81E-20	-32.94059	-32.10517	-32.63875
25	45677.48	24.34076	5.83E-20	-32.93814	-32.06827	-32.62386
26	45695.03	33.77015	5.82E-20	-32.93927	-32.03495	-32.61254
27	45714.31	37.02357	5.80E-20	-32.94166	-32.00289	-32.60248
28	45739.36	48.04690	5.77E-20	-32.94824	-31.97502	-32.59662
29	45756.32	32.48238*	5.76E-20*	-32.94894*	-31.94127	-32.58487
30	45767.84	22.01408	5.78E-20	-32.94568	-31.90356	-32.56916

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Table 4-1 VAR Lag Order Selection Criteria

Vector Autoregression Estimates				
Sample(adjusted): 13 2780				
Included observations: 2768 after adjusting endpoints				
Standard errors in ( ) & t-statistics in [ ]				
	D(OPEN)	D(HIGH)	D(LOW)	D(CLOSE)
C	5.36E-05	0.000360	0.000276	0.000502
	(5.0E-05)	(0.00012)	(0.00014)	(0.00018)
	[1.06416]	[3.10321]	[2.02756]	[2.78581]
R-squared	0.923343	0.414037	0.408144	0.035187
Adj. R-squared	0.922104	0.404569	0.398580	0.019597
Sum sq. resids	0.018278	0.097285	0.133390	0.234118
S.E. equation	0.002591	0.005977	0.006999	0.009272
F-statistic	745.4264	43.72851	42.67681	2.256990
Log likelihood	12580.66	10266.67	9829.832	9051.269
Akaike AIC	-9.057559	-7.385601	-7.069965	-6.507420
Schwarz SC	-8.961221	-7.289263	-6.973627	-6.411081
Mean dependent	0.000504	0.000507	0.000504	0.000506
S.D. dependent	0.009283	0.007746	0.009025	0.009365
Determinant Residual Covariance		6.06E-20		
Log Likelihood (d.f. adjusted)		45531.59		
Akaike Information Criteria		-32.76849		
Schwarz Criteria		-32.38314		

Table 4-2 LS 1 11 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C[1]

The VAR(11,4) model estimates are shown in Table 4-2. The open, high and low log-returns are moderately to strongly captured by the model, indicating that lag variables strongly influence these variables. However, the model apart from reducing the serial correlations in the close log-return residuals, does not account much for the variances in close log-returns. This is probably due to the non-synchronous nature of the sampling process, whereby the closing prices are the most current prices within a day, thus rendering the closing log-returns unaccountable in the traditional VEC model.

[1] Whenever possible E-Views notation is retained for the table headings. This is to enable tractability of the models.

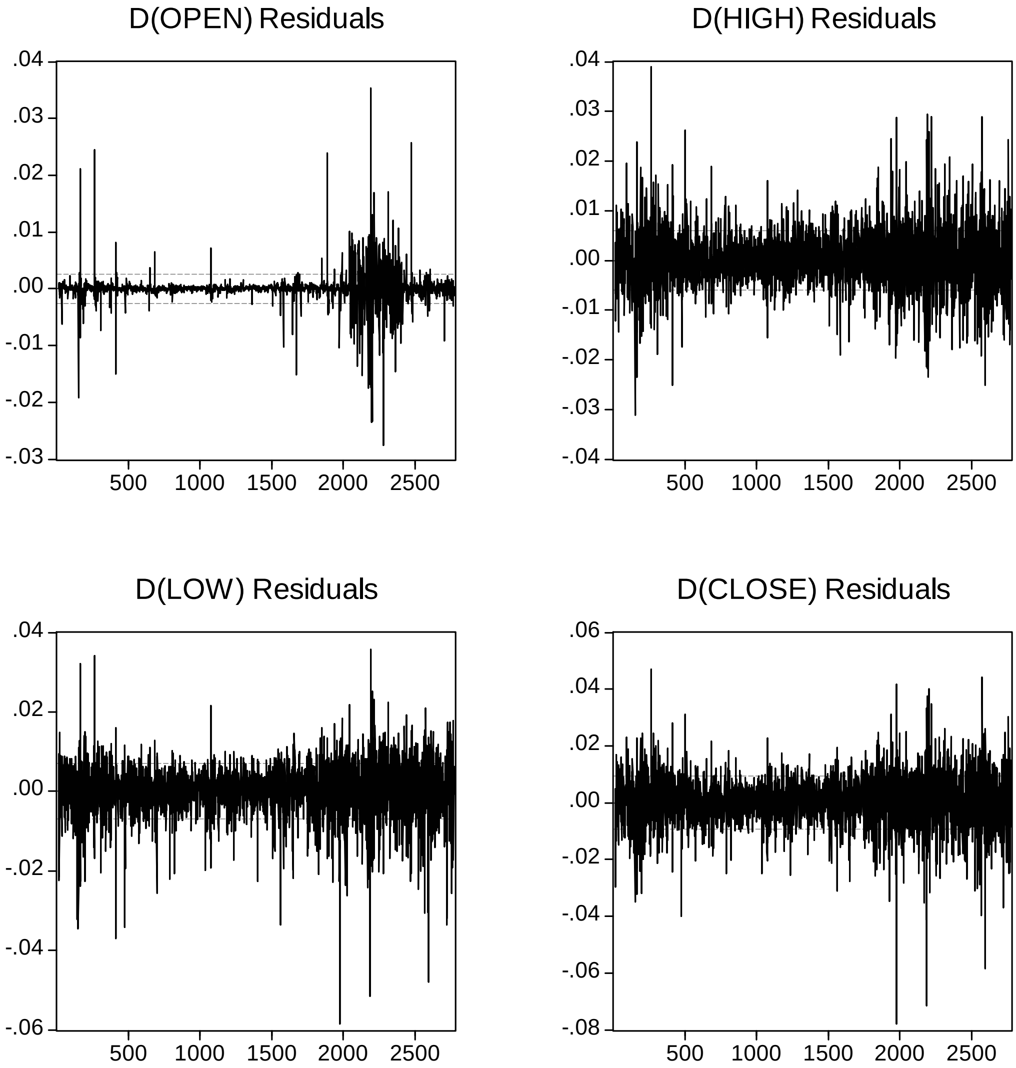


Figure 4-1 VAR RESIDUALS (LS 1 11 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C)

The VAR(11,4) residuals for all log-return variables are shown in Figure 4-1. The D(OPEN) residuals appear “abnormally” distributed whereas the high, low and close residuals appear normal or continuously distributed. At this point, one can only state that the D(OPEN) returns appear to be fully captured by the model with non-typical residuals.

Further, the high and low residuals are asymmetrically distributed and the close residuals are symmetrically distributed.

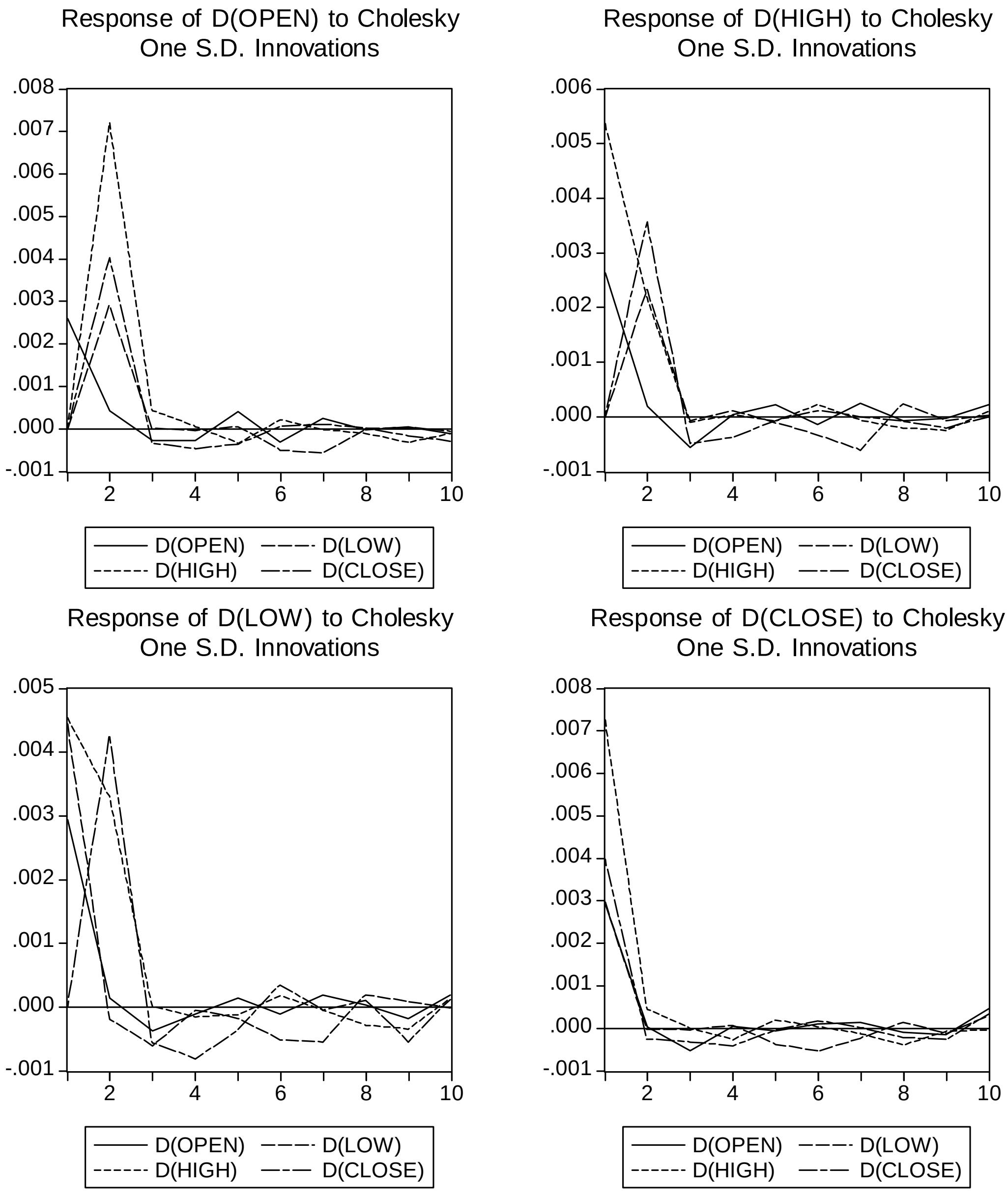


Figure 4-2 IRF (LS 1 11 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C)

The impulse response functions are depicted in Figure 4-2. The open log-returns are influenced by all the other (high, low and close) lag 1 variables, thus explaining the strong fit of the VAR(11,4) model. The high log-returns are influenced by the current shock and open log-returns and the lag 1 close and low variables. The low log-returns are influenced by the current shock and open and high log-returns and the lag 1 close log-returns. The close log-returns are influenced primarily by the current shocks and variables.

5 The VECM model

A natural progression from a VAR representation is the VECM model, especially when the level series are non-stationary. We initially test for the rank of the cointegration using the methodology by Johansen (1988).

Sample(adjusted): 12 2780
Included observations: 2769 after adjusting endpoints
Trend assumption: Linear deterministic trend
Series: CLOSE HIGH LOW OPEN
Lags interval (in first differences): 1 to 10

Unrestricted Cointegration Rank Test

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.081218	500.2015	47.21	54.46
At most 1 **	0.075362	265.6481	29.68	35.65
At most 2 **	0.017423	48.68849	15.41	20.04
At most 3	6.92E-06	0.019169	3.76	6.65

*(**) denotes rejection of the hypothesis at the 5%(1%) level
Trace test indicates 3 cointegrating equation(s) at both 5% and 1% levels

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.081218	234.5534	27.07	32.24
At most 1 **	0.075362	216.9596	20.97	25.52
At most 2 **	0.017423	48.66932	14.07	18.63
At most 3	6.92E-06	0.019169	3.76	6.65

*(**) denotes rejection of the hypothesis at the 5%(1%) level
Max-eigenvalue test indicates 3 cointegrating equation(s) at both 5% and 1% levels

Table 5-1 Cointegration Tests (EC(C,3) 1 10 OPEN HIGH LOW CLOSE)

From Table 5-1 the max-eigenvalue test indicates 3 cointegrating equation(s) at both 5% and 1% levels. Table 5-2 details the 3 cointegrating equations and their adjustment coefficients.

3 Cointegrating Equation(s):		Log likelihood	45840.60
Normalized cointegrating coefficients (std.err. in parentheses)			
CLOSE	HIGH	LOW	OPEN
1.000000	0.000000	0.000000	-0.999681
			(8.7E-05)
0.000000	1.000000	0.000000	-1.005055
			(0.00070)
0.000000	0.000000	1.000000	-0.994943
			(0.00071)
Adjustment coefficients (std.err. in parentheses)			
D(CLOSE)	-1.046382	0.248668	0.139272
	(0.30621)	(0.13469)	(0.12627)
D(HIGH)	0.411063	-0.285535	-0.214568
	(0.19758)	(0.08690)	(0.08147)
D(LOW)	0.414077	-0.394900	-0.559414
	(0.23016)	(0.10124)	(0.09491)
D(OPEN)	0.947824	0.012107	-0.008975
	(0.08340)	(0.03668)	(0.03439)

Table 5-2 Cointegrating Equation(s

Vector Error Correction Estimates				
Sample(adjusted): 12 2780				
Included observations: 2769 after adjusting endpoints				
Standard errors in ( ) & t-statistics in [ ]				
Cointegrating Eq:	CointEq1	CointEq2	CointEq3	
CLOSE(-1)	1.000000	0.000000	0.000000	
HIGH(-1)	0.000000	1.000000	0.000000	
LOW(-1)	0.000000	0.000000	1.000000	
OPEN(-1)	-0.999681	-1.005055	-0.994943	
	(8.7E-05)	(0.00070)	(0.00071)	
	[-11442.0]	[-1429.36]	[-1403.40]	
C1	-0.003125	0.032196	-0.032239	
Error Correction:	D(CLOSE)	D(HIGH)	D(LOW)	D(OPEN)
CointEq1	-1.046382	0.411063	0.414077	0.947824
	(0.30621)	(0.19758)	(0.23016)	(0.08340)
	[-3.41721]	[2.08053]	[1.79910]	[11.3654]
CointEq2	0.248668	-0.285535	-0.394900	0.012107
	(0.13469)	(0.08690)	(0.10124)	(0.03668)
	[1.84627]	[-3.28563]	[-3.90080]	[0.33005]
CointEq3	0.139272	-0.214568	-0.559414	-0.008975
	(0.12627)	(0.08147)	(0.09491)	(0.03439)
	[1.10300]	[-2.63366]	[-5.89435]	[-0.26100]
C2	0.000117	0.000427	0.000202	0.000503
	(0.00022)	(0.00014)	(0.00016)	(5.9E-05)
	[0.53692]	[3.03402]	[1.23311]	[8.46998]
R-squared	0.036410	0.413329	0.413488	0.927231
Adj. R-squared	0.021205	0.404071	0.404233	0.926082
Sum sq. resids	0.233984	0.097413	0.132192	0.017356
S.E. equation	0.009266	0.005979	0.006965	0.002524
F-statistic	2.394556	44.64764	44.67693	807.4902
Log likelihood	9055.829	10269.05	9846.380	12657.38
Akaike AIC	-6.509085	-7.385376	-7.080087	-9.110424
Schwarz SC	-6.414916	-7.291207	-6.985918	-9.016255
Mean dependent	0.000501	0.000509	0.000505	0.000507
S.D. dependent	0.009366	0.007745	0.009024	0.009282
Determinant Residual Covariance		5.23E-20		
Log Likelihood		45840.60		
Log Likelihood (d.f. adjusted)		45751.89		
Akaike Information Criteria		-32.91000		
Schwarz Criteria		-32.50764		

Table 5-3 EC(C,3) 1 10 OPEN HIGH LOW CLOSE

From Table 5-2 and Table 5-3, we make a number of observations. The normalised cointegrating coefficients only load on the OPEN series with negative coefficients. Thus we have:

(2)
$$\begin{aligned} &1.000000 \bullet C_t \quad -0.999681 \bullet O_t \\ &1.000000 \bullet H_t \quad -1.005055 \bullet O_t \\ &1.000000 \bullet L_t \quad -0.994943 \bullet O_t \end{aligned}$$

This resembles the futures-spot parity equations, with H_t, L_t, C_t as the futures prices and the O_t as the spot price. Thus, one can say that the error correction process is a no-arbitrage process. The cointegrating coefficients measure the long-run cost of carry for the H_t, L_t, C_t prices. The C1 values reflect the log-run price of immediacy embedded in the cointegrating vectors. C2 reflect the long-run risk premiums for the various series.

The VECM model is based on 10 lags. Table 5-3 does not display the coefficients for the lag logarithmic returns. There are 3 cointegrating vectors and hence 1 stochastic trend. However, the R-squared for the close logarithmic returns is still low (0.036410), indicating a possible under-specification in as far explaining the closing returns are concerned.

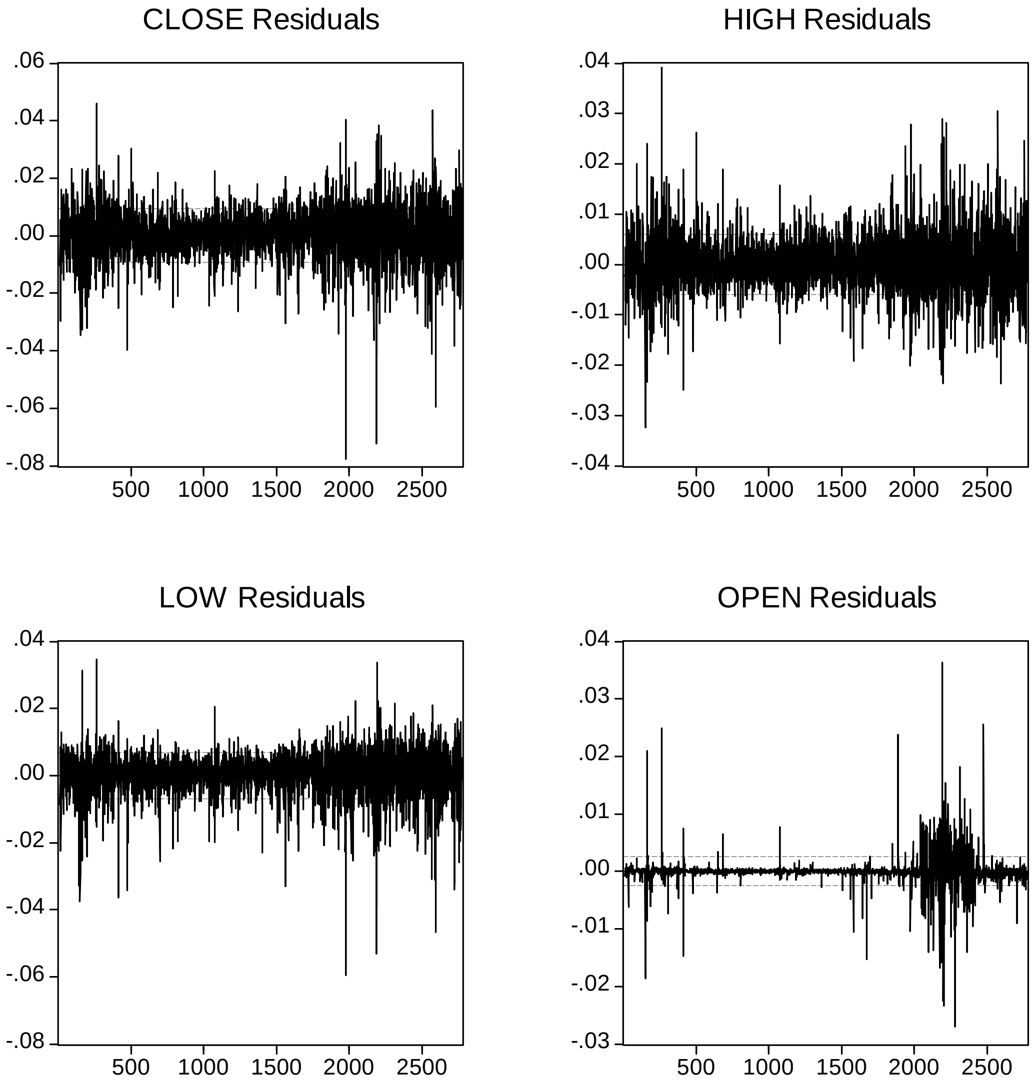


Figure 5-1 EC(C,3) 1 10 OPEN HIGH LOW CLOSE

The residuals from the VECM(C,3) model also exhibit marked differences from each other. The D(OPEN) residuals appear to be unrelated to the other variables. This is a direct result of the non-synchronous sampling and the opening prices do not capture the current disturbances over the current trading day.

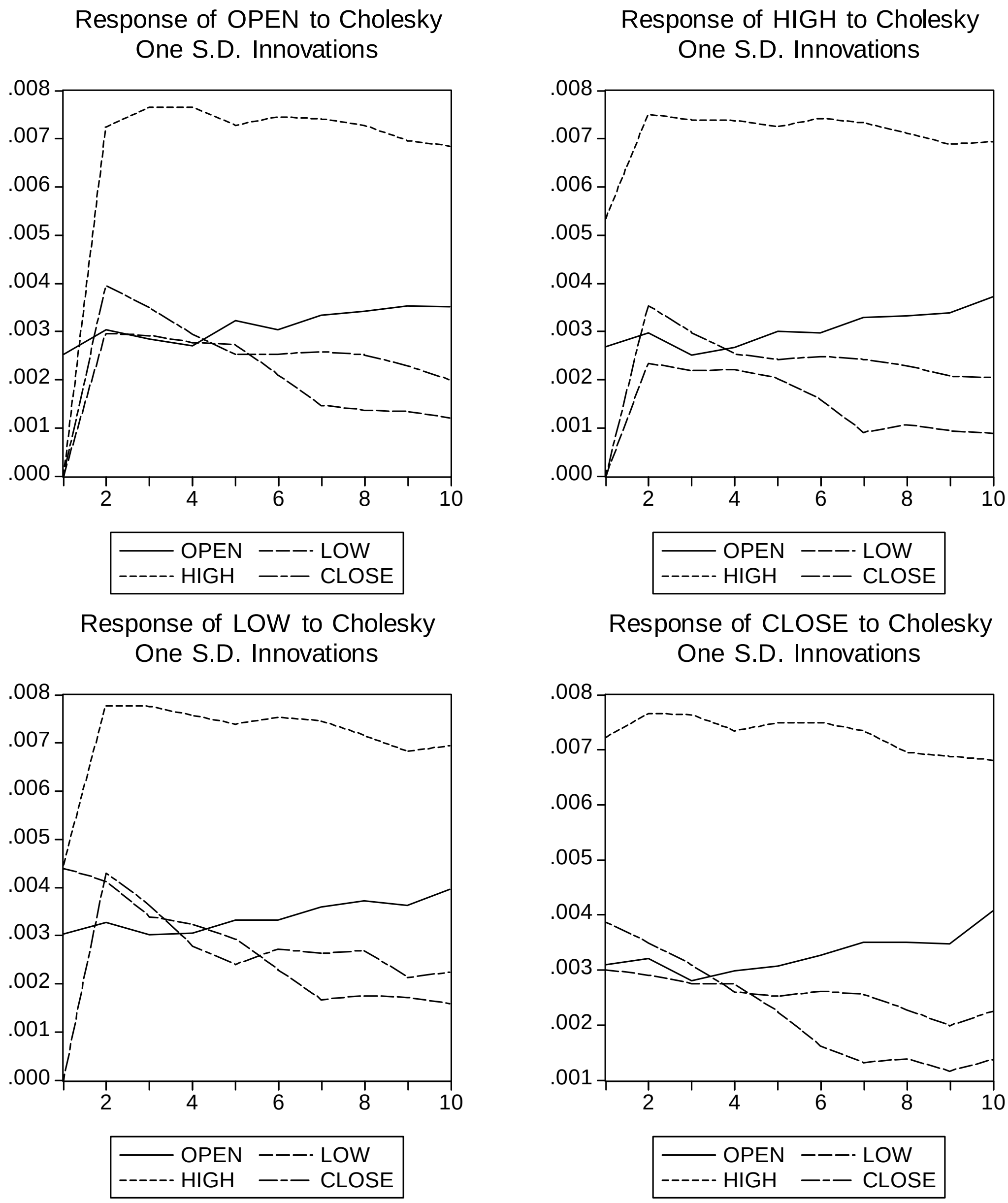


Figure 5-2 EC(C,3) 1 10 OPEN HIGH LOW CLOSE

The responses reflect the non-stationarity of the level series.

6 The VECM-lead(CointEq1) model

As the VECM Model is under-specified, we fit an augmented VECM model called the VECM-lead(CointEq1) model, where the first cointegrating vector is treated as a leading exogenous variable.

Vector Error Correction Estimates
Sample(adjusted): 12 2779
Included observations: 2768 after adjusting endpoints
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1	CointEq2	CointEq3	
CLOSE(-1)	1.000000	0.000000	0.000000	
HIGH(-1)	0.000000	1.000000	0.000000	
LOW(-1)	0.000000	0.000000	1.000000	
OPEN(-1)	-0.999681	-1.005033	-0.994954	
	(1.5E-09)	(0.00071)	(0.00072)	
	[-6.6E+08]	[-1418.77]	[-1383.94]	
C1	-0.003123	0.032016	-0.032146	
Error Correction:	**D(CLOSE)**	**D(HIGH)**	**D(LOW)**	**D(OPEN)**
CointEq1	-0.033182	0.931238	1.025439	0.967127
	(0.08341)	(0.12685)	(0.14555)	(0.08344)
	[-0.39780]	[7.34137]	[7.04549]	[11.5908]
CointEq2	0.007596	-0.409922	-0.541365	0.007599
	(0.03663)	(0.05570)	(0.06391)	(0.03664)
	[0.20739]	[-7.35903]	[-8.47022]	[0.20739]
CointEq3	-0.011864	-0.292089	-0.650514	-0.011868
	(0.03432)	(0.05220)	(0.05989)	(0.03433)
	[-0.34565]	[-5.59596]	[-10.8618]	[-0.34565]
C2	0.000510	0.000626	0.000437	0.000508
	(5.9E-05)	(9.0E-05)	(0.00010)	(5.9E-05)
	[8.62087]	[6.95725]	[4.22640]	[8.58043]
COINTEQ01(1)	1.019146	0.524929	0.617817	0.019152
	(0.00551)	(0.00838)	(0.00962)	(0.00552)
	[184.835]	[62.6034]	[64.2158]	[3.47228]
R-squared	0.928869	0.759498	0.766734	0.927540
Adj. R-squared	0.927720	0.755611	0.762965	0.926370
Sum sq. resids	0.017268	0.039934	0.052574	0.017279
S.E. equation	0.002518	0.003830	0.004394	0.002519
F-statistic	808.1513	195.4347	203.4174	792.1950
Log likelihood	12659.33	11499.02	11118.43	12658.44
Akaike AIC	-9.114399	-8.276028	-8.001030	-9.113761
Schwarz SC	-9.018060	-8.179690	-7.904692	-9.017422
Mean dependent	0.000504	0.000508	0.000505	0.000505
S.D. dependent	0.009367	0.007746	0.009025	0.009283
Determinant Residual Covariance		0.000000		

Table 6-1 EC(C,3) 1 10 OPEN HIGH LOW CLOSE @ COINTEQ01(1)

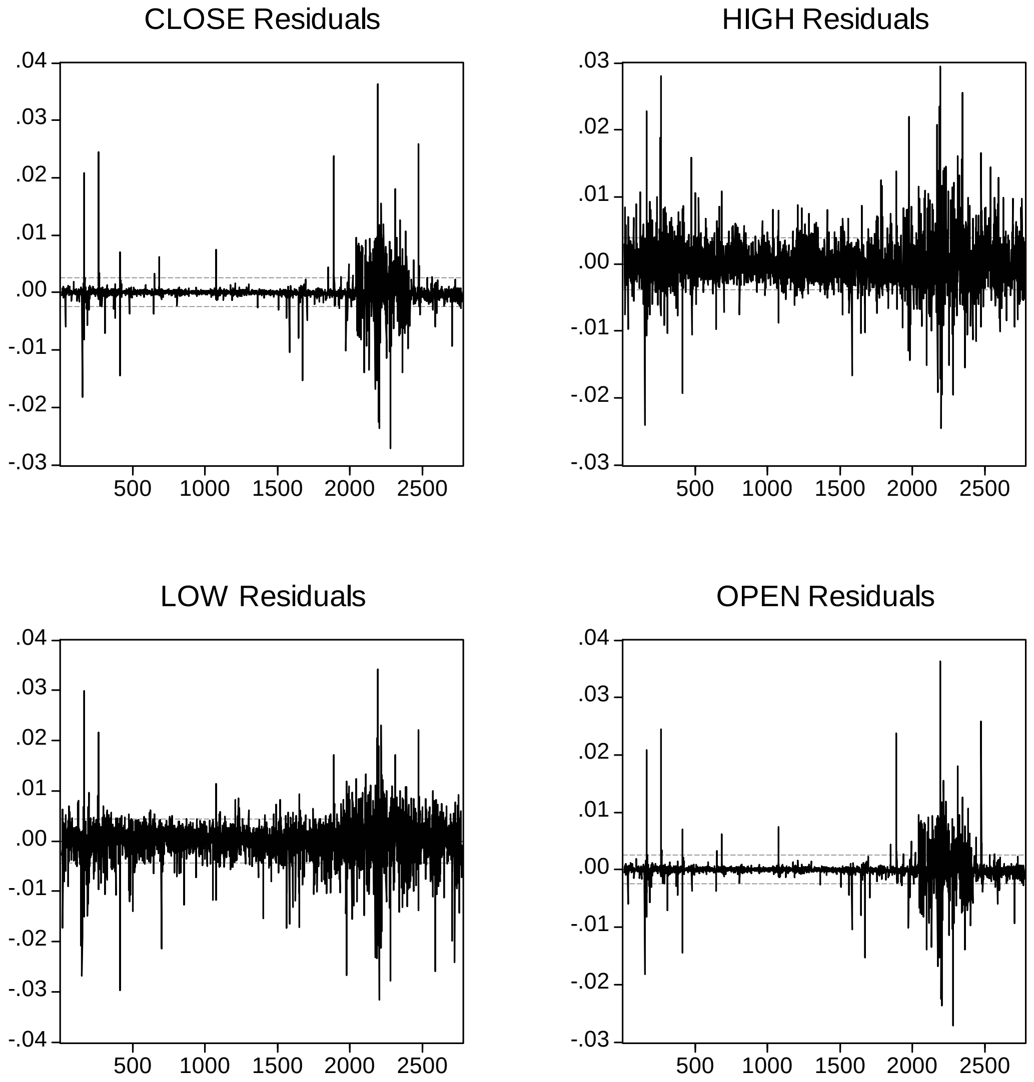


Figure 6-1 EC(C,3) 1 10 OPEN HIGH LOW CLOSE @ COINTEQ01(1)

We can see from Table 6-1 and Figure 6-1 that the D(CLOSE) fully captured by the lead(CointEq1) augmentation. The lead and current cointegrating vector 1 is common to both the D(OPEN) and D(CLOSE) process. CointEq Vectors 2 and 3 are common to the D(HIGH) and D(LOW). The error correction process can be said to mirror the data generation process for D(OPEN) and D(CLOSE) series. However, the D(HIGH) and D(LOW) series can be further fitted. We do this by considering a VECM-lead(CointEq1,2,3) model in Section 7.

7 The VECM-lead(CointEq1,2,3) model

When we attempt to fit a VECM model with the 3 cointegrating vectors as exogenous variable, we face a "near singular matrix" problem. However, in as much as the VAR can be modelled as a VECM model, we can also model the VECM as a VAR with exogenous variable, where the current and the immediate-past cointegrating vectors are the exogenous terms.

Vector Autoregression Estimates
Sample(adjusted): 12 2779
Included observations: 2768 after adjusting endpoints
Standard errors in ( ) & t-statistics in [ ]

	D(OPEN)	D(HIGH)	D(LOW)	D(CLOSE)
C2	0.000518	0.000521	0.000516	0.000518
	(5.9E-05)	(6.0E-05)	(5.9E-05)	(5.9E-05)
	[8.74851]	[8.74851]	[8.74851]	[8.74851]
COINTEQ01(1)	0.017605	0.017694	0.017516	1.017599
	(0.01246)	(0.01252)	(0.01240)	(0.01246)
	[1.41275]	[1.41275]	[1.41275]	[81.6863]
COINTEQ02(1)	-0.043179	0.956602	-0.042961	-0.043165
	(0.01613)	(0.01621)	(0.01605)	(0.01612)
	[-2.67711]	[59.0109]	[-2.67711]	[-2.67711]
COINTEQ03(1)	0.039050	0.039247	1.038853	0.039038
	(0.01386)	(0.01393)	(0.01379)	(0.01385)
	[2.81762]	[2.81762]	[75.3384]	[2.81762]
COINTEQ01	0.962884	0.967751	0.958015	-0.037423
	(0.08326)	(0.08369)	(0.08284)	(0.08324)
	[11.5642]	[11.5642]	[11.5642]	[-0.44959]
COINTEQ02	0.054188	-0.945539	0.053914	0.054170
	(0.03865)	(0.03884)	(0.03845)	(0.03863)
	[1.40214]	[-24.3434]	[1.40214]	[1.40214]
COINTEQ03	-0.038065	-0.038257	-1.037872	-0.038053
	(0.03497)	(0.03514)	(0.03479)	(0.03496)
	[-1.08856]	[-1.08856]	[-29.8315]	[-1.08856]
R-squared	0.927912	0.895421	0.924497	0.929234
Adj. R-squared	0.926693	0.893654	0.923221	0.928038
Sum sq. resids	0.017190	0.017364	0.017017	0.017179
S.E. equation	0.002513	0.002526	0.002501	0.002513
F-statistic	761.4048	506.4724	724.2934	776.7348
Log likelihood	12665.56	12651.61	12679.60	12666.45
Akaike AIC	-9.117459	-9.107375	-9.127598	-9.118097
Schwarz SC	-9.016839	-9.006755	-9.026978	-9.017477
Mean dependent	0.000505	0.000508	0.000505	0.000504
S.D. dependent	0.009283	0.007746	0.009025	0.009367
Determinant Residual Covariance		0.000000		

Table 7-1 LS 1 10 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C COINTEQ01(1) COINTEQ02(1) COINTEQ03(1) COINTEQ01 COINTEQ02 COINTEQ03

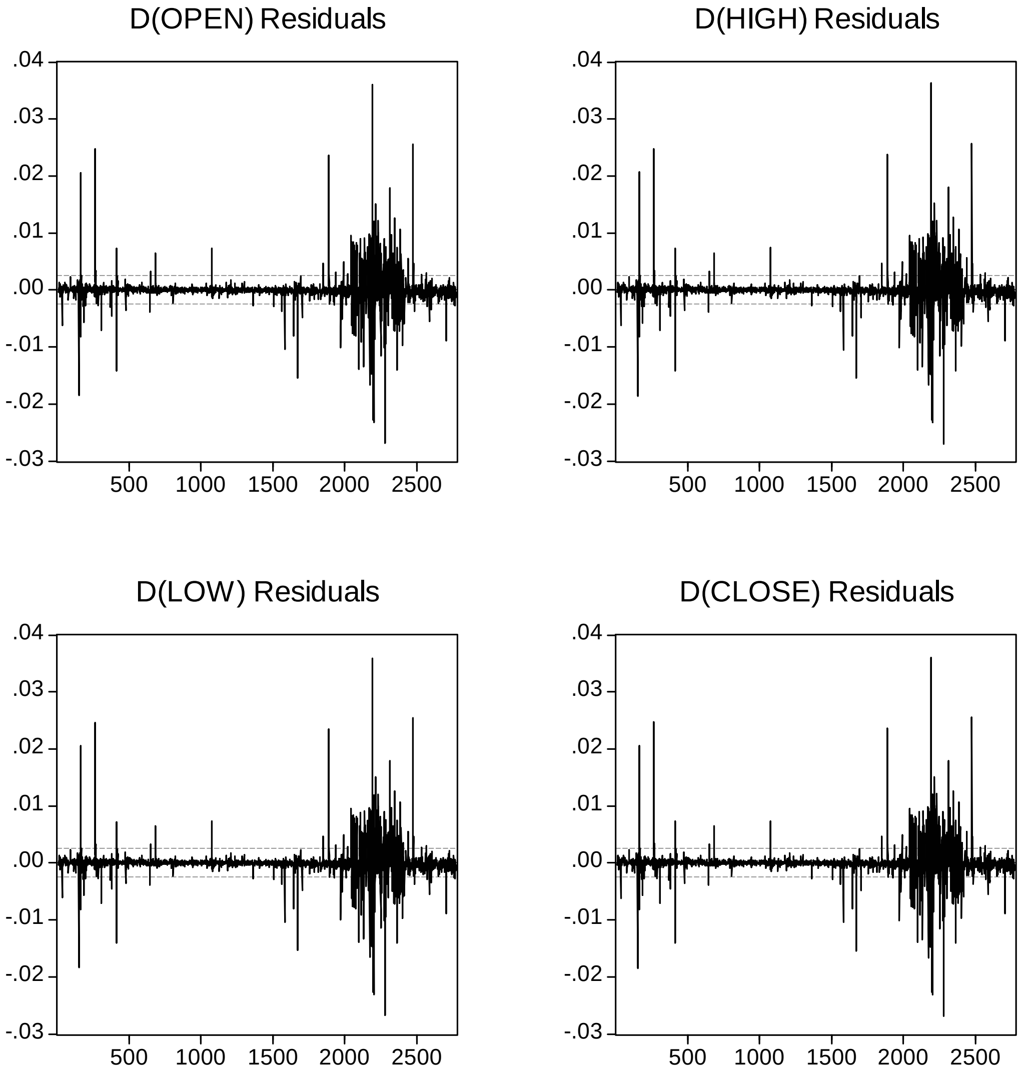


Figure 7-1 LS 1 10 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C COINTEQ01(1) COINTEQ02(1) COINTEQ03(1) COINTEQ01 COINTEQ02 COINTEQ03

Note that the residuals are identical with perfect positive correlations as listed in Table 7-2 and shown in Figure 7-1.

	D(OPEN)	D(HIGH)	D(LOW)	D(CLOSE)
D(OPEN)	1.000000	1.000000	1.000000	1.000000
D(HIGH)	1.000000	1.000000	1.000000	1.000000
D(LOW)	1.000000	1.000000	1.000000	1.000000
D(CLOSE)	1.000000	1.000000	1.000000	1.000000

Table 7-2 LS 1 10 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C COINTEQ01(1) COINTEQ02(1) COINTEQ03(1) COINTEQ01 COINTEQ02 COINTEQ03

Hence we are able to specify a model with a number of common disturbances, the normal disturbances (3 cointegrating residuals) and abnormal disturbances.

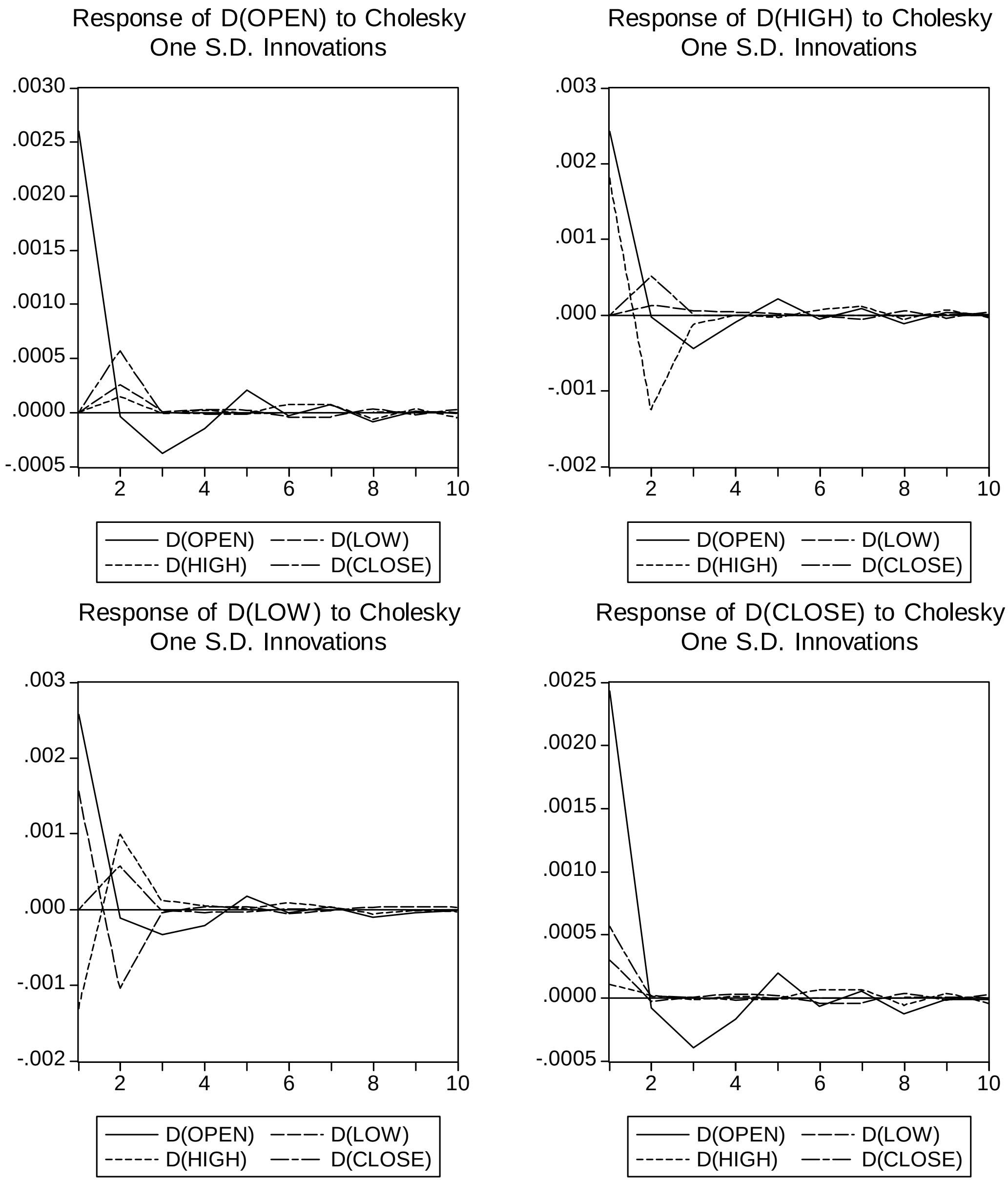


Figure 7-2 LS 1 10 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C COINTEQ01(1) COINTEQ02(1) COINTEQ03(1)

From Table 7-1 we can see that all the logarithmic returns considered a nearly fully explained (R-squares of 0.927912, 0.895421,0.924497 and 0.929234 for the open, high, low and close logarithmic returns) by the assumed model. Further the C2 values are nearly identical, highlighting a common risk premium and implying a good fit of the VECM-lead(CointEq1,2,3) model.

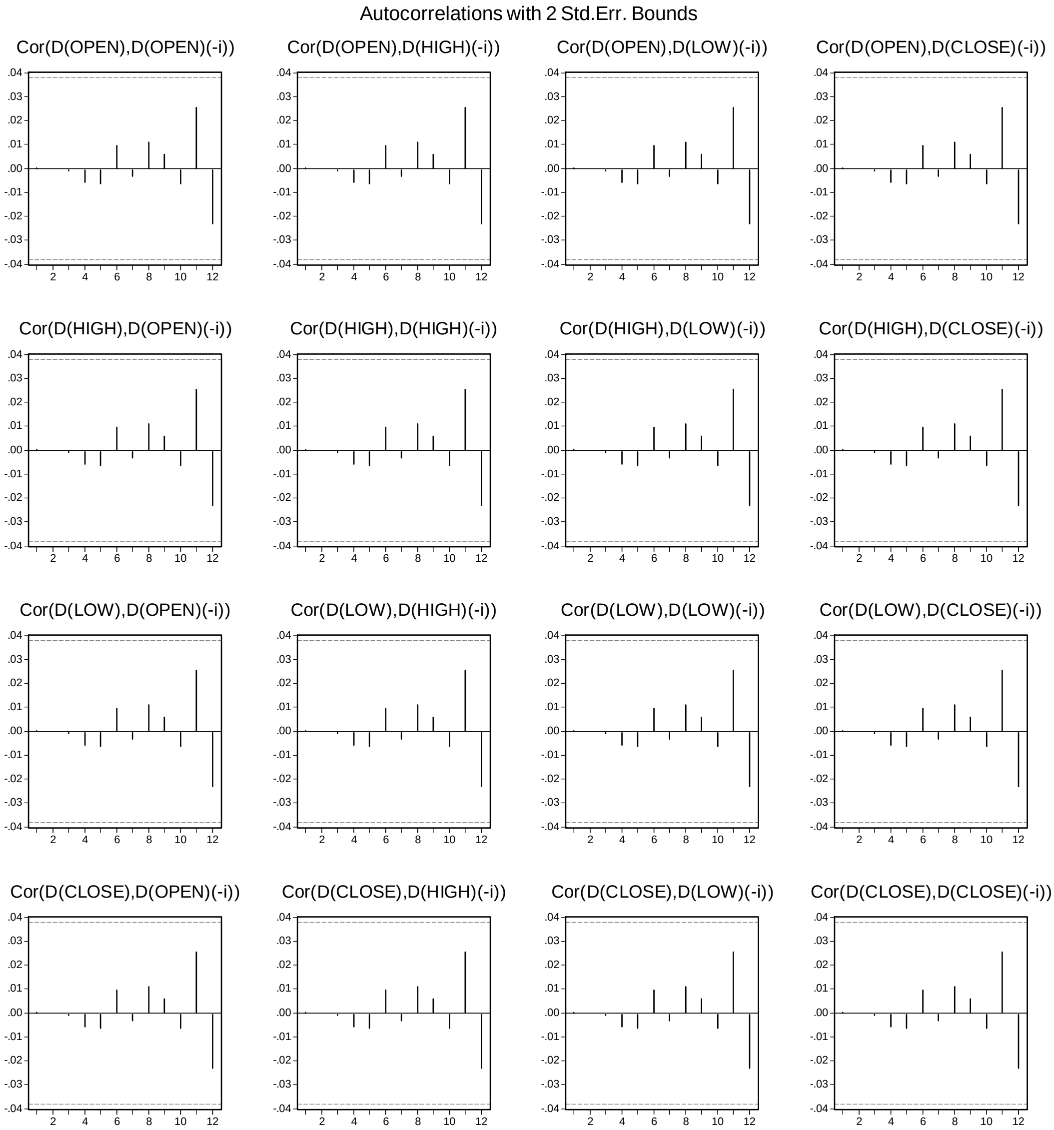


Figure 7-3 LS 1 10 D(OPEN) D(HIGH) D(LOW) D(CLOSE) @ C COINTEQ01(1) COINTEQ02(1) COINTEQ03(1) COINTEQ01 COINTEQ02 COINTEQ03

Figure 7-3 indicates that the residuals (which are perfectly positively correlated) are also not serially correlated. The D(CLOSE) significantly loads on COINTEQ01(1), the D(OPEN) significantly loads on COINTEQ01(0), the D(HIGH) and D(LOW) significantly loads on COINTEQ02(1) and COINTEQ02(0). In words, today's change in close price is dependent on today's COINTEQ01(1) residual which is estimated to be $\Delta C_t \approx C_t - 0.999681 O_t$. Today's change in open price is dependent on yesterday's COINTEQ01(0) residual which is estimated to be $\Delta O_t \approx C_{t-1} - 0.999681 O_{t-1}$.

Today's change in high prices are dependent on today's COINTEQ02(1) and yesterday's COINTEQ02(0). Similarly, today's change in low prices are dependent on today's COINTEQ03(1) and yesterday's COINTEQ03(0). Both the changes in high and low prices also load significantly on yesterday's COINTEQ01(0). Hence the cointegrating residuals play a paramount role on the process captured. In fact one can say that the cointegrating vectors explain all the "normal" variance in the logarithmic returns.

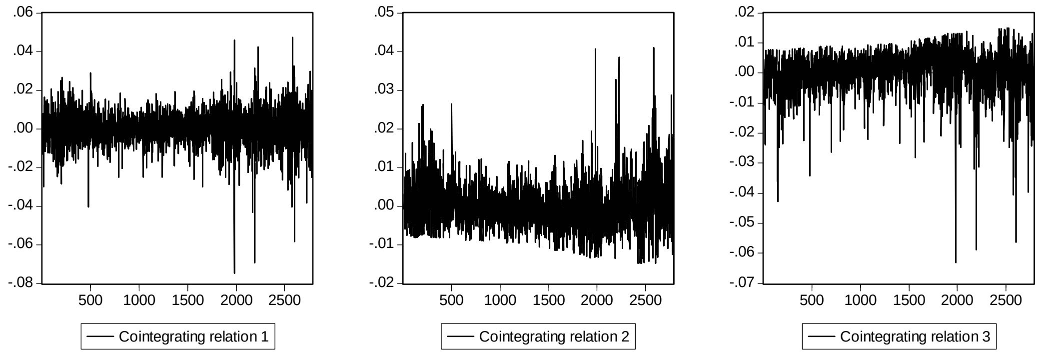


Figure 7-4 EC(C,3) 1 10 CLOSE HIGH LOW OPEN

Hence the error correction process cum the return generation process has three components, each of which reflect a specific uncertainty in prices. Moreover, the cointegrating components are the "normal" disturbances that significantly capture and reflect the logarithmic returns. The residuals of the VAR-lead-lag(CointEq1,2,3) model can be said to capture the "abnormal" disturbances. Since the cointegrating vectors reflect the normal disturbances and since the COINTEQ01 vector is symmetric, we assume this to be a reflection of the "true" normal information set. Similarly, the asymmetric COINTEQ02 and COINTEQ03 vectors are assumed to refect the sell-side and buy-side expectations.

8 The Cointegrating Vectors

The cointegrating vectors themselves exhibit some stylised facts. We now investigate and illustrate these facts in detail. Table 8-1 lists the summary statistics for the 3 cointegrating residuals. COINTEQ01 is moderately and negatively skewed. COINTEQ02 is strongly positively skewed and COINTEQ03 is strongly negatively skewed. All cointegrating residuals exhibit high kurtosis and consequently all reject the Jarque-Bera test for normality.

	COINTEQ01	COINTEQ02	COINTEQ03
Mean	3.45E-18	-6.53E-17	1.72E-17
Median	5.89E-05	-0.000661	0.000958
Maximum	0.047107	0.040885	0.014908
Minimum	-0.074819	-0.014805	-0.063071
Std. Dev.	0.008835	0.005925	0.006763
Skewness	-0.527120	1.149337	-1.951248
Kurtosis	8.445128	7.022164	12.74227
Jarque-Bera	3549.032	2476.144	12707.55
Probability	0.000000	0.000000	0.000000
Sum	9.25E-15	-1.82E-13	4.55E-14
Sum Sq. Dev.	0.216058	0.097189	0.126608
Observations	2769	2769	2769

Table 8-1 LS 0 0 COINTEQ01 COINTEQ02 COINTEQ03 @ C

The cointegrating vectors also exhibit some serial and cross-correlations. COINTEQ01 is not serially correlated, but the other two are strongly serially correlated. This indicates the COINTEQ01 is probably proxies the informational disturbances and COINTEQ02 and COINTEQ03 proxy the expectational disturbances[2]. This lends support to the "expectation" and "information" components of the marginal or normal disturbances. Consequently we make the following identifications:

$$\begin{aligned} \zeta_{t-1} &\approx COINTEQ01 \\ \xi^{+}_{t-1} &\approx COINTEQ02 \\ \xi^{-}_{t-1} &\approx COINTEQ03 \end{aligned} \qquad (3)$$

[2] Expectational disturbances tend to be biased.

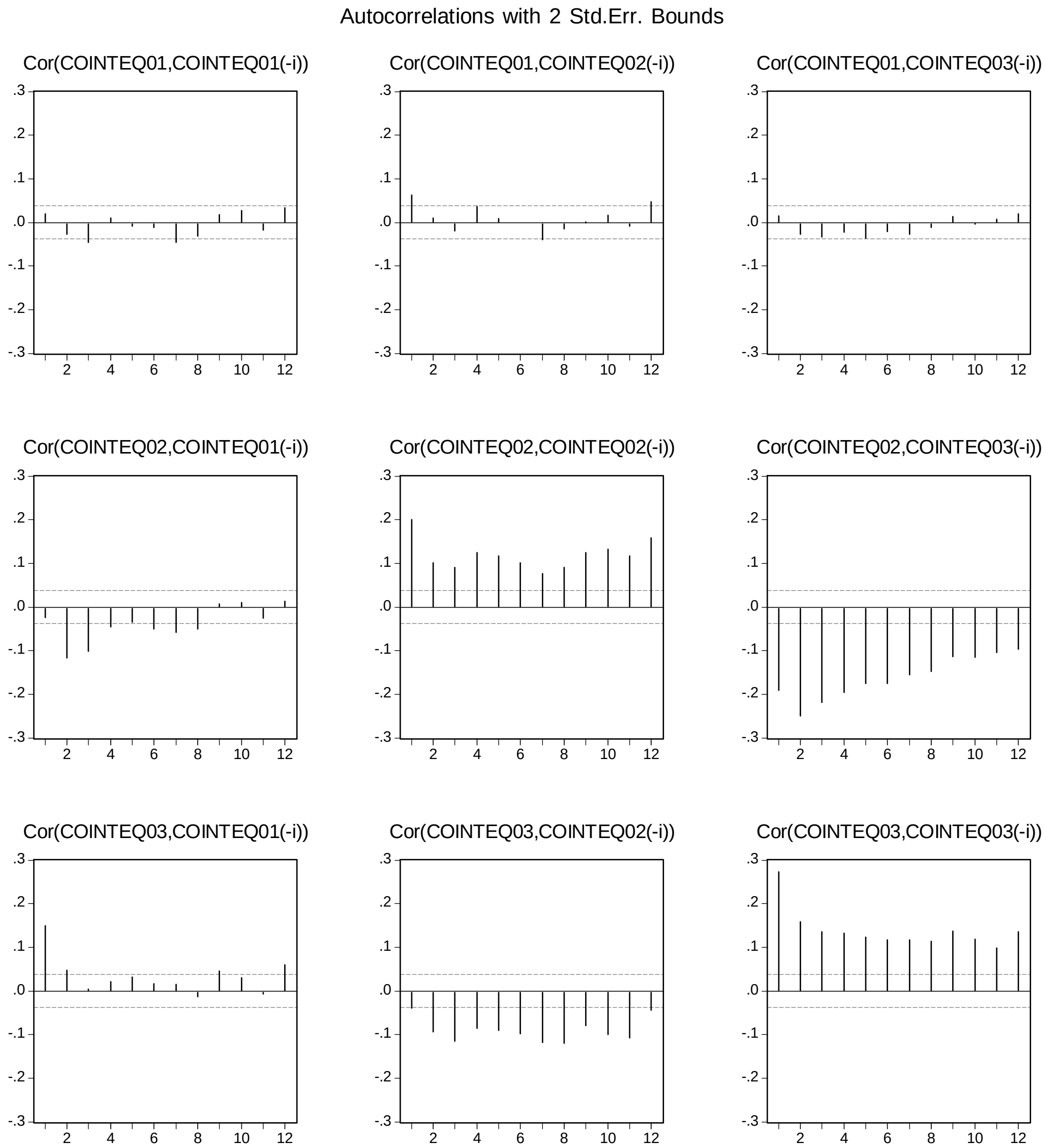


Figure 8-1 LS 0 0 COINTEQ01 COINTEQ02 COINTEQ03 @ C

As the cointegrating residuals are stationary, we fit a VAR model to remove the serial and cross-correlations. We use the HQ selection criteria and obtain an optimal lag of 4 as shown in Table 8-2.

VAR Lag Order Selection Criteria Endogenous variables: COINTEQ01 COINTEQ02 COINTEQ03 Exogenous variables: C Sample: 1 2780 Included observations: 2739						
Lag	LogL	LR	FPE	AIC	SC	HQ
0	31593.59	NA	1.93E-14	-23.06724	-23.06076	-23.06490
1	32480.72	1771.678	1.01E-14	-23.70845	-23.68253	-23.69908
2	32608.66	255.2269	9.30E-15	-23.79530	-23.74995	-23.77891
3	32658.84	99.99253	9.02E-15	-23.82537	-23.76058	-23.80196
4	32696.82	75.58758	8.83E-15	-23.84652	-23.76230*	-23.81609*
5	32711.33	28.85280	8.80E-15	-23.85055	-23.74688	-23.81309
6	32731.25	39.57007	8.73E-15	-23.85853	-23.73542	-23.81404
7	32745.50	28.27553	8.70E-15	-23.86236	-23.71982	-23.81085
8	32758.68	26.12332	8.67E-15	-23.86541	-23.70344	-23.80688
9	32768.63	19.69502	8.66E-15	-23.86611	-23.68469	-23.80055
10	32785.92	34.18190	8.61E-15	-23.87216	-23.67131	-23.79958
11	32800.08	27.97540	8.58E-15	-23.87593	-23.65564	-23.79632
12	32811.05	21.63606	8.57E-15	-23.87736	-23.63764	-23.79074
13	32819.98	17.59147	8.57E-15	-23.87731	-23.61815	-23.78366
14	32829.32	18.39037	8.56E-15*	-23.87756*	-23.59896	-23.77688
15	32836.70	14.51921	8.57E-15	-23.87638	-23.57834	-23.76868
16	32843.27	12.89629	8.59E-15	-23.87460	-23.55713	-23.75988
17	32848.77	10.78776	8.61E-15	-23.87204	-23.53514	-23.75030
18	32858.36	18.80728	8.61E-15	-23.87248	-23.51613	-23.74371
19	32868.81	20.45917	8.60E-15	-23.87354	-23.49776	-23.73774
20	32881.90	25.58885	8.57E-15	-23.87652	-23.48130	-23.73370
21	32886.63	9.249371	8.60E-15	-23.87341	-23.45875	-23.72357
22	32893.06	12.53773	8.62E-15	-23.87153	-23.43744	-23.71466
23	32902.21	17.82962	8.62E-15	-23.87164	-23.41811	-23.70775
24	32907.05	9.422218	8.64E-15	-23.86860	-23.39563	-23.69769
25	32917.15	19.63469	8.63E-15	-23.86940	-23.37700	-23.69146
26	32920.72	6.943790	8.67E-15	-23.86544	-23.35360	-23.68048
27	32927.13	12.42906	8.69E-15	-23.86355	-23.33227	-23.67156
28	32935.35	15.93884	8.69E-15	-23.86298	-23.31226	-23.66397
29	32949.75	27.87253	8.66E-15	-23.86692	-23.29677	-23.66089
30	32959.97	19.76056*	8.65E-15	-23.86781	-23.27822	-23.65476

* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Table 8-2 VAR Lag Order Selection Criteria [COINTEQ01 COINTEQ02 COINTEQ03]

The lag order for the cointegrating residuals is reduced to 4 from 11, indicating that information based adjustments is much shorter than adjustments due to lagged logarithmic returns. Since the explanatory power of the lagged log-returns is much smaller than the cointegrating returns, this is a welcomed result.

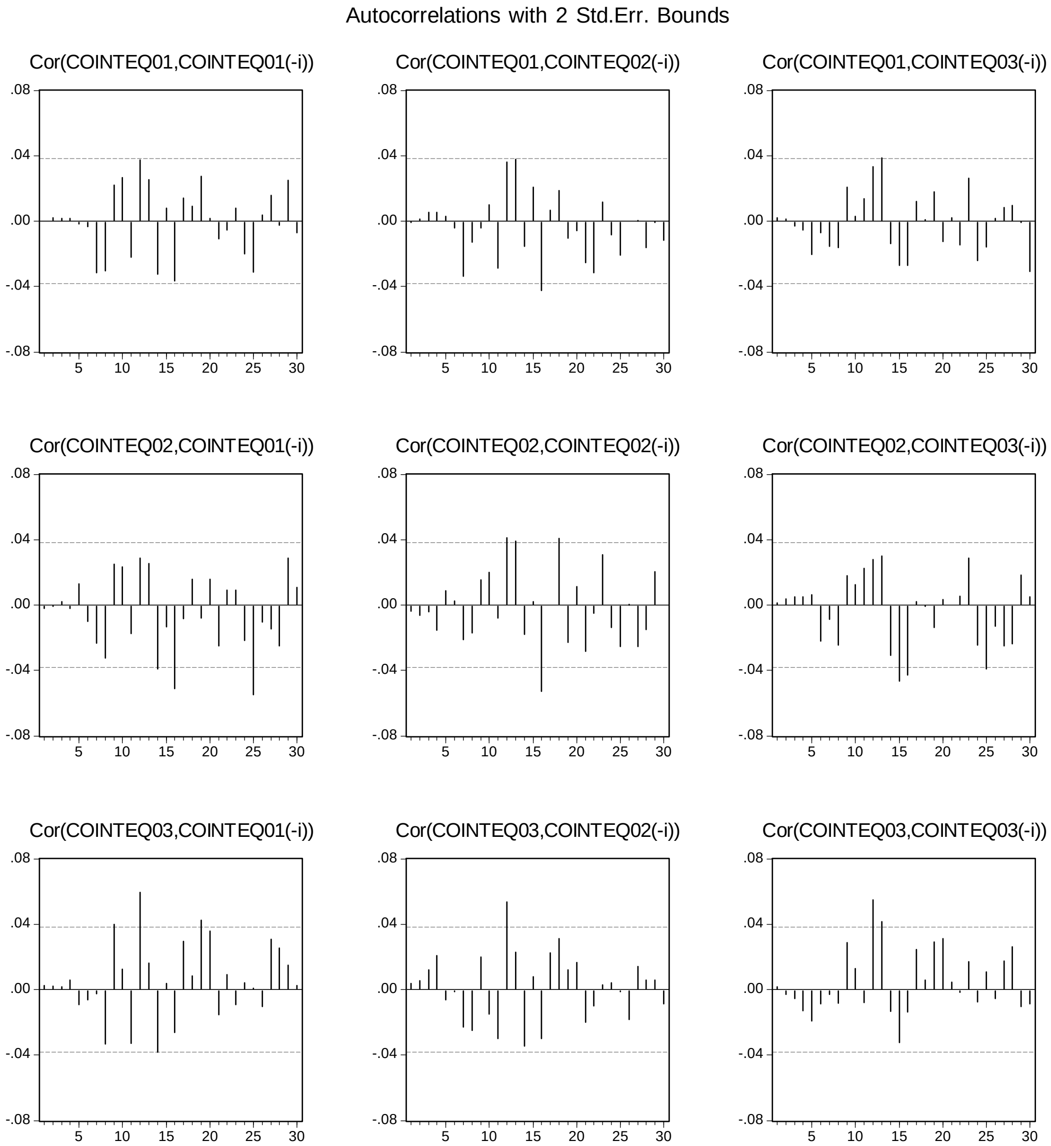


Figure 8-2 LS 1 4 COINTEQ01 COINTEQ02 COINTEQ03 @ C

The residuals of the cointegrating VAR model have no or weak serial correlations.

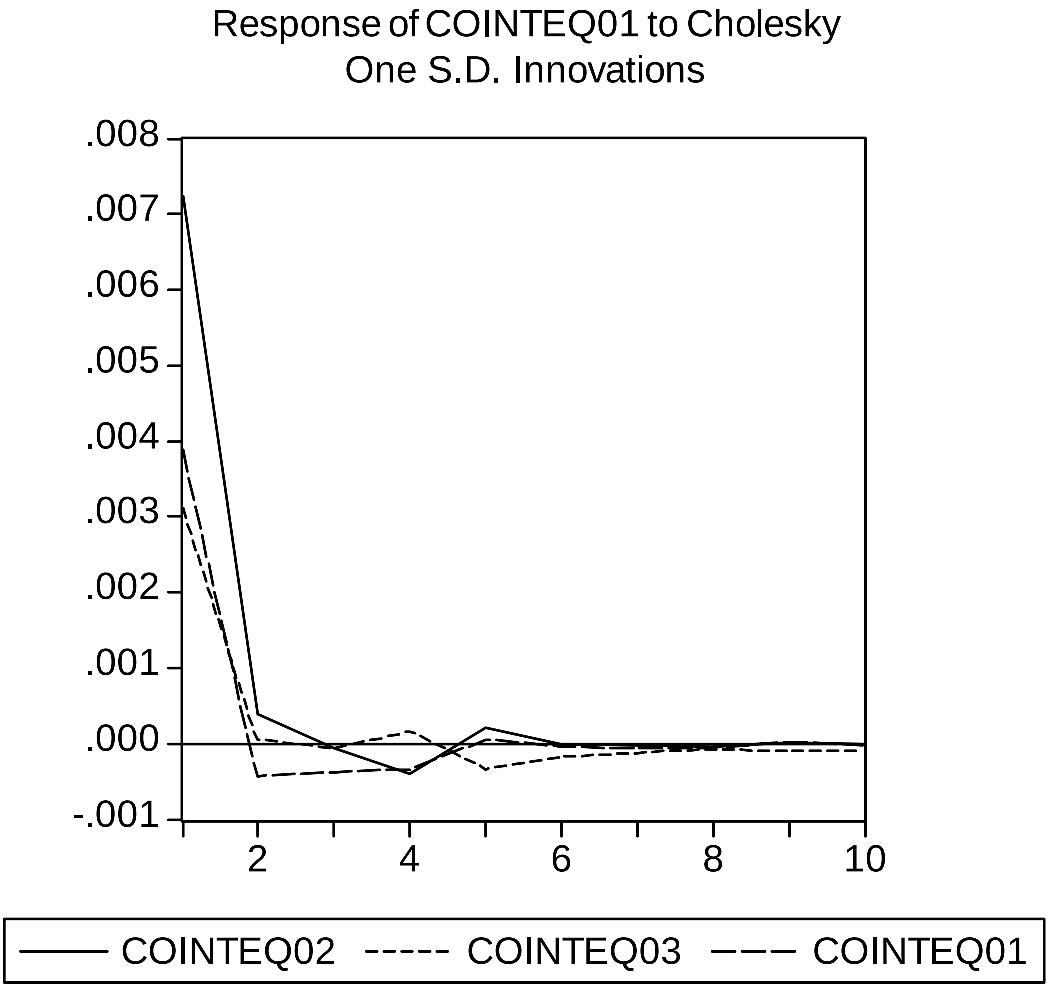


Figure 8-3 LS 1 4 COINTEQ02 COINTEQ03 COINTEQ01 @ C

An impulse response function plot is illustrated in Figure 8-3. One standard deviation innovations to all three cointegrating vectors is felt by COINTEQ01. This lends support to the hypothesis that COINTEQ01 is information and expectations driven.

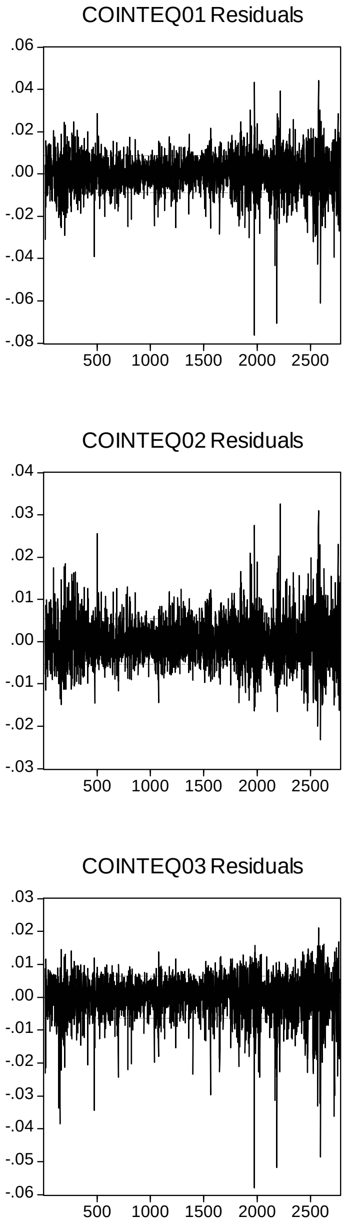


Figure 8-4 LS 1 4 COINTEQ01 COINTEQ02 COINTEQ03 @ C

Figure 8-4 depicts the cointegrating VAR residuals, which have no serial correlations. The overall characteristics of the timeseries plots have not been altered drastically. This is further confirmed by the summary statistics as per Table 8-3.

	CIRESID01	CIRESID02	CIRESID03
Mean	3.44E-19	2.28E-19	1.06E-18
Median	0.000146	-0.000400	0.000627
Maximum	0.044121	0.032590	0.021092
Minimum	-0.076093	-0.023280	-0.057901
Std. Dev.	0.008775	0.005353	0.006318
Skewness	-0.674100	0.555998	-1.621308
Kurtosis	8.846206	5.320504	11.28696
Jarque-Bera	4147.012	762.8261	9123.142
Probability	0.000000	0.000000	0.000000
Sum	4.76E-16	4.26E-16	3.23E-15
Sum Sq. Dev.	0.212825	0.079211	0.110316
Observations	2765	2765	2765

Table 8-3 LS 1 4 COINTEQ01 COINTEQ02 COINTEQ03 @ C

The results for the cointegrating vectors indicate a strong possibility of non-normal unconditional distributions for all three vectors. Though it is not the primary goal of this paper, we fit three descriptive distributions to the three cointegrating vectors and find that they are not rejected at the 0.05% level by the Cramer-von Mises (W2), Watson (U2) and Anderson-Darling (A2) statistics. The three distributions fitted are the logistic, the maximum extreme value and the minimum extreme value distributions. (The normal is rejected at the 0.001% level for all cointegrating vectors.)

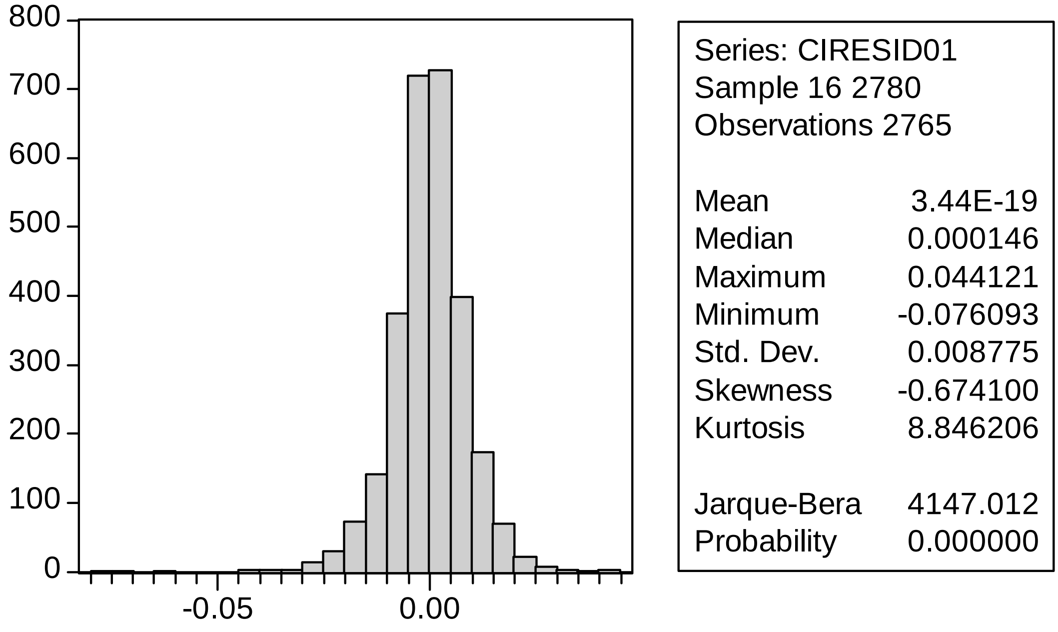


Figure 8-5 Histogram and Stats for CIRESID01

Empirical Distribution Test for CIRESID01
Hypothesis: Logistic
Sample(adjusted): 16 2780
Included observations: 2765 after adjusting endpoints

Method	Value	Adj. Value	Probability
Cramer-von Mises (W2)	0.242252	0.242311	< 0.005
Watson (U2)	0.241804	0.241863	< 0.005
Anderson-Darling (A2)	1.665480	1.665630	< 0.005

Method: Maximum Likelihood (Marquardt)
Estimation settings: tol= 0.10000
Initial Values: C(1)=3.4E-19, C(2)=0.00484
Convergence achieved after 1 iteration
Covariance matrix computed using second derivatives

Parameter	Value	Std. Error	z-Statistic	Prob.
MU	0.000164	0.000156	1.049977	0.2937
S	0.004580	8.19E-05	55.89010	0.0000
Log likelihood	9314.235	Mean dependent var.		3.44E-19
No. of Coefficients	2	S.D. dependent var.		0.008775

Table 8-4 Empirical Distribution Test for CIRESID01

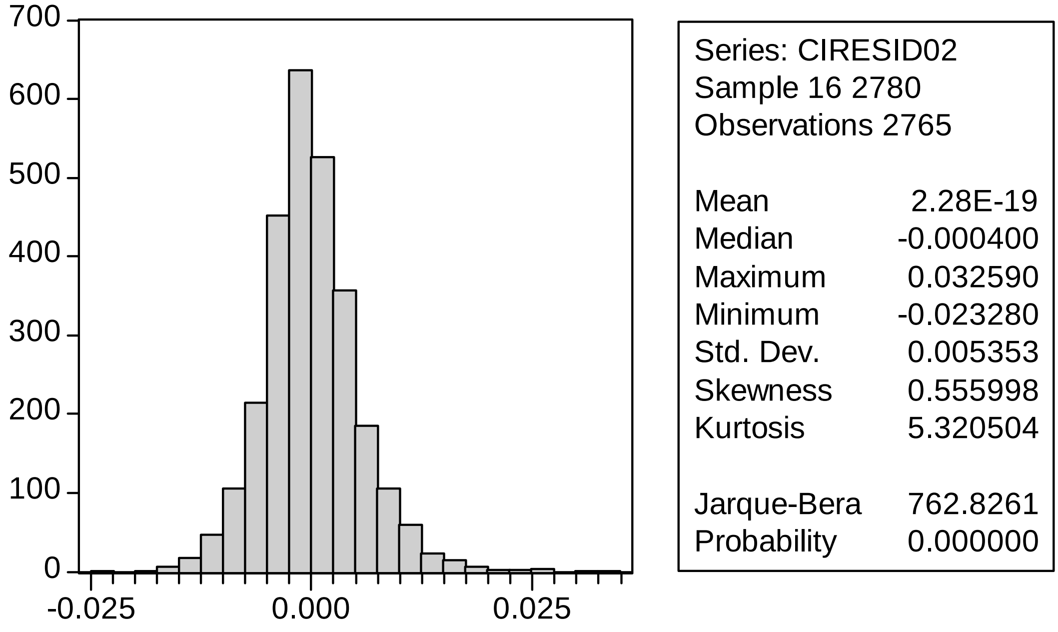


Figure 8-6 Histogram and Stats for CIRESID02

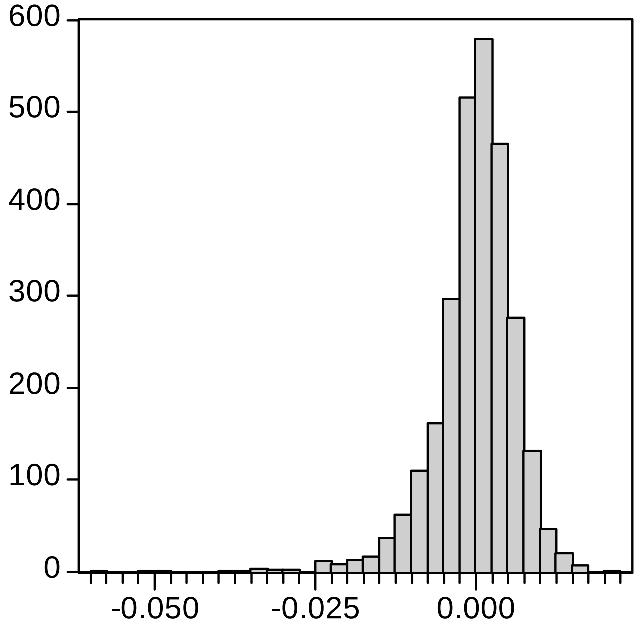
Empirical Distribution Test for CIRESID02
Hypothesis: Extreme Value Max
Sample(adjusted): 16 2780
Included observations: 2765 after adjusting endpoints

Method	Value	Adj. Value	Probability
Cramer-von Mises (W2)	6.481856	6.506510	< 0.01
Watson (U2)	6.407398	6.431768	< 0.01
Anderson-Darling (A2)	41.85633	42.01553	< 0.01

Method: Maximum Likelihood (Marquardt)
Estimation settings: tol= 0.10000
Initial Values: C(1)=-0.00241, C(2)=0.00417
Convergence achieved after 3 iterations
Covariance matrix computed using second derivatives

Parameter	Value	Std. Error	z-Statistic	Prob.
M	-0.002568	0.000103	-24.89224	0.0000
S	0.005283	6.15E-05	85.91097	0.0000
Log likelihood	10382.65	Mean dependent var.		2.28E-19
No. of Coefficients	2	S.D. dependent var.		0.005353

Table 8-5 Empirical Distribution Test for CIRESID02



Series: CIRESID03	
Sample 16 2780	
Observations 2765	
Mean	1.06E-18
Median	0.000627
Maximum	0.021092
Minimum	-0.057901
Std. Dev.	0.006318
Skewness	-1.621308
Kurtosis	11.28696
Jarque-Bera	9123.142
Probability	0.000000

Figure 8-7 Histogram and Stats for CIRESID03

Empirical Distribution Test for CIRESID03
Hypothesis: Extreme Value Min
Sample(adjusted): 16 2780
Included observations: 2765 after adjusting endpoints

Method	Value	Adj. Value	Probability
Cramer-von Mises (W2)	3.535262	3.548708	< 0.01
Watson (U2)	3.490771	3.504048	< 0.01
Anderson-Darling (A2)	21.18180	21.26236	< 0.01

Method: Maximum Likelihood (Marquardt)
Estimation settings: tol= 0.10000
Initial Values: C(1)=0.00284, C(2)=0.00493
Convergence achieved after 1 iteration
Covariance matrix computed using second derivatives

Parameter	Value	Std. Error	z-Statistic	Prob.
M	0.002818	9.92E-05	28.39861	0.0000
S	0.005265	5.88E-05	89.50017	0.0000
Log likelihood	10239.83	Mean dependent var.		1.06E-18
No. of Coefficients	2	S.D. dependent var.		0.006318

Table 8-6 Empirical Distribution Test for CIRESID03

9 ARCH Effects

Note we have not allowed for ARCH effects in our model. Our intent is to consider unconditional distributions. Further, we do not wish to dilute the focus of this paper, this is the error correction process is the return generation process. However, one can take a number of approaches, either a univariate GARCH model or a mutlivariate GARCH model to remove the ARCH effects [we do not quote the references for GARCH models in this paper].

10 Conclusion

In this paper we take a “naive” view of the return generation process by allowing all the price data to “econometrically” speak for themselves. As returns are serially correlated we first attempt a VAR model. As the prices are cointegrated we then apply the VECM model. From the VECM model we are able to estimate the cointegrating vectors. We find that the cointegrating vectors (both lead and lag) explain the logarithmic returns nearly completely and thus propose a VECM-lead(CointEq1,2,3) model.

The findings indicate that the return generation process can be modelled as an error correction process. Abnormal information or “news” only plays a marginal role in the process. Normal information and expectations play a significant role in the process. Our model does not contradict the EMH hypothesis. However, we make a distinction between “expectations” over “information” in the return generation process.

The model also supports the view that asset price dynamics comprise of normal and abnormal shocks [see Merton (1976)]:

(1) The normal shocks can be due to “temporary imbalance between supply and demand” Merton (1976), changes in the price of risk or in the

economic outlook, or other new information that causes marginal changes in the asset value.

(2) The abnormal shocks are due to “the arrival of new important information about the asset that has more than a marginal effect on value” Merton (1976).

We further found that the COINTEQ01 residuals are more “normal” than the D(CLOSE) logarithmic returns. The results for the VECM-lead(CointEq1,2,3) model indicate that COINTEQ01 (or $C_t - 0.999681 O_t$) is a good proxy for ΔC_t logarithmic returns or the change in close logarithmic prices is equivalent to the difference between the current close and current open logarithmic prices.

We also find that the cointegrating logarithmic returns are non-normally distributed, with the logistic and extreme-value distributions being able to describe the serially uncorrelated cointegrating residuals. Whilst this may be pre-emptive in this paper, it sets a possible direction for further research into the unconditional distributions of financial asset returns.

11 References

Engle, R F, and C W J Granger, 1987, Co-integration and error correction: Representation, estimation and testing, *Econometrica* 55, 251-76.

Johansen, S, 1988, Statistical analysis of cointegrating vectors, *Journal of Economic Dynamics and Control* 12, 231-254.

Lo, A W, and A C MacKinlay, 1999. *A Non-Random Walk Down Wall Street* (Princeton University Press, Princeton).

Markellos, R N, 2002, Nonlinear dynamics in economics and finance, Working Paper 8/02 (Dept. of Management Science and Technology, Athens University of Economics and Business, Athens).

Merton, RC, 1976, Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics* 3, 125-144.

Samuelson, P, 1965, Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review* 6, 41-49.