Journal of Korean
Data & Information Science Society
2004, Vol. 15, No. 2, pp. 413~422

Box-Cox Transformation for Conditional Heteroscedasticity in Domestic Financial Time Series

S.Y. Hwang[1] · J.H. Lee[2]

Abstract

Box-Cox power transformation is employed for analyzing volatilities in Korean financial time series such as KOSPI, KOSDAQ index and interest rates. Statistical procedures for Box-Cox transformed ARCH models are presented. For illustration, diverse financial time series data are analyzed and appropriate power transformations are suggested for each data.

Keywords : ARCH, Box-Cox transformation, financial time series

1. 서　론

주가지수, 선물지수, 환율, 금리 등과 같은 금융 시계열자료들의 공통적인 특징은 시간의 추이에 따라 변동성이 매우 심하다는 것이다. 따라서 이들 자료에는 오차항이 동일한 분산을 따른다는 가정보다는 이분산을 따른다는 가정이 타당하다. 이분산성을 고려한 자기회귀조건부 이분산(autoregressive conditional heteroscedasticity ; ARCH) 모형은 Engle(1982)에 의해 처음 제시되었다. 이후 이를 확장시킨 다양한 형태의 비선형 ARCH류 모형들, 예를 들어 beta-ARCH, Threshold-ARCH, Box-Cox transformed ARCH, Box-Cox transformed threshold GARCH 모형등이 제안되었다. 이들 모형들에 대해서는 Hwang and Kim(2004)의 참고문헌을 참조하기 바란다. 앞으로 표현을 간략히 하기 위해 Box-Cox는 BC로 표기하도록 한다.

BC-변환 ARCH 모형은 기존의 Engle 의 ARCH 모형을 정규성(또는 대칭성)을 확보하고 이분산의 함수구조를 좀더 정교하게 모형화 하고자 제안된 모형으로 아직까지 국내 시계열에의 적합은 적절히 이루어지지 않은 상태이다. 본 연구에서는 BC-변환 ARCH 에 대해 알아본 후, 이분산성이 존재하는 국내의 대표적인 금융 시계열 자료 6개에 모형을 적합 시키고 각 시계열에 적절한 변환의 형태를 제시하고자 한다.

---

1) First Author : Professor, Department of Statistics, Sookmyung Women's Univ., Seoul, 140-742, Korea. E-mail : shwang@sookmyung.ac.kr
2) Graduate student, Department of Statistics, Sookmyung Women's Univ., Seoul, Korea.

2. BC-변환 ARCH 모형의 소개

BC-변환 ARCH는 Higgins와 Bera(1992)에 의해 처음 소개되었다. 이들은 BC-변환을 통해서 오차항의 정규성과 대칭성을 만족시킬 수 있음을 주장하였다. BC-변환 ARCH 모형을 따르는 { ε_t }의 모형식은 다음과 같다.

$$\varepsilon_t = \sqrt{h_t}\, e_t$$
$$h_t^{\delta} = \alpha_0 + \alpha_1(\varepsilon_{t-1}^2)^{\delta} + \alpha_2(\varepsilon_{t-2}^2)^{\delta} + \cdots + \alpha_q(\varepsilon_{t-q}^2)^{\delta} \tag{2.1}$$

여기서 $\alpha_i \geq 0$, $\delta > 0$ 이며 e_t 는 평균이 0이고 분산이 1인 iid 확률변수열이다.

BC-변환 ARCH 모형은 보다 유연한 형태의 조건부 분산 모형을 제공한다. $\delta=1$ 은 ARCH(q)모형이며 $\delta=1/2$ 은 절대값 ARCH 모형을 제공한다. 멱변환 δ 가 1보다 작으면 분산함수 h_t 가 과거값(ε_{t-1}, ε_{t-2}, ...)의 제곱에 대해 오목(concave) 함수이며 δ 가 1보다 크면 볼록(convex)함수를 그리고 δ가 1 이면 직선함수를 나타낸다. BC-변환 ARCH 인 식 (2.1)의 { ε_t } 는 마팅게일 차(martingale difference) 이므로 계열상관이 없는 넓은 의미에서의 백색잡음이다. 여기에 자기상관구조를 부여하기 위하여 다음과 같은 AR 모형을 고려해보자.

관측된 자료 { y_t}가 다음과 같은 수식을 따를 때 자료 { y_t}는 BC-변환된 AR(m)-ARCH(q) 모형을 따른다고 정의한다.

$$y_t = -\phi_1 y_{t-1} - \cdots - \phi_m y_{t-m} + \varepsilon_t \tag{2.2}$$

여기서 { ε_t}는 수식 (2.1)의 BC-변환 ARCH(q)를 따르며 자기 회귀계수 ϕ에 음수를 붙인 것은 SAS/ETS 의 Proc autoreg 의 정의식과 보조를 맞추기 위함이다.

3. 국내 금융시계열을 위한 BC-변환 AR-ARCH 모형

국내 금융 시계열 자료의 변동성(조건부 분산 ; h_t)이 가지는 특징은 1998년을 기점으로 하여 그 변동성이 더욱 심하다는 것을 알 수 있다. 이것은 1998년에 일어난 외환위기(IMF 구제 금융)의 영향 때문이다. IMF 구제 금융의 직접적 영향에서 벗어나기까지 1년 이상이 소요되었고, 이 기간 동안의 금융 시계열 자료는 특히 변동성이 심하고 불안한 모습을 보인다. 외환 위기 이후 국내의 금융 시계열자료 - 주가지수, 코스닥 지수, 환율 , 금리 등 - 의 결정은 정부의 개입에서 벗어나 시장 자율에 맡겨지면서 국내·외 정세에 민감하게 반응하게 되었다. 2001년 9.11 테러, 이라크 전쟁, 걸프만 전쟁 등과 같은 불안한 국제 정세 및 불안한 남북관계 등은 금융 시계열 자료에 다양한 형태로 영향을 미쳤다. 이처럼 국내 금융 시계열 자료는 경제적 요소 뿐만 아니라 정치적 요소 등 다양한 형태의 요소들에 영향을 받기 때문에 일반적인 시계열

자료들에 비교하여 변동성이 매우 심하다는 특징을 가진다.

BC-변환된 AR-ARCH 모형의 적합을 위해 다음 단계를 적용시켰다.

1단계 : 로그변환 , 차분 등을 통해서 데이터를 정상화 시킨 후 정상화된 자료를 $\{ y_t \}$ 로 하여 분석을 시작한다.

2단계 : SAS/ETS 의 Proc autoreg를 이용하여(여기서는 $\delta=1$ 인 Engle 의 모형에 해당) AR-차수 $m=m_0$ 과 ARCH-차수 $q=q_0$ 를 정한 후 backstep 옵션을 통해 모형의 윤곽을 확정한다. 또한 여기서 얻은 AR 계수인 $\widehat{\phi_1},\ldots,\widehat{\phi_{m_0}}$ 를 다음 단계를 위해 저장한다. 이에 대한 자세한 내용은 박유성과 송석헌(1998, 5장)을 참조하기 바란다.

3단계 : 2단계에서 얻은 m_0 , q_0 그리고 $\widehat{\phi_1},\ldots,\widehat{\phi_{m_0}}$ 를 이용하여 다음 식을 고려한다.

$$y_t = -\widehat{\phi_1} y_{t-1} - \cdots - \widehat{\phi_{m_0}} y_{t-m_0} + \varepsilon_t \tag{3.1}$$

$$h_t^{\delta} = \alpha_0 + \alpha_1(\varepsilon_{t-1}^2)^{\delta} + \alpha_2(\varepsilon_{t-2}^2)^{\delta} + \cdots + \alpha_p(\varepsilon_{t-q_0}^2)^{\delta} \tag{3.2}$$

이제 $\alpha_0, \alpha_1, \cdots, \alpha_q$ 값과 δ 를 추정하기 위해 수식 (3.1)로부터 얻은 잔차 $\widehat{\varepsilon_t}$, 즉

$$\widehat{\varepsilon_t} = y_t - [-\widehat{\phi_1} y_{t-1} - \cdots - \widehat{\phi_{m_0}} y_{t-m_0}] \qquad \text{.3)}$$

로 부터 식(3.2)의 h_t 를 $\widehat{\varepsilon}_t^2$ 로 대치 시킨 다음의 회귀식을 생각한다.

$$\widehat{\varepsilon_t}^{2\delta} = \alpha_0 + \alpha_1(\widehat{\varepsilon}_{t-1}^2)^{\delta} + \alpha_2(\widehat{\varepsilon}_{t-2}^2)^{\delta} + \cdots + \alpha_{q_0}(\widehat{\varepsilon}_{t-q_0})^{\delta} + \text{오차}. \tag{3.4}$$

4단계 : 이제 멱변환 δ 를 0.1에서 2.0까지 0.1씩 증가시키면서 회귀모형 (3.4)의 잔차제곱합 SSE_δ를 구한다. δ변환을 통한 단위의 척도가 각각 틀리기 때문에 SSE_δ 의 직접 비교는 불가능하므로 표준화된 척도로서 다음과 같은 로그우도함수를 사용한다.

$$L(\delta) = n\ln(\delta) - \frac{n}{2}(SSE_\delta) + n(\delta-1)\ln(GM(\widehat{\varepsilon}^2)) \tag{3.5}$$

여기서, $GM(\cdot)$은 $\widehat{\varepsilon_1}^2, \ldots, \widehat{\varepsilon_n}^2$ 의 기하평균값이다(cf. 강명욱외 3인(1996, 5장)).

4. 실증 분석

이 장에서는 국내의 대표적인 금융 시계열 자료 6가지 (종합주가지수, 주가지수선물, 코스닥지수, 코스닥지수선물, 대미환율, 회사채수익률)에 대해서 BC 변환된 AR-ARCH 모형을 적합시켜 보도록 하자. 자료는 모두 일별 자료이며 자료의 범위는 다음과 같다.

자료1 : **종합주가지수**(92. 1. 1 - 03. 8.31) ; 자료2 : **KOSPI 선물** (92. 1. 1 - 03. 8.31)

자료3 : **코스닥지수**(97. 1. 1 - 03. 8.31) ; 자료4 : **코스닥 지수 선물** (99. 1. 1 - 03. 8.31)
자료5 : **회사채 수익률**(98. 9. 1 - 03. 9.30) ; 자료6 : **대미 환율** (98. 5. 1 - 03. 8.31).

먼저 자료를 정상화 시켜 얻은 입력자료인 $\{y_t\}$의 시계열도는 그림 4.1-4.6 에 수록하였다. 이분산 h_t 를 위한 멱변환 δ 를 제시하기 위하여 앞장에서 서술한 4단계를 수행한 결과 다음의 결과를 얻을 수 있다.

종합주가지수(δ=0.8)

$$h_t^{0.8} = 47.5403 + 0.0789\,\varepsilon_{t-1}^{1.6} + 0.1859\,\varepsilon_{t-2}^{1.6} + 0.1445\,\varepsilon_{t-3}^{1.6} + 0.1827\,\varepsilon_{t-4}^{1.6} + 0.1621\,\varepsilon_{t-5}^{1.6}$$

KOSPI 선물 (δ=0.3)

$$h_t^{0.3} = 0.6508 + 0.0868\,\varepsilon_{t-1}^{0.6} + 0.2028\,\varepsilon_{t-2}^{0.6} + 0.1550\,\varepsilon_{t-3}^{0.6} + 0.1741\,\varepsilon_{t-4}^{0.6} + 0.1641\,\varepsilon_{t-5}^{0.6}$$

코스닥지수(δ=0.4)

$$h_t^{0.4} = 0.8795 + 0.2455\,\varepsilon_{t-1}^{0.8} + 0.2698\,\varepsilon_{t-2}^{0.8} + 0.2547\,\varepsilon_{t-3}^{0.8} + 0.2960\,\varepsilon_{t-4}^{0.8}$$

코스닥 지수 선물 (δ=0.7)

$$h_t^{0.7} = 4.4313 + 0.1428\,\varepsilon_{t-1}^{1.4} + 0.1073\,\varepsilon_{t-2}^{1.4} + 0.1531\,\varepsilon_{t-3}^{1.4} + 0.1226\,\varepsilon_{t-4}^{1.4}$$

회사채 수익률(δ=0.3)

$$h_t^{0.3} = 0.5279 + 0.5358\,\varepsilon_{t-1}^{0.6} + 0.3759\,\varepsilon_{t-2}^{0.6}$$

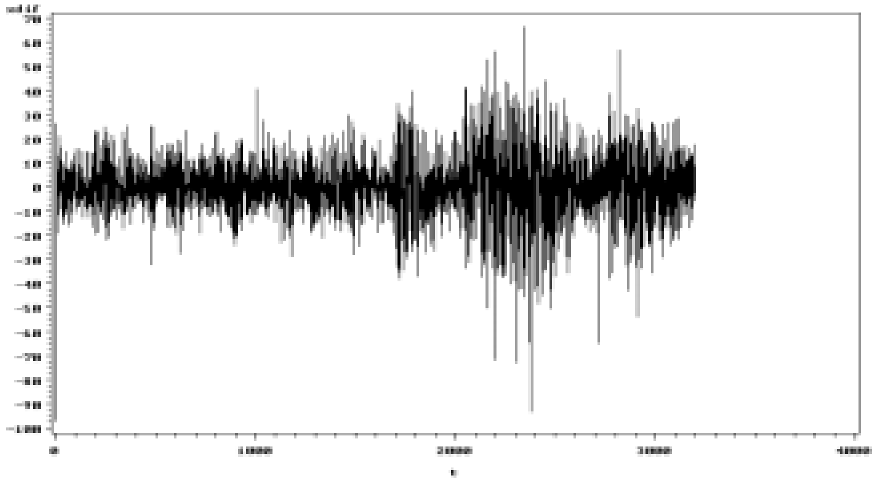
대미 환율 (δ=0.6)

$$h_t^{0.6} = 6.74 + 0.3627\,\varepsilon_{t-1}^{1.2} + 0.2768\,\varepsilon_{t-2}^{1.2} + 0.2124\,\varepsilon_{t-3}^{1.2}$$

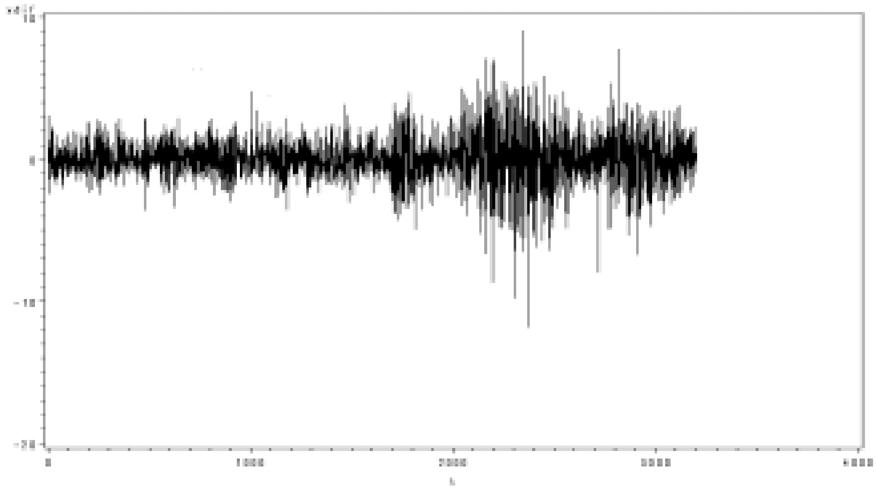
대미환율자료의 적합과정 4단계를 자세히 살펴보도록 하자. 다른 자료들은 유사한 방법으로 수행 가능하므로 생략하기로 한다.

1단계 : 정상자료 $\{y_t\}$을 얻기 위해 원자료 $\{w_t\}$ 에 로그변환과 차분을 실시하였다.

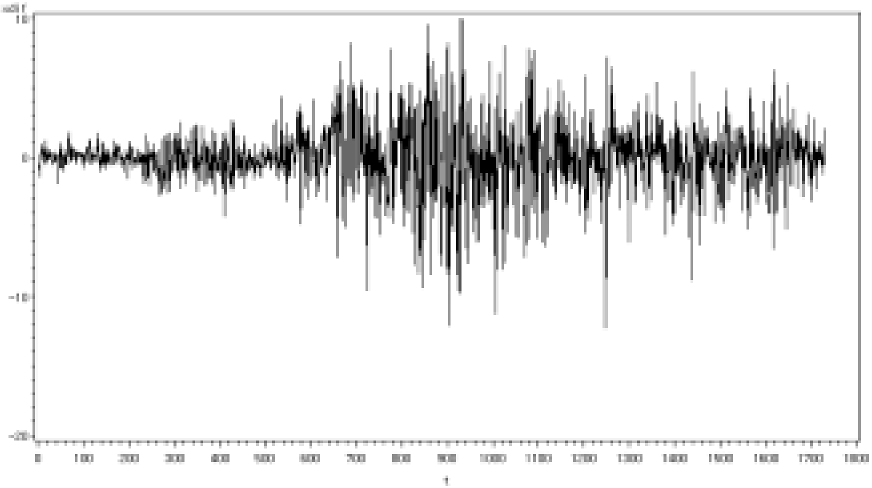
$$y_t = \{\log(w_t) - \log(w_{t-1})\} \times 100$$



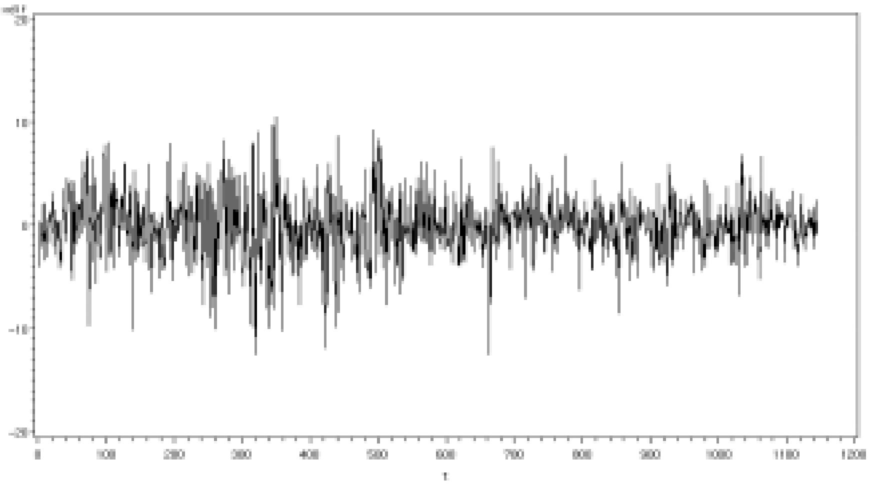
<그림 4.1> 종합주가지수(일차 차분)



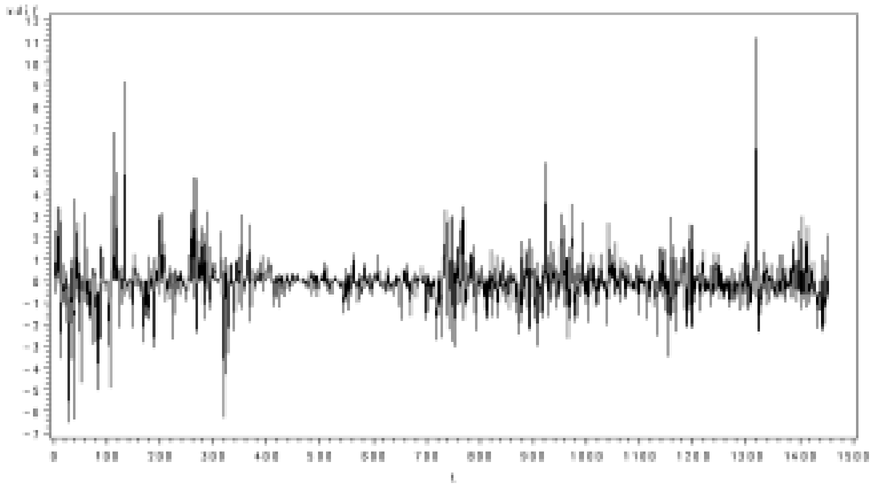
<그림 4.2> KOSPI 선물(일차 차분)



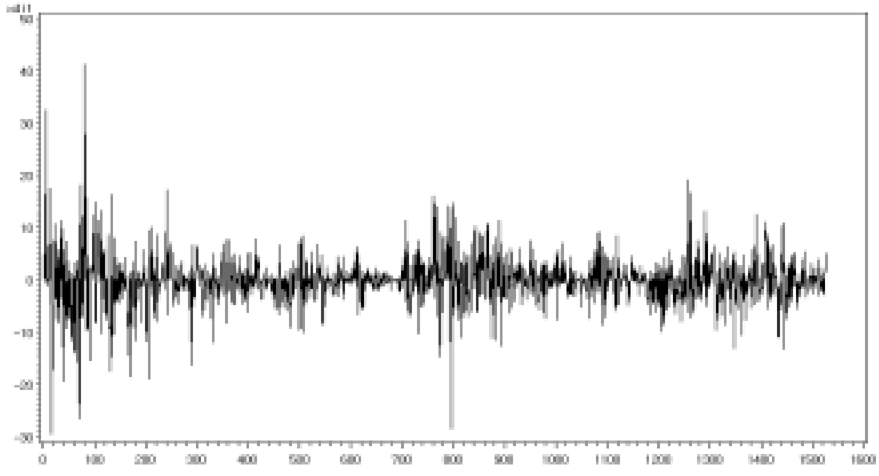
<그림 4.3> 코스닥지수(로그변환 후 차분*100)



<그림 4.4> 코스닥 지수 선물(로그변환 후 차분*100)



<그림 4.5> 회사채수익률(로그변환 후 차분*100)



<그림 4.6> 대미 환율(로그변환 후 차분*100)

2단계 : SAS/Proc autoreg를 이용하여 자료의 이분산 여부와 자기상관 여부의 검정을 실시하였다.

<표 4.1> 자기상관 검정 결과 및 이분산성 검정 결과

Order	DW	Pr < DW	Pr > DW	Order	Q	Pr > Q	LM	Pr > LM
	Durbin-Watson Statistics				Q and LM Tests for ARCH Disturbances			
1	1.7558	<.0001	1.0000	1	43.6075	<.0001	43.5508	<.0001
2	1.9422	0.1348	0.8652	2	70.8192	<.0001	60.7580	<.0001
3	1.8796	0.0106	0.9894	3	89.6189	<.0001	69.0946	<.0001
4	1.9884	0.4402	0.5598	4	97.4878	<.0001	70.3011	<.0001
5	1.9559	0.2237	0.7763	5	108.2729	<.0001	73.7913	<.0001
6	1.8851	0.0171	0.9829	6	124.7753	<.0001	80.5541	<.0001
7	1.8803	0.0144	0.9856	7	164.2226	<.0001	101.4659	<.0001
8	1.9797	0.4139	0.5861	8	167.7459	<.0001	102.0671	<.0001
9	1.9392	0.1627	0.8373	9	205.0480	<.0001	120.4811	<.0001
10	2.0068	0.6422	0.3578	10	223.5499	<.0001	123.3600	<.0001
11	2.0138	0.7006	0.2994	11	265.2668	<.0001	140.5577	<.0001
12	1.9198	0.0992	0.9008	12	309.3753	<.0001	152.4179	<.0001

<표 4.1>의 출력결과를 보면 lag 1부터 12까지의 모든 lag에서 Q와 LM검정이 유의하게 나타나고 있어 ARCH 효과가 있음을 알 수 있다. Durbin-Watson 통계량을 살펴보면 자기상관이 있음을 알 수 있다. Backstep의 옵션을 이용해서 y_t 의 AR-모형을 결정하였다.

$$y_t = -\widehat{\phi}_1 y_{t-1} - \widehat{\phi}_3 y_{t-3} + \varepsilon_t$$

ARCH(q)에서 q는 모수 절약의 기준에 따라 $LM_{k+1} - LM_k$가 근사적으로 $\chi^2(1)$ 분포를 따른다는 사실을 이용하여 선택하게 된다. <표 4.1>에서 LM 통계량을 살펴보면 유의수준5%에서 ARCH 모형의 차수는 $q_0 = 3$ 이 적당함을 알 수 있다.

3단계 : 이제 h_t 모형으로부터

$$h_t^{\delta} = \alpha_0 + \alpha_1 \varepsilon_{t-1}^{2\delta} + \alpha_2 \varepsilon_{t-2}^{2\delta} + \alpha_3 \varepsilon_{t-3}^{2\delta}$$

δ 추정을 위한 다음의 회귀식을 고려한다.

$$\widehat{\varepsilon}_t^{2\delta} = \alpha_0 + \alpha_1 \widehat{\varepsilon}_{t-1}^{2\delta} + \alpha_2 \widehat{\varepsilon}_{t-2}^{2\delta} + \alpha_3 \widehat{\varepsilon}_{t-3}^{2\delta} + \text{오차} \tag{4.1}$$

여기서 $\widehat{\varepsilon}_t$ 은 잔차로서 $\widehat{\varepsilon}_t = y_t - [-\widehat{\phi}_1 y_{t-1} - \widehat{\phi}_3 y_{t-3}]$ 이다.

4단계 : 이제 멱변환 δ 를 0.1에서 2.0까지 0.1씩 증가시키면서 식(3.5)의 로그우도함수를 계산한다.

$$L(\delta) = n\ln(\delta) - \frac{n}{2}(SSE_{\delta}) + n(\delta - 1)\ln(GM(\widehat{\varepsilon}^2)).$$

<표 4.2>로부터 $\delta = 0.6$ 일 때 $L(\delta)$ 의 값이 최대값을 가진다는 것을 알 수 있으며 이를 식 (4.1)에 대입하여 $h_t^{0.6} = 6.74 + 0.3627\varepsilon_{t-1}^{1.2} + 0.2768\varepsilon_{t-2}^{1.2} + 0.2124\varepsilon_{t-3}^{1.2}$

을 얻는다.

<표 4.2> 대미 환율자료의 δ 값과 δ 값에 따른 $L(\delta)$

DELTA	$L(\delta)$	DELTA	$L(\delta)$	DELTA	$L(\delta)$	DELTA	$L(\delta)$
0.1	-13065.72	0.6	-10157	1.1	-12899.42	1.6	-16980.73
0.2	-11885	0.7	-10367.37	1.2	-13674.04	1.7	-17842.39
0.3	-11149.32	0.8	-10840.55	1.3	-14473.87	1.8	-18713.77
0.4	-10618.78	0.9	-11460.94	1.4	-15293.93	1.9	-19593.5
0.5	-10264.82	1	-12157.3	1.5	-16130.5	2	-20480.51

5. 결론 및 분계점(threshold) 모형의 소개

지금까지 시간에 따라 변동성이 큰 자료들을 분석하기에 적합한 BOX-COX (BC) 변환 ARCH 모형에 대해서 알아보고 6가지 우리 나라 금융 시계열 자료에 적합시켜 보았다. 한가지 특징적인 사항은 변환 δ 가 모두 1 보다 작게 나온 사실인데 이는 국내 금융시계열이 Engle(δ = 1)의 모형에 비해 과거 잔차(ε_{t-1})제곱의 입장에서 오목(concave)한 함수형태를 띄고 있다는 것을 의미한다. 이는 잔차의 제곱 ε_{t-1}^2 이 아주 커지게 되면 분산 h_t 가 Engle(δ = 1)의 모형에 비해 작아지는 경향이 있으며 이는 국내 금융시장에 규제가 존재한다는 것을 암시해주고 있다고 해석할 수 있다. 예를 들어 주가와 연관된 시계열의 경우 상하한($\pm 15\%$)이 규제를 하고 있고 환율 연관 시계열의 경우 정부 개입이 이런 현상을 가져온 것으로 판단된다.

이분산 h_t 를 비대칭 함수로 모형화 하는 것도 추후 과제로 해봄직 하다. 비대칭 모형중 분계점(threshold)을 도입한 BC 변환 분계점 ARCH 모형은 다음과 같다(cf. Hwang and Kim(2004)).

$$h_t^{\delta} = \alpha_0 + \alpha_{11}(\varepsilon_{t-1}^{+2})^{\delta} + \alpha_{12}(\varepsilon_{t-1}^{-2})^{\delta} + \cdots + \alpha_{q1}(\varepsilon_{t-q}^{+2})^{\delta} + \alpha_{q2}(\varepsilon_{t-q}^{-2})^{\delta}$$

여기서 $x^+ = \max(x, 0)$, $x^- = \max(-x, 0)$ 으로서 분계점 함수를 표시한다. 만일 뚜렷한 비대칭 변동성이 우리나라 금융 시계열 자료들에 존재 한다면 BC 변환 ARCH 모형 뿐만 아니라 BC 변환 분계점 ARCH 모형을 적합시켜 보는 것도 의미 있는 일이 될 것이다.

참 고 문 헌

1. 강명욱, 김영일, 안철환, 이용구(1996). *회귀분석: 모형개발과 진단*. 율곡출판사.

2. 박유성, 송석헌 (1998). *SAS/ETS를 이용한 경영경제자료분석*. 정일출판사.
3. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50,* 987-1007.
4. Higgins, M.L. and Bera, A.K (1992). A class of nonlinear ARCH models. *International Econometrics Review*, *33,* 137-158.
5. Hwang, S.Y. and Kim, T.Y.(2004). Power transformation and threshold modeling for ARCH innovations with applications to tests for ARCH structure. *Stochastic Processes and their Applications, 110*, 295-314.
6. SAS/ETS(1995). SAS institue inc., N.C. Cary.

[2004년 2월 접수, 2004년 5월 채택]