The Korean Journal of Applied Statistics (2017) DOI: https://doi.org/10.5351/KJAS.2017.30.5.747
30(5), 747–757

Choice of frequency via principal component in high-frequency multivariate volatility models

M. K. Jin[a] · J. E. Yoon[a] · S. Y. Hwang[a,1]

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

(Received August 3, 2017; Revised September 9, 2017; Accepted September 10, 2017)

Abstract

We investigate multivariate volatilities based on high frequency time series. The PCA (principal component analysis) method is employed to achieve a dimension reduction in multivariate volatility. Multivariate realized volatilities (RV) with various frequencies are calculated from high frequency data and "optimum" frequency is suggested using PCA. Specifically, RVs with various frequencies are compared with existing daily volatilities such as Cholesky, EWMA and BEKK after dimension reduction via PCA. An analysis of high frequency stock prices of KOSPI, Samsung Electronics and Hyundai motor company is illustrated.

Keywords: high frequency, multivariate volatility, principal component

1. 서론

변동성(volatility)은 금융시계열의 대표적인 특징이며 금융현상과 관련한 위험(risk)을 측정하는 척도로 이용된다. 변동성은 관측이 불가능하며, 비슷한 값들끼리 일정기간 클러스터를 이루는 변동성 집중(volatility cluster)과 비대칭 레버리지(leverage) 현상 등이 나타난다는 특징이 있다. 금융시계열에서 변동성은 일반적으로 Engle (1982)의 ARCH 모형 또는 ARCH 모형의 일반화된 모형인 Bollerslev (1986)의 GARCH 모형으로 모형화한다.

다변량 수익률 변동성을 다룰 때에는 각각의 수익률들끼리 동적인 관계(dynamic correlation)가 존재하므로 특수한 모형화가 필요하다. 대표적인 예로 exponential weighted moving average (EWMA) 모형, Baba-Enble-Kraft-Kroner (BEKK) 모형, constant conditional correlation (CCC) 모형, dynamic conditional correlation (DCC) 모형 등이 있다 (Lee와 Hwang, 2017; Hwang 등, 2009). 또한 촐레스키 분해를 이용하여 변동성을 모델링 하는 Cholesky decomposition and volatility modeling (Cholesky 모형)이 있다 (Tsay, 2014).

최근 데이터에 대한 접근이 용이해지면서 일별 종가뿐 아니라 1분 빈도 자료, 5분 빈도 자료 등과 같은 고빈도 자료(high-frequency data)를 이용하여 변동성을 추정하는 연구가 활발히 진행되고 있다. 변동성을 추정하기 위해 실현변동성(realized volatility; RV), 조정된 실현변동성(RV$^{(a)}$), 융합(hybrid) 방법 등을 이용할 수 있다 (Xiao, 2013; Yoon과 Hwang, 2015).

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

본 연구에서는 다변량 고빈도 금융시계열자료에서 변동성을 추정하고자 할 때 몇 분 빈도의 실현변동성을 이용하는 것이 효과적일지 알아보고자 한다. 이를 위해 고빈도 자료에서 1분에서 5분까지로 빈도 단위를 다르게 하여 실현변동성을 구한다. 일별 종가에 대해 5가지 모형(Cholesky, EWMA, BEKK, DCC, CCC)을 이용하여 변동성을 추정한 후 주성분 분석을 통해 다변량 변동성들의 차원 축소를 실시하였으며 제1주성분 간 비교를 통해 실현변동성을 구성하는데 있어서 가장 적절한 시간 단위가 몇 분인지 알아보았다.

2. 다변량 변동성 모형

본 절에서는 대표적인 다변량 변동성 모형(multivariate volatility model)에 대해 살펴보고자 한다. t일에서의 k개의 로그 수익률을 $r_t = (r_{1t}, \ldots, r_{kt})^T$로 나타내었으며 다음과 같이 모형화 한다.

$$r_t = \mu_t + a_t,$$
$$a_t = \Sigma_t^{\frac{1}{2}} e_t,$$

여기서 r_t는 k개인 수익률 벡터이고 μ_t는 평균 수익률 벡터, $a_t = (a_{1t}, \ldots, a_{kt})^T$는 오차항 벡터를 나타내며 Σ_t는 $k \times k$ 양정치 행렬이고 e_t는 $E(e_t) = 0$이고 $\mathrm{Var}(e_t) = I_k$인 벡터이다. 여기서 Σ_t는 다변량 변동성이라 지칭하는 a_t의 조건부 분산 공분산 행렬 $\Sigma_t = \mathrm{Cov}(a_t|F_{t-1})$이며, F_{t-1}은 $t-1$시점까지 주어진 정보를 나타낸다.

Σ_t을 모형화하는 Cholesky 모형, EWMA, BEKK, CCC 및 DCC 모형에 대해 살펴보도록 하자 (cf., Choi, 2009). 최근에 Lee와 Hwang (2017)은 다변량 변동성 모형에서 정준상관분석을 통한 연구를 수행하였다.

2.1. Cholesky 모형

이 방법은 촐레스키 분해(Cholesky decomposition)를 이용하는 방법으로서 자세한 내용은 Tsay (2014, 2010)를 참고하기 바란다. 오차항 a_t에 대해 $b_{i,t}$라는 새로운 변수를 이용하여 다음과 같은 선형식으로 나타내며 회귀분석모형으로 접근한다.

$$a_{kt} = \beta_{k1,t} a_{1t} + \cdots + \beta_{k,k-1,t} a_{k-1,t} + b_{kt}$$

최소제곱법에 의해 다음이 성립한다.

$$\beta_{kj,t} = \frac{\mathrm{Cov}(a_{kt}, b_{jt}|F_{t-1})}{\mathrm{Var}(b_{jt}|F_{t-1})}$$
$$\mathrm{Var}(b_{kt}|F_{t-1}) = \mathrm{Var}(a_{kt}|F_{t-1}) - \beta_{k,k-1,t}^2 \mathrm{Var}(b_{k-1,t}|F_{t-1})$$

즉, 다음 행렬방정식을

$$\begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -\beta_{21,t} & 1 & 0 & \ldots & 0 & 0 \\ -\beta_{31,t} & -\beta_{32,t} & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\beta_{k1,t} & -\beta_{k2,t} & -\beta_{k3,t} & \ldots & -\beta_{k,k-1,t} & 1 \end{pmatrix} \begin{pmatrix} a_{1t} \\ a_{2t} \\ a_{3t} \\ \vdots \\ a_{kt} \end{pmatrix} = \begin{pmatrix} b_{1t} \\ b_{2t} \\ b_{3t} \\ \vdots \\ b_{kt} \end{pmatrix}$$

$\beta_t a_t = b_t$으로 표현하고 $a_t = \beta_t^{-1} b_t$을 얻을 수 있다. b_t의 변동성 행렬로 다시 정리하면 다음과 같다. 여기서 $\Sigma_{b,t}$는 대각행렬이다.

$$\Sigma_t = \beta_t^{-1} \Sigma_{b,t} \left(\beta_t^{-1}\right)^T$$

Σ_t의 행렬식은 다음과 같이 구할 수 있으며 여기서 $\sigma_{bi,t}$는 정보 F_{t-1}하에 b의 변동성이다.

$$|\Sigma_t| = |\Sigma_{b,t}| = \prod_{i=1}^{k} \sigma_{bi,t}$$

위의 행렬식을 이용하여 다음과 같은 다섯 가지 과정을 거쳐 다변량 변동성을 모델링 한다.

(1) 순환최소제곱(recursive least squares; RLS) 계산을 적용한다. i번째 선형회귀식에서 추정값을 $\hat{\beta}_{i,t}$라고 하고 i는 2부터 k까지이다. $\hat{\beta}_{i,t}$는 $i-1$차원 벡터이다.

(2) β_t의 평활추정치(smoothed estimate)를 얻기 위해 λ가 0.94인 EWMA과정을 적용한다. 특히 $\hat{\mu}_i$는 $\hat{\beta}_{i,t}$의 표본평균이고 편차벡터는 $\hat{\beta}^*_{i,t}$이다. $\beta_{i,t}$의 평활추정치 $\tilde{\beta}_{i,t}$를 계산한다.

$$\begin{aligned}
\hat{\beta}^*_{i,t} &= \hat{\beta}_{i,t} - \hat{\mu}_i,\\
\tilde{\beta}_{i,t} &= \tilde{\beta}^*_{i,t} + \hat{\mu}_i,\\
\tilde{\beta}^*_{i,t} &= \lambda \tilde{\beta}^*_{i,t-1} + (1-\lambda)\hat{\beta}^*_{i,t}, \quad t = 2, 3, \ldots.
\end{aligned}$$

평활(smoothing)의 초기값은 $\tilde{\beta}^*_{i,1} = \hat{\beta}^*_{i,1}$이다.

(3) $\hat{b}_{1t} = a_{1t}$라고 하자. 잔차계열(residual series)을 구하면 다음과 같다.

$$\begin{aligned}
\hat{b}_{it} &= a_{it} - \mathbf{a}_{\mathbf{i,t}}^T \tilde{\beta}_{i,t}, \quad i = 2, \ldots, k\\
\mathbf{a}_{\mathbf{i,t}} &= (a_{1t}, \ldots, a_{i-1,t})^T
\end{aligned}$$

(4) 각각의 $\hat{b}_{1t}$ series에 단변량 GARCH를 적합 시키고 i는 1부터 k까지의 조건부 분산 $\hat{\sigma}^2_{b,i,t}$를 얻는다.

(5) $\tilde{\beta}_{i,t}$를 사용해서 적합된 변동성 행렬 $\hat{\Sigma}_t$을 계산한다.

촐레스키 분해를 이용한 접근은 양정치 행렬 Σ_t를 위한 모수가 필요 없다는 장점이 있다. 또한 추정 시 이용하는 우도함수(likelihood function)가 단순하기 때문에 직교변환을 이용하여 재매개변수화(re-parameterization) 과정을 할 수 있다는 점 또한 장점이 된다. 통계 프로그램 R에서 구현할 수 있으며 구체적인 방법은 Tsay (2014)를 참고하기 바란다.

2.2. Exponential weighted moving average 모형

다음은 EWMA모형을 이용한 변동성 추정 방법이다.

$$\hat{\Sigma}_t = (1-\lambda) a_{t-1} a_{t-1}^T + \lambda \hat{\Sigma}_{t-1},$$

여기서 λ는 0과 1사이의 평활 상수이다. 일반적으로 λ값은 J.P.Morgan 사의 RiskMetrics에서 사용하는 0.94를 이용하며 직전 변동성인 $\hat{\Sigma}_{t-1}$에 더 큰 가중치를 주는 것을 볼 수 있다.

2.3. Baba-Enble-Kraft-Kroner 모형

BEKK 모형의 가장 단순한 형태인 BEKK$(1,1)$은 다음과 같다.

$$\Sigma_t = CC^T + A\left(a_{t-1}a_{t-1}^T\right)A^T + B\Sigma_{t-1}B^T,$$

여기서 C는 하삼각행렬이고, A와 B는 $k \times k$ 행렬이다. CC^T가 양정치 행렬이면 Σ_t는 항상 양정치 행렬이 되며 조건부 분산 공분산행렬이 양정치 행렬이라는 제약조건을 만족시킨다.

2.4. Constant conditional correlation 모형

CCC 모형은 다음과 같다.

$$\Sigma_t = D_tRD_t = \rho_{ij}\sqrt{\Sigma_{iit}\Sigma_{jjt}},$$

여기서 $D_t = \mathrm{diag}(\Sigma_{11t}^{1/2}, \ldots, \Sigma_{kkt}^{1/2})$이며 $R = \rho_{ij}$로 상관계수 행렬 R을 상수로 고정했다는 특징이 있다. t시점에서 i번째 수익률의 조건부 분산은 Σ_{iit}이고 단변량 GARCH로 모형화 된다. CCC모형은 상관계수 행렬을 상수로 고정함으로 추정할 모수의 개수를 줄였다는 장점이 있으나 시간의 흐름에 따른 변화를 반영하지 못하고 고정되어 있다는 단점이 있다 (Choi 등, 2009).

2.5. Dynamic conditional correlation 모형

Engle (2002)은 다음과 같은 DCC 모형을 제안하였다.

$$\Sigma_t = D_tRD_t,$$

여기서 $D_t = \mathrm{diag}(\Sigma_{11t}^{1/2}, \ldots, \Sigma_{kkt}^{1/2})$이고 Σ_{iit}는 단변량 GARCH로 모형화된다는 점은 위의 CCC모형과 동일하나 상관계수를 다음과 같이 나타내었다는 차이점이 있다.

$$R = J_tQ_tJ_t.$$

조건부 상관계수 행렬 R을 구성하는 J_t와 Q_t는 다음과 같다. Q_t는 $k \times k$양정치 행렬, $\bar{Q}$는 비조건부 분산-공분산 행렬이며 ϵ_t의 원소는 표준화된 오차벡터이다.

$$\begin{aligned} J_t &= \mathrm{diag}\left(q_{11t}^{-\frac{1}{2}}, \ldots, q_{kkt}^{-\frac{1}{2}}\right) \\ Q_t &= q_{ij,t} = (1-\theta_1-\theta_2)\bar{Q} + \theta_1 Q_{t-1} + \theta_2\epsilon_{t-1}\epsilon_{t-1}^T \\ \epsilon_{i,t} &= a_{i,t}/\sqrt{\sigma_{ii,t}}, \end{aligned}$$

여기서 θ_1과 θ_2는 음이 아닌 모수로 $0 < \theta_1 + \theta_2 < 1$을 만족한다 (cf., Tsay, 2010, Ch.10). DCC모형은 조건부 상관계수 행렬의 시간의 따른 변화를 반영한다는 장점이 있지만 그 역할을 하는 계수는 θ_1과 θ_2로 고정되어 있어 모든 변수에 대해 동일하게 적용된다는 한계점이 있다.

3. 고빈도 다변량 변동성

금융시계열에서 이용하는 일별 자료는 대부분의 경우 하루의 종가자료이므로 일간(intra-day)변동성에 대한 많은 정보가 누락되어 있다고 볼 수 있다. 본 절에서는 고빈도 자료를 활용하여 변동성을 추정하는 방법에 대해 알아보도록 한다. 선행연구로는 Andersen과 Bolerslev (1997), Lee와 Hwang (2017), Yoon과 Hwang (2015), Oh와 Shin (2012)이 있다.

3.1. 실현변동성

일간 로그 수익률(daily log return) 벡터 $r_t = (r_{1t}, \ldots, r_{kt})^T$는 t일에서 k개의 로그 수익률로 구성된 벡터를 나타내며, 일중 로그 수익률(intra-day log return) $r_{t,i}$벡터는 t일에서 일정한 간격으로 n개 추출한 t일의 i번째 관측시점 로그수익률로 다음과 같다.

$$r_{t,i} = (r_{1t,i}, \ldots, r_{kt,i})^T, \quad i = 1, \ldots, n.$$

일간 로그 수익률 r_t는 일중 로그 수익률 $r_{t,i}$ 벡터의 n개의 합으로 계산한다.

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

r_t의 조건부 분산 공분산 행렬은 다음과 같다 (Tsay, 2010, Chapter 3).

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

일중 로그 수익률 벡터 $r_{t,i}$가 백색잡음과정(white noise series)라고 가정하면 조건부 분산-공분산 행렬은 다음과 같다.

$$\mathrm{Var}(r_t|F_{t-1}) = n\mathrm{Var}(r_{t,i}) = n\left[E\left(r_{t,i}r_{t,i}^T\right) - E(r_{t,i})E(r_{t,i})^T\right],$$

여기서 $E(r_{t,i}) = 0$으로 가정하면, 일간 로그 수익률 r_t의 실현변동성(RV_t)은 다음과 같다.

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

이러한 실현변동성은 시장미시구조 잡음이 없다는 가정 하에 수익률의 "제곱 합"으로 계산된 방법이다. 만약 n이 무한대로 커진다면, 변동성의 일치추정량이 된다는 것이 알려져 있다 (Andersen 등, 2003). 하지만 이러한 실현변동성은 수익률이 자기상관을 갖지 않는다는 가정 하에 성립한 것이며 실제 고빈도 자료는 시장미시구조 잡음이 존재하므로 고빈도 자료가 자기상관(autocorrelation) 구조를 가지게 되어 실현변동성에서 편의(bias)가 발생한다.

3.2. 수정 실현변동성

실현변동성(RV_t)에서 생성된 편의를 제거하면서 데이터의 정보손실 또한 막는 방법으로 Zhou (1996)는 1차 자기상관(MA(1))을 고려한 실현변동성을 제안하였다. $r_{t,i} = (r_{1t,i}, \ldots, r_{kt,i})^T$이 MA(1) 구조를 따르며 $E(r_t) = E(\sum_{i=1}^{n} r_{t,i}) = 0$이라고 가정 하에 r_t의 조건부 분산-공분산 행렬은 다음과 같다.

$$\begin{aligned}
\mathrm{Var}(r_t|F_{t-1}) &= \mathrm{Cov}\left(\sum_{i=1}^{n} r_{t,i}, \sum_{i=1}^{n} r_{t,i}|F_{t-1}\right) = E\left[\left(\sum_{i=1}^{n} r_{t,i}\right)\left(\sum_{i=1}^{n} r_{t,i}\right)^T |F_{t-1}\right] \\
&= \sum_{i=1}^{n} E\left(r_{t,i}r_{t,i}^T\right) + \sum_{i=2}^{n} E\left(r_{t,i}r_{t,i-1}^T\right) + \sum_{i=1}^{n-1} E\left(r_{t,i}r_{t,i+1}^T\right) \\
&= \sum_{i=1}^{n} r_{t,i}r_{t,i}^T + \frac{n}{n-1}\sum_{i=2}^{n} r_{t,i}r_{t,i-1}^T + \frac{n}{n-1}\sum_{i=1}^{n-1} r_{t,i}r_{t,i+1}^T.
\end{aligned}$$

이를 통해 편의 수정 실현변동성은 다음과 같이 구할 수 있다.

$$\sum_{i=1}^{n} r_{t,i} r_{t,i}^T + \frac{n}{n-1} \sum_{i=2}^{n} r_{t,i} r_{t,i-1}^T + \frac{n}{n-1} \sum_{i=1}^{n-1} r_{t,i} r_{t,i+1}^T.$$

일중 로그 수익률이 음의 자기 상관을 따른다면 위 식이 음의 값을 가질 수 있다는 문제가 발생한다. 이를 해결하기 위해 Hansen과 Lunde (2006)는 Bartlett 커널을 사용하여 음이 아닌(nonnegative) 실현변동성을 제안하였다. 이 방법은 일중 로그 수익률 $r_{t,i}$가 MA(1)을 가지며 양의 변동성만을 갖는 수정 실현변동성($\mathrm{RV}_{\mathrm{AC},t}$)이다 (cf., Lee와 Hwang, 2017).

$$\mathrm{RV}_{\mathrm{AC},t} \equiv \sum_{i=1}^{n} r_{t,i} r_{t,i}^T + \frac{n}{n-1} \sum_{i=1}^{n-1} r_{t,i} r_{t,i+1}^T.$$

4. 주성분을 통한 RV선택

다변량 자료의 변동성 분석에 있어서, 예를 들어, 3차원 자료에서의 변동성을 모델링하는 경우 조건부 분산 공분산 행렬에서 각 변수의 분산 뿐 아니라 두 변수들 간의 공분산도 등장하므로 공통된 공분산을 제외하고 6개의 변수로 분산 공분산 행렬이 구성된다. 6차원의 변수들을 효과적으로 비교하기 위하여 주성분 분석방법을 이용하여 차원 축소할 수 있다. 주성분에 대한 기본 개념은 Seong (1997)을 참고하기 바란다.

4.1. 주성분(principal component)

상관행렬을 이용한 주성분 분석을 실시한다. 확률벡터 X에 대응하는 표준화 벡터를 Z라 하자. 표준화 벡터 Z의 공분산 행렬은 X의 상관행렬이다. X의 상관행렬에서 i번째 주성분을 구할 때 고유값을 λ_i, 고유벡터를 e_i라고 할 때 이에 대응하는 i번째 주성분 C_i는 다음과 같다.

$$C_i = e_i^T Z = e_{1i} Z_1 + \cdots + e_{ip} Z_p, \quad i = 1, 2, \ldots, p.$$

전체분산(p)에 대한 주성분 C_i의 설명비율은 λ_i/p이다. 본 논문에서는 가장 높은 설명력을 보이는 주성분인 "제1주성분"을 이용하였다.

4.2. RV의 선택 절차

예를 들어, 3차원 자료에서의 변동성 행렬은 다음과 같다.

$$\Sigma_t = \begin{pmatrix} \sigma_{11,t} & \sigma_{12,t} & \sigma_{13,t} \\ \sigma_{21,t} & \sigma_{22,t} & \sigma_{23,t} \\ \sigma_{31,t} & \sigma_{32,t} & \sigma_{33,t} \end{pmatrix}.$$

Σ_t는 대칭행렬(symmetric)이므로 다음과 같은 총 6개의 변수를 생성한 후

$$\sigma_{11,t} = X_1, \qquad \sigma_{22,t} = X_2, \qquad \sigma_{33,t} = X_3,$$
$$\sigma_{12,t} = \sigma_{21,t} = X_4, \qquad \sigma_{13,t} = \sigma_{31,t} = X_5, \qquad \sigma_{23,t} = \sigma_{32,t} = X_6.$$

X의 표준화벡터 $Z^T = [Z_1, \ldots, Z_6]$의 분산 공분산 행렬 ρ에서 주성분을 구성한다.

Table 5.1. Volatility estimates for various methods

Model	Parameter estimates
Cholesky	$E[r_t] = E\left[(r_{kt}, r_{st}, r_{ht})^T\right] = (0.000148, 0.00033, 0.000098)^T$
	$\sigma^2_{kt} = 0.000001 + 0.066122 b^2_{k,t-1} + 0.918231 \sigma^2_{k,t-1}$
	$\sigma^2_{st} = 0.000006 + 0.046443 b^2_{s,t-1} + 0.918088 \sigma^2_{s,t-1}$
	$\sigma^2_{ht} = 0.000009 + 0.042684 b^2_{h,t-1} + 0.926651 \sigma^2_{h,t-1}$
EWMA	$\hat{\Sigma}_t = (1-0.94) a_{t-1} a^T_{t-1} + 0.94 \hat{\Sigma}_{t-1}$
BEKK	$E[r_t] = (0.000147956, 0.000330439, 0.0000981848)^T$
	$\begin{pmatrix} \sigma_{kk,t} & \sigma_{ks,t} & \sigma_{kh,t} \\ \sigma_{sk,t} & \sigma_{ss,t} & \sigma_{sh,t} \\ \sigma_{hk,t} & \sigma_{hs,t} & \sigma_{hh,t} \end{pmatrix} = \begin{pmatrix} 0.010281 & 0 & 0 \\ 0.0120711 & 0.0126898 & 0 \\ 0.0123161 & -0.0016708 & 0.0172950 \end{pmatrix} \begin{pmatrix} 0.010281 & 0.0120711 & 0.0123161 \\ 0 & 0.0126898 & -0.0016708 \\ 0 & 0 & 0.0172950 \end{pmatrix}$
	$+ \begin{pmatrix} 0.1 & 0.02 & 0.02 \\ 0.02 & 0.1 & 0.02 \\ 0.02 & 0.02 & 0.1 \end{pmatrix} \begin{pmatrix} a^2_{k,t-1} & a_{k,t-1}a_{s,t-1} & a_{k,t-1}a_{h,t-1} \\ a_{s,t-1}a_{k,t-1} & a^2_{s,t-1} & a_{s,t-1}a_{h,t-1} \\ a_{h,t-1}a_{k,t-1} & a_{h,t-1}a_{s,t-1} & a^2_{h,t-1} \end{pmatrix} \begin{pmatrix} 0.1 & 0.02 & 0.02 \\ 0.02 & 0.1 & 0.02 \\ 0.02 & 0.02 & 0.1 \end{pmatrix}$
	$+ \begin{pmatrix} 0.8 & 0.02 & 0.02 \\ 0.02 & 0.8 & 0.02 \\ 0.02 & 0.02 & 0.8 \end{pmatrix} \begin{pmatrix} \sigma_{kk,t-1} & \sigma_{ks,t-1} & \sigma_{kh,t-1} \\ \sigma_{sk,t-1} & \sigma_{ss,t-1} & \sigma_{sh,t-1} \\ \sigma_{hk,t-1} & \sigma_{hs,t-1} & \sigma_{hh,t-1} \end{pmatrix} \begin{pmatrix} 0.8 & 0.02 & 0.02 \\ 0.02 & 0.8 & 0.02 \\ 0.02 & 0.02 & 0.8 \end{pmatrix}$
CCC	$R = \begin{pmatrix} 1 & 0.69179 & 0.55929 \\ 0.69179 & 1 & 0.32540 \\ 0.55929 & 0.32540 & 1 \end{pmatrix}$
	$\sigma_{kk,t} = 0 + 0.5246 a^2_{k,t-1} + 0.91297 \sigma_{kk,t-1}$
	$\sigma_{ss,t} = 0.00001 + 0.02585 a^2_{s,t-1} + 0.94128 \sigma_{ss,t-1}$
	$\sigma_{hh,t} = 0.00004 + 0.04475 a^2_{h,t-1} + 0.87329 \sigma_{hh,t-1}$
DCC	$E[r_t] = E\left[(r_{kt}, r_{st}, r_{ht})^T\right] = (0.0001479562, 0.0003304391, 0.00009818485)^T$
	$\sigma_{kk,t} = 0.000001 + 0.066174 a^2_{k,t-1} + 0.918508 \sigma_{kk,t-1}$
	$\sigma_{ss,t} = 0.000006 + 0.029519 a^2_{s,t-1} + 0.950839 \sigma_{ss,t-1}$
	$\sigma_{hh,t} = 0.000021 + 0.041653 a^2_{h,t-1} + 0.910347 \sigma_{hh,t-1}$
	$Q_t = (1 - 0.92 - 0.04391025)\bar{Q} + 0.92 Q_{t-1} + 0.04391025 \epsilon_{t-1}\epsilon^T_{t-1}$

Cholesky = Cholesky decomposition and volatility modeling; EWMA = exponential weighted moving average; BEKK = Baba-Enble-Kraft-Kroner; CCC = constant conditional correlation; DCC = dynamic conditional correlation.

본 연구에서는 2절에서 말한 5가지 모형에서 얻은 주성분, 실현변동성(RV)과 수정 실현변동성($\mathrm{RV}_{\mathrm{AC}}$)에서의 주성분을 비교하고자 한다. 5가지 모형, Cholesky 모형, EWMA 모형, BEKK 모형, CCC 모형 및 DCC 모형의 주성분을 차례대로 $C_{\mathrm{Ch}}, C_E, C_B, C_C, C_D$라고 하고, 1분에서 5분까지 각 분 단위로 계산한 실현변동성과 수정 실현변동성에서의 주성분을 각각 $C_{\mathrm{RV}i}$와 $C_{\mathrm{RV_{AC}}i}$로 나타내자. 여기서 i는 1부터 5까지의 값을 갖는다. 제1주성분을 통한 차원 축소를 하였으므로 $C_{\mathrm{Ch}}, C_E, C_B, C_C, C_D$와 높은 상관계수를 가진 $C_{\mathrm{RV}i}$와 $C_{\mathrm{RV_{AC}}i}$이 몇 분 단위의 고빈도 자료를 이용한 것인지 살펴보면 실현 변동성을 구성하는데 있어서 “최적” 시간 간격(frequency)을 선택할 수 있을 것이다.

5. 예제를 통한 분석

KOSPI 자료와 KOSPI의 시가총액 상위 종목인 삼성전자, 현대차 자료를 이용해서 3변량 실증분석을

Table 5.2. First principal component for various methods

Model	Explained proportion	First PC
Cholesky	95.86%	$C_{\mathrm{Ch}} = 0.414032Z_{\mathrm{Ch}kk} + 0.405790Z_{\mathrm{Ch}ss} + 0.391663Z_{\mathrm{Ch}hh}$ $+0.414892Z_{\mathrm{Ch}ks} + 0.411085Z_{\mathrm{Ch}kk} + 0.411566Z_{\mathrm{Ch}sh}$
EWMA	82.45%	$C_E = 0.430943Z_{Ekk} + 0.387871Z_{Ess} + 0.332678Z_{Ehh}$ $+0.433195Z_{Eks} + 0.433769Z_{Ekh} + 0.421136Z_{Esh}$
BEKK	94.99%	$C_B = 0.415463Z_{Bkk} + 0.410150Z_{Bss} + 0.385428Z_{Bhh}$ $+0.414087Z_{Bks} + 0.416624Z_{Bkh} + 0.406903Z_{Bsh}$
CCC	82.21%	$C_C = 0.418996Z_{Ckk} + 0.388144Z_{Css} + 0.338870Z_{\mathrm{Ch}h}$ $+0.429515Z_{Cks} + 0.440334Z_{Ckh} + 0.424944Z_{Csh}$
DCC	80.14%	$C_D = 0.438309Z_{Dkk} + 0.397768Z_{Dss} + 0.283249Z_{Dhh}$ $+0.444249Z_{Dks} + 0.440858Z_{Dkh} + 0.421572Z_{Dsh}$

Cholesky = Cholesky decomposition and volatility modeling; EWMA = exponential weighted moving average; BEKK = Baba-Enble-Kraft-Kroner; CCC = constant conditional correlation; DCC = dynamic conditional correlation.

Table 5.3. First principal component for RV with various frequencies

RV	Explained proportion	First PC
1분	51.46%	$C_{\mathrm{RV1}} = 0.538132Z_{\mathrm{RV1}kk} + 0.534800Z_{\mathrm{RV1}ss} + 0.222568Z_{\mathrm{RV1}hh}$ $+ 0.099669Z_{\mathrm{RV1}ks} + 0.332669Z_{\mathrm{RV1}kh} + 0.504246Z_{\mathrm{RV1}sh}$
2분	77.17%	$C_{\mathrm{RV2}} = 0.442299Z_{\mathrm{RV2}kk} + 0.429815Z_{\mathrm{RV2}ss} + 0.265178Z_{\mathrm{RV2}hh}$ $+ 0.428217Z_{\mathrm{RV2}ks} + 0.431719Z_{\mathrm{RV2}kh} + 0.423746Z_{\mathrm{RV2}sh}$
3분	79.39%	$C_{\mathrm{RV3}} = 0.440795Z_{\mathrm{RV3}kk} + 0.421600Z_{\mathrm{RV3}ss} + 0.290587Z_{\mathrm{RV3}hh}$ $+ 0.428942Z_{\mathrm{RV3}ks} + 0.433564Z_{\mathrm{RV3}kh} + 0.414177Z_{\mathrm{RV3}sh}$
4분	79.56%	$C_{\mathrm{RV4}} = 0.433437Z_{\mathrm{RV4}kk} + 0.404105Z_{\mathrm{RV4}ss} + 0.309480Z_{\mathrm{RV4}hh}$ $+ 0.433968Z_{\mathrm{RV4}ks} + 0.436633Z_{\mathrm{RV4}kh} + 0.417226Z_{\mathrm{RV4}sh}$
5분	79.72%	$C_{\mathrm{RV5}} = 0.427798Z_{\mathrm{RV5}kk} + 0.406126Z_{\mathrm{RV5}ss} + 0.328965Z_{\mathrm{RV5}hh}$ $+ 0.431954Z_{\mathrm{RV5}ks} + 0.436096Z_{\mathrm{RV5}kh} + 0.408741Z_{\mathrm{RV5}sh}$

실시하였다. 2010년 1월 2일부터 2015년 6월 30일까지의 총 1,360개의 일별 자료 KOSPI, 삼성전자, 현대차의 일일 종가자료와 더불어 같은 기간 1분 단위로 측정된 고빈도 자료를 이용하였다. KOSPI, 삼성전자, 현대차의 로그수익률을 각각 r_{kt}, r_{st}, r_{ht}라 하자.

$$r_t = (r_{kt}, r_{st}, r_{ht})^T = \text{세 개의 수익률로 이루어진 벡터 수익률}$$
$$r_{t,i} = (r_{kt,i}, r_{st,i}, r_{ht,i})^T = t\,\text{일에서의 일중 로그수익률}$$

9시부터 15시까지 운영되는 국내 주식시장은 1분 단위로 관측되었다면 $n = 360$이지만 마감 전 10분 동안의 동시호가(유가증권 매매 거래 시 동시에 접수된 호가 또는 시간의 선후가 분명하지 않은 호가)를 제외하고 관측된 수익률의 개수는 $n = 350$이다. 2분단위로 관측되었다면 $n = 175$, 3분단위일 경우 $n = 116$, 4분단위일 경우 $n = 87$, 5분 단위일 경우 $n = 70$이 된다. Table 5.1은 2절에서 소개한 모형을 추정한 결과이다.

Cholesky 모형이 앞 시간 구간에 36개의 결측치를 제외한 2010년 2월 25일부터 2015년 6월 30일까지 자료를 이용하였으므로 주성분을 구할 때는 다른 모든 모형들(EWMA, BEKK, CCC, DCC) 그리고 RV와 $\mathrm{RV_{AC}}$에서도 동일한 구간의 자료를 이용하였다. KOSPI, 삼성전자와 현대차의 변동성을 표준화

Table 5.4. First principal component for $\mathrm{RV_{AC}}$ with various frequencies

$\mathrm{RV_{AC}}$	Explained proportion	First PC
1분	79.84%	$C_{\mathrm{RV_{AC}}1} = 0.442182Z_{\mathrm{RV_{AC}}1kk} + 0.426551Z_{\mathrm{RV_{AC}}1ss} + 0.262267Z_{\mathrm{RV_{AC}}1hh} + 0.431823Z_{\mathrm{RV_{AC}}1ks} + 0.433460Z_{\mathrm{RV_{AC}}1kh} + 0.423541Z_{\mathrm{RV_{AC}}1sh}$
2분	82.23%	$C_{\mathrm{RV_{AC}}2} = 0.433732Z_{\mathrm{RV_{AC}}2kk} + 0.409666Z_{\mathrm{RV_{AC}}2ss} + 0.311305Z_{\mathrm{RV_{AC}}2hh} + 0.430246Z_{\mathrm{RV_{AC}}2ks} + 0.432604Z_{\mathrm{RV_{AC}}2kh} + 0.418188Z_{\mathrm{RV_{AC}}2sh}$
3분	81.34%	$C_{\mathrm{RV_{AC}}3} = 0.433295Z_{\mathrm{RV_{AC}}3kk} + 0.403731Z_{\mathrm{RV_{AC}}3ss} + 0.323322Z_{\mathrm{RV_{AC}}3hh} + 0.430457Z_{\mathrm{RV_{AC}}3ks} + 0.434678Z_{\mathrm{RV_{AC}}3kh} + 0.412999Z_{\mathrm{RV_{AC}}3sh}$
4분	80.40%	$C_{\mathrm{RV_{AC}}4} = 0.435610Z_{\mathrm{RV_{AC}}4kk} + 0.394299Z_{\mathrm{RV_{AC}}4ss} + 0.320602Z_{\mathrm{RV_{AC}}4hh} + 0.432384Z_{\mathrm{RV_{AC}}4ks} + 0.436337Z_{\mathrm{RV_{AC}}4kh} + 0.417900Z_{\mathrm{RV_{AC}}4sh}$
5분	78.97%	$C_{\mathrm{RV_{AC}}5} = 0.434873Z_{\mathrm{RV_{AC}}5kk} + 0.394440Z_{\mathrm{RV_{AC}}5ss} + 0.324893Z_{\mathrm{RV_{AC}}5hh} + 0.432343Z_{\mathrm{RV_{AC}}5ks} + 0.439167Z_{\mathrm{RV_{AC}}5kh} + 0.412261Z_{\mathrm{RV_{AC}}5sh}$

Table 5.5. Pearson's correlations between first principal components

	Cholesky	EWMA	BEKK	CCC	DCC
RV1	0.5366	0.5205	0.4492	0.5313	0.5040
RV2	0.5165	0.5034	0.4269	0.5122	0.4873
RV3	0.5253	0.5176	0.4360	0.5242	0.5001
RV4	0.5153	0.5086	0.4286	0.5163	0.4903
RV5	0.4848	0.4783	0.3980	0.4892	0.4620
$\mathrm{RV_{AC}}1$	0.5149	0.5028	0.4255	0.5116	0.4872
$\mathrm{RV_{AC}}2$	0.5025	0.4935	0.4152	0.5024	0.4767
$\mathrm{RV_{AC}}3$	0.5164	0.5101	0.4276	0.5180	0.4928
$\mathrm{RV_{AC}}4$	0.4922	0.4860	0.4050	0.4930	0.4687
$\mathrm{RV_{AC}}5$	0.4816	0.4776	0.3949	0.4846	0.4602

Cholesky = Cholesky decomposition and volatility modeling; EWMA = exponential weighted moving average; BEKK = Baba-Enble-Kraft-Kroner; CCC = constant conditional correlation; DCC = dynamic conditional correlation.

시킨 후 구한 제 1 주성분(PC)은 다음과 같이 추정되었다.

Tables 5.2, 5.3, 5.4는 각각 5가지 모형, RV, $\mathrm{RV_{AC}}$의 주성분 분석을 통해 얻은 제1주성분과 제1주성분의 설명력을 구한 결과이다. 1분 실현변동성(RV)의 첫 번째 주성분의 설명력이 51.46%로 다른 값들보다 낮은 편이나, 전체적으로 설명력이 70%에서 80%로 제1주성분이 충분한 설명력을 지닌다고 판단된다. Table 5.5는 각 모형과 실현변동성의 제1주성분 간의 피어슨 상관계수를 구한 결과이다. 각 모형의 주성분과 높은 상관관계를 갖는 RV의 주성분은 1분 데이터이고 다음으로 3분 데이터이다. 또한 각 모형의 주성분과 가장 높은 상관관계를 갖는 $\mathrm{RV_{AC}}$의 주성분은 3분 데이터이다. 이를 바탕으로 3분 단위로 계산한 실현현동성의 주성분이 다른 단위의 실현변동성의 주성분보다 높은 상관관계를 가진다는 것을 알 수 있다. 따라서, 국내 주가 일중 수익률이 상관 구조가 있다면 $\mathrm{RV_{AC}}$ 계산에 3분 주기를 이용하고 상관구조가 없다면 RV에서 1분 주기를 이용하는 것이 좋을 것이다. 일중 수익률의 상관 구조에 관계없이 실현변동성을 계산하려면 3분 주기의 고빈도 자료 사용이 합리적으로 판단된다.

본 논문에서는 조건부 분산(변동성) 및 조건부 공분산으로 구성된 다변량 벡터를 일차원 차원 축소를 통해 얻은 제1주성분으로 비교하였다. 고차의 주성분(제2 또는 제3주성분)을 사용하는 경우 설명력은 높일 수 있으나 비교 통계량인 피어슨 상관계수 대신 정준상관계수(canonical correlation)를 고려해야 하므로 다시 한 번 차원축소를 해야 하는 번거로움이 생기게 된다. 주성분을 만들 때 벡터 구성변수들의

표준편차가 많이 다른 경우 차원축소 결과인 제1주성분은 표준편차가 큰 변수에 크게 의존하는 단점이 있을 수 있으므로 이를 피하기 위해서 본 논문에서 예시한 표준화 시킨 후 주성분을 구하는 상관행렬 주성분 분석을 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., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modelling and forecasting realized volatility, *Econometrics*, **71**, 579–625.

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

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.

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

Engle, R. F. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics*, **20**, 339–350.

Lee, G. J. and Hwang, S. Y. (2017). Multivariate volatility for high-frequency financial series, *The Korean Journal of Applied Statistics*, **30**, 169–180.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise, *Journal of Business & Economic Statistics*, **24**, 127–161.

Hwang, S. Y., Choi, M. S., and Do, J. D. (2009). Assessments for multivariate-GARCH models using backtesting: case study, *Korean Journal of Applied Statistics*, **22**, 261–270.

Oh, R. and Shin, D. W. (2012). Market microstructure noise and optimal sampling frequencies for the realized variances of stock prices of four leading Korean companies, *Korean Journal of Applied Statistics*, **25**, 15–27.

Seong, W. H. (1997). *Applied Multivariate Analysis: Theory, Methods, SAS Application*, Tamjin, Seoul.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed), John Wiley, New York.

Tsay, R. S. (2014). *Multivariate Time Series Analysis With R and Financial Application*, John Wiley, 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. (2015). Volatility computations for financial time series: high frequency and hybrid method, *Korean Journal of Applied Statistics*, **28**, 1163–1170.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates, *Journal of Business & Economic Statistics*, **14**, 45–52.

주성분을 이용한 다변량 고빈도 실현 변동성의 주기 선택

진민경[a] · 윤재은[a] · 황선영[a,1]

[a]숙명여자대학교 통계학과

(2017년 8월 3일 접수, 2017년 9월 9일 수정, 2017년 9월 10일 채택)

요 약

본 논문은 다변량 실현 변동성 계산에서 주기 선택 방안에 대해 연구하고 있다. 고빈도(high frequency) 시계열 자료에 기초한 일간 변동성인 실현변동성을 계산하고 차원 축소 방법인 주성분을 도입하였다. Cholesky 모형을 포함한 다양한 다변량 변동성모형을 주성분을 통해 비교하였으며 KOSPI/삼성전자/현대차 고빈도 수익률 자료를 이용하여 예시하였다.

주요용어: 고빈도 시계열, 다변량 변동성, 주성분

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