

연속형-GARCH 시계열의 범주형화(Clipping)를 통한 분석

최문선¹ · 백지선² · 황선영³

¹숙명여자대학교 통계학과, ²통계청 통계개발원 조사연구실, ³숙명여자대학교 통계학과
(2010년 6월 접수, 2010년 6월 채택)

요약

본 논문에서는 연속형-GARCH 시계열 자료인 금융 시계열 자료에 대해서 클리핑(clipping)을 통해 얻은 이항(binary) 범주형 시계열을 분석하고 응용하는 방법에 대해 연구하고 있다. 모수추정 방법을 소개하고 있으며 이를 이용하여 이분산 시계열과 연관된 확률을 추정하는 방법을 제시하였다.

주요어: 클리핑(clipping), 범주형(categorical) 시계열, 이분산(heteroscedastic) 시계열.

1. 서론

일반적으로 시계열 자료는 연속형 변수로 얻어지며 이에 대한 모형으로는 Box-Jenkins (1970)의 ARIMA 모형을 비롯한 실용적인 분석 방법들이 방대하게 연구 응용되고 있다. 그러나 모든 시계열 자료들이 연속형은 아니며 이산형 변수의 형태를 갖는 경우도 많다. Kedem (1980), Kaufmann (1987), Fokianos와 Kedem (1998, 2003) 외의 많은 연구자들이 범주형 시계열 자료에 대한 분석을 연구하였다. 본래 자료가 연속형 시계열 자료라 할지라도 이를 범주형으로 변환하거나 또는 관측한 범주형 시계열 자료를 관측할 수 있는 다른 연속형 시계열 자료로부터 얻었다고 가정한다면 분석을 통해 간단하지만 직관적이고 유용한 결과를 제시해 줄 수도 있다. 예를 들어, 강수량의 실제 값도 중요하지만 어떤 일정 강수량을 기준으로 이 값의 초과 여부에만 관심을 갖고 모형화하거나, 주식 수익률 자체의 예측보다는 단순히 수익률의 증감에 초점을 맞추어 분석을 할 수 있다. 강수량에 대한 예측 또는 주식 수익률의 예측 값을 계산하는 것보다 강수 확률이나 주가의 등락 확률을 이용하여 자료의 움직임 설명하는 것이 실무자에게 더 필요한 정보가 될 수도 있다.

Zhen과 Basawa (2009)는 이항(binary) 모형을 이용하여 이러한 분석 방법을 보여주고자 하였으며 이 목적을 위해 연속형 시계열을 범주형으로 만드는 작업을 클리핑(Clipping)이라 지칭하였다. 이 용어는 원래 연속형 자료로부터 합목적적인 이산형 변환자료를 생성(클리핑)한다는 의미일 것이다. 이들은 AR(p) 모형을 따르는 연속형 시계열 자료를 클리핑을 통해 이항(binary) 자료로 변환하고 이러한 경우에 가정한 AR(p) 모형을 추정하는 여러 가지 방법을 제시하였으며 2005-2006년도 IBM 일별주가자료 분석을 통하여 클리핑의 가능한 유용성에 대해 연구하였다.

본 논문에서는 조건부 이분산 시계열의 클리핑에 대해 연구하도록 한다. 즉, 변동성 시계열의 대표적인 모형인 AR(p)-GARCH(m, s) 과정을 따르는 연속형-GARCH 시계열 변수가 클리핑을 통해 이항 변수로서 다루어지는 경우를 살펴보고자 한다. 2절에서는 이항 시계열 모형과 이에 대한 몇 가지 추정 방법을 알아보고 3절에서는 이를 이용하여 연속형 자료인 세 종류의 국내 주식 자료(KOSPI, 현대차, 대우증권)에 적합 시켜 보았으며 그 응용으로서, 클리핑된 모형으로부터 이항 변수에 대한 1시차 후 예측 확률을 구하였다.

³교신저자: (140-742) 서울시 용산구 효창원길 52, 숙명여자대학교 통계학과, 교수. E-mail: shwang@sm.ac.kr

2. 연속형-GARCH 모형으로부터 클리핑(Clipping)된 이항(Binary) 모형

시계열 $\{Z_t\}$ 가 정상성(stationarity)을 만족하는 확률 과정이라고 할 때, 클리핑(clipping)을 통해서 다음과 같은 이항(binary) 확률 과정 $\{Y_t\}$ 을 정의한다.

$$Y_t = \begin{cases} 1, & Z_t > c, \\ 0, & Z_t \leq c, \end{cases} \quad (2.1)$$

여기서 c 는 상수로서 분석 목적에 따라 분석자에 의해 미리 정해지는 값이며, 확률 과정 $\{Z_t\}$ 에 대해서는 이분산변동성(heteroscedastic volatility)을 갖는 금융 시계열 자료의 성질을 고려하여 AR(p)-GARCH(m, s) 모형을 가정한다. AR(p)-GARCH(m, s) 모형은 다음과 같다.

$$\begin{aligned} Z_t &= \phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} + \epsilon_t, \\ \epsilon_t | F_{t-1} &\sim i.i.d. \bullet(0, h_t), \\ h_t &= \alpha_0 + \sum_{i=1}^s \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^m \beta_j h_{t-j}, \end{aligned}$$

위의 식에서 F_{t-1} 은 ($t-1$)시점까지의 모든 정보를 나타내며 $\bullet(0, h_t)$ 은 기대값이 0이고 분산이 h_t 인 임의의 분포를 의미한다. 금융 시계열 분야에서는 주로 정규분포나 t -분포, 또는 GED(generalized error distribution)를 가정한다.

Zhen과 Basawa (2009)는 $\{Z_t\}$ 에 대해 등분산(homoscedastic) 모형인 AR(p) 모형을 고려하였을 때 모수를 추정하는 5가지 방법을 제시하였다 [(i)이항 변수 $\{Y_t\}$ 의 정보만을 이용한 방법; (ii) Markov chain 가정을 이용한 방법; (iii) $\{Z_t\}$ 의 정보만을 이용한 방법; (iv) Partial likelihood을 이용한 방법; (v) Extended partial likelihood을 이용한 방법]. 본 논문에서는 Zhen과 Basawa (2009)가 제시한 방법들 중에서 이분산 모형인 AR(p)-GARCH(m, s) 모형의 경우에 적용될 수 있는 $\{Z_t\}$ 의 정보만을 이용하는 방법과 partial likelihood 그리고 extended partial likelihood를 이용하여 모수들을 추정하는 방법에 대해서 간단히 알아보기로 한다.

2.1. 연속형 시계열 $\{Z_t\}$ 의 정보만을 이용한 방법

가장 간단한 방법으로서 $\{Y_t\}$ 의 정보는 반영하지 않고 $\{\epsilon_t\}$ 의 조건부분포를 가정한 후, 관측한 자료 $\{Z_t\}$ 만을 이용하여 최대우도추정량을 구하는 방법이며 우도함수는 다음과 같다.

$$L_2(\theta) = \prod_{t=1}^n f_\epsilon \left[\left(Z_t - \phi_0 - \sum_{i=1}^p \phi_i Z_{t-i} \right) / \sqrt{h_t} \right], \quad (2.2)$$

여기서 $\theta = (\phi_0, \phi_1, \dots, \phi_p, \alpha_0, \alpha_1, \dots, \alpha_s, \beta_1, \dots, \beta_m)'$ 이고, $f_\epsilon(\cdot)$ 은 평균이 0이고 분산이 1인 표준화된 오차항의 분포이다. 우도함수가 주어지지 않은 경우에는 준우도함수(quasi-likelihood function) 기법을 이용할 수 있다.

2.2. Partial likelihood을 이용한 방법

이항(binary) 변수 $\{Y_t\}$ 의 조건부확률 $p_t(\theta) = P_\theta(Y_t = 1 | Z_{(t-1)})$ 와 조건부확률밀도함수 $p_t(y_t; \theta) =$

$P_\theta(Y_t = y_t | Z_{(t-1)})$ 은 다음과 같이 계산할 수 있다. 여기서 $Z_{(t-1)} = (Z_{t-1}, \dots, Z_{t-p})$ 을 의미한다.

$$p_t(\theta) = P\left(\phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} + \epsilon_t > c\right) = P\left(\epsilon_t > c - \phi_0 - \sum_{i=1}^p \phi_i Z_{t-i}\right)$$

$$p_t(y_t; \theta) = [p_t(\theta)]^{y_t} [1 - p_t(\theta)]^{(1-y_t)}.$$

오차항에 대해 정규분포를 가정하고, 조건부분산 h_t 를 고려하여 조건부확률 $p_t(\theta)$ 을 다시 쓰면 다음과 같다. 표준정규분포의 확률분포함수를 $\Phi(\cdot)$ 로 표시할 때

$$p_t(\theta) = \Phi\left[\left(\phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} - c\right) / \sqrt{h_t}\right]. \tag{2.3}$$

식 (2.3)의 조건부확률을 이용하여 관측된 자료 $\{z_1, y_1, z_2, y_2, \dots, z_n, y_n\}$ 에서 이항 변수 $\{Y_t\}$ 에 대한 partial likelihood을 다음과 같이 나타낼 수 있으며, 이 함수를 최대화함으로써 MPLE(maximum partial likelihood estimator)을 구할 수 있다.

$$PL_1(\theta) = \prod_{t=1}^n \Phi\left[\left(\phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} - c\right) / \sqrt{h_t}\right]^{y_t} \left[1 - \Phi\left[\left(\phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} - c\right) / \sqrt{h_t}\right]\right]^{(1-y_t)}. \tag{2.4}$$

Partial likelihood에 대한 자세한 내용은 Wong (1986)과 Slud와 Kedem (1994)을 참고하기 바란다.

2.3. Extended partial likelihood

$\{Z_{t-1}, Y_t\}$ 의 결합밀도함수 $p_\theta(Z_{t-1}, Y_t) = p_\theta(Y_t | Z_{t-1})p_\theta(Z_{t-1})$ 을 이용하여 extended partial likelihood을 다음과 같이 나타낼 수 있다 (cf. Zhen과 Basawa, 2009).

$$PL_2(Z, Y; \theta) = PL_1(\theta)L_z(\theta)$$

$$= \prod_{t=1}^n \Phi\left[\left(\phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} - c\right) / \sqrt{h_t}\right]^{y_t} \left[1 - \Phi\left[\left(\phi_0 + \sum_{i=1}^p \phi_i Z_{t-i} - c\right) / \sqrt{h_t}\right]\right]^{(1-y_t)}$$

$$\times f_\epsilon\left[\left(Z_t - \phi_0 - \sum_{i=1}^p \phi_i Z_{t-i}\right) / \sqrt{h_t}\right],$$

여기서 $L_z(\theta)$ 은 식 (2.2)이며 $PL_1(\theta)$ 은 식 (2.4)를 나타낸다.

3. 사례분석

본 장에서는 세 가지 국내 주식 자료에 클리핑을 통해 이항 모형을 적합 시켜 보았다. Zhen과 Basawa (2009)에서 수행한 시뮬레이션 결과를 보면 5가지 추정 방법 중에서 $\{Z_t\}$ 을 이용하여 추정하는 방법이 효율성 측면에서 가장 뛰어난 것을 확인할 수 있다. 따라서 2.1절의 방법을 사용하여 모수를 추정하고 추정된 결과를 바탕으로 1시차 후 예측 확률을 계산하기로 한다. 자료는 2000년 1월 4일부터 2010년 2월 26일까지의 총 2,509개의 일별 KOSPI, 현대차, 대우증권의 주가 자료를 차분 또는 로그차분하여 얻은 수익률이다. 각 주식 자료 $\{Z_t\}$ 을 다음과 같은 클리핑을 통하여 이항확률변수 $\{Y_t\}$ 를 얻는다.

$$Y_t = \begin{cases} 1, & Z_t > 0, \\ 0, & Z_t \leq 0. \end{cases} \tag{3.1}$$

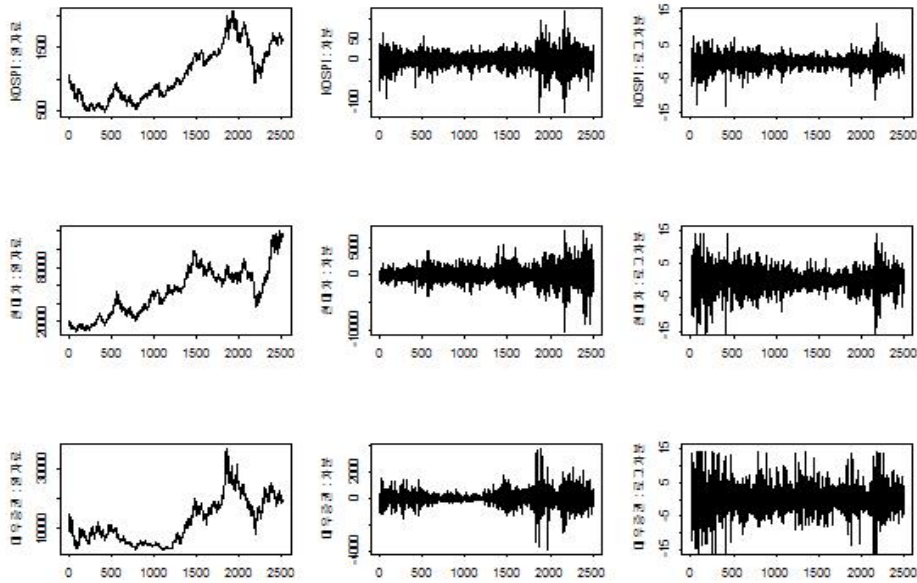


그림 3.1. 원자료, 차분자료, 로그차분자료의 시도표

표 3.1. 추정된 GARCH(1,1) 계수

AR(p)-GARCH(1, 1)		KOSPI		현대차		대우증권	
		차분	로그차분	차분	로그차분	차분	로그차분
		AR(6)	AR(9)	AR(5)	AR(5)	AR(10)	AR(5)
정규분포	$\hat{\alpha}_0$	2.1519	0.0233	8239.32	0.0320	247.20	0.2445
	$\hat{\alpha}_1$	0.0603	0.0736	0.0648	0.0587	0.0729	0.0663
	$\hat{\beta}_1$	0.9331	0.9215	0.9332	0.9392	0.9299	0.9184
t-분포	$\hat{\alpha}_0$	1.7511	0.0228	8406.22	0.0335	79.67	0.1546
	$\hat{\alpha}_1$	0.0566	0.0641	0.0633	0.0573	0.0719	0.0602
	$\hat{\beta}_1$	0.9389	0.9302	0.9347	0.9402	0.9335	0.9300

이는 앞의 식(2.1)에서 정해진 상수값을 $c = 0$ 으로 둬으로써 전일대비 주가의 상승 및 하락을 나타내고자 하는 것이다.

그림 3.1은 각 자료의 원자료 시도표와 차분, 로그차분자료의 시도표를 차례대로 그린 것이다. 주가지수와 두 개의 주가 모두 전체적으로 오름세를 보이며 KOSPI의 변동성에 비해 현대차와 대우증권 주가의 변동성이 매우 큼을 볼 수 있다. 모형의 간편성을 위해 각 자료마다 AR(p)-GARCH(1, 1) 모형을 적합시켰으며 대부분 AR 차수가 5-10 사이에서 추정되었다. 표 3.1은 각 자료의 추정된 모수를 정리한 표이며 (AR(p) 모형의 계수들은 수록을 생략하였다), 여기서 추정된 GARCH(1, 1) 모형의 계수들은 모두 유의수준 $\alpha = 0.05$ 에서 유의하였다. 오차항의 분포로는 정규분포와 t-분포를 모두 고려하였다. 추정된 모수를 이용하여 전체 자료 중에서 2009년 3월 2일부터 2010년 2월 26일까지 253개의 자료에 대하여 다음의 식 (3.2)를 통해 1시차 후 예측 확률을 계산하였다.

$$p_{t+1}(\hat{\theta}) = P_{\hat{\theta}}(Y_{t+1} = 1|Z_{(t)}) = F \left[\left(\hat{\phi}_0 + \sum_{i=1}^p \hat{\phi}_i Z_{t+1-i} \right) / \sqrt{\hat{h}_t} \right], \quad (3.2)$$

표 3.2. KOSPI: 1시차 후 예측 확률($p_{t+1}(\hat{\theta})$)과 실제 관측값(y_{t+1})

차분		로그차분		y_{t+1}
정규분포	t-분포	정규분포	t-분포	
0.5082	0.5022	0.4981*	0.4872*	0
0.4834	0.4941	0.4797	0.5061*	1
0.5096*	0.5083*	0.4754	0.4731	1
0.5221	0.5114	0.5522	0.5333	0
0.5027	0.5006	0.4922*	0.4886*	0
0.5054*	0.5123*	0.4990	0.5110*	1
0.4977	0.4872	0.4977	0.4774	1
0.5021*	0.4966	0.5303*	0.5171*	1
0.5204*	0.5190*	0.5049*	0.4937	1
0.4898*	0.4887*	0.4719*	0.4635*	0
0.4858*	0.4894*	0.5062	0.5088	0
0.4853	0.4866	0.4973	0.4971	1
0.5181*	0.5122*	0.5193*	0.5005*	1
0.5001	0.5007	0.5020	0.4974*	0
0.4913	0.4979	0.4868	0.4946	1
0.4976	0.4955	0.5261*	0.5208*	1
0.5081*	0.4988	0.5302*	0.5095*	1
0.5130*	0.5128*	0.5175*	0.5100*	1
0.4997	0.5009*	0.4827	0.4785	1
0.4992*	0.4957*	0.5245	0.5136	0
0.4798*	0.4807*	0.5027	0.5020	0
0.4652	0.4795	0.4501	0.4744	1
0.5103*	0.5138*	0.5004*	0.5031*	1
0.5187*	0.5078*	0.5357*	0.5148*	1
0.5359*	0.5302*	0.5560*	0.5410*	1
0.5005*	0.4992	0.5061*	0.5015*	1
0.4932	0.4890	0.5114*	0.5005*	1
0.4810*	0.4795*	0.5048	0.4997*	0
0.4601	0.4719	0.4312	0.4484	1
0.5360*	0.5357*	0.5290*	0.5181*	1
0.5031*	0.4908	0.5114*	0.4893	1
0.5040*	0.5121*	0.5225*	0.5326*	1
0.4954*	0.4970*	0.5047	0.5062	0
0.4728	0.4665	0.4865	0.4765	1
0.5047	0.5125	0.5206	0.5306	0
0.4945	0.4970	0.4571	0.4576	1
0.5090*	0.5091*	0.5205*	0.5188*	1
0.5009*	0.4998	0.5224*	0.5232*	1
0.5149*	0.5102*	0.5223*	0.5131*	1
0.5061	0.5019	0.5164	0.5079	0

여기서 $F(\cdot)$ 은 평균이 0이고 분산이 1인 표준화된 오차항 분포의 확률분포함수이다. 표 3.2은 KOSPI의 결과이다. 차분수익률과 로그차분수익률에 오차항 분포로서 정규분포와 t-분포를 이용한 4가지 경우 각각에 대해 계산한 1시차 후 예측 확률과 실제 관측한 수익률에 대한 이항 변수 y_{t+1} 을 나타내고 있다.

표 3.3. 현대차: 1시차 후 예측 확률($p_{t+1}(\hat{\theta})$)과 실제 관측값(y_{t+1})

차분		로그차분		y_{t+1}
정규분포	t-분포	정규분포	t-분포	
0.4543*	0.4593*	0.4630*	0.4636*	0
0.4884	0.4940	0.4851	0.4942	1
0.5432*	0.5371*	0.5449*	0.5342*	1
0.5150	0.5060	0.5112	0.5038	0
0.4686	0.4725	0.4722	0.4766	1
0.5222*	0.5178*	0.5191*	0.5155*	1
0.4809*	0.4727*	0.4833*	0.4731*	0
0.4405	0.4461	0.4479	0.4527	1
0.5280*	0.5285*	0.5260*	0.5255*	1
0.4946	0.4920	0.4937	0.4921	1
0.5225*	0.5189*	0.5233*	0.5169*	1
0.5392*	0.5277*	0.5321*	0.5227*	1
0.4945*	0.4838*	0.4950*	0.4839*	0
0.4521*	0.4473*	0.4580*	0.4541*	0
0.4080	0.4161	0.4189	0.4265	1
0.5152*	0.5146*	0.5173*	0.5124*	1
0.5165*	0.5110*	0.5151*	0.5089*	1
0.5221	0.5215	0.5203	0.5191	0
0.4885	0.4919	0.4854	0.4935	1
0.5213*	0.5163*	0.5213*	0.5135*	1
0.5085	0.4965*	0.5069	0.4945*	0
0.4121	0.4218	0.4218	0.4324	1
0.5866*	0.5780*	0.5815*	0.5696*	1
0.5154*	0.4952	0.5101*	0.4912	1
0.4954	0.4823	0.4972	0.4818	1
0.4837	0.4701	0.4826	0.4731	1
0.4382	0.4275	0.4472	0.4335	1
0.4506*	0.4402*	0.4584*	0.4461*	0
0.3934	0.4007	0.4105	0.4166	1
0.5977	0.5846	0.5931	0.5739	0
0.4365	0.4281	0.4381	0.4355	1
0.5068*	0.5071*	0.5119*	0.5071*	1
0.5005	0.4895*	0.4947*	0.4891*	0
0.4272*	0.4254*	0.4415*	0.4350*	0
0.4836	0.4869	0.4841	0.4888	1
0.5125*	0.5109*	0.5133*	0.5094*	1
0.5101	0.5072	0.5079	0.5057	0
0.4722	0.4794	0.4754	0.4833	1
0.5550*	0.5517*	0.5485*	0.5440*	1
0.5193	0.5065	0.5155	0.5035	0

기호 *가 표시된 곳은 예측 확률이 실제 주가의 상승 및 하락을 올바르게 예측한 경우이다. 표 3.3은 현대차에 대한 결과이며, 표 3.4은 대우증권에 대한 결과이다. 표 3.2, 3.3, 3.4 모두 253개 예측 확률들 중 최근 두 달(40일)의 결과만을 제시하였다. 표들을 전체적으로 보았을 때, 예측 확률은 대부분 0.4-0.6

표 3.4. 대우증권: 1시차 후 예측 확률($p_{t+1}(\hat{\theta})$)과 실제 관측값(y_{t+1})

차분		로그차분		y_{t+1}
정규분포	t-분포	정규분포	t-분포	
0.4774*	0.4921*	0.4891*	0.5073	0
0.4990	0.4941	0.5116*	0.5014*	1
0.4916	0.4921	0.4778	0.4832	1
0.4925*	0.4863*	0.5048	0.5002	0
0.5515	0.5522	0.5356	0.5398	0
0.4890	0.4889	0.5071*	0.4994	1
0.4893	0.4843	0.4759	0.4767	1
0.4986	0.5099*	0.5107*	0.5186*	1
0.5126*	0.5202*	0.5228*	0.5257*	1
0.4865*	0.4806*	0.4943*	0.4886*	0
0.4718*	0.4802*	0.4717*	0.4811*	0
0.4790	0.4903	0.4753	0.4816	1
0.5054*	0.5090*	0.4965	0.5040*	1
0.5392	0.5332	0.5372	0.5346	0
0.5090*	0.5179*	0.5035*	0.5123*	1
0.4711	0.4671	0.4615	0.4537	1
0.4824	0.4978	0.4715	0.4868	1
0.5098*	0.5141*	0.5123*	0.5165*	1
0.5291	0.5211	0.5097	0.5093	0
0.4876*	0.4916*	0.4906*	0.4920*	0
0.5039	0.5052	0.4750*	0.4757*	0
0.4707	0.4809	0.4678	0.4706	1
0.5005*	0.4984	0.4925	0.4937	1
0.5487*	0.5441*	0.5316*	0.5333*	1
0.5540	0.5495	0.5407	0.5379	0
0.5106*	0.5117*	0.5062*	0.5029*	1
0.4641	0.4755	0.4723	0.4760	1
0.4701*	0.4769*	0.4790*	0.4831*	0
0.4707	0.4731	0.4888	0.4905	1
0.4817	0.4798	0.4797	0.4845	1
0.5409*	0.5266*	0.5175*	0.5104*	1
0.5368*	0.5500*	0.5264*	0.5383*	1
0.5164	0.5041	0.5057	0.4909*	0
0.4769	0.4893	0.4788	0.4874	1
0.4684	0.4771	0.4866	0.4892	1
0.5178*	0.5133*	0.4990	0.5014*	1
0.4893	0.4931	0.5124*	0.5141*	1
0.5232*	0.5145*	0.5007*	0.5008*	1
0.4940	0.4992	0.4881	0.4888	1
0.4993*	0.5058	0.4955*	0.5004	0

사이의 값을 갖는다. 이는 오차항의 분포로 가정한 정규분포와 t-분포 모두 0을 기준으로 좌우대칭의 모양을 갖기 때문에 0 근처에서 가장 빈번하게 오차항들이 발생하는 현상의 영향을 받았기 때문이라고 생각된다. 4가지 경우 비슷하게 수익률의 오르고 내림을 예측하는 것을 볼 수 있다. 차이가 나는 곳들을

표 3.5. 1시차 후 예측 확률 결과표

	KOSPI		현대차		대우증권	
	차분	로그차분	차분	로그차분	차분	로그차분
정규분포	134 (0.5296)	138 (0.5455)	131 (0.5178)	132 (0.5217)	128 (0.5059)	129 (0.5099)
t -분포	118 (0.4664)	134 (0.5296)	133 (0.5257)	133 (0.5257)	131 (0.5178)	138 (0.5455)

보면 차분자료가 로그차분자료에 비해 수익률의 증감이 바뀌는 곳에서 비교적 예측의 정확성이 조금 더 높으며 로그차분자료는 차분자료에 비해 수익률의 증감이 지속적으로 이어지는 경우에 조금 더 잘 예측함을 볼 수 있다. 그러나 두 자료가 예측의 정확성에 있어서 뚜렷한 차이를 보이지는 않는다. 표 3.5는 253개(1년)의 자료에 대한 예측 확률의 정확성을 정리한 것이다. 표 안의 숫자는 253개의 예측 확률 중 실제 수익률의 증감을 맞춘 횟수이며 괄호 안의 숫자는 그 비율이다. KOSPI의 경우 로그차분수익률에서 정규분포를 가정한 경우 가장 예측 확률의 정확성이 높으며, 현대차는 차분자료와 로그차분자료 모두에서 t -분포를 가정한 경우가 정규분포를 가정한 경우보다 조금 더 정확성이 높다. 대우증권은 로그차분자료에서 t -분포를 가정한 경우 정확성이 높은 편이다. 대우증권의 경우 다른 두 주식 자료에 비해 변동성이 매우 크기 때문에 꼬리가 더 두꺼운 t -분포를 가정하였을 때 예측 확률의 정확성이 더 높게 나온 것으로 생각된다. 만일 $\{Z_t\}$ 의 모형인 $AR(p)$ -GARCH(1,1)에서 $AR(p)$ 모형이 유의하지 않으면 각 주가 자료는 조건부 이분산 임의보행(random walk)과정을 따르며 이 경우 표 3.5의 예측 확률들은 모두 0.5로 예상될 것이다. 분석 결과 $AR(p)$ -GARCH(1,1) 모형의 설명력인 결정계수는 KOSPI 자료의 경우에 (차분-정규분포: 0.99799, 차분- t 분포: 0.99739, 로그차분-정규분포: 0.99806, 로그차분- t 분포: 0.99693)으로 모두 99% 이상의 설명력을 보여주고 있다. 또한 현대차와 대우증권에 대해서도 계산하였을 때 모두 0.99 정도의 큰 값들이 나왔다. 이처럼 분석모형이 유의하여 자료가 임의보행(random walk)이 아님이 명백함에도 불구하고 표 3.5의 예측 확률들이 0.5를 많이 뛰어넘지 않고 약간만 상회하는 결과를 보이는 점은 특기할 만하다 하겠다.

4. 결론

본 논문에서는 연속형 시계열 변수를 범주형, 특히 간단한 이항(binary) 변수로 변환하여 얻어진 정보들을 이용하여 분석하는 경우를 살펴보았다. Zhen과 Basawa (2009)에서는 연속형 시계열 변수가 $AR(p)$ 를 따른다고 가정하고 분석하였으나 본 논문에서는 일반적으로 금융 시계열 자료가 갖는 특성인 이산형 변동성을 고려하여 $AR(p)$ -GARCH(m, s)을 따른다는 가정하에 모형을 추정하는 방법들을 알아 보았다. 또한 국내 주식 자료에 모형을 적용하여 모수를 추정하고 추정된 모수를 이용하여 1시차 예측 확률을 계산하였다. 관측된 연속형 시계열 자료는 대부분의 시계열 분석에서 원하는 모든 정보를 가지고 있다고 볼 때, 이것을 클리핑을 통해 이항 변수로 바꿈으로써 정보의 손실이 일어나는 것은 사실이다. Zhen과 Basawa (2009)가 수행한 시뮬레이션에서 여러 가지 모수 추정 방법들 중 연속형 자료 $\{Z_t\}$ 만을 이용하여 추정하는 경우가 가장 효율적인 것도 이러한 사실을 말해준다. 그러나 정보의 손실은 일어남에도 불구하고, 연속형 자료의 점 예측치보다 단순한 주가의 등락, 수익률의 증감에 대한 확률이 복잡하고 방대한 자료를 다루어야 하는 실무자에게는 더 직관적이고 편리한 정보가 될 수 있을 것이다. 본 논문에서는 두 가지 범주로만 나누는 경우만을 고려하였으나 조금 더 세분화된 결과를 원하는 경우 클리핑을 통해 여러 개의 범주로 나누는 것도 생각해 볼 수 있다.

참고문헌

- Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Fokianos, L. and Kedem, B. (1998). Prediction and classification of non-stationary categorical time series, *Journal of Multivariate Analysis*, **67**, 277–296.
- Fokianos, L. and Kedem, B. (2003). Regression theory for categorical time series, *Statistical Science*, **18**, 357–376.
- Kaufmann, H. (1987). Regression models for nonstationary categorical time series: Asymptotic estimation theory, *The Annals of Mathematical Statistics*, **1**, 79–98.
- Kedem, B. (1980). *Binary Time Series*, Marcel Dekker, New York.
- Slud, E. and Kedem, B. (1994). Partial likelihood analysis of logistic regression and autoregression, *Statistica Sinica*, **4**, 89–106.
- Wong, W. H. (1986). Theory of partial likelihood, *The Annals of Mathematical Statistics*, **14**, 88–123.
- Zhen, X. and Basawa, I. V. (2009). Estimation for binary models generated by Gaussian autoregressive processes, *Journal of Statistical Computation and Simulation*, DOI: 10.1080/00949650902922711.

An Analysis of Categorical Time Series Driven by Clipping GARCH Processes

M.S. Choi¹ · J.S. Baek² · S.Y. Hwang³

¹Department of Statistics, Sookmyung Women's University

²Methodology Division, Statistical Research Institute, Statistics Korea

³Department of Statistics, Sookmyung Women's University

(Received June 2010; accepted June 2010)

Abstract

This short article is concerned with a categorical time series obtained after clipping a heteroscedastic GARCH process. Estimation methods are discussed for the model parameters appearing both in the original process and in the resulting binary time series from a clipping (*cf.* Zhen and Basawa, 2009). Assuming AR-GARCH model for heteroscedastic time series, three data sets from Korean stock market are analyzed and illustrated with applications to calculating certain probabilities associated with the AR-GARCH process.

Keywords: Clipping, categorical time series, heteroscedastic GARCH.

³Corresponding author: Professor, Department of Statistics, Sookmyung Women's University, 52 Hyochangwon-gil, Yongsan-gu, Seoul 140-742, Korea. E-mail: shwang@sm.ac.kr