

Functional ARCH analysis for a choice of time interval in intraday return via multivariate volatility

D. H. Kim^a · J. E. Yoon^{a,1} · S. Y. Hwang^a

^aDepartment of Statistics, Sookmyung Women's University

(Received February 4, 2020; Revised March 11, 2020; Accepted April 1, 2020)

Abstract

We focus on the functional autoregressive conditional heteroscedasticity (fARCH) modelling to analyze intraday volatilities based on high frequency financial time series. Multivariate volatility models are investigated to approximate fARCH(1). A formula of multi-step ahead volatilities for fARCH(1) model is derived. As an application, in implementing fARCH(1), a choice of appropriate time interval for the intraday return is discussed. High frequency KOSPI data analysis is conducted to illustrate the main contributions of the article.

Keywords: fARCH, high frequency time series, multivariate volatility

1. 서론

환율, 주가와 같은 금융 시계열에서는 수익률의 평균 수준보다는 불확실성(uncertainty)을 설명하는 수익률의 변동성(volatility)에 집중하며, 위험(risk)을 측정하는 수단으로 변동성을 사용한다. 금융 시계열 분석에서는 변동성을 직접 관측할 수는 없으나, 일정 기간 큰 값들이 지속된 후 작은 값들이 일정 기간 지속되는 변동성 집중(volatility cluster) 현상이나 가격의 큰 인상과 큰 하락에 대하여 다른 반응을 보이는 비대칭 효과(leverage effect)와 같은 특징을 가지고 있다 (Lee와 Hwang, 2017). 수익률의 변동성은 조건부 분산(conditional variance)을 의미하며 Engle (1982)은 변동성을 추정하기 위하여 조건부 분산이 이전 시차까지의 오차항들에 영향을 받는 자기회귀 조건부 이분산(autoregressive conditional heteroscedasticity; ARCH) 모형을 제안하였고, Bollerslev (1986)는 ARCH 모형을 일반화하여 일반화된 자기회귀 조건부 이분산(generalized ARCH; GARCH) 모형을 제안하였다.

다변량 변동성(multivariate volatility) 분석에서는 변동성 간의 동적인 상관관계(dynamic correlation)가 존재하므로 각각의 수익률을 개별로 모형화하기보다는 상관관계를 고려하여 모형화가 수행되어야 한다. 이러한 다변량 변동성 모형에는 exponential weighted moving average (EWMA) 모형과 단변량 GARCH 모형을 확장시킨 모형인 Baba-Engle-Kraft-Kroner (BEKK) 모형 등이 있다. 이러한 모형에는 모형의 차원이 증가하면 추정할 모수의 수가 급격히 증가하는 단점이 있다. 이를 보완

This research was supported by a grant from the National Research Foundation of Korea (NRF-2018R1A2B2004157).

¹Corresponding author: Sookmyung Women's University, Cheongpa-ro, 47-gil 100, Yongsan-gu, Seoul 04310, Korea. E-mail: yoenje@sookmyung.ac.kr

하는 모형으로는 조건부 상관계수를 시간에 관계없이 일정한 상수로 가정하여 모수를 줄인 constant conditional correlation (CCC) 모형과 조건부 상관계수가 시간에 따라 변하는 dynamic conditional correlation (DCC) 모형 등이 있다 (Tsay, 2010, Chapter 10). 빅데이터와 연관하여 데이터의 실시간 수집 및 전송 기법의 발달로 하루에도 수많은 시점에서 수많은 빈도의 금융 시계열 데이터를 얻을 수 있게 되었다. 이를 고빈도 시계열 자료(high-frequency time series data)라 부른다. 고빈도 일중-관측값(intraday observation)을 이용한 변동성 추정에 대한 연구가 최근 활발히 이루어지고 있다. 즉, 고빈도 자료(high-frequency data)를 이용한 함수적 데이터 분석(functional data analysis; FDA) 관점에서 변동성 모델링에 대한 연구가 이루어지고 있으며 이는 하루의 수많은 일중 로그 수익률(intraday log returns)들을 연속적인 흐름(continuous flow)으로 보고 이 함수(function)를 (하나의) 관측치로 보는 방법이다. 함수형 변동성 분석은 연속적인 일중 수익률 변동성을 모형화한 것으로 Hörmann 등 (2013)의 functional-ARCH(1) (fARCH(1)) 모형과 fARCH(1) 모형을 확장시킨 Aue 등 (2017)의 functional-GARCH(1, 1) (fGARCH(1, 1)) 모형이 있다. Yoon 등 (2017, 2018)은 국내 고빈도 시계열 자료에 fARCH(1) 모형과 fGARCH(1, 1) 모형을 적합하여 예시하였다.

본 연구의 두 가지 기여는 다음과 같다. 첫째, 함수적 변동성은 다변량 변동성에서 무한차수(infinite dimensional)에 해당하는 연속개념이므로 이론적으로는 하루의 거래 시간 m 개를 선택해서 m 개의 일중 로그 수익률을 벡터로 모아 놓은 m 변량 변동성 분석을 한 후 m 을 무한대로 보냄으로써 함수적 변동성을 근사(approximation)할 수 있다. 하지만 다변량 변동성 분석의 단점인 차원이 증가함에 따른 모형 모수의 급격한 증가(이를 curse of dimensionality라 부른다)를 피하기 위해 저차원(4절 예제에서는 $m = 3$ 차원) 다변량 분석을 일중 시간대를 이동하면서 여러 번 분석하는 방식으로 해결하고자 하였다. 둘째, 함수형 변동성 분석에서 중요한 요소인 일중 로그 수익률의 시간 간격에 대해 연구하였다. Hansen과 Lunde (2006)는 일중 로그 수익률의 시간 간격 결정을 실현변동성(realized volatility) (Lee와 Hwang, 2017) 관점에서 5분으로 권장하였다. 국내 금융시계열 연구로서 Lee와 Hwang (2017) 및 Jin 등 (2017)은 실현변동성 입장에서 1분 및 5분 간격 일중 로그 수익률 사용을 제시한 바 있다. 본 연구에서는 시간 간격을 함수형 변동성 관점에서 정하고자 한다. 이를 위해 고빈도 자료에서 시간 간격이 다른 일중 로그 수익률(들)을 구한 후 다변량 GARCH 모형과 fARCH(1) 모형을 각각 적합하여 변동성을 추정하고, 다변량 GARCH 모형과 다양한 시간 간격의 fARCH(1) 모형 적합 결과를 비교하여 적절한 시간 간격을 제시해 보았다.

2. 다변량 변동성 모형

본 절에서는 대표적인 다변량 변동성 모형에 대해 소개하고자 한다. 먼저 특정 k 번째 day(일)에서의 m 개 일중 로그 수익률로 이루어진 벡터 $\mathbf{r}_k = (r_{1k}, r_{2k}, \dots, r_{mk})^T$ 를 다음과 같이 모형화한다.

$$\mathbf{r}_k = \boldsymbol{\mu}_k + \mathbf{a}_k,$$

여기서 $\boldsymbol{\mu}_k = E(\mathbf{r}_k | \mathbf{F}_{k-1})$ 는 $k-1$ 시점(일)까지의 정보 \mathbf{F}_{k-1} 이 주어질 때의 \mathbf{r}_k 의 조건부 기댓값 벡터이며, 일반적으로 $\boldsymbol{\mu}_k$ 는 vector autoregressive moving average (VARMA) 모형을 이용하여 적합 시킨다 (Hwang 등, 2009). 오차항 벡터 $\mathbf{a}_k = (a_{1k}, a_{2k}, \dots, a_{mk})^T$ 는 다음과 같이 모형화 한다.

$$\mathbf{a}_k = \mathbf{H}_k^{\frac{1}{2}} \mathbf{e}_k$$

$\mathbf{H}_k = \text{Cov}(\mathbf{a}_k | \mathbf{F}_{k-1})$ 는 $m \times m$ 양정치 행렬이며 $k-1$ 시점(일)까지의 정보 \mathbf{F}_{k-1} 이 주어질 때의 \mathbf{a}_k 의 조건부 분산-공분산 행렬이다. 랜덤(iid) 벡터인 $\mathbf{e}_k = (e_{1k}, e_{2k}, \dots, e_{mk})^T$ 에 대해서는 $E(\mathbf{e}_k) = \mathbf{0}$ 이고 $\text{Var}(\mathbf{e}_k) = \mathbf{I}$ 라 가정한다.

2.1. Exponential weighted moving average 모형

EWMA 모형으로 추정하는 조건부 분산-공분산 행렬 \mathbf{H}_k 는 다음과 같다.

$$\mathbf{H}_k = (1 - \lambda)\mathbf{a}_{k-1}\mathbf{a}_{k-1}^T + \lambda\mathbf{H}_{k-1}.$$

현재 시점과 가까운 최근 자료일수록 현재의 변동성에 더 많은 관련이 있음을 반영하기 위해서 EWMA 모형에서는 최근 자료일수록 더 많은 가중치를 주는 지수평활법을 이용한다. 평활 상수 λ 는 0과 1 사이의 값을 가지며, 일반적으로 RiskMetrics에서는 λ 값으로 0.94를 사용한다. 오차항 벡터 \mathbf{a}_k 의 분포를 가정하는 경우, λ 에 대해 추정할 수 있다 (Tsay, 2010, Chapter 10).

2.2. Baba-Engle-Kraft-Kroner 모형

Engle과 Kroner (1995)는 조건부 분산-공분산 행렬 \mathbf{H}_k 가 양정치 행렬이라는 조건을 항상 만족시킬 수 있도록 BEKK 모형을 제안하였다. BEKK(1, 1) 모형은 다음과 같다.

$$\mathbf{H}_k = \mathbf{C}\mathbf{C}^T + \mathbf{A} \left(\mathbf{a}_{k-1}\mathbf{a}_{k-1}^T \right) \mathbf{A}^T + \mathbf{B}\mathbf{H}_{k-1}\mathbf{B}^T.$$

위 모형에서 행렬 \mathbf{C} 는 하삼각행렬이며, 행렬 \mathbf{A} 와 \mathbf{B} 는 $m \times m$ 행렬이다. $\mathbf{C}\mathbf{C}^T$ 가 양정치 행렬로 주어 진다면 어떤 행렬 \mathbf{A} , \mathbf{B} 에 대해서도 \mathbf{H}_k 는 항상 양정치 행렬이 된다.

2.3. Constant conditional correlation 모형

모수의 수를 줄이고 추정 모형을 간단하게 만들기 위해서 Bollerslev (1990)는 조건부 상관계수를 시간에 무관한 상수로 고정하는 CCC 모형을 제안하였고 모형은 다음과 같다.

$$\mathbf{H}_k = \mathbf{D}_k\mathbf{R}\mathbf{D}_k = \left(\rho_{ij} \sqrt{h_{ii,k}h_{jj,k}} \right),$$

여기서 $\mathbf{D}_k = \text{diag}(h_{11,k}^{1/2}, \dots, h_{mm,k}^{1/2})$ 는 $m \times m$ 대각행렬이고, $h_{ii,k}$ 는 k 시점에서 i 번째 수익률의 조건부 분산을 의미한다. $\mathbf{R} = (\rho_{ij})$ 는 대각원소가 1이고 대칭인 조건부 상관계수 행렬로 시간에 따라 변하지 않는 상수로 고정된다.

2.4. Dynamic conditional correlation 모형

조건부 상관계수를 상수로 고정한 CCC 모형에서 나아가 여러 저자들은 조건부 상관계수 행렬이 시간에 의존하여 변화하는 모형들을 제시하였다. 이러한 모형을 DCC 모형이라고 하며, 그 중에서 Engle (2002)이 제안한 DCC 모형은 다음과 같다.

$$\mathbf{H}_k = \mathbf{D}_k\mathbf{R}_k\mathbf{D}_k, \quad \mathbf{D}_k = \text{diag} \left(h_{11,k}^{1/2}, \dots, h_{mm,k}^{1/2} \right), \quad \mathbf{R}_k = \mathbf{J}_k\mathbf{Q}_k\mathbf{J}_k,$$

여기서 $\mathbf{J}_k = \text{diag}(q_{11,k}^{-1/2}, \dots, q_{mm,k}^{-1/2})$ 는 대각행렬이며, $\mathbf{Q}_k = (q_{ij,k})$ 는 $m \times m$ 양정치 행렬이다. \mathbf{Q}_k 는 다음의 식을 만족한다.

$$\mathbf{Q}_k = (1 - \theta_1 - \theta_2)\bar{\mathbf{Q}} + \theta_1\boldsymbol{\epsilon}_{k-1}\boldsymbol{\epsilon}_{k-1}^T + \theta_2\mathbf{Q}_{k-1},$$

여기서 $\boldsymbol{\epsilon}_k$ 는 원소로 $\epsilon_{i,k} = a_{i,k}/\sqrt{h_{ii,k}}$ 를 가지는 표준화된 오차 벡터이고 $\bar{\mathbf{Q}}$ 는 $\boldsymbol{\epsilon}_k$ 의 비조건부 분산-공분산 행렬이다. θ_1 과 θ_2 는 $0 < \theta_1 + \theta_2 < 1$ 을 만족하고 음이 아닌 값을 갖는다.

3. Functional ARCH 모형

표준적인 일별(daily) 수익률 변동성 분석을 위해 GARCH 형태의 금융 시계열 모형들이 제시되어 왔다. 최근에는 자료를 수집하고 처리하는 능력이 발전하면서 하루의 일중 수익률이 곡선 형태의 연속적인 흐름으로 표현되는 금융 자료를 얻을 수 있게 되었다. 이런 곡선 형태의 자료를 하나의 함수로 생각할 수 있고 함수적 자료를 얻을 수 있음에 따라 GARCH 형태의 금융 시계열 모형을 함수적 형태로 확장시키는 방안이 제시되었다. 즉, 고빈도 자료를 이용한 FDA (Hörmann과 Kokoszka, 2012) 관점에서 접근하는 함수적 변동성 모형 적합을 고려해볼 수 있다. Hörmann 등 (2013)이 제안한 fARCH(1) 모형은 Engle (1982)의 ARCH 모형을 함수적으로 확장시킨 모형이다.

일반적인 변동성 모형은 $\{y_k, 1 \leq k \leq T\}$ 로 나타내는 1일부터 T 일까지의 종가를 기준으로 계산한 일간 로그 수익률(daily log return)을 이용한다. 함수적 변동성 모형에서는 고빈도 자료에서 적절한 시간 간격을 잡아 계산한 일중 로그 수익률을 시간 t 에 대한 함수로써 $\{y_k(t), 1 \leq k \leq T, 0 \leq t \leq S\}$ 로 이용한다. k 는 특정일을 나타내며 t 는 일중 시간을 나타낸다. 또한 $S = 1$ 이라고 가정하여 구간 $[0, 1]$ 이 하루를 나타내는 것으로 조정하여 0은 시작시간을 1은 종료시간을 표시한다. 고빈도 자료에서 적절한 시간 간격을 잡아 일중 로그 수익률을 계산할 수 있으며, 본 연구에서는 1분, 3분, 5분, 7분, 10분 간격으로 로그 수익률을 계산하여 이용하였다. Hansen과 Lunde (2006)는 5분 간격 사용을 제안하였으며 Yoon 등 (2017, 2018)에서도 5분 간격 일중 로그 수익률을 이용하였다.

3.1. fARCH(1) 모형

기존의 ARCH(1) 모형이 상수와 과거 수익률의 제곱의 선형결합으로 나타내지는 것과 유사하게 fARCH(1)에서도 과거 수익률의 제곱들과 그것의 변동성 함수와의 함수적 선형결합(functional linear combination) 형태로 현재 관측값의 변동성 함수를 나타낸다 (Yoon 등, 2017). 필요한 모형의 수식 및 개요를 Hörmann 등 (2013) 및 Yoon 등 (2017)을 기반으로 정리하도록 한다. 함수적 변동성 모형을 다루기 위해서는 먼저 적절한 함수공간을 정의해야 한다. 함수공간 F 는 정의역 $[0, 1]$ 에서 실수값을 갖는 함수들로 구성된 일반적인 함수공간(generic function space)이다. 함수공간의 예로는 힐버트 공간(Hilbert space)이 되는 $H = L^2([0, 1])$ 이 있고 $\sup\text{-norm}\|x\|_\infty = \sup_{t \in [0, 1]} |x(t)|$ 인 연속함수들로 이루어진 공간 $F = C[0, 1]$ 이 있다. Hörmann 등 (2013)의 기호를 사용해서, 함수공간 F 에서 음이 아닌 함수(non-negative function)들의 집합을 F^+ 로 나타낸다. 기호 $\delta(t)$ 를 함수공간 F^+ 의 음이 아닌 함수, $\{\epsilon_k\}$ 를 F 에 있는 iid 랜덤 함수의 수열, $\beta(t, s)$ 를 음이 아닌 커널 함수(kernel function)라 하자. 함수공간 F 에서 정의된 랜덤 함수(random function)의 수열 $\{y_k(t), k \in Z, t \in [0, 1]\}$ 에 대해 fARCH(1)을 다음과 같이 정의한다.

$$\begin{aligned} y_k(t) &= \epsilon_k(t)\sigma_k(t), \quad t \in [0, 1], \\ \sigma_k^2(t) &= \delta(t) + \beta(y_{k-1}^2)(t). \end{aligned}$$

어떠한 함수공간을 선택하느냐에 따라 fARCH(1) 프로세스가 존재하기 위한 조건이 달라진다. Hörmann 등 (2013)은 fARCH(1)에서 강정상 솔루션(strictly stationary solution)이 존재하기 위한 충분조건과 이론적 증명을 제시하였다. 본 논문에서는 함수공간 F 가 힐버트 공간 $H = L^2([0, 1])$ 인 경우만 소개하겠다 (Yoon 등, 2017, 2018).

유계한(bounded) 커널 연산자 $\beta(t, s)$ 를 다음과 같이 정의한다.

$$\beta(y_{k-1}^2)(t) = \int_0^1 \beta(t, s)y_{k-1}^2(s) ds$$

또한, $K(\epsilon_1^2) = (\iint \beta^2(t, s) \epsilon_1^4(s) ds dt)^{1/2}$ 을 정의한다. $E\{K(\epsilon_1^2)\}^\alpha < 1$ 을 만족하는 $\alpha > 0$ 이 존재한다면 위의 fARCH(1) 프로세스는 함수공간 $F = H$ 에서 유일한 강정상 솔루션을 갖게 된다 (Hörmann 등, 2013). 모형 fARCH(1)에서의 함수 $\delta(t)$ 와 연산자 $\beta(t)$ 의 추정에 대해 살펴보겠다. 추정에 대한 이론적 가정과 증명은 Hörmann 등 (2013)을 참고하기 바람이며 여기서는 주요한 식 중심으로 다루고자 한다. 유한한 표본(크기 N)에서 평균함수에 대해서는 $\mu(t) = E[y(t)] = 0$ 을 가정하여 다음의 공분산 연산자와 교차 공분산 연산자를 이용한다.

$$\hat{C}(y) = \frac{1}{N} \sum_{k=1}^N \langle Z_k, y \rangle Z_k, \quad y \in H,$$

$$\hat{C}_1(y) = \frac{1}{N} \sum_{k=1}^N \langle Z_k, y \rangle Z_{k+1}, \quad y \in H.$$

연산자 β 는 다음과 같이 추정된다.

$$\hat{\beta}(y; K) = \pi_K \hat{C}_1 \hat{C}^{-1}(y; K),$$

$$\hat{C}^{-1}(y; K) = \sum_{j=1}^K \hat{\lambda}_j^{-1} \langle \hat{e}_j, y \rangle \hat{e}_j,$$

여기서 $(\hat{\lambda}_j, \hat{e}_j)$ 는 큰 값부터 내림차순으로 정렬한 고윳값과 대응되는 \hat{C} 의 고유함수이다. 이를 바탕으로 β 는 다음과 같이 구할 수 있으며 $\hat{m}_2 = (1/N) \sum_{k=1}^N y_k^2$ 임을 이용하여 δ 를 구한다.

$$\hat{\beta}(t, s; K) = \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{j=1}^K \sum_{i=1}^K \hat{\lambda}_j^{-1} \langle Z_k, \hat{e}_j \rangle \langle Z_{k+1}, \hat{e}_i \rangle \hat{e}_j(s) \hat{e}_i(t),$$

$$\hat{\delta} = \hat{m}_2 - \hat{\beta}(\hat{m}_2),$$

여기서 N 은 자료의 수이며 K 는 이용하는 고유함수의 개수를 나타낸다. K 는 분석자가 선택할 수 있는데 Hörmann 등 (2013)은 $K = 2, 3, 4$ 인 경우가 최상의 결과를 제공한다고 제시한 바 있다. R의 fda 패키지를 이용하여 추정의 모든 과정을 수행할 수 있으며 추정 시 고려하는 정규직교 함수들의 기저로 푸리에 기저(Fourier bases)나 B-스플라인 기저(B-spline bases)가 있다. 푸리에 기저는 보통 일정한 주기로 관측된 주기적인 데이터(periodic data)나 주기적인 데이터에 가까운 데이터, B-스플라인 기저는 비주기적인 데이터(non-periodic data)에 사용한다. B-스플라인 기저는 flexible system을 생성하며 상대적으로 작은 K 에서 좋은 근사를 하는 것으로 알려져 있다. Yoon 등 (2018)의 실증 분석에서는 B-스플라인 기저를 이용하여 추정하였다.

3.2. 미래 l -시차 후 변동성 예측

ARCH(1) 모형에서와 같이 fARCH(1) 모형을 이용하여 미래 시차의 변동성을 점화식 구조를 통하여 예측할 수 있다. 일중 로그 수익률 $\{y_k(t), 1 \leq k \leq T, 0 \leq t \leq S\}$ 은 함수공간 $F = H$ 에서의 fARCH(1) 프로세스이고 $\{\epsilon_k\}$ 에 대해서 $E(\epsilon_k(t)) = 0$, $E(\epsilon_k^2(t)) = 1$ 이라고 가정한다. 이때, 다음과 같이 T 시점에서 1-시차 후 변동성은 다음과 같다 (Hörmann 등, 2013).

$$\sigma_{T+1}^2 = \text{Var}(y_{T+1}|F_T) = \delta + \beta(y_T^2)$$

임의의 미래시차 $2, 3, \dots, l$ 에 대한 공식은 아직 보고된 바 없으므로 점화식 공식으로 유도한다면 후속 연구에 유용할 것이다. 함수공간 $F = H$ 에 포함되는 랜덤 함수 $\{\epsilon_k^2\}$ 의 기댓값의 연산이 유계한 연산

Table 4.1. Estimated results of multivariate GARCH

Model	Parameter estimates
EWMA	$\hat{\lambda} = 0.94$
CCC(1, 1)	$h_{11,k} = 0.00002 + 0.02239a_{1,k-1}^2 + 0.97088h_{11,k-1}$ $h_{22,k} = 0.00002 + 0.04818a_{2,k-1}^2 + 0.94074h_{22,k-1}$ $h_{33,k} = 0.00005 + 0.05992a_{3,k-1}^2 + 0.90529h_{33,k-1}$ $\hat{\mathbf{R}} = \begin{bmatrix} 1 & -0.03550 & 0.05362 \\ -0.03550 & 1 & -0.04409 \\ 0.05362 & -0.04409 & 1 \end{bmatrix}$
DCC(1, 1)	$h_{11,k} = 0.00256 + 0.03757a_{1,k-1}^2 + 0.00025h_{11,k-1}$ $h_{22,k} = 0.00131 + 0.11278a_{2,k-1}^2 + 0.00553h_{22,k-1}$ $h_{33,k} = 0.00124 + 0.08886a_{3,k-1}^2 + 0.00472h_{33,k-1}$ $\mathbf{Q}_k = (1 - 0 - 0.0961) \begin{bmatrix} 1 & -0.02156 & 0.02812 \\ -0.02156 & 1 & -0.05080 \\ 0.02812 & -0.05080 & 1 \end{bmatrix}$ $+ 0 \times \boldsymbol{\epsilon}_{k-1} \boldsymbol{\epsilon}_{k-1}^T + 0.0961 \mathbf{Q}_{k-1}$
BEKK(1, 1)	$\hat{\mathbf{C}} = \begin{bmatrix} 0.0027 & & \\ -0.0001 & 0.0005 & \\ -0.0002 & -0.0006 & 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} 0.0214 & 0.1443 & 0.1297 \\ -0.0859 & 0.3254 & -0.1134 \\ 0.0331 & -0.1960 & 0.1875 \end{bmatrix}$ $\hat{\mathbf{B}} = \begin{bmatrix} -0.0453 & 0.0339 & 0.1261 \\ 0.1803 & 0.5694 & 0.8595 \\ -0.0367 & -0.4117 & 0.2801 \end{bmatrix}$

EWMA = exponential weighted moving average; CCC = constant conditional correlation; DCC = dynamic conditional correlation; BEKK = Baba-Engle-Kraft-Kroner.

자 β 와 함께 이루어지면, 다음이 성립한다 : $E(\beta(\epsilon_k^2)) = \beta(E(\epsilon_k^2))$. 따라서 다음과 같이 T 시점에서의 2-시차 후 변동성을 예측할 수 있다.

$$\begin{aligned}
 \sigma_{T+2}^2 &= \text{Var}(y_{T+2}|F_T) = E(y_{T+2}^2|F_T) = E(\sigma_{T+2}^2 \epsilon_{T+2}^2|F_T) \\
 &= E(\delta \epsilon_{T+2}^2 + \beta(y_{T+1}^2) \epsilon_{T+2}^2|F_T) = \delta + E(\beta(y_{T+1}^2)|F_T) \\
 &= \delta + E(\beta(\sigma_{T+1}^2 \epsilon_{T+1}^2)|F_T) = \delta + \beta(E(\sigma_{T+1}^2 \epsilon_{T+1}^2|F_T)) \\
 &= \delta + \beta(\sigma_{T+1}^2).
 \end{aligned}$$

유사한 방식으로 T 시점에서의 l -시차 후 변동성을 다음과 같은 점화식으로 구할 수 있다.

$$\sigma_{T+l}^2 = \text{Var}(y_{T+l}|F_T) = \delta + \beta(\sigma_{T+l-1}^2).$$

다음 절의 자료 분석 예제에서는 $l = 1$ 및 $l = 2$ 시차 후 변동성 σ_{T+l}^2 계산이 예시되어 있다.

4. 자료 분석 예제

본 절에서는 앞에서 소개한 모형을 국내 KOSPI 지수 자료에 적합한 결과를 예시하고 이를 기반으로 fARCH(1) 모형 적합 시 이용할 적절한 로그 수익률 시간 간격을 제시하고자 한다. 분석 대상인 고빈도 자료는 2010년 1월 5일부터 2015년 6월 30일까지 1,349일간의 개장 시간부터 마감 시간까지 1분 간격으로 조사된 고빈도 KOSPI 자료이다. 특정 k 일의 시간 t 에서의 주가를 $P_k(t)$ 라고 할 때, $y_k(t) =$

$\log P_k(t) - \log P_k(t-h)$ 로 일중 로그 수익률을 정의한다. 보통 $h = 5$ 로 하여 5분 간격 로그 수익률을 계산하여 (Hansen과 Lunde, 2006) 이용하며 Yoon 등 (2017)에서도 고빈도 KOSPI 지수 자료에 fARCH(1) 모형 적합 시 $h = 5$ 로 하여 5분 간격 로그 수익률을 계산하여 이용하였다. 본 연구에서는 5분 간격 로그 수익률($h = 5$) 뿐 아니라 1분 간격($h = 1$), 3분 간격($h = 3$), 7분 간격($h = 7$), 10분 간격($h = 10$)으로 총 5개의 일중 로그 수익률을 고려하였으며 이 중 최적 시간 간격 h 를 찾고자 하였다. 참고로, Hörmann 등 (2013)의 공동저자인 Professor Ron Reeder (University of Utah, USA)가 보내 준 R 코드를 수정/조정하여 fARCH(1) 분석을 수행했음을 밝혀둔다.

4.1. 다변량 GARCH 모형 적합

다변량 GARCH 모형 분석에서는 위의 로그 수익률 중 12개의 시점 t 에서의 로그 수익률을 추출하여 분석에 이용하였다. 9시 15분, 9시 45분, 10시 15분, 10시 45분, ..., 14시 15분, 14시 45분 이렇게 각 시간대별로 두 시점을 30분 간격으로 추출하여(총 12시점) 순서대로 세 시점씩 묶어 총 4개의 3변량 벡터 시계열 $\mathbf{r}_{1k}, \mathbf{r}_{2k}, \mathbf{r}_{3k}, \mathbf{r}_{4k}$ 을 생성하였다. Hwang 등 (2009)은 국내 금융 시계열 분석에 있어서 BEKK 및 CCC 모형이 다른 다변량 GARCH 모형들에 비해 우수함을 보여주었다. 본 연구에서는 Hwang 등 (2009)의 결과를 토대로 자료에 BEKK, CCC 모형을 적합시키고 이와 더불어 비모수적인 모형인 EWMA 모형과 CCC 모형을 확장시킨 모형인 DCC 모형을 적합하였다. 모든 벡터 시계열 \mathbf{r}_k 에 공통적인 모형을 적합시키기 위하여 \mathbf{r}_k 의 평균 벡터를 영으로 설정하고 EWMA, CCC(1, 1), DCC(1, 1), BEKK(1, 1) 모형을 적합하였다. Table 4.1은 시간 간격 h 가 1분인 로그 수익률로 이루어진 3변량 벡터 시계열 \mathbf{r}_{1k} 에 다변량 변동성 모형을 적합한 결과이다. 마찬가지로 방식으로 $\mathbf{r}_{2k}, \mathbf{r}_{3k}, \mathbf{r}_{4k}$ 에 대한 결과와 다른 시간 간격인 $h = 3, 5, 7, 10$ 에 대한 결과도 얻을 수 있었으며 지면 관계상 제시는 생략하였다.

4.2. fARCH(1) 모형 적합

KOSPI 자료에서 5개의 시간 간격 h 를 따라 구한 일중 로그 수익률 $y_k(t)$ 에 fARCH(1) 모형을 적합한 결과를 소개하겠다. Figure 4.1은 추정된 커널 함수 $\hat{\beta}(t, s)$ 의 그래프이며 Figure 4.2는 추정된 $\hat{\delta}(t)$ 의 그래프이다.

Figures 4.1과 4.2를 보면 전체적으로 t 가 0이나 1에 가까워질 때 $\hat{\beta}(t, s)$ 의 값이 커지는 것으로 보이고, 시간 간격 h 가 1분일 때를 제외하곤 $\hat{\beta}(t, s)$ 의 그래프가 비슷한 형태를 가지는 것으로 보인다. 또한 $\hat{\delta}$ 를 보면 t 가 0이나 1에 가까워질수록 $\hat{\delta}$ 의 값이 커지는 것으로 보이고, 시간 간격 h 가 큰 값을 가질수록 $\hat{\delta}$ 의 값이 커지는 것으로 보인다. 또한 fARCH(1) 모형을 이용하여 l -시차 후의 변동성을 예측할 수 있는데, Figure 4.3은 3장 2절에서 유도한 식을 이용하여 KOSPI 자료 2015년 6월 22일 기준 (1분 간격 로그 수익률에 대한) 1-시차, 2-시차 후의 변동성을 예측한 그래프이다. 즉, 6월 23일과 24일의 변동성을 추정하여 나타낸 그래프이며, 각각의 그래프에는 각 날짜에 해당하는 1분 간격 로그 수익률의 제공 값을 함께 표시하였다.

4.3. 일중 로그 수익률 시간 간격 선택

다변량 GARCH 모형의 차원을 증가시키면 무한차원의 개념으로 확장시킬 수 있고, 이는 무한차원을 고려한 함수적 변동성과 비교하기에 적합하다고 판단된다. 따라서 다변량 변동성 모형으로 추정된 변동성을 기준으로 fARCH(1) 모형으로 추정된 변동성을 비교하기로 한다. fARCH(1) 모형을 적합하여 변동성을 추정하면 거래일 하루 동안의 움직임을 나타내는 곡선의 형태의 변동성이 추정된다. 추정된 변동성 $\sigma_k^2(t)$ 중에서 다변량 변동성 모형 적합 시 이용했던 12개의 시점 t 에서의 변동성을 고려하여 다변량 변동성 모형을 적합한 후 추정된 변동성과 비교하였다. Tables 4.2와 4.3은 다변량 변동성 모형으로

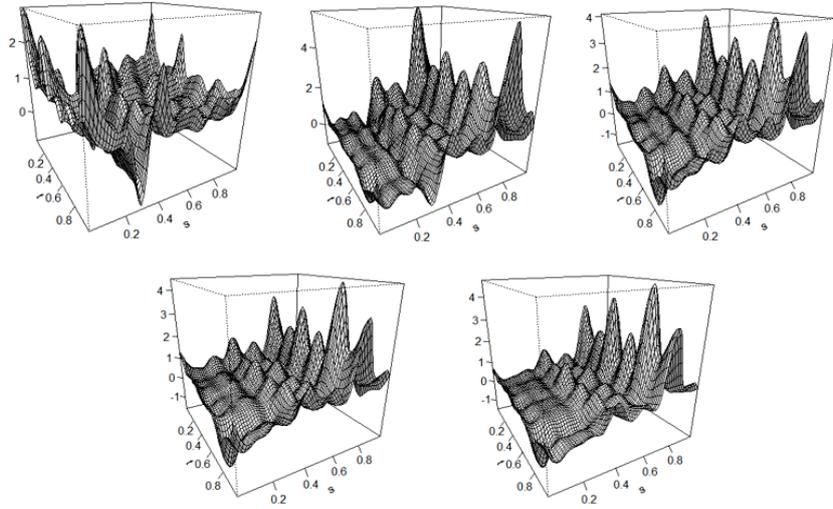


Figure 4.1. Plot of $\hat{\beta}(t, s)$ ($h = 1, 3, 5, 7, 10$ in order).

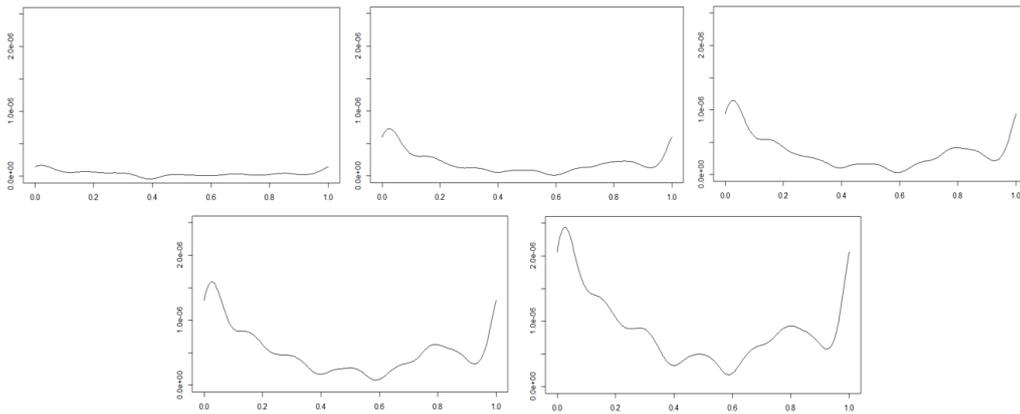


Figure 4.2. Plot of $\hat{\delta}(t)$ ($h = 1, 3, 5, 7, 10$ in order).

추정한 변동성을 기준으로 하여 시간 간격 h 를 달리하며 fARCH(1) 모형으로 추정된 변동성의 mean squared error (MSE)와 mean absolute error (MAE)를 구한 결과이다. 다변량 변동성 모형으로부터 12개 시점에서 추정된 변동성은 fARCH(1) 변동성의 근사값으로 간주할 수 있다. 따라서 어느 시간 간격이 좋을지 하는 논의는 다변량 변동성 모형으로부터 추정된 변동성과 fARCH(1)으로부터 추정된 변동성과의 차이가 어떤 시간 간격에서 최소인가를 찾는 문제로 귀결된다.

각 MSE와 MAE의 값은 동일한 시간 간격 h 를 가지는 로그 수익률에 fARCH(1) 모형과 다변량 변동성 모형을 적합하여 추정된 변동성을 이용하여 구한 값이다. EWMA, CCC, DCC, BEKK 모형을 기준으로 fARCH(1) 모형으로 추정된 변동성의 MSE와 MAE 값을 볼 때, 모든 기준에서 시간 간격 h 가 1분일 때의 값이 가장 작은 것으로 나타난다. 또한 시간 간격 h 가 커질수록 MSE와 MAE 값이 커지는 것을 볼 수 있다. 측정 단위와 무관한 통계량으로써 mean absolute percentage error (MAPE)를 고려하였다. Table 4.4는 fARCH(1) 모형으로 추정된 변동성의 MAPE를 구한 결과이다. EWMA, CCC,

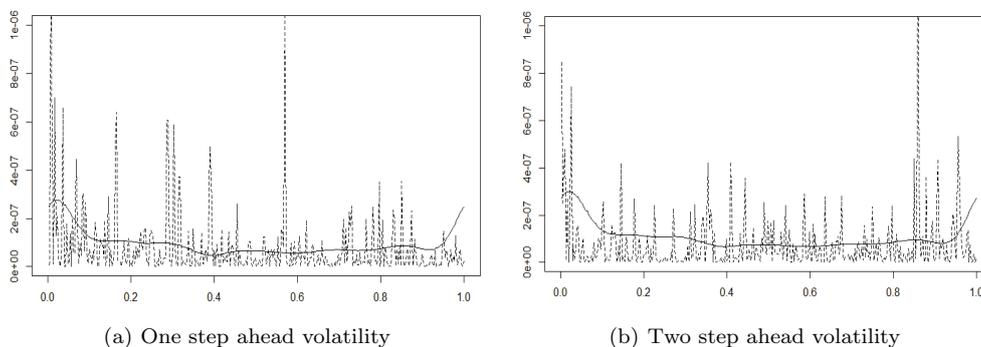


Figure 4.3. Squared return (dotted line) and estimated intraday volatility (solid line).

Table 4.2. MSE of volatility estimated by fARCH(1)

Model	MSE			
	EWMA	CCC	DCC	BEKK
fARCH(1) : $h = 1$	$2.6139e^{-06}$	$1.6258e^{-05}$	$1.7011e^{-06}$	$2.0776e^{-06}$
fARCH(1) : $h = 3$	$3.1031e^{-05}$	$2.9465e^{-05}$	$2.9404e^{-05}$	$2.7446e^{-05}$
fARCH(1) : $h = 5$	$1.0340e^{-04}$	$1.1108e^{-04}$	$1.0469e^{-04}$	$8.5465e^{-05}$
fARCH(1) : $h = 7$	$2.0725e^{-04}$	$1.9948e^{-04}$	$2.3835e^{-04}$	$2.3312e^{-04}$
fARCH(1) : $h = 10$	$3.8395e^{-04}$	$3.4347e^{-04}$	$2.4847e^{-04}$	$3.0921e^{-04}$

MSE = mean squared error; fARCH = functional autoregressive conditional heteroscedasticity; EWMA = exponential weighted moving average; CCC = constant conditional correlation; DCC = dynamic conditional correlation; BEKK = Baba-Engle-Kraft-Kroner.

Table 4.3. MAE of volatility estimated by fARCH(1)

Model	MAE			
	EWMA	CCC	DCC	BEKK
fARCH(1) : $h = 1$	$6.4969e^{-04}$	$6.2738e^{-04}$	$5.4912e^{-04}$	$5.5090e^{-04}$
fARCH(1) : $h = 3$	$2.2360e^{-03}$	$1.9354e^{-03}$	$1.6735e^{-03}$	$1.7964e^{-03}$
fARCH(1) : $h = 5$	$3.9804e^{-03}$	$3.5499e^{-03}$	$2.9837e^{-03}$	$3.1449e^{-03}$
fARCH(1) : $h = 7$	$5.6999e^{-03}$	$4.9953e^{-03}$	$5.1693e^{-03}$	$4.8346e^{-03}$
fARCH(1) : $h = 10$	$8.3482e^{-03}$	$7.3517e^{-03}$	$5.4574e^{-03}$	$6.7038e^{-03}$

MAE = mean absolute error; fARCH = functional autoregressive conditional heteroscedasticity; EWMA = exponential weighted moving average; CCC = constant conditional correlation; DCC = dynamic conditional correlation; BEKK = Baba-Engle-Kraft-Kroner.

Table 4.4. MAPE of volatility estimated by fARCH(1)

Model	MAPE			
	EWMA	CCC	DCC	BEKK
fARCH(1) : $h = 1$	54.9847	44.7128	41.6422	40.3806
fARCH(1) : $h = 3$	77.6101	56.0878	51.7754	48.6323
fARCH(1) : $h = 5$	76.4221	53.6674	44.9503	44.5932
fARCH(1) : $h = 7$	80.6188	55.7525	54.4026	44.4701
fARCH(1) : $h = 10$	82.5556	55.5225	38.2043	43.6580

MAPE = mean absolute percentage error; fARCH = functional autoregressive conditional heteroscedasticity; EWMA = exponential weighted moving average; CCC = constant conditional correlation; DCC = dynamic conditional correlation; BEKK = Baba-Engle-Kraft-Kroner.

BEKK 모형을 기준으로 fARCH(1) 모형으로 추정된 변동성의 MAPE 값을 볼 때 시간 간격 h 가 1분일 때의 값이 가장 작은 것으로 나타난다. EWMA와 CCC 모형에서는 5분 간격이 두 번째로 좋으며 BEKK 모형에서는 10분 간격이 두 번째로 나타난다. DCC 모형을 기준으로 MAPE 값을 보면 h 가 10분일 때의 값이 가장 작고, 그다음으로는 h 가 1분일 때와 5분이 뒤따른다. 따라서, 변동성 모형 입장에서 보면 고빈도 KOSPI의 일중 로그 수익률 계산을 위한 적절한 시간 간격은 $h = 1$ (분)이라고 생각되며 5분 시간 간격 선택도 무난해 보인다.

5. 연구의 한계점 및 추후 연구

본 논문에서는 고빈도 금융 시계열의 변동성에 초점을 맞추고 이를 추정하는 함수형 모형인 fARCH(1)에 대해 연구하였다. 함수형 변동성 모형의 근사 모형으로써 (다양한) 다변량 변동성 모형을 고려하였다. 또한 다변량 변동성 모형이 함수형 변동성 모형을 성공적으로 근사시킨다는 전제 아래 fARCH(1) 모형 적합 자료를 구성하는 일중 로그 수익률 계산에 필요한 시간 간격의 선택에 대해 알아보았다. 본 연구의 기여는 엄밀한 수리통계학적 측면보다는 기존의 다변량 변동성과 함수형 변동성(fARCH(1) 모형)의 연계성을 예시하는 데 있으며 이에 따른 본 연구의 한계점과 추후 연구 방향에 대해 논의하고자 한다. 첫째, 함수형 변동성 모형은 함수적 점화식 형태로서 fGARCH(p, q)로 확장될 수 있으나 (현재로서는) 자료 분석 프로그램으로서의 R-code는 fARCH(1)과 fGARCH(1, 1)만이 가능하며 fARCH(1) R-code는 안정적이나 fGARCH(1, 1)은 아직 불안정적으로 판단되어 (Yoon 등, 2017, 2018) 본 연구에서는 fARCH(1)만을 고려하였다. 다변량 변동성 모형으로는 Lee와 Hwang (2017), Hwang 등 (2009)에서 국내 KOSPI 자료 변동성 모형으로 사용한 EWMA, CCC(1, 1), DCC(1, 1), BEKK(1, 1) 모형을 분석하였다. 추후 고차 모형이나 다른 다변량 변동성 분석을 시도해 볼 수 있을 것이다. 둘째, 하루에 12시점에서 다변량 변동성을 추정하기 위해 4개의 3변량 벡터($r_{1k}, r_{2k}, r_{3k}, r_{4k}$) 분석을 4개의 모형(EWMA, CCC(1, 1), DCC(1, 1), BEKK(1, 1)) 각각에 대해 수행하였다. 또한 5개의 시간 간격 일중 로그 수익률에 대해 동일한 모형을 적합하였다. 따라서 총 $12 \times 4 \times 5 = 240$ 개의 다변량 변동성 모형 분석이 이루어졌으며 이들을 동일 잣대로 비교하기 위해서 엄밀한 통계적 추론과 모형 진단 과정은 생략하였으며 이를 추후 연구로 남겨두고자 한다. 또한 12변량 변동성 분석을 피하기 위해 4개의 3변량 분석을 수행하였으나 “curse of dimensionality”를 피할 수 있는 간결한 모형이 제안된다면 12변량 변동성 분석을 통한 fARCH(1) 함수형 변동성 근사값을 제시할 수 있을 것이다. 이를 확장할 수 있다면 24변량 혹은 36변량 변동성 모형으로 좀 더 “개선된” 근사-변동성을 얻을 수 있을 것이다. 2절에서 소개된 CCC 모형에서 $R = (\rho_{ij}) : m \times m$ 을 등(equal)-조건부 상관계수인 ($\rho_{ij} = \rho$, 모든 i, j 에서) $R = (\rho) : m \times m$ 로 모형화 하면 다변량 변동성 모형에서의 모수 수가 급격하게 줄어드는 효과를 볼 수 있다. 이러한 모형을 Equi-CCC 모형이라 칭하며 이에 대한 이론 및 응용 연구가 별개로 진행되고 있음을 밝혀둔다. 본 연구에서는 다변량 변동성을 연속형 함수형 변동성의 근사 개념으로 추정한 후, 응용으로서 일중 로그 수익률 계산의 시간 간격 선택 과정을 예시하기 위해 2010년 1월 5일부터 2015년 6월 30일까지 총 1,349일간의 1분 간격 고빈도 KOSPI 자료 분석을 수행하였다. 예시된 분석의 안정성(stability) 및 강건성(robustness)을 위해서 자료의 기간을 두 부분 혹은 세 부분으로 나누어서 세부 분석을 하는 것도 좋을 것이다. 이와 함께 다른 고빈도 금융 시계열에의 적용을 추후 연구과제로 남겨두고자 한다.

References

- Aue, A., Horváth, L., and Pellatt, D. F. (2017). Functional generalized autoregressive conditional heteroskedasticity, *Journal of Time Series Analysis*, **38**, 3–21.

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Review of Economics and Statistics*, **72**, 498–505.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.
- Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH, *Econometric Theory*, **11**, 122–150.
- Engle, R. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business & Economic Statistics*, **20**, 339–350.
- Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise, *Journal of Business & Economic Statistics*, **24**, 127–161.
- Hörmann, S. and Kokoszka, P. (2012). Functional Time Series. In *Handbook of Statistics*, (Vol. 30, pp. 157–186), Elsevier.
- Hörmann, S., Horvath, L., and Reeder, R. (2013). A functional version of the ARCH model, *Econometric Theory*, **29**, 267–288.
- Hwang, S. Y., Choi, M. S., and Do, J. D. (2009). Assessments for MGARCH models using back-testing: Case study, *Korean Journal of Applied Statistics*, **22**, 261–270.
- Jin, M. K., Yoon, J. E., and Hwang, S. Y. (2017). Choice of frequency via principal component in high-frequency multivariate volatility models, *The Korean Journal of Applied Statistics*, **30**, 747–757.
- Lee, G. J. and Hwang, S. Y. (2017). Multivariate volatility for high-frequency financial series, *The Korean Journal of Applied Statistics*, **30**, 169–180.
- Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed), John Wiley & Sons, New York.
- Yoon, J. E., Kim, J. M., and Hwang, S. Y. (2017). Functional ARCH (fARCH) for high-frequency time series: illustration, *Korean Journal of Applied Statistics*, **30**, 983–991.
- Yoon, J. E., Kim, J. M., and Hwang, S. Y. (2018). The fGARCH (1, 1) as a functional volatility measure of ultra high frequency time series, *Korean Journal of Applied Statistics*, **31**, 667–675.

함수형 ARCH 분석 및 다변량 변동성을 통한 일중 로그 수익률 시간 간격 선택

김다희^a · 윤재은^{a,1} · 황선영^a

^a숙명여자대학교 통계학과

(2020년 2월 4일 접수, 2020년 3월 11일 수정, 2020년 4월 1일 채택)

요약

본 논문에서는 고빈도 함수적 ARCH 모델을 소개하고 근사모형으로써 다변량 변동성 모델을 고려하였다. 이를 기반으로 함수형 변동성 분석에서 중요한 요소인 일중 로그 수익률의 적절한 시간 간격을 찾아보았다. 또한 함수적 ARCH 모형에서 l -시차 후 변동성 예측식을 제시하고 고빈도 KOSPI 자료에 적합하여 예시하였다.

주요용어: 함수적 ARCH 모형, 고빈도 시계열, 다변량 변동성

본 연구는 한국연구재단 기초연구과제 사업의 지원을 받았습니다 (과제번호: NRF-2018R1A2B2004157).

¹교신저자: (04310) 서울특별시 용산구 청파로47길 100, 숙명여자대학교 통계학과.

E-mail: yoonje@sookmyung.ac.kr