*The Korean Journal of Applied Statistics* (2015) DOI: http://dx.doi.org/10.5351/KJAS.2015.28.6.1163
**28**(6), 1163–1170

# Volatility Computations for Financial Time Series: High Frequency and Hybrid Method

J.E. Yoon[a] · S.Y. Hwang[a,1]

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

(Received October 1, 2015; Revised November 13, 2015; Accepted November 26, 2015)

**Abstract**

Various computational methods for obtaining volatilities for financial time series are reviewed and compared with each other. We reviewed model based GARCH approach as well as the data based method which can essentially be regarded as a smoothing technique applied to the squared data. The method for high frequency data is focused to obtain the realized volatility. A hybrid method is suggested by combining the model based GARCH and the historical volatility which is a data based method. Korea stock prices are analysed to illustrate various computational methods for volatilities.

Keywords: GARCH, high frequency data, hybrid volatility

## 1. 서론

금융시계열에서 수익률의 변동성(volatility, 조건부 분산)은 큰 변동성이 어느 정도 지속되고 이어서 작은 변동성 또한 어느 정도 지속되는 변동성 집중(volatility cluster) 현상, 같은 수익률이더라도 부호에 따라 변동성이 다르게 움직이는 비대칭 레버리지(leverage) 현상 등 다른 일반 시계열 자료에서 잘 나타나지 않는 특징을 보이고 있다. 또한 금융시장의 위험관리, 옵션 가격결정 등에 있어 변동성 산출은 중요한 역할을 하고 있으며 이러한 변동성을 효과적으로 추정하고 예측하는 것은 금융시계열에서 핵심적인 요소가 된다.

수익률의 변동성을 추정하는 대표적인 방법으로는 수익률의 조건부 분산을 수리 모형화 하여 모수를 추정하는 모형 기반(model based) 방법과 과거 수익률 자료를 제곱을 이용하여 평활하는 자료 기반(data based) 방법이 있다. 전자는 GARCH 모형이나 비대칭모형인 T-GARCH 모형과 같은 조건부 이분산(conditionally heteroscedastic) 수식 모형을 설정하고 모수를 추정하여 조건부 분산 모형식을 얻는 모형 기반 모수적인 방법이다. 후자는 과거 수익률의 제곱 자료에 적절한 가중치를 부여하여 평활하여 변동성을 추정하는 자료 기반 비모수적인 방법으로 대표적인 예로는 지수가중이동평균(exponentially weighted moving average; EWMA)과 역사적 변동성(historical volatility) 방법이 있다. 최근 들어, 이러한 전통적인 방법 외에 고빈도 자료(high-frequency data)를 이용하는 방법, 종가와 시가, 하한가와 상한가 등을 이용하는 방법 등이 개발되고 있다. 오늘날 대용량 데이터 처리 능력의 향상 등으로 인하여 변동성 분석에 있어 일별 주가지수 종가 자료 뿐 아니라 1분 빈도의 자료 등과 같은 고빈도 자료를

[1]Corresponding author: Department of Statistics, Sookmyung Women's University, Yongsan-Gu, Seoul 04310, Korea. E-mail: shwang@sookmyung.ac.kr

사용한 분석도 가능해졌다. 고빈도 자료를 사용하여 매일의 실현변동성(realized volatility; RV)을 계산할 수 있으며 Andersen과 Bollerslev (1997)는 5분 단위의 고빈도 자료를 이용하여 실현변동성을 추정하는 방법론을 소개한 바 있다.

본 연구에서는 다양한 변동성 산출방법을 소개하고 국내주가자료를 통해 비교 분석하고 있다. 1분 단위로 조사된 주가지수 고빈도 자료를 이용하여 실현변동성을 산출해 보았으며, GARCH 모형 및 비대칭 T-GARCH 모형을 이용하여 변동성을 추정하는 모형 기반 방법과, 역사적 변동성 및 지수가중이동평균을 이용하는 자료 기반 방법을 예시하고, 이들을 융합한 융합(Hybrid) 방법을 제안해 보았다.

## 2. 고빈도 자료 기반 방법

일별 종가의 로그 수익률을 이용하여 변동성을 추정하는 기존의 방법과는 달리 하루 동안 매 분 단위로 관측된 고빈도 자료를 활용하여 변동성을 추정하는 방법을 생각해 보자. 기호 $r_t$를 시점 $t$($t$일)의 일간 로그 수익률(daily log return)이라고 하자. 모형 또는 자료에 기반하여 변동성을 추정하는 경우에는 이 일간 로그 수익률($r_t$)을 이용한다. 하지만 고빈도 자료에 기반하여 추정하는 경우에는 일중 로그 수익률(intradaily log return)이라는 것을 고려한다. 일중 로그 수익률은 $t$일 중 일정한 간격으로 $n$개가 조사되었다고 가정하여 $\{r_{t,i}\}$, $i = 1, 2, \ldots, n$로 나타내며 $t$일의 $i$번째 관측시점의 로그 수익률을 의미한다. $t$일의 일간 로그 수익률 $r_t$는 다음과 같이 $n$개의 일중 로그 수익률의 합으로 나타낼 수 있다.

$$r_t = \sum_{i=1}^{n} r_{t,i}.$$

일간 로그 수익률의 조건부 분산은 다음과 같다.

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

$F_{t-1}$은 $t-1$시점까지 주어진 정보를 나타낸다. 일간 로그 수익률의 조건부 분산은 $n$개의 일중 로그 수익률의 분산과 공분산의 합으로 구할 수 있다. 위의 조건부 분산은 일중 로그 수익률 $\{r_{t,i}\}$이 백색잡음 과정(white noise series)이라 가정하면 다음과 같이 구할 수 있다.

$$\mathrm{Var}(r_t|F_{t-1}) = n\mathrm{Var}(r_{t,1}) = \frac{n}{n-1}\sum_{i=1}^{n}(r_{t,i} - \bar{r}_t)^2 \approx \sum_{i=1}^{n} r_{t,i}^2.$$

이러한 식에 기반하여 일간 로그 수익률 $r_t$의 실현변동성($\mathrm{RV}_t$)은 다음과 같이 정의한다.

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

$t$일의 실현변동성 $\mathrm{RV}_t$는 $t$일의 일중 로그 수익률의 제곱합으로 정의되며 $n$개의 일중 로그 수익률 $r_{t,i}$ 각각은 평균이 0이고 유한한 분산을 갖는 iid 과정이라 가정한다. 고빈도 자료를 이용하는 경우에는 모형에 대한 가정 없이 단순하게 변동성을 추정할 수 있으며 일중 로그 수익률 정보를 활용한다는 장점이 있다. 하지만 관측간격이 작을 경우 매입-매도 진동과 같은 시장 미시구조의 영향을 받을 수 있으며 $t$일의 변동성을 추정할 때 $t-1$일의 종가와 $t$일의 시가 사이의 야간 수익률을 고려하지 않기 때문에 변동성이 과소 추정될 수 있다는 단점이 있다 (Hansen과 Lunde, 2005).

## 3. 자료 기반 및 모형 기반 방법

### 3.1. 자료 기반 방법: 역사적 변동성과 지수가중이동평균

모형에 대한 가정 없이 과거의 변동성과 수익률 자료를 이용하여 변동성을 추정하는 자료에 기반한 비모수적인 방법으로는 역사적 변동성(historical volatility)과 지수가중이동평균(EWMA)이 있다. 역사적 변동성은 과거 수익률의 변화 형태가 현재에도 비슷하게 반복될 것이라 생각하여 과거 일정기간 동안의 로그 수익률을 이용하여 $t$시점에서의 변동성을 산출하는 방법이며 다음과 같이 계산한다.

$$h_t = \frac{1}{k}\sum_{i=1}^{k}(r_{t-i} - \bar{r})^2.$$

$t$시점에서 $k$시점 전까지의 로그 수익률 $k$개를 이용하여 계산하며 $\bar{r}$은 로그 수익률의 표본평균이다. $k$값을 어떻게 하는지에 따라 고려하는 과거 기간이 달라진다. 일반적으로 일주일 단위($k = 5$), 한 달 단위($k = 22$), 세 달 단위($k = 66$)를 많이 이용한다. 역사적 변동성의 경우 고려하는 과거 기간에 따라 값이 달라지게 된다. 고려하는 과거 기간이 길면 정확성은 높아지나 최근의 변화 경향을 잘 반영하지 못할 수 있고, 고려하는 과거 기간이 짧아지면 최근의 경향은 잘 반영하지만 편의가 발생할 가능성이 있다.

지수가중이동평균은 역사적 변동성과 마찬가지로 과거의 로그 수익률 뿐 아니라 과거의 변동성도 같이 고려하는 방법이며 다음과 같이 구할 수 있다.

$$h_t = \lambda h_{t-1} + (1-\lambda) r_{t-1}^2.$$

$t$시점의 변동성은 한 시점 전의 변동성과 한 시점 전의 로그 수익률의 제곱의 합으로 계산한다. 여기서 $\lambda$는 평활상수이며 0과 1사이의 값을 갖는다. 일반적으로 $\lambda$값은 J.P.Morgan사의 RiskMetrics에서 사용하는 0.94를 이용하며 바로 전 시점의 변동성에 더 큰 비중을 둔다. 이 방법은 현재의 변동성을 추정할 때 과거의 변동성과 수익률을 동시에 고려하되 가중치를 지수적으로 감소시켜 최근 자료에 더 많은 비중을 주는 방법으로 변동성을 추정한다. 최근 자료에 더 많은 비중을 주기 때문에 과거의 모든 값에 동일한 비중을 주는 역사적 변동성에 비해 상대적으로 변동성의 움직임을 잘 설명하는 것으로 알려져 있다. 여기서 초기값은 역사적 변동성 등 자료에서 구한 값을 이용한다. 이 방법 또한 모형을 가정하지 않으며 과거의 변동성과 로그 수익률 자료만을 이용하여 변동성을 산출하여 모형을 가정하는 경우보다 비교적 단순하지만 미래 시점의 변동성을 예측하는 데에는 한계가 있다는 단점이 있다 (Tsay, 2010).

### 3.2. 모형 기반 방법: GARCH 모형과 T-GARCH 모형

Bollerslev (1986)가 제안한 대표적인 시계열 변동성 모형인 GARCH$(p,q)$ 모형은 다음과 같다.

$$\epsilon_t = \sqrt{h_t}e_t, \quad h_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i\epsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j h_{t-j},$$

여기서 $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$이며, $\{e_t\}$는 평균이 0이고 분산이 1인 iid 과정이고 $\sum_{i=1}^{\max(p,q)}(\alpha_i + \beta_i) < 1$이다. 순수 GARCH 모형인 $\epsilon_t$는 무상관 백색잡음이므로 로그 수익률 $r_t$와의 관계식은 $r_t = E(r_t|F_{t-1}) + \epsilon_t$로 구성된다. $t$시점에서의 변동성은 오차항과 과거의 변동성에 의해 계산된다. 오차항이 제곱의 형태로 반영되므로 현재의 변동성에 오차항의 부호는 영향을 미치지 않으며 변동성에 대한 레버리지 효과(leverage effect)를 효과적으로 나타내지 못한다는 단점이 있다. 일차모형인 GARCH$(1,1)$ 모형은 다음과 같다.

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

$t$시점에서 변동성은 한 시점 전의 오차항의 제곱과 한 시점 전의 변동성의 합으로 구할 수 있다. 변동성의 비대칭성을 효과적으로 반영하기 위해 Rabemananjara와 Zakoian (1993)은 다음과 같은 T-GARCH(Threshold-GARCH) 모형을 제안하였다.

$$\epsilon_t = \sqrt{h_t} e_t, \quad h_t = \alpha_0 + \sum_{i=1}^{p} \left[ \alpha_{i1} \left(\epsilon_{t-i}^{+}\right)^2 + \alpha_{i2} \left(\epsilon_{t-i}^{-}\right)^2 \right] + \sum_{j=1}^{q} \beta_j h_{t-j},$$

여기서 $\alpha_{i1} \geq 0$, $\alpha_{i2} \geq 0$, $\beta_j \geq 0$, $\epsilon^+ = \max(\epsilon, 0)$, $\epsilon^- = \max(-\epsilon, 0)$이다. $\alpha_{i1}$와 $\alpha_{i2}$가 다른 값을 가짐으로써 수익률의 상승과 수익률의 하락에 따라 변동성에 미치는 영향을 다르게 반영할 수 있게 되며 비대칭성을 반영할 수 있게 된다. 일반적으로 $\alpha_{i2}$가 $\alpha_{i1}$보다 큰 값을 가지는 경우가 많은데 이는 수익률의 상승보다 수익률의 하락이 변동성에 미치는 영향이 큼을 의미하는 레버리지 효과를 반영하는 것으로 볼 수 있다. 여기서 $\alpha_{i1} = \alpha_{i2}$면 T-GARCH 모형은 GARCH 모형과 같게 된다. 일차모형인 T-GARCH$(1,1)$ 모형은 다음과 같다 (Hwang과 Basawa, 2004).

$$\epsilon_t = \sqrt{h_t} e_t, \quad h_t = \alpha_0 + \alpha_{11} \left(\epsilon_{t-1}^{+}\right)^2 + \alpha_{12} \left(\epsilon_{t-1}^{-}\right)^2 + \beta_1 h_{t-1}.$$

이와 같이 모형을 이용하여 변동성을 추정하는 경우에는 설정한 모형에 대한 모수를 추정한 후 추정된 식을 이용하여 $t$시점에서의 변동성을 계산한다. 추정한 모형을 이용하여 몇 시점 후의 변동성을 예측할 수 있다는 장점이 있다.

### 3.3. 융합(Hybrid) 방법

이번에는 앞에서 소개한 자료에 기반한 방법과 모형에 기반한 방법을 동시에 사용하여 변동성을 추정하는 방법을 제안하고자 한다.

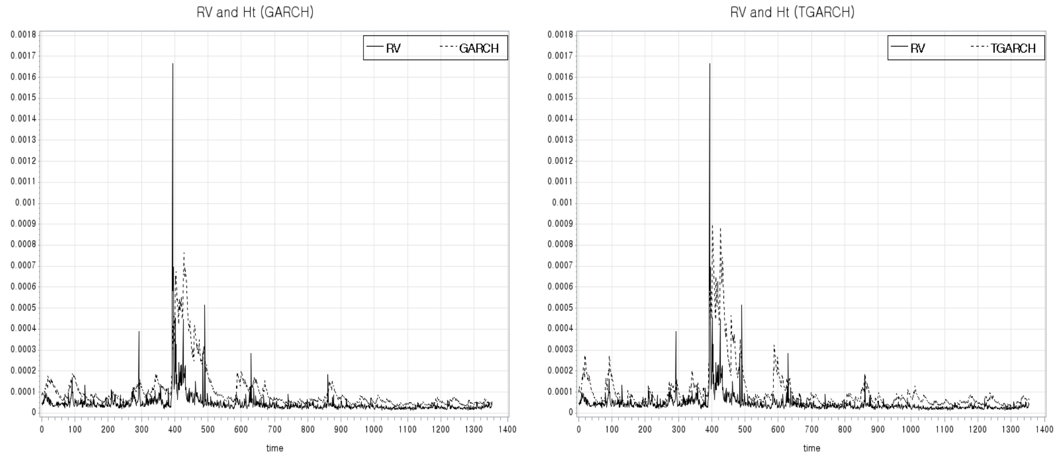
역사적 변동성을 구할 때 과거의 수익률을 이용한다는 점과 모형에 기반을 둔 방법에서 미래의 변동성을 예측할 수 있다는 점을 이용하여 변동성을 추정하는 방법을 생각해 보았다. 기존의 역사적 변동성을 구할 때는 과거 $k$개의 로그 수익률만을 이용하여 값을 추정하였는데 새로 제안하는 방법에서는 과거의 로그 수익률과 미래 변동성 예측값을 동시에 이용하여 $t$시점에서의 변동성을 추정한다. $k = 5$로 하였고 다음과 같이 $t$시점에서의 변동성을 계산한다.

$$h_t = \frac{1}{5} \left\{ r_{t-2}^2 + r_{t-1}^2 + h_t(1) + h_t(2) + h_t(3) \right\}$$

$t$시점에서의 변동성은 바로 전 시점과 두 시점 전의 과거 수익률의 제곱과 $t-1$시점까지의 정보가 주어져 있는 상황에서 모형에 기반한 방법으로 구한 1시점 후, 2시점 후, 3시점 후의 변동성 예측값을 반영하여 추정한 것으로 여기서 $h_t(1)$, $h_t(2)$, $h_t(3)$는 각각 $E(h_t|F_{t-1})$, $E(h_{t+1}|F_{t-1})$, $E(h_{t+2}|F_{t-1})$을 의미한다. 기존의 방법들이 과거 값들로만 현재의 변동성을 추정했다면 이 방법은 과거 자료와 $t-1$시점까지 주어진 정보를 이용하여 예측한 미래의 값까지 고려하여 일종의 중심화(centering)하는 개념으로 $t$시점에서의 변동성을 예측한 것으로 생각할 수 있다. 과거 자료와 예측되는 미래의 변동성까지 함께 고려한다는 점에서 변동성의 움직임을 잘 반영하는데 유용할 것이라 생각된다.

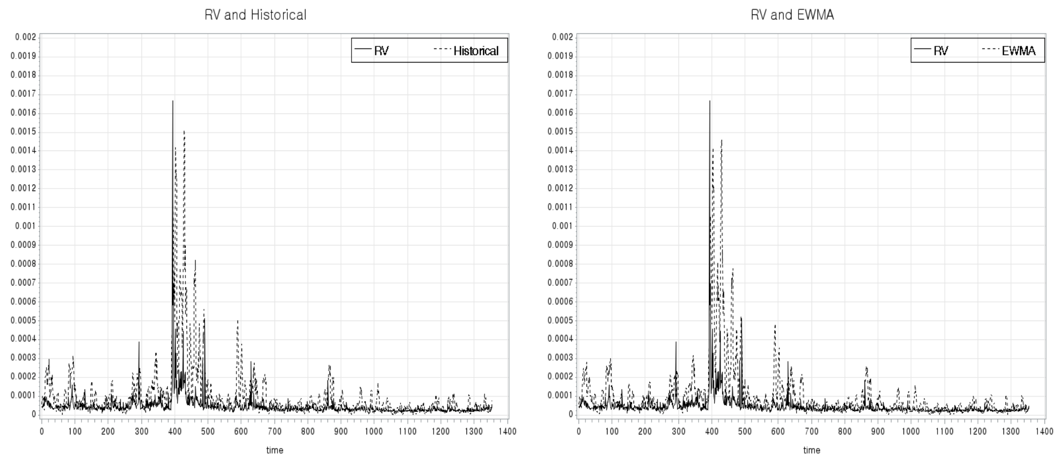
## 4. Data 분석: KOSPI 200

본 절에서는 실제 금융시계열 자료에 앞에서 소개한 방법들로 변동성을 구한 후 비교해 보았다. 2010년 1월 2일부터 2015년 6월 30일까지의 총 1360개의 KOSPI 200자료를 이용하였고 같은 기간 동안 매일 1분 단위로 조사된 고빈도 자료를 이용하였다.




(a) GARCH(1, 1) (b) T-GARCH(1, 1)

**Figure 4.1.** Model based volatility and realized volatility.



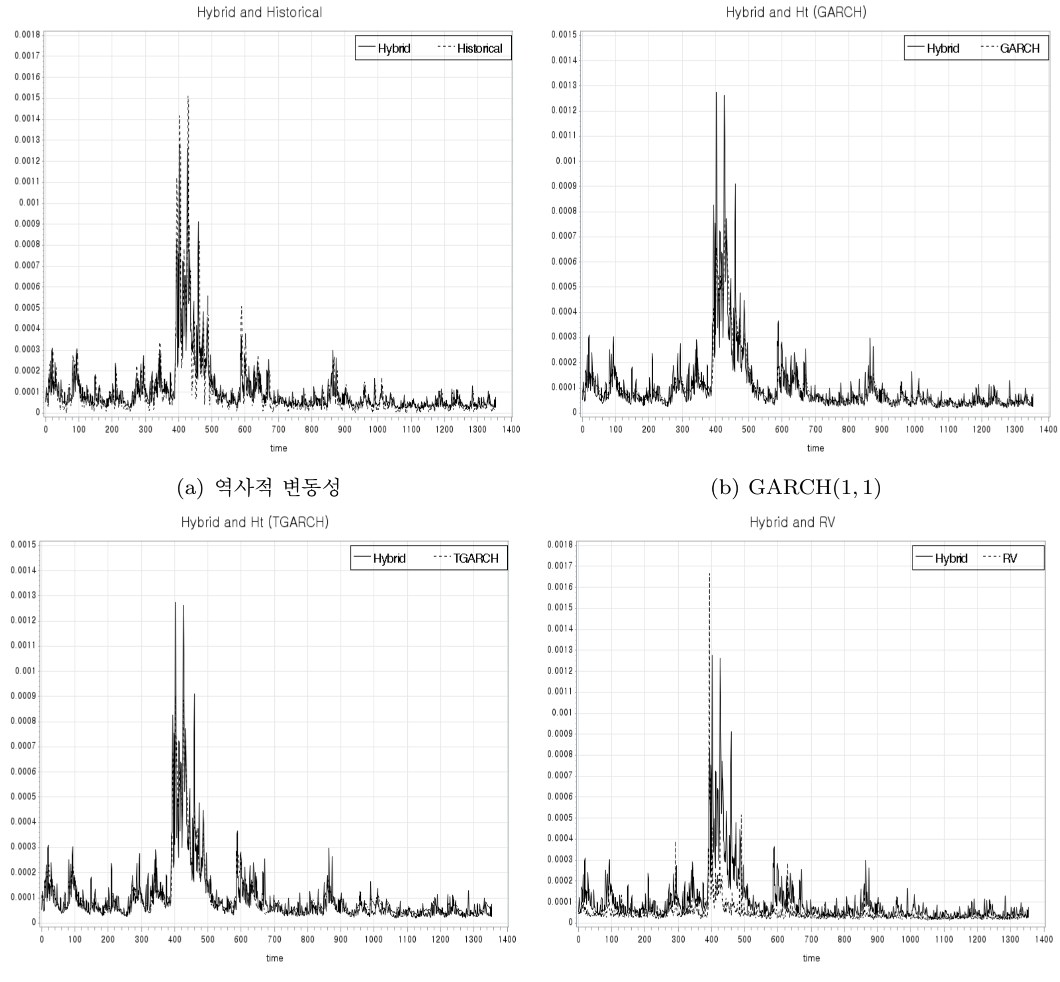

(a) 역사적 변동성 (b) 지수가중이동평균

**Figure 4.2.** Data based volatility and realized volatility.

먼저 로그 수익률을 계산한 후 GARCH(1,1) 모형과 T-GARCH(1,1) 모형을 적합하여 얻은 결과는 다음과 같다. 여기서 자료 $\{r_t\}$의 평균 $\bar{r} = 0$이고 자기상관성을 보이지 않으므로 조건부 평균 $E(r_t|F_{t-1}) = 0$으로서 $r_t = \epsilon_t = \sqrt{h_t}e_t$이다.

$$\text{GARCH}: \ h_t = (1.3165)10^{-6} + 0.0658\epsilon_{t-1}^2 + 0.9194h_{t-1},$$
$$\text{T-GARCH}: \ h_t = (1.972)10^{-6} + \left[0.02291\left(\epsilon_{t-1}^+\right)^2 + 0.115786\left(\epsilon_{t-1}^-\right)^2\right] + 0.92924h_{t-1}.$$

역사적 변동성은 $k = 5$로 하여 일주일 단위로 하였으며 지수가중이동평균에서는 평활상수 $\lambda = 0.94$로 하여 각각 변동성을 계산하였다. 실현변동성은 매일 1분 단위로 조사된 KOSPI자료(360/1일)를 이용




(a) 역사적 변동성

(b) GARCH(1, 1)

(c) T-GARCH(1, 1)

(d) RV (실현변동성)

**Figure 4.3.** Comparisons with hybrid volatility.

하여 1분 단위의 로그 수익률을 계산한 후 제곱합하는 방법으로 매일의 실현변동성을 구하였으며 Hybrid 방법에서는 GARCH(1, 1) 모형을 이용하여 예측값을 구하였다.

Figure 4.1에서는 GARCH(1, 1) 모형, T-GARCH(1, 1) 모형을 이용하여 구한 변동성(점선)보다 실현변동성(실선)이 더 작은 값을 갖는 것을 볼 수 있다. 또한 모형을 가정하고 계산한 변동성이 조금 더 평활된 움직임을 보이고 있다. 관측기간 중 급격한 변화를 보이는 구간은 2011년 8월에서 2012년 1월에 해당되는 구간이며 이는 미국의 신용등급 강등에 의한 영향으로 추정된다. Figure 4.2에서도 역사적 변동성, 지수가중이동평균을 이용하여 구한 변동성(점선)보다 실현변동성(실선)이 더 작은 값을 갖는 것을 볼 수 있으며 모형을 가정하여 구한 변동성보다는 움직임이 더 비슷한 것으로 볼 수 있다. Figure 4.3에서는 새로 제안한 방법을 이용하여 계산한 변동성(실선)과 역사적 변동성(점선), T-GARCH(1, 1) 모형을 이용하여 구한 변동성(점선)을 비교해 보았으며 움직임이 상당히 유사한 것을 볼 수 있다.

**Table 4.1.** Correlations between various volatilities

| | RV | GARCH | T-GARCH | Historical | EWMA | Hybrid |
|---|---|---|---|---|---|---|
| RV | 1 | 0.66601 | 0.69320 | 0.67919 | 0.66488 | 0.68171 |
| GARCH | 0.66601 | 1 | 0.93912 | 0.87308 | 0.88596 | 0.92217 |
| T-GARCH | 0.69320 | 0.93912 | 1 | 0.87155 | 0.87609 | 0.89858 |
| Historical | 0.67919 | 0.87308 | 0.87155 | 1 | 0.94817 | 0.90329 |
| EWMA | 0.66488 | 0.88596 | 0.87609 | 0.94817 | 1 | 0.85979 |
| Hybrid | 0.68171 | 0.92217 | 0.89858 | 0.90329 | 0.85979 | 1 |

RV = realized volatility; EWMA = exponentially weighted moving average.

Table 4.1에서는 각 방법으로 구한 변동성의 양의 제곱근(즉, 표준편차)간의 상관계수를 제시하였다. 고빈도 자료를 이용하여 계산한 실현변동성은 T-GARCH$(1,1)$ 모형을 이용하여 계산한 변동성과 가장 높은 상관관계를 보이고 있고 그 다음으로는 역사적 변동성과 높은 상관관계를 보이고 있다. 새롭게 제안한 Hybrid 방법으로 계산한 변동성은 계산 공식에 사용한 GARCH$(1,1)$ 모형 변동성과 역사적 변동성과 높은 상관관계를 보이고 있다. Hybrid 변동성은 계산 시 이용한 두 방법과 가장 상관관계가 높으나 다른 방법과도 고루 높은 상관관계를 보이고 있다. 또한 Table 4.1을 보면 같은 방법으로 분류되는 방법으로 추정한 변동성 간에 상관관계가 높은 것을 볼 수 있다.

결과를 종합하여 살펴보면 고빈도 자료를 활용하여 산출한 변동성 역시 약간의 차이는 있으나 기존의 모수적인 방법, 비모수적인 방법에 의해 계산된 변동성과 같은 움직임을 보임을 볼 수 있으며 따라서, 변동성을 추정하는데 있어서 일중 수익률을 활용하는 방법도 의미가 있을 것이라 생각된다. 융합방법인 $h_t = (1/5)\{r_{t-2}^2 + r_{t-1}^2 + h_t(1) + h_t(2) + h_t(3)\}$는 역사적 변동성과 GARCH 변동성을 결합한 단순중심이동평균법과 유사한 개념이다. 가중치를 도입한 가중이동평균법인 $h_t^w = w_{-2}r_{t-2}^2 + w_{-1}r_{t-1}^2 + w_0 h_t(1) + w_1 h_t(2) + w_3 h_t(3)$, $\sum w_i = 1$에 대한 연구(예를 들어, 최적 $w_i$의 결정)를 현재 진행 중에 있다.

## References


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

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

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.

Hwang, S. Y. and Basawa, I. V. (2004). Stationarity and moment structure for Box-Cox transformed threshold GARCH$(1,1)$ processes, *Statistics & Probability Letters*, **68**, 209–220.

Rabemananjara, R. and Zakoian, J. M. (1993). Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics*, **8**, 31–49.

Tsay, R. S. (2010). *Analysis of Financial Time Series*, 3rd edition, John Wiley & Sons.

# 금융시계열 변동성 측정 방법의 비교 분석: 고빈도 자료 및 융합 방법

윤재은[a] · 황선영[a,1]

[a]숙명여자대학교 통계학과

(2015년 10월 1일 접수, 2015년 11월 13일 수정, 2015년 11월 26일 채택)

**요 약**

본 연구에서는 금융시계열 변동성 측정을 위한 다양한 방법들을 소개하고 비교분석 하였다. 최근 들어 활발한 연구가 이루어지고 있는 고빈도(high frequency) 자료에 기초한 변동성 측정방법을 국내 주가에 적용시켜 1분 단위 고빈도 주가로부터 일별 변동성을 계산하였다. 또한, 모형 기반 방법인 GARCH와 자료 기반 방법인 역사적 변동성(historical volatility)을 융합하여 새로운 변동성 측정법을 제안하였다.

주요용어: GARCH, 고빈도 자료 기반 방법, 융합 방법.

[1](04310) 서울특별시 용산구 청파로47길 100, 숙명여자대학교 통계학과. E-mail: shwang@sookmyung.ac.kr