The Korean Journal of Applied Statistics (2016) DOI: http://dx.doi.org/10.5351/KJAS.2016.29.1.001
29(1), 1–12

A recent overview on financial and special time series models

S.Y. Hwang[a,1]

[a]Department of Statistics, Sookmyung Women's University

(Received January 18, 2016; Revised January 21, 2016; Accepted January 21, 2016)

Abstract

Contrasted with the standard linear ARMA models, financial time series exhibits non-standard features such as fat-tails, non-normality, volatility clustering and asymmetries which are usually referred to as "stylized facts" in financial time series context (Terasvirta, 2009). We are accordingly led to *ad hoc* models (apart from ARMA) to accommodate stylized facts (Andersen *et al.*, 2009). The paper aims to give a contemporary overview on financial and special time series models based on the recent literature and on the author's publications. Various models are illustrated including asymmetric models, integer valued models, multivariate models and high frequency models. Selected statistical issues on the models are discussed, bringing some perspectives to the future works in this area.

Keywords: financial time series, GARCH type models, stylized facts

1. 금융 및 특수시계열 모형의 필요성

시계열 분석이란 시간의 흐름에 따라 정리 관측된 서로 상관되어 있는 자료 $(X_1, X_2, \ldots, X_n)$를 분석하는 통계 방법론이다. 시계열 자료는 크게 정상시계열과 비정상 시계열로 구분되며 비정상 시계열은 대부분 적절한 변환을 통해 정상 시계열로 바꿀 수 있으므로 본 논문에서는 시계열 모형으로서 정상시계열을 중심으로 논의하도록 한다. 표준적인 시계열 분석 모형은 선형모형인 자기회귀이동평균(ARMA; Box 등, 1994) 모형으로서 모형의 특징은 시계열의 분포가 (다변량) 정규분포이며 따라서 분포가 시간에 대해 가역적인 성질(time reversible in distribution), 즉, $(X_1, X_2, \ldots, X_n)$의 분포가 $(X_n, X_{n-1}, \ldots, X_1)$의 분포와 같다는 점이다. 가장 간단한 ARMA 모형은 가장 널리 쓰이는 AR(1)이다.

$$X_t = \theta_0 + \theta_1 X_{t-1} + e_t, \tag{1.1}$$

여기서 e_t는 평균이 0이고 분산이 σ_e^2인 정규분포를 따르는 iid 확률변수열이다. Tong (1990)은 선형 ARMA 모형의 단점을 지적하고 다양한 비선형 시계열 모형을 제안하였다. 그 중에서 분계점(threshold)을 가진 AR(1) 모형 (threshold AR(1); TAR(1))은 다음과 같다.

$$X_t = \theta_0 + \theta_{11} X_{t-1}^+ + \theta_{12} X_{t-1}^- + e_t, \tag{1.2}$$

This research was supported by Basic Science Research Program through the NRF of Korea funded by the Ministry of Education (2015–057031).

[1]Department of Statistics, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-Gu, Seoul 04310, Korea. E-mail: shwang@sookmyung.ac.kr

여기서 $x^+ = \max(x, 0)$와 $x^- = \max(-x, 0)$는 $x = 0$에서의 분계 함수를 의미한다. 확률계수 AR(1) 모형 (random coefficient AR(1); RCAR(1))은 계수가 확률변수인 모형이다.

$$X_t = \theta_{0t} + \theta_{1t}X_{t-1} + e_t,$$

여기서 $(\theta_{0t}, \theta_{1t})$는 iid 확률벡터이다. 정수 값을 갖는 계수 시계열(count time series)을 위한 AR 모형 (integer valued AR(1); INAR(1)) 모형도 흥미로운 모형일 것이다.

$$X_t = \theta_1 \circ X_{t-1} + e_t, \tag{1.3}$$

여기서 기호 $\circ$는 $\theta_1 \circ X_t = \sum_{i=1}^{X_t} B_i$로서 binomial thinning operator를 나타내며 $\{B_i\}$는 성공확률이 θ_1인 iid 베르누이 확률변수열이다. 오차항 $\{e_t\}$는 평균이 θ_0인 계수(count)형 iid 확률변수(예, 포아송)이다. 표준적인 모형 (1.1)의 비선형으로의 확장 모형은 이외에도 특정자료의 필요에 맞게 연구되고 있으며 Tong (1990) 및 Grunwald 등 (2000)을 참고하기 바란다. 금융시계열은 일반적인 시계열에 비해서 다음과 같은 새로운 네 가지 특징들(stylized facts; cf., Terasvirta, 2009)을 보이는 경향이 있다. (i) 주변분포가 정규분포에 비해 꼬리가 두꺼운 fat tail 또는 leptokurtic 현상을 보인다. 즉, 극단적인 값이 나올 가능성이 정규분포에 비해 높다. (ii) 주변분포가 정규분포가 아닌 non-normal 분포를 따르는 경우가 많다. (iii) 시계열 도표를 살펴보면 변동성으로 측정되는 진폭이 큰 시간구간과 작은 시간구간으로 확연히 분리되는 변동성 집중, 즉, volatility clustering 현상이 보인다. (iv) 변동성이 과거정보에 대해 비대칭적인 모습(asymmetry in volatility: leverage effect)을 보이는 경우가 많다. 시계열 모형의 조건부 평균(μ_t)과 조건부 분산(h_t)

$$\mu_t = E(X_t|F_{t-1}), \qquad h_t = \mathrm{Var}(X_t|F_{t-1}) \tag{1.4}$$

은 시계열을 특징짓는 가장 중요한 두 가지 요소이다. 이 논문에서 F_{t-1}은 $t-1$시점까지 포함된 정보 집합(σ-field)을 나타낸다. 따라서 식 (1.4)에서 μ_t와 h_t는 F_{t-1} 정보 집합에 속하는 확률변수이다. 조건부 평균은 시계열의 자기 상관을 결정하는 요소이며 따라서 자기상관이 없는 경우에는 조건부 평균을 상수 $\mu_t = \mu$로 선택한다. 조건부 분산 h_t는 변동성(volatility)이며 대부분의 시계열 모형에서는 h_t가 상수인 모형이지만 금융시계열에서는 통상적으로 변동성을 확률과정 정상시계열로 간주하며 조건부 이분산(conditionally heteroscedastic) 모형 또는 시간에 따라 변하는 모형(time varying conditional variance model)이라 부른다. 조건부 이분산 모형은 Engle (1982)에 의해 ARCH(autoregressive conditional heteroscedastic) 모형으로 공식적으로 연구되기 시작하였으며 이 후 약 30년간 계량경제학, 재무학 및 시계열분석 등의 분야에서 방대한 이론 및 응용연구가 누적되어 왔다. 본 논문에서는 통계학(시계열 분석)적인 관점에서 이들 금융시계열 및 특수모형들에 대해 알아보고자 한다.

2. 금융 및 특수시계열 모형의 소개

시점 t에서 특정 자산의 가격을 P_t라 하자. 로그수익률(백분율) X_t는 다음과 같이 정의된다.

$$X_t = 100(\log P_t - \log P_{t-1}), \tag{2.1}$$

여기서 로그는 자연로그이며 X_t는 (백분율)수익률인 $\widetilde{X_t} = 100((P_t - P_{t-1})/P_{t-1})$의 근사값이며 연속복리이자(continuously compounded interest)로서 다기간(multi-period) 수익율 계산이 간편하고 합의 형태이므로 대표본 통계이론 적용에 유리하다 (Tasy, 2010, Ch.1). 금융시계열의 경우, $\{X_t\}$는 무상관, 즉 자기상관이 없는 경우가 많으며, 대신에 $\{X_t^2\}$ 또는 $\{|X_t|\}$는 어느 정도의 자기상관을 보이는 경

우가 대부분이다. 잘 알려진 분석 경험(empirical)에 의하면 $\{|X_t|^m\}$ 시계열 중에서 $m = 1$인 경우가 가장 큰 자기상관을 제공하는 것으로 알려져 있으며 이 현상을 금융시계열의 테일러효과(Taylor effect; Terasvirta (2009), Chung과 Hwang (2016a))라 부른다. 즉, $\{X_t\}$는 무상관이지만 독립은 아닌 시계열이며 따라서 이런 성질을 가진 시계열 모형의 개발 필요성이 대두하게 되었다. 전통적인 ARMA 모형은 평균수준인 μ_t의 분석 및 예측에 관심을 두고 있으나 금융시계열에서는 μ_t보다는 수익률의 진폭, 즉, 조건부 분산인 h_t를 파악하기 위한 모형 개발에 주력하고 있다.

조건부 이분산 GARCH 모형: Bollerslev (1986)에 의해 고안된 ARCH 모형의 일반화 형태인 GARCH 모형 중에서 다음과 같은 일차모형인 GARCH$(1,1)$ 모형 $\{\epsilon_t\}$는 다음과 같다.

$$\epsilon_t = \sqrt{h_t} e_t, \quad h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{2.2}$$

여기서 e_t는 평균이 0이고 분산이 1로 표준화 된 iid 확률변수열이며 h_t는 조건부 분산(변동성)이다. 변동성 $h_t = \mathrm{Var}(\epsilon_t|F_{t-1})$는 F_{t-1}에 속하는 확률변수이다. 일차모형 (2.2)는 고차모형인 (p,q) 모형으로 확장할 수 있으며, 대부분의 경우 간단한 GARCH$(1,1)$으로도 좋은 분석을 수행할 수 있다는 점을 Hansen과 Lunde (2005)는 주장한 바 있다. 로그수익률 $\{X_t\}$에 자기 상관이 없는 경우에는 두 가지 모형화가 가능하다. (i) $X_t = \mu + \epsilon_t$: 여기서 μ는 로그수익률의 기댓값이며 (ii) 표본기간 평균 로그수익률이 $\overline{X}$일 때 평균수정한 $X_t - \overline{X}$를 ϵ_t로 간주하여 $X_t - \overline{X} = \epsilon_t$에 모형 (2.2)를 적용 할 수도 있다. 수익률 $\{X_t\}$에 자기상관이 존재하는 경우에는 $X_t = \mu_t + \epsilon_t$로 모형화 한다. 예를 들어 수익률의 자기상관계수가 지수적으로 감소하는 경우 다음과 같은 AR(1)-GARCH$(1,1)$ 모형을 사용해도 좋을 것이다.

$$X_t = \theta_0 + \theta_1 X_{t-1} + \epsilon_t : \quad \epsilon_t = \sqrt{h_t} e_t,\ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}. \tag{2.3}$$

같은 양이면 수익률의 상승보다 수익률의 하락이 변동성에 미치는 영향이 큼을 의미하는 레버리지(leverage, 비대칭)효과를 반영하는 모형은 분계점 GARCH 모형으로서 (Glosten 등, 1993) 일차모형인 TGARCH$(1,1)$ 모형은 다음과 같다.

$$h_t = \alpha_0 + \alpha_{11} \left(\epsilon_{t-1}^{+}\right)^2 + \alpha_{12} \left(\epsilon_{t-1}^{-}\right)^2 + \beta_1 h_{t-1}, \tag{2.4}$$

여기서 $\alpha_{11} \le \alpha_{12}$이다. 변동성의 레버리지 효과를 위한 모형으로는 Exponential GARCH(EGARCH) 모형, 멱변환 분계점 모형(APGARCH) 등 다양한 변형 모형들이 있으며 Hansen과 Lunde (2005), Andersen 등 (2009) 그리고 Francq와 Zakoian (2013)을 참고하기 바란다. 변동성 h_t는 관측불가능하며 이를 GARCH 형태의 수식을 이용하여 추정하는 모형 기반 방법과는 달리 수익률 제곱 자료를 평활하여 변동성을 추정하는 "비모수적" 방법도 가능하다. 수익률의 제곱 자료에 적절한 가중치를 부여하여 평활하여 변동성 추정하는 방법으로 지수가중이동평균(exponentially weighted moving average; EWMA)과 역사적 변동성(historical volatility) 방법이 있으며 (Yoon과 Hwang, 2015b), 제곱 자료의 시계열적 상관 구조를 고려한 종속구조 KDE(kernel density estimation) 방법을 이용한 비모수적 방법 연구도 활발하다 (Andersen 등, 2009, 3장).

Integer valued GARCH(INGARCH) 모형: 시계열이 0을 포함한 계수(count)값을 갖는 경우에 GARCH 모형 (2.2)는 약간의 수정이 필요하다. 예를 들어 수익률의 범위에 따라 $0, 1, 2, \ldots$ 등의 값으로 수익률을 재 표현하는 경우나, 수익률이 상승하는 경우는 1 하락하는 경우는 0으로 표현된 금융시계열이 이 경우에 포함된다. 의학시계열 데이터의 경우 특정 질병의 월별 발생건수 시계열이 계수시계

열의 범주에 속한다. 계수형 시계열 자료를 위한 Ferland 등 (2006)이 제안한 integer-valued GARCH, 즉, INGARCH(p,q) 모형은 $t-1$시점까지의 정보가 주어진 조건하에서 계수 시계열 X_t의 조건부 분포가 평균과 분산이 λ_t인 포아송 분포를 따르며 변동성 λ_t가 다음의 수식을 따른다.

$$\lambda_t = \alpha_0 \; + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}. \tag{2.5}$$

이 모형은 조건부 분포로 포아송분포를 가정하고 있으므로 평균과 분산이 동일하여 과산포(over-dispersion) 문제를 설명하지 못하는 단점이 있으며 이를 보완하기 위해 Zhu (2011)는 조건부 분포로 음이항(negative binomial) 분포를 이용하였다. 최근 계수시계열의 중요한 현상인 영과잉(zero inflation; ZI)을 고려한 영과잉-INGARCH 연구도 활발히 연구되고 있다 (Zhu, 2012; Yoon과 Hwang, 2015a).

Multivariate GARCH(MGARCH) 모형: 다변량 시계열은 전통적으로 벡터-ARMA, 즉, VARMA 모형을 통한 분석이 주를 이루고 있다. VARMA 분석은 변동성은 상수로 간주하고 (조건부) 평균 벡터의 분석에 초점을 맞추고 있다. VARMA 분석의 공적분(cointegration)과 그랜져-인과성(Granger causality)은 계량경제학의 중요한 토픽이다. 다변량 시계열을 구성하는 각각의 일변량 시계열들은 단위근을 가진 비정상 시계열이지만 이들 단위근 시계열들의 선형결합이 정상시계열이 되는 경우 다변량 벡터시계열은 "공적분 되어 있다(cointegrated)"라고 한다. 정상시계열을 이루는 선형결합을 장기균형방정식이라 부른다. 공적분에 대한 자세한 내용은 Wei (2006, Ch.17)을 참고하기 바란다. 특정시계열(x_t)가 다른 시계열(y_t)를 그랜져 인과한다는 의미는 y_t를 예측할 때 x의 과거시차 변수 $x_{t-1}, x_{t-2}, \ldots$을 사용하는 것이 사용하지 않는 것 보다 (통계적으로 유의하게) 더 잘하는 경우를 의미 한다 (Li, 2004, p.27). 실용적이면서도 간편한 그랜져 인과성 검정은 x_t와 y_t를 벡터 시계열로 묶은 후 벡터-AR, 즉, VAR 분석으로부터 Wald-검정을 수행하는 것이다 (Kim과 Lee, 2005).

다변량 변동성을 분석하는 모형인 MGARCH 모형의 장점은 다변량 수익률 각각의 변동성과 더불어 변동성들 간의 동적인 관계(조건부 상관계수)를 동시에 모델링 할 수 있다는 데 있으며 이를 이용하여 VaR(Value at Risk)를 계산함으로 해서 재정분야의 포트폴리오 및 자산 배분(asset allocation)에 중요한 역할을 한다. 백색잡음벡터로서의 k차원 벡터수익률 $\{X_t\}$ 를 다음과 같이 모형화 한다: $X_t = \mu + \epsilon_t$, 여기서 μ는 평균수익률 벡터이며, $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{kt})$는 MGARCH 모형으로서 다음과 같다.

$$\epsilon_t = H_t^{\frac{1}{2}} e_t, \tag{2.6}$$

여기서 ϵ_t의 조건부 분산-공분산행렬인 $H_t = \mathrm{Cov}(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \ldots)$이고 $H_t^{1/2}$는 $k \times k$ 양정치행렬이다. 또한, e_t는 $k \times 1$벡터로 다음의 조건을 만족한다: $E(e_t) = 0$, $\mathrm{Var}(e_t) = I_k$. 조건부 분산-공분산행렬 H_t를 모형화 하기 위한 방법론으로 EWMA(Exponentially Weighted Moving-average) 모형, DVEC(Diagonal VEC) 모형 및 BEKK 모형 등이 있다 (Tsay, 2010, Ch.10; Song 등, 2008). MGARCH 모형의 단점은 모수 개수가 차원 k가 증가하면서 너무 많아진다는 점이다. 추정해야할 모수 개수를 확연히 줄이기 위해 조건부 상관계수가 시점에 상관없이 일정하다고 가정한 모형, 즉, CCC(Constant Conditional Correlation) 모형이 개발되었다. 모수절약형 CCC 모형은 다음과 같다.

$$\begin{aligned} H_t &= D_t R D_t = \left(\left(\rho_{ij} \sqrt{h_{iit} h_{jjt}} \right) \right), \\ D_t &= \mathrm{diag}\left(h_{11t}^{\frac{1}{2}} \cdots h_{kkt}^{\frac{1}{2}} \right), \qquad R = ((\rho_{ij})), \end{aligned} \tag{2.7}$$

여기서 h_{iit} $(i = 1, \ldots, k)$는 각각 단변량 GARCH로 모형화되고, R은 상관계수 행렬이다. 예로서, GARCH$(1,1)$을 따르는 CCC 모형의 D_t행렬 각각의 대각원소들은 다음과 같다.

$$h_{iit} = \omega_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i h_{ii,t-1}, \quad i = 1, \ldots, k.$$

CCC 모형은 조건부 상관계수 행렬이 시간에 따라 변하지 않는다는 가정을 통해 추정의 어려움을 극복하였지만 시간에 따른 변화라는 다변량 시계열의 속성을 고려하지 못한 단점이 있다. 시간에 의존하는 조건부 상관계수 행렬을 고려한 CCC 모형의 일반화된 형태가 연구되었는데 이 모형이 dynamic conditional correlation(DCC) 모형이다. DCC에서는 CCC와 달리 조건부 상관계수도 GARCH 모형화 하여 시간변화(time-varying) 요소를 추가한 모형이라 할 수 있다 (Tsay, 2010, p.531; Choi 등, 2009). 차원 수 증가에 따른 추정 모수 수가 급격히 많아진다는 문제점은 차원축소(dimension reduction)를 통해서도 극복할 수 있다. 전통적인 다변량 인자분석 개념을 이용해 다변량 변동성 기저에 있는 인자를 파악하고 주요 소수 인자만을 선택해서 차원 축소를 수행한다. 자산수익률 다변량 벡터의 차원축소 방법으로는 macroeconomic factor model, BARRA factor model로 알려져 있는 fundamental factor model 및 statistical factor models의 세 가지 유형이 있다. 자세한 내용은 Connor (1995), Tsay (2010, 9장), Song 등 (2008)을 참고하기 바란다.

High Frequency(고빈도)자료 기반 실현변동성(realized volatility; RV): 일반적으로 금융시계열의 변동성은 일별(daily) 변동성을 의미하며 일별종가자료로부터 계산하고 있다. 하지만, 빅데이터와 연관된 실시간 자료로 부터의 고빈도 자료(high-frequency data)를 이용해서 일별 변동성을 계산하는 방법에 대한 연구도 활발하게 진행되고 있다. 대용량 데이터 처리 능력의 향상 등으로 인하여 변동성 분석에 있어 일별 주가지수 종가 자료 뿐 아니라 1분 빈도의 자료 등과 같은 고빈도 자료를 사용한 분석도 가능해졌고 고빈도 자료를 사용하여 일별 변동성을 계산하는 방법이 연구되고 있다. Andersen과 Bollerslev (1997)는 5분 단위의 고빈도 자료를 이용하여 실현변동성을 추정하는 방법을 소개한 바 있다. 고빈도 자료를 이용해 계산한 일별 변동성을 실현변동성(realized volatility)이라 하고 RV로 나타낸다. 여기서 "실현"이란 원래 관측 불가능한 변동성을 고빈도 자료로부터 "관측(실현)"한 것으로 간주할 수 있다는 의미이다. 일별 종가의 로그 수익률을 이용하여 변동성을 추정하는 기존의 방법과는 달리 하루 동안 매 분 단위로 관측된 고빈도 자료를 활용하여 변동성을 추정하는 방법을 생각해 보자 (Yoon과 Hwang, 2015b; Tsay, 2010, Ch.5). 기호 r_t를 시점 t(t일)의 일간 로그 수익률이라고 하자. 모형에 의해 변동성을 추정하는 경우에는 이 일간 로그 수익률(r_t)을 이용한다. 하지만 고빈도 자료에 기반하여 추정하는 경우에는 일중 로그 수익률(intra-daily log return)이라는 것을 고려한다. 일중 로그 수익률은 t일 중 일정한 간격으로 n개가 조사되었다고 가정하여 $\{r_{t,i}\}$, $i = 1, 2, \ldots, n$로 나타내며 t일의 i번째 관측시점의 로그 수익률을 의미한다. t일의 일간 로그 수익률 r_t는 n개의 일중 로그 수익률의 합 $r_t = \sum_{i=1}^n r_{t,i}$으로 나타낼 수 있다(예를 들어 국내 주식시장의 경우 1분 단위로 조사된 고빈도 자료는 $n = 360$이다). 일간 로그 수익률의 조건부 분산은 다음과 같이 계산한다.

$$\mathrm{Var}(r_t | F_{t-1}) = \sum_{i=1}^n \mathrm{Var}(r_{t,i} | F_{t-1}) + 2 \sum_{i<j}^n \mathrm{Cov}[(r_{t,i}, r_{t,j} | F_{t-1}), \tag{2.8}$$

여기서 F_{t-1}은 $(t-1)$일까지 주어진 정보를 나타낸다. 일간 로그 수익률의 조건부 분산은 n개의 일중 로그 수익률의 분산과 공분산의 합으로 구할 수 있다. 위의 조건부 분산은 일중 로그 수익률 $\{r_{t,i}\}$이 F_{t-1}에 조건부적 백색잡음과정이라 가정하면 다음과 같이 구할 수 있다.

$$\mathrm{Var}(r_i | F_{t-1}) = n \mathrm{Var}(r_{t,1}) = \frac{n}{n-1} \sum_{i=1}^n (r_{t,i} - \overline{r_t})^2,$$

여기서 $\overline{r_t}$를 0으로 간주하면 다음과 같이 정의된 일간 로그 수익률 r_t의 실현변동성(RV_t)은 타당성을 부여 받는다.

$$\mathrm{RV}_t = \sum_{i=1}^{n} r_{t,i}^2. \tag{2.9}$$

즉, 실현변동성 RV_t는 t일의 일중 로그 수익률의 제곱합으로 정의된다. 일중 로그 수익률 $r_{t,i}$가 독립임을 가정하지 않고 종속구조를 도입할 수도 있다. Tsay (2010, p.160)는 n개의 $\{r_{t,i}\}$, $i = 1, 2, \ldots, n$가 일차종속구조인 MA(1) 모형을 따르는 경우를 고려하여 식 (2.8)에 변화를 주었다. 실현변동성 RV_t의 다양한 변형에 대해서는 Xiao (2013)을 참고하기 바란다. 고빈도 금융자료에서 수익률이 아닌 거래사이 시간간격(time between transactions)을 분석하는 모형도 GARCH 모형의 변형으로 연구되어 왔다. Engle과 Russell (1998)은 고빈도 자료에서 거래들 사이의 시간간격인 듀레이션에 대한 분석기법으로 ACD(autoregressive conditional duration) 모형을 소개하였다. ACD 모형은 GARCH 모형과 비슷한 수리 구조를 가지고 있다. 기호 t_i는 i번째 "거래"가 일어나는 시간이고, "거래"들 사이의 시간 간격 $X_i = t_i - t_{i-1}$을 듀레이션(duration)이라고 한다. ACD 모형에서 조건부 기대 듀레이션을 ψ_i로 표현하자. 즉, $\psi_i = E(X_i|F_{i-1})$이다. 듀레이션 X_i는 조건부 기대 듀레이션 ψ_i와 승법 형태의 오차 ϵ_i로 구성된다고 가정 한다 (Chung과 Hwang, 2016b).

$$X_i = \psi_i \epsilon_i, \quad \epsilon_i \sim \mathrm{iid}\,(1, \sigma_\epsilon^2)$$

듀레이션은 양수이며, ϵ_i도 양의 값만을 가지며 평균은 1이고 분산이 σ_ϵ^2인 확률분포를 가정하는데 보통 지수분포와 와이블 분포를 이용한다. 간단한 모형인 ACD$(1,1)$은 다음과 같다.

$$X_i = \psi_i \epsilon_i, \quad \psi_i = \omega + \alpha_1 X_{i-1} + \beta_1 \psi_{i-1}.$$

ACD 모형에 대한 자세한 내용은 Engle과 Russell (1998), Tsay (2010, p.255), Chung과 Hwang (2016b)을 참고하기 바란다.

3. 통계적 이슈

금융 시계열 자료의 실제 우도가 관심모수 $\phi(p \times 1$벡터)의 함수(indexed)일 때 ϕ에 대한 통계적 추론과 연관된 주요 이슈들에 대해 알아보도록 한다.

최우추정(ML): 자료 $(X_1, X_2, \ldots, X_n)$의 우도가 알려진 경우에 모수 ϕ에 관해서 로그우도함수를 최대화 시키는 방법을 최우추정(ML)이라 한다. 최우추정량은 적절한 조건하에서 (Hwang과 Basawa, 2014) 마팅게일 추정함수 형태의 추정량 중에서 가장 "작은" 극한 분산을 제공하는 최적(asymptotic optimum) 추정량이지만 우도를 알고 있다는 강한 가정이 필요하다는 단점이 있다. 또한 대부분의 경우 시계열의 초기값을 상수로 $X_0 = x_0$ 로 하는 조건부 우도함수이므로 정확하게는 조건부 최우추정량이다. 정상 시계열인 경우 초기값의 영향력은 표본크기 n이 커짐에 따라 없어지므로 문제가 되지 않지만 비정상 또는 비에르고딕 시계열의 경우는 초기값의 영향력이 유지되므로 초기값 문제에 주의를 기울여야 한다 (Hwang 등, 2014a).

준-최우추정(Quasi ML; QML): 오차항의 분포가 정규분포라 가정하고 자료의 우도를 구해서 최우추정하는 방법이다. 예를 들어, 식 (2.3)의 AR(1)-GARCH$(1,1)$ 모형에서 e_t의 분포를 $N(0,1)$으로 구한 우도를 최대화 시켜 얻은 추정량이 QML이다. 오차가 정규분포가 아닌 경우에 최대화 목적

함수가 실제 우도가 아니므로 준-최우추정이라 부른다. 따라서 컴퓨터 팩키지를 통해 얻는 것은 대부분 QML이라 할 수 있다. 금융시계열 GARCH 분야에서 QML이 유용한 이유는 (적절한 조건하에서) 실제 우도가 무엇이든 관계없이 QML은 모수 ϕ의 일치추정량이기 때문이다(이를 universal consistency라 한다 (Francq와 Zakoian, 2013; Chung과 Hwang, 2016a).

유사-최우추정(Pseudo-ML; PML): 금융시계열의 특징인 stylized fact에 따르면 가정된 도구우도(instrumental likelihood)를 만들 때 정규분포가 아닌 꼬리가 두꺼운 분포나 혹은 비대칭 분포를 이용하는 것이 타당해 보인다. 예를 들어 AR(1)-GARCH(1, 1)에서 도구우도를 만들기 위해 e_t의 분포를 자유도가 ν인 (분산이 1로) 표준화된 t-분포 또는 표준화된 일반화 오차분포(GED)를 사용할 수 있다 (pdf 수식은 Tsay (2010, p.121)를 참고하기 바란다). 비대칭 오차분포를 사용하는 경우도 있다. Fernandez와 Steel (1998)은 기존의 대칭분포를 (꼬리를 두껍게 하면서) 비대칭 화하는 방법을 고안하였다. 알려진 대칭분포 $f(e_t)$로부터 다음과 같은 비대칭 분포를 만들 수 있다.

$$f_s(e_t) = \frac{2}{\xi + \xi^{-1}} \left[f\left(\frac{e_t}{\xi}\right) I(e_t \geq 0) + f(\xi e_t) I(e_t < 0) \right], \tag{3.1}$$

여기서 $\xi > 0$는 비대칭 모수로서 1에서 멀어질수록 식 (3.1)의 비대칭 정도는 심화된다. 따라서 식 (3.1)을 이용하여 비대칭 표준정규분포 혹은 비대칭 표준화된 t-분포를 e_t의 분포로 사용할 수 있다. 도구우도를 이용한 PML은 적절한 조건하에서 모수 ϕ의 일치 추정량 및 극한정규분포 추정량을 제공 한다 (Straumann, 2005).

준-우도(Quasi likelihood; QL): 금융시계열 모형 $\{X_t\}$가 식 (1.4)에서 정의된 두 개의 조건부 적률인 조건부 평균(μ_t)과 조건부 분산(h_t)으로만 정의되는 준모수적(semi-parametric)경우를 고려하자. 모수 ϕ는 $\mu_t(\phi)$와 $h_t(\phi)$에 동시에 나타난다. 다음과 같은 이차원 벡터로부터

$$g_t(\phi) = \begin{pmatrix} X_t - \mu_t(\phi) \\ (X_t - \mu_t(\phi))^2 - h_t(\phi) \end{pmatrix}. \tag{3.2}$$

준-우도 QL은 다음과 같이 정의된다.

$$Q_n(\phi) = \sum_{t=1}^{n} \left(\frac{\partial \mu_t(\phi)}{\partial \phi}, \frac{\partial h_t(\phi)}{\partial \phi} \right) V_t^{-1}(\phi) g_t(\phi), \tag{3.3}$$

여기서 $V_t(\phi)$는 $g_t(\phi)$의 조건부 공분산 행렬이며

$$V_t(\phi) = E\left[g_t(\theta) g_t^T(\theta) | F_{t-1} \right] = \begin{pmatrix} h_t(\phi) & \mu_{3t}(\phi) \\ \mu_{3t}(\phi) & \mu_{4t}(\phi) - h_t^2(\phi) \end{pmatrix}, \tag{3.4}$$

여기서 $\mu_{3t}(\phi)$와 $\mu_{4t}(\phi)$는 3차 및 4차 조건부 적률, 즉, $\mu_{3t}(\phi) = E[(X_t - \mu_t(\theta))^3 | F_{t-1}]$이고 $\mu_{4t}(\phi) = E[(X_t - \mu_t(\theta))^4 | F_{t-1}]$이다. 준-우도의 장점은 금융시계열이 점화수식으로 정의되지 않고, 처음의 두 개의 조건부 적률로만 정의되는 큰 모형 족에 속하는 다양한 모형들 모두에 적용할 수 있는 광범위성에 있다. 준-우도 (3.3)을 구성할 때 (3.4)를 통해 실제 우도로부터 잉여모수(nuisance parameter) η가 생겨 이를 추정해야 하는 번거로운 단점이 있으며 일반적인 준-우도에 대해서 Heyde (1997)를, GARCH 분야 준-우도는 Hwang 등 (2014b)를 참고하기 바란다.

고담베 최적 추정함수(Godambe optimum estimating function): 관심모수 $\phi(p \times 1$벡터)의 추론을 위한 추정함수(estimating function) $U_n(\phi)$는 자료와 ϕ의 함수로서 $p \times 1$벡터이다. 추정방정식

$U_n(\phi) = 0$는 ϕ의 추정량을 제공한다. 따라서 추정함수의 선택은 ϕ의 추론에 매우 중요하며 가능한 최적(optimum) 추정함수를 선택하고자 노력해야 한다. Godambe (1985)는 확률과정에서 "선형" 추정함수 집합을 정의하고 이 집합에서 최적인 고담베 최적 추정함수 $U_n^O(\phi)$를 구하는 공식을 제시하였다. 금융시계열 분야에서도 이 분야에 대한 많은 연구가 이루어졌으며 GARCH 분야의 최근 연구인 Chung과 Hwang (2016a)을 참고하기 바란다. 계량경제학의 일반화 적률추정(Generalized Method of Moment; GMM)은 도구변수(instrumental variable)를 이용한 고담베 최적 추정함수의 일종이라 할 수 있다 (Lai와 Xing, 2008, 9장).

준-모수(semi-parametric) GARCH: 식 (2.2)에서 e_t의 분포가 알려져 있지 않고 변동성 함수 h_t 형태는 모수 ϕ를 포함하면서 지정되어 있는 경우에 e_t의 밀도함수를 비모수방법론(예, KDE)으로 추정해서 분석하고 이를 이용해서 ϕ를 모수적으로 추정하는 방식을 모수와 비모수의 중간 형태라 하여 준-모수(semi-parametric) GARCH 분석이라 한다. 이 경우 모수 ϕ의 추정은 적응추정(adaptive estimation)과정을 거칠 수 있다. 모수 ϕ의 추정에만 관심이 있는 경우에는 준-우도(Quasi likelihood; QL) 추정을 이용할 수 있으며 이 때 잉여모수(nuisance parameter)의 처리에 대한 주의를 기울여야 한다. 비모수(non-paramtric) GARCH 분석은 변동성 함수 h_t를 비모수 함수 추정하는 것을 의미 한다 (Linton, 2009).

Stochastic Volatility(SV) 모형: 변동성 h_t는 $t-1$시점까지의 정보(F_{t-1})에 의존하는 함수이다. 여기에 랜덤성이 더해져서 확률적으로 복잡해진 모형이 Stochastic Volatility(SV) 모형이다. 비대칭 변동성 모형인 EGARCH(exponential GARCH)에 stochastic component가 더해지면 SV 모형을 얻을 수 있다. 다음의 간단한 SV 모형에서

$$\epsilon_t = \sqrt{h_t} e_t, \quad \ln h_t - \beta_1 \ln h_{t-1} = \alpha_0 + \nu_t, \tag{3.5}$$

여기서 $\{\nu_t\}$는 iid $N(0, \sigma_\nu^2)$이며 $\{e_t\}$와는 독립인 새로운 stochastic component이다. SV 모형의 추정은 보통 MCMC 방법을 통해 수행 된다 (Taylor, 1994).

Value at Risk(VaR): 자산(포트폴리오) 위험관리에 유용한 개념인 VaR은 '주어진 신뢰수준(confidence level)하에서 목표기간(target horizon)동안 정상적인 시장(normal market)을 전제로 할 때 발생 가능한 최대 손실'로 정의된다 (Jorion, 1997; Tsay, 2010, Ch.7; Choi 등, 2009). l기간 동안의 자산(포트폴리오)의 가치변동을 $\triangle V(l)$라고 할 때 신뢰수준 $100 \times (1-\alpha)\%$에서의 VaR를 다음과 같이 정의한다.

$$\Pr[\triangle V(l) > \mathrm{VaR}] = 1 - \alpha, \tag{3.6}$$

VaR 계산을 위한 프로그램은 J.P. Morgan사에서 개발한 RiskMetrics가 있으며 정확한 VaR를 계산하기 위해서는 포트폴리오를 구성하는 자산들의 MGARCH 분석이 성공적으로 이루어져야 하므로 MGARCH에 대한 연구가 개선됨에 따라 VaR 분야도 같이 발전해 나가고 있다. 자세한 내용은 Tsay (2010)의 7장(Extreme Values, Quantiles and Value at Risk)을 참고하기 바란다.

News Impact Curve(NIC): Engle과 Ng (1993)은 시점 t에서의 변동성 h_t와 가장 최근의 수익률 ϵ_{t-1}과의 함수관계를 News Impact Curve(NIC)라 명명하고 변동성의 비대칭성을 그래프를 통해 알아보았다. NIC에서는 $t-2$까지의 정보 F_{t-2}는 상수로 간주하여 기댓값으로 대치한다. 다양한 비대칭

NIC 수식은 Zivot (2009)에 수록되어 있다. Lee 등 (2013)은 다양한 변동성 모형들의 일차원 NIC를 이차원 NIC로 확장하고, 주성분(principal component)을 통한 NIC를 제안한 바 있다. 예를 들어, 식 (2.2) GARCH$(1,1)$의 이차원 NIC는 다음과 같다.

$$h_t = (1+\beta_1)\alpha_0 + \alpha_1\epsilon_{t-1}^2 + \alpha_1\beta_1\epsilon_{t-2}^2 + \frac{\alpha_0\beta_1^2}{1-\alpha_1-\beta_1}. \tag{3.7}$$

비모수적 변동성 함수추정을 통한 NIC 연구에 대해서는 Linton (2009)을 참고하기 바란다.

지속 및 불안정 변동성(persistent and unstable volatility): 정상 ARMA 모형에서 l-시차 후 예측 값은 $l \to \infty$에 따라 모형의 (무조건부) 평균인 μ로 접근한다. 이를 정상 ARMA 모형에서 평균으로의 회귀(mean reversion)라 한다. 정상 GARCH에서도 l-시차 후 변동성 예측 값은 $l \to \infty$에 따라 (무조건부) 분산인 $E(\epsilon_t^2)$로 접근한다. 이는 변동성 평균으로의 회귀성질이라 할 수 있으나 그러하지 않은 경우도 발생한다. 식 (2.2)의 GARCH$(1,1)$의 경우에 (시점 n에서의) l-시차 후 변동성 예측 값 $h_n(l) = \mathrm{Var}(\epsilon_{n+l}|F_n)$은 다음과 같은 점화식을 따른다 (Tsay, 2010, p.133).

$$h_n(l) = \alpha_0 + (\alpha_1+\beta_1)h_n(l-1), \quad l \geq 2. \tag{3.8}$$

따라서 $\alpha_1+\beta_1 < 1$이면 변동성 평균으로의 회귀에 따라 $h_n(l) \to \alpha_0/(1-\alpha_1-\beta_1)$이 성립하지만 $\alpha_1+\beta_1 = 1$일 때는 $h_n(1) = (l-1)\alpha_0 + h_n(1)$이 되어 $h_n(1) = h_{n+1}$이 미래의 변동성에 지속적으로 남아있게 되며 이를 GARCH$(1,1)$의 지속(persistent) 변동성이라 부른다. 특별히 $\alpha_1+\beta_1 = 1$인, 즉, 지속변동성을 가진 GARCH 모형을 IGARCH(integrated GARCH) 모형이라 부른다. 또한, $\alpha_1+\beta_1 > 1$인 경우는 식 (3.8)에서 보듯이 미래 변동성이 폭발적으로 증가한다. 따라서 $\alpha_1+\beta_1 > 1$ 인 경우를 불안정 변동성이라 부른다. 비대칭 모형인 분계점-GARCH 모형에서의 지속 및 불안정 변동성 연구는 Park 등 (2009)와 Hwang 등 (2010)을 참고하기 바란다.

4. 결언

본 연구에서는 금융시계열 및 특수 모형에 관한 개요 및 조망(overview)을 다루었다. 가능한 최근의 문헌과 저자의 논문을 토대로 내용을 작성하였고 본 논문이 향후 금융시계열 모형 분야 연구를 시작하려는 시계열 통계학자들에게 도움이 되기를 바란다. 여기서 다루지 않은 내용들 중에는 현재 연구가 활발하게 이루어지고 있는 분야인 계산 중심(computational) GARCH, 장기억(long-memory) GARCH, 비모수(non-parametric) GARCH 및 베이지안(Bayesian) GARCH 분석 등이 있으며 이들 및 미래연구에 대한 참고문헌으로 Lai와 Xing (2008), Andersen 등 (2009) 그리고 Tsay (2010)를 추천한다.

References

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance*, **4**, 115–158.

Andersen, T. G., Davis, R. A., Kreiss, J.-P., and Mikosch, T. (2009). *Handbook of Financial Time Series*, Springer, Berlin.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd Ed., Prentice Hall, New Jersey.

Choi, S. M., Hong, S. Y., Choi, M. S., Park, J. A., Baek, J. S., and Hwang, S. Y. (2009). Analysis of multivariate-GARCH via DCC modeling, *Korean Journal of Applied Statistics*, **22**, 995–1005.

Chung, S. and Hwang, S. Y. (2016a). A profile Godambe information of power transformations for ARCH time series, to appear in *Communications in Statistics-Theory and Methods*.

Chung, S. and Hwang, S. Y. (2016b). Stock return volatility based on intraday high frequency data: double-threshold ACD-GARCH Model, to appear in *Korean Journal of Applied Statistics*.

Connor, G. (1995). Three types of factor models: a comparison of their explanatory power, *Financial Analysis Journal*, **51**, 42–46.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility, *Journal of Finance*, **48**, 1749–1778.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data, *Econometrica*, **66**, 1127–1162.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued GARCH process, *Journal of Time Series Analysis*, **27**, 923–942.

Fernandez, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness, *Journal of the American Statistical Association*, **93**, 359–371.

Francq, C. and Zakoian, J. M. (2013). Optimal predictions of powers of conditionally heteroscedastic processes, *Journal of Royal Statistical Society B*, **75**, 345–367.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance*, **48**, 1779–1801.

Godambe, V. P. (1985). The foundation of finite sample estimation in stochastic processes, *Biometrika*, **72**, 419–428.

Grunwald, G. K., Hyndman, R. J., Tedesco, L., and Tweedie, R. L. (2000). Non-Gaussian conditional linear AR(1) models, *Australian and New Zealand Journal of Statistics*, **42**, 479–495.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH $(1,1)$?, *Journal of Applied Econometrics*, **20**, 873–889.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application*, Springer, New York.

Hwang, S. Y., Baek, J. S., Park, J. A., and Choi, M. S. (2010). Explosive volatilities for threshold-GARCH processes generated by asymmetric innovations, *Statistics & Probability Letters*, **80**, 26–33.

Hwang, S. Y. and Basawa, I. V. (2014). Martingale estimating functions for stochastic processes : A review toward a unifying tool, *Contemporary Developments in Statistical Theory*, edited by Lahiri *et al.*, Springer, Switzerland, 9–28.

Hwang, S. Y., Basawa, I. V., Choi, M. S., and Lee, S. D. (2014a). Non-ergodic martingale estimating functions and related asymptotics, *Statistics*, **48**, 487–507.

Hwang, S. Y., Choi, M. S., and Yeo, I.-K. (2014b). Quasilikelihood and quasi maximum likelihood for GARCH-type processes: Estimating function approach, *Journal of the Korean Statistical Society*, **43**, 631–641.

Jorion, P. (1997). *Value at Risk: The New Benchmark for Controling Market Risk*, McGraw-Hill, Chicago.

Kim, H. and Lee, M. (2005). *Econometric and Financial Time Series*, Kyungmunsa, Seoul.

Lai, T. L. and Xing, H. (2008). *Statistical Models and Methods for Financial Markets*, Springer, New York.

Lee, J. W., Yoon, J. E., and Hwang, S. Y. (2013). A graphical improvement in volatility analysis for financial series, *Korean Journal of Applied Statistics*, **26**, 785–796.

Li, W. K. (2004). *Diagnostic Checks in Time Series*, Chapman & Hall, New York.

Linton, O. B. (2009). Semi-parametric and nonparametric ARCH modeling, in *Handbook of Financial Time Series*, 157–168, Eds., Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T., Springer, Berlin.

Park, J. A., Baek, J. S., and Hwang, S. Y. (2009). Persistent threshold-GARCH processes: model and application, *Statistics & Probability Letters*, **79**, 907–914.

Song, E., Choi, M. S., and Hwang, S. Y. (2008). Volatility analysis for multivariate time series via dimension reduction, *Communications of the Korean Statistical Society*, **15**, 825–835.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*, LNS No. 181, Springer, Berlin.

Taylor, S. J. (1994). Modeling stochastic volatility: a review and comparative study, *Mathematical Finance*,

4, 183–204.

Terasvirta, T. (2009). An introduction to univariate GARCH models, In *Handbook of Financial Time Series*, 17–42, Eds., Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T., Springer, Berlin.

Tong, H. (1990). *Nonlinear Time Series*, Oxford University Press, Oxford.

Tsay, R. S. (2010). *Analysis of Financial Time Series*, Third Ed. Wiley, New York.

Wei, W. W. S. (2006). *Time Series Analysis*, 2nd Ed., Pearson, New York

Xiao, L. (2013). Realized volatility forecasting: empirical evidence from stock market indices and exchange rates, *Applied Financial Economics*, **23**, 57–69.

Yoon, J. E. and Hwang, S. Y. (2015a). Zero-inflated INGARCH using conditional Poisson and negative binomial: data application, *Korean Journal of Applied Statistics*, **28**, 583–592.

Yoon, J. E. and Hwang, S. Y. (2015b). Volatility computations for financial time series: high frequency and hybrid method, *Korean Journal of Applied Statistics*, **28**, 1163–1170.

Zhu, F. (2011). A negative binomial integer-valued GARCH model, *Journal of Time Series Analysis*, **32**, 54–67.

Zhu, F. (2012). Zero-inflated Poisson and negative binomial integer-valued GARCH models, *Journal of Statistical Planning and Inference*, **142**, 826–839.

Zivot, E. (2009). Practical issues in the analysis of univariate GARCH models, In *Handbook of Financial Time Series*, 113–155, Eds., Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T., Springer, Berlin.

금융 및 특수시계열 모형의 조망

황선영[a,1]

[a]숙명여자대학교 통계학과

(2016년 1월 18일 접수, 2016년 1월 21일 수정, 2016년 1월 21일 채택)

요 약

금융시계열은 일반 시계열과는 차별적으로 stylized facts로 불리는 특징을 가지고 있다. 이 특징들은 급첨 성질, 비정규분포, 변동성 집중 및 비대칭성을 포함한다. 이러한 특징들을 설명하기 위해서는 기존의 선형 ARMA 모형에서 벗어난 특수한 모형이 필요하게 되었다. 본 논문은 변동성 모형인 GARCH 형태의 모형을 중심으로 특수 금융시계열 모형들을 소개하고 연관된 통계적 이슈들에 대해 가능한 최근 연구를 중심으로 폭 넓게 조망하고 있다.

주요용어: 금융시계열, GARCH 형태의 모형, 금융시계열의 특징(stylized facts)

본 연구는 한국연구재단의 기초연구과제 사업의 지원을 받았습니다 (과제번호: 2015-057031).
[1](04310) 서울특별시 용산구 청파로47길 100, 숙명여대 통계학과. E-mail: shwang@sookmyung.ac.kr