The Korean Journal of Applied Statistics (2020)
33(6), 805–812
DOI: https://doi.org/10.5351/KJAS.2020.36.6.805

FPCA for volatility from high-frequency time series via R-function

Jae Eun Yoon[a] · Jong-Min Kim[b] · Sun Young Hwang[a,1]

[a]Department of Statistics, Sookmyung Women's University;
[b]Statistics Discipline, University of Minnesota-Morris

(Received September 17, 2020; Revised October 17, 2020; Accepted October 22, 2020)

Abstract

High-frequency data are now prevalent in financial time series. As a functional data arising from high-frequency financial time series, we are concerned with the intraday volatility to which functional principal component analysis (FPCA) is applied in order to achieve a dimension reduction. A review on FPCA and R function is made and high-frequency KOSPI volatility is analysed as an application.

Keywords: functional PCA, high-frequency time series, R-functions

1. 서론

최근 자료를 수집하고 처리하는 기술이 발전하면서 이산적으로 관측되지만 연속적인 특성을 갖는 변수에 대한 고빈도 자료(high-frequency data)를 얻을 수 있게 되었다. 강우량, 기온 등 시간에 따라 연속적인 특성을 가지고 있지만 실제 자료는 이산형 자료(discrete data)로 얻어지는 경우가 많이 있다. 이러한 자료를 함수형 자료(functional data)로 변환하여 분석하는 방법을 함수형 자료 분석(functional data analysis; FDA)이라고 한다. Ramsay와 Silverman (2005), Horvath 등 (2012), Wang 등 (2016)은 다양한 함수형 자료 분석기법과 추론에 관해 연구하였다.

본 연구에서는 특별히 고빈도 금융시계열 자료의 변동성(volatility)에 대해 생각해 보고자 한다. 과거 금융시계열 자료가 정해진 몇 시점에서만 관측이 되었던 반면에 최근에는 고빈도 시계열 자료(high-frequency time series data)를 쉽게 얻을 수 있게 되었다. 이를 이용하여 연속에 가까운 하루의 일중 로그 수익률(intraday log return)들을 구할 수 있으며 이들을 시간 t에 대한 함수로 보고 함수형 자료 분석을 실시할 수 있다. 이러한 함수형 접근은 함수형 시계열이 등분산(homoskedastic)인 경우 뿐 아니라 고빈도 변동성(high-frequency volatility) 분석과 관련되어서도 적용되고 있다. Hormann 등 (2013)의 functional ARCH(줄여서 fARCH) 모형, Aue 등 (2017)의 functional GARCH(줄여서 fGARCH) 모형 등이 있으며 Yoon 등 (2017)은 fARCH 모형과 fGARCH 모형을 국내 금융시계열 자료에 적합한 결과를 제시하였다. 최근에 Kim과 Hwang (2020)은 코퓰라를 이용한 고빈도 금융시계열의 변동성 연구를 수행하였다.

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

함수형 자료 분석에서는 연속적인 특성을 지닌 이산형 자료를 함수로 바꾸어서 분석한다. 시간 $t_1, t_2, \ldots, t_n$과 이에 따른 관측치 $\{x(t_j), j = 1,2,\ldots,n\}$를 생각해보자. 시간 간격을 0에 가깝게 매우 촘촘한 시간 간격을 갖도록 하고 관측치의 차원을 무한대로 확장한다면 $\{x(t_j)\}$를 $\{x(t)\}$라는 시간에 대한 하나의 함수로 나타낼 수 있다. 함수형 자료는 본질적으로 무한차원(infinite dimension)을 가지게 된다. 함수형 자료 분석은 이러한 함수를 분석하는 것이며 금융시계열 자료에서는 하루의 일중 수익률 또는 일중 수익률 제곱을 함수로 나타내게 된다. 개별적인 관측치들을 이용하여 하나의 함수로 나타내기 위해서는 얻어진 자료를 함수 형태로 변환하는 작업이 필요하다. 이를 위해 적절한 평활화 방법(smoothing method)을 이용한다. 많이 사용되는 방법은 K개의 기저함수(basis function) ϕ_k의 선형결합으로 함수를 나타내는 방법으로 다음과 같이 표현된다.

$$x_i(t) = \sum_{k=1}^{K} c_{ik}\phi_k(t), \quad i = 1,2,\ldots,N.$$

K는 기저함수의 수이고 N은 자료 수이며, 예를 들어 N일의 금융시계열 자료라면 N개의 함수 $x_1(t)$, $x_2(t),\ldots,x_N(t)$를 갖게 된다. 기저함수의 수 K가 큰 값이면 평활정도가 작아 편의(bias)가 줄어들게 되고, 작은 값이면 평활정도가 커서 분산을 작게 하는 경향이 있다 (Ramsay와 Silverman, 1997). 선형결합 계수 c_{ik}를 모아 놓은 계수행렬 $C = ((c_{ik})): N \times K$ 행렬을 구성하는 상수들은 최소제곱법을 이용하여 얻을 수 있다. 기저함수는 퓨리에(Fourier) 함수, 스플라인(Spline), 웨이브릿(wavelets) 등이 있는데 일반적으로 많이 사용되는 것은 퓨리에 함수이다. 퓨리에 함수는 일정한 주기로 관측된 주기적인 자료(periodic data)나 이에 가까운 자료에 유용하게 적용된다고 알려져 있다 (Ramsay와 Silverman, 1997). 이산형 형태의 고빈도 금융시계열 자료를 퓨리에 함수로 근사시켜 변환하면 완만한 곡선으로 바꾸어 나타낼 수 있다.

2. 함수형 주성분분석

주성분분석(principal component analysis)은 대표적인 다변량 분석 방법으로 서로 독립적인 몇 개의 주성분(principal component)으로 전체의 변동(variation)을 설명하는 방법이다. 주성분은 변수들의 선형결합으로 표시되며 소수의 주성분으로 전체 변동을 설명한다는 점에서 차원축소의 의미를 가지고 있다. 함수형 주성분분석(functional principal component analysis; FPCA) 또한 마찬가지로 차원축소의 의미를 가지고 있으며 함수형 회귀분석(functional regression analysis), 함수형 군집분석(functional cluster analysis) 등의 다양한 함수형 자료 분석을 위한 차원축소 방법으로 많이 활용되고 있다. FPCA의 자세한 내용은 Ramsay와 Silverman (1997), Ramsay와 Silverman (2002), Ramsay 등 (2009), Wang 등 (2016)을 참고하기 바라며 여기서는 FPCA의 기초 개념만을 소개하고자 한다.

앞으로 L^2-공간은 구간 $I = [0,1]$에서 정의된 $\int_0^1 x^2(t)dt < \infty$를 만족하는 실수값 함수(measurable real-valued function) x의 집합으로 정의한다. L^2-공간은 분할가능(separable)한 힐버트 공간(Hilbert space)이며 내적(inner product)이 다음과 같다.

$$< x, y > = \int x(t)y(t)dt,$$

여기서와 이후의 적분영역은 $\int = \int_0^1$이다. L^2-공간의 구간 I에서의 함수형 자료를 $x_i(s),\ i = 1,2,\ldots,N$라고 하자. 함수 $x_i(s)$에 대해서 다음의 식을 보겠다.

$$f_i = \int \xi(s)x_i(s)ds = < \xi, x_i >,$$

여기서 f_i를 주성분점수(principal component score)라고 한다. FPCA의 첫 단계는 $\| \xi_1 \|^2 = \int \xi_1(s)^2 ds = 1$이라는 제약조건 아래 $N^{-1} \sum_i f_{i1}^2 = N^{-1} \sum_i < \xi_1, x_i >^2$을 최대화하는 함수 ξ_1을 찾는 것이다. 이 함수 $\xi_1(s)$가 첫 번째 주성분함수가 되며 이를 첫 번째 조화성분(harmonics)이라고 부르기도 한다. 그 이후의 주성분함수 $\xi_m(s)$는 다변량 주성분분석에서와 마찬가지로 앞의 주성분함수들과 직교성을 만족, 즉 $< \xi_k, \xi_m > = 0, k < m$을 만족한다는 조건 아래 K개의 주성분함수 $\xi_i(s), i = 1,2, \dots, K$를 추정한다. 다변량 주성분분석에서 주성분을 추정하기 위해 공분산행렬 또는 상관행렬의 고유분석(eigen-analysis)을 실시한다. 마찬가지로 함수형 주성분분석에서도 주성분함수를 추정하기 위해 공분산함수 또는 공분산연산자에 대한 고유분석을 실시한다. 먼저 각 함수들 $x_i(t)$에 대해 $N^{-1} \sum_i x_i(t) = 0$이며 주성분함수에 대해 $\int \xi^2(s) ds = 1$이며 $< \xi_k, \xi_m > = 0$을 만족한다고 가정한다. 함수형 주성분분석에서는 다음의 고유방정식(eigen equation)을 얻을 수 있다.

$$\int v(s,t)\xi(t)dt = \lambda\xi(s) ,$$

여기서 $v(s,t)$는 공분산함수 $v(s,t) = (N-1)^{-1} \sum_{i=1}^{N} x_i(s) x_i(t)$이며 이를 커널(kernel)로 하는 적분 연산자(integral operator)를 공분산연산자 V라고 하자. 공분산연산자 V는 힐버트-슈미트 연산자(Hilbert-Schmidt operator)이다 (Conway, 1994). 공분산연산자를 이용하면 위의 식을 다음과 같이 나타낼 수 있다.

$$V\xi = \lambda\xi ,$$

여기서 ξ는 고유함수, λ는 고유값을 의미하며 공분산연산자 V에 대한 고유분석을 실시하여 고유함수와 고유값을 추정한다. 공분산연산자는 대칭(symmetric)이며 비음정치(nonnegative definite)이므로 음이 아닌 고유값을 갖는다.

위의 고유방정식에서 고유함수 ξ를 추정하기 위한 실제적인 방법 중 한 가지는 연속성을 갖고 있는 함수를 행렬로 바꾸어 접근하는 것이다. 함수형 자료를 이산화(discretizing)하여 행렬로 만들어 고유분석을 실시한다. 이러한 방법론은 함수형 주성분분석에 대한 초창기 연구에서 소개된 바 있다 (Rao, 1958, 1987; Tucker, 1958). 고유함수 ξ를 추정하기 위한 다른 방법은 서론에서 언급한 평활화 방법을 이용하는 것이다. 각 함수 x_i들을 알고 있는 적절한 K개의 기저함수 ϕ_k들의 선형결합으로 다음과 같이 나타낸다고 가정하자.

$$x_i(t) = \sum_{k=1}^{K} c_{ik}\phi_k(t) .$$

N개의 함수 $x_1, x_2, \dots, x_N$과 K개의 기저함수 $\phi_1, \phi_2, \dots, \phi_k$에 대해 다음과 같이 나타낼 수 있다.

$$x = C\phi$$

C는 $N \times K$ 계수행렬이다. K차 대칭행렬 $W = \int \phi^T \phi$을 만들어서 $\xi(s) = \phi(s)^T b$로 나타내고 $u = W^{1/2} b$로 정의하면 앞의 고유방정식을 다음과 같이 행렬의 식으로 나타낼 수 있게 된다.

$$N^{-1} W^{\frac{1}{2}} C^T C W^{\frac{1}{2}} u = \lambda u$$

기저함수가 정규직교함수라면 W가 단위행렬 E가 되며 위의 식은 K차 대칭행렬 $N^{-1} C^T C$에 대한 고유분석이 된다. 대칭행렬에 대한 고유분석을 실시함으로 주성분을 추정한다. 자세한 내용은 Ramsay와 Silverman (1997)을 참고하기 바란다.

함수형 주성분분석에서 주성분의 수를 정하는 것은 중요한 문제이다. 주성분의 수를 p라고 하면 p개의 주

성분이 $X(t)$의 전체 변동 중에서 설명하는 비율인 cumulative percentage of total variance (CPV)는 다음과 같이 계산한다.

$$\mathrm{CPV}(p) = \frac{\sum_{k=1}^{p} \hat{\lambda}_k}{\sum_{k=1}^{n} \hat{\lambda}_k}$$

원하는 수준의 $\mathrm{CPV}(p)$에 도달하게 하는 p값을 선택하면 되고 보통 80~90% 수준을 이용한다. 스크리 도표(scree plot), pseudo-AIC, pseudo-BIC 등을 이용하여 주성분의 수를 정할 수도 있다. 자세한 내용은 Horvath와 Kokoszka (2012)를 참고하기 바란다.

3. 함수형 주성분분석 수행을 위한 R의 함수

R의 fda 패키지, fdapace 패키지, Matlab의 PACE 패키지 등을 이용하여 함수형 주성분분석을 포함한 여러 함수형 자료 분석을 수행할 수 있다. R에서 함수형 주성분분석을 수행하는 두 함수를 비교해보고자 한다.

R에서는 fda 패키지의 "pca.fd" 함수, fdapace 패키지의 "FPCA" 함수를 이용하여 함수형 주성분분석을 수행할 수 있다. 두 함수의 가장 큰 차이점은 적용하는 자료의 형태와 주성분점수를 추정하는 데 있다. pca.fd 함수는 수치적분을 이용하고 FPCA 함수는 조건부 기댓값을 이용하여 주성분점수를 추정한다.

3.1. pca.fd 함수

fda 패키지의 pca.fd 함수는 함수형 주성분분석을 위해 일반적으로 많이 사용되는 함수이며 일정한 시간 간격으로 관측되어 정칙성(regularity)을 만족하며 각 개체의 반복수가 많은 조밀한 자료(densely observed data)를 대상으로 한다. 이러한 자료는 비교적 간단한 과정을 거쳐 고유값과 고유함수, 주성분점수 등을 계산할 수 있다 (Horvath와 Kokoszka, 2012). 함수형 주성분분석을 위해 pca.fd 함수를 이용하여 표본공분산연산자의 고유함수와 고유값, 주성분점수를 계산할 수 있으며 이 함수를 적용하기 위해서는 함수형 개체(functional object)여야 한다.

pca.fd 함수는 다음의 과정을 수행한다. 먼저 각 시점에 대한 평균(cross-sectional mean) $\hat{\mu}$을 계산한다. 공분산연산자는 각 시점의 그리드(grid) 위에서 추정한 공분산 표면(autocovariance surface) $\mathrm{Cov}(X(s), X(t))$으로 근사한다. 이에 대한 고유분석을 수행하여 고유함수 $\hat{\xi}$와 고유값 $\hat{\lambda}$를 추정한다. 수치적분을 이용하여 다음과 같이 주성분점수 $\hat{f}$를 추정한다.

$$\hat{f}_{ik} = \int [x(t) - \hat{\mu}(t)]\xi_i(t)dt$$

pca.fd 함수를 적용한 분석의 예시로는 Ramsay와 Silverman (1997), Horvath와 Kokoszka (2012)를 참고하기 바란다.

3.2. FPCA 함수

일반적인 함수형 자료 분석에서는 정칙적으로 관측되고 각 개체의 반복수가 많은 자료를 대상으로 하는데 각 개체의 반복측정 수가 작으며 비정칙적으로(irregularly) 관측된 경시적 자료(longitudinal data)인 경우 기존의 함수형 자료 분석 방법을 적용하는 데 무리가 있다. Yao 등 (2005)은 principal components analysis through conditional expectation (PACE) 알고리즘을 이용한 "성긴" 경시적 자료(sparse longitudinal data)에

대한 함수형 주성분분석을 연구하였다. 성긴 자료(sparse data)에 대한 엄격한 정의는 없으나 관측된 시점이 규칙적이지 않고 각 개체의 수가 평균적으로 20개보다 작은 경우 성긴 자료로 볼 수 있다. 이러한 경우 기존의 함수형 주성분분석 방법을 적용하고자 하면 수치적분을 이용하여 주성분점수를 계산하는 단계에서 문제가 생길 수 있다. 평활화된(smoothed) 값을 이용하여 분석하게 되며 평활화된 값을 구하는 것이 함수형 주성분분석을 수행하기 위한 중요한 단계가 된다.

FPCA 함수는 다음의 과정을 수행한다. 먼저 평활화된 평균(smoothed mean) $\hat{\mu}$를 계산한다. 표본공분산의 비대각원소를 이용하여 평활화된 공분산(smoothed covariance)을 추정한다. 이렇게 얻은 평활화된 공분산에 대한 고유분석을 수행하여 추정된 고유함수 $\hat{\xi}$와 고유값 $\hat{\lambda}$를 얻은 후 평활화된 공분산을 양반정치 표면(positive semi-definite surface)에 사영시킨다. 조건부 기댓값을 이용한 다음의 식으로 주성분점수 $\hat{f}$를 추정한다.

$$\hat{f}_{ik} = \hat{E}[\hat{f}_{ik}|X_i] = \hat{\lambda}_k \xi_{ik}^T \, \Sigma_{X_i}^{-1}(X_i - \hat{\mu}_i)$$

fdapace 패키지의 FPCA 함수를 이용하여 위의 과정을 수행할 수 있다. PACE 알고리즘에 대한 자세한 내용은 Yao 등 (2005)을 참고하기 바란다.

4. 자료분석 예시

이 절에서는 국내 KOSPI 고빈도 자료의 변동성에 함수형 주성분분석을 수행하는 두 함수를 적용하여 얻은 결과를 비교해 보고자 한다. 2010년 1월 5일부터 2015년 6월 30일까지 1,349일의 기록으로 개장시간 동안 1분 간격으로 조사된 KOSPI 고빈도 자료를 이용해서 로그수익률의 변동성에 대해 알아보도록 한다. 기호 $P_k(t)$를 k일의 시간 t에서의 주가라고 하면 $y_k(t) = \log P_k(t) - \log P_k(t-h)$로 하고 시간 간격을 다르게 하여 로그 수익률을 계산할 수 있다. 본 논문에서는 $h = 5$로 하여 5분 간격 로그 수익률을 계산하여 이용하였다. 로그 수익률을 제곱한 값 $y_k^2(t)$에 대해 평활화 방법을 이용해 함수형 자료로 변환한 후 pca.fd 함수와 FPCA 함수를 이용하여 함수형 주성분 분석을 수행하였다. KOSPI 자료는 촘촘하고도 규칙적인 시간 간격으로 얻어진 자료로 일반적으로 함수형 자료 분석을 위한 자료로서의 조건을 잘 만족하고 있다. 이런 자료에 pca.fd 함수를 적용한 결과와 PACE 알고리즘을 이용하는 FPCA 함수를 적용한 결과를 비교해보고자 한다.

퓨리에 기저전개를 이용하여 관측치들을 함수형 자료로 변환한 후 변환된 함수형 자료에 pca.fd 함수를 적용하였다. FPCA 함수는 가우시안 커널(Gaussain kernel)을 이용하여 평활화하며 Epanechnikov kernel, rectangular kernel 등의 다른 커널도 이용가능하다.

분석 시 이용한 코드는 다음과 같다.

(i) 먼저 pca.fd 함수를 이용한 경우이다.

```
number_of_basis_functions = 400
basis < − create.fourier.basis(rangeval = c(0,10), nbasis = number_of_basis_functions, period = fourier_period)
functional_data_squared < − Data2fd(t(data), argvals=seq(from = 0, to = 10, length.out = L), basis)
principal_components < − pca.fd(fdobj = functional_data_squared, nharm = 2)
eigenfunctions < − principal_components$harmonics
```

함수형 개체로 만들기 위해 “create.fourier.basis” 함수와 “Data2fd” 함수를 이용하였다. 제곱된 수익률 자료

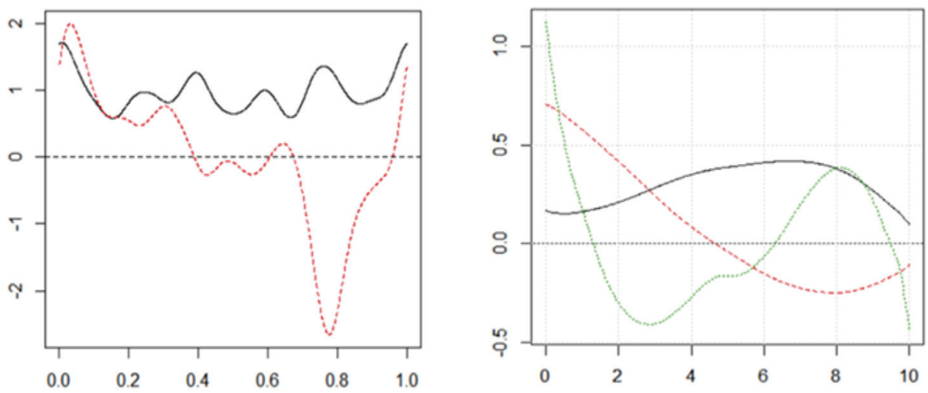

Figure 4.1. Eigen functions : pca.fd (left) and FPCA (right).

Table 4.1. CPV values

	CPV(pca.fd)	CPV(FPCA)
FPC 1	0.5834	0.87
FPC 2	0.6954	0.99

를 data로 저장하였고 이를 함수형 자료로 변환 후 “pca.fd” 함수를 이용하여 함수형 주성분분석을 실시하였다. 옵션 “nharm”을 이용하여 주성분함수의 수를 정할 수 있다. 여기서 L은 하루 중 관측된 자료의 수, 345이다.

(ii) 같은 자료에 대해 FPCA 함수를 다음과 같이 적용하였다.

```
s < − seq(0,10, length.out=N)
functional_data_squared < − MakeFPCAInputs(IDs = rep(1:L, each=N), tVec=rep(s,L), t(data))
fpcaObj < − FPCA(functional_data_squared$Ly, functional_data_squared$Lt, list(plot = TRUE, methodMuCovEst='smooth', userBwCov=5))
eigenfunctions1 < − fpcaObj$phi[,1]
```

“MakeFPCAInputs” 함수를 이용하여 함수형 자료로 변환 후 “FPCA” 함수를 적용하였다. 여기서 L은 하루 중 관측된 자료의 수인 345, N은 전체 관측일수인 1,349이다.

Figure 4.1은 pca.fd 함수(좌)와 FPCA 함수(우)를 이용하여 구한 고유함수 그래프이다. 검은색 실선이 첫 번째 고유함수이며 빨간색 점선은 두 번째 고유함수로 각각 첫 번째 주성분(FPC 1)과 두 번째 주성분(FPC 2)에 해당된다. 그래프의 y축은 값을, x축은 시간을 의미한다. 이산적으로 관측된 자료를 연속형인 함수형 자료로 변환하는 과정에서 하루를 나타내는 시간 구간을 pca.fd에서는 $[0, 1]$, FPCA에서는 $[0, 10]$으로 하였다. pca.fd 함수를 적용 시 두 개의 고유함수를 얻었으며 원하는 개수를 지정할 수 있다. FPCA 함수에서는 전체 변동의 99%를 설명하기 위해 필요한 수만큼 고유함수를 구해준다. KOSPI 자료 분석 결과 비교에서는 두 개의 함수형 주성분 기준으로 살펴보고자 한다. 첫 번째 주성분은 전체에 걸쳐 0보다 큰 양의 값을 가지고 있는데 수익률을 제곱한 값이 갖고 있는 시계열 자료의 변동성으로써의 특징을 반영하고 있는 성분으로 보인다. 두 번째 주성분은 후반부로 갈수록 감소하는 경향을 보인다.

Table 4.1은 각각의 방법에서 얻은 2개의 함수형 주성분의 누적설명비율인 cumulative percentage of total variance (CPV) 값을 보여주고 있다. pcd.fd 함수를 이용한 결과보다 FPCA 함수를 이용한 결과에서 첫 번째 함수형 주성분의 설명 비율이 월등히 높고 전체 변동 중 두 개의 함수형 주성분이 설명하는 누적 비율이 높은 것을 알 수 있다. 이러한 큰 차이가 발생하는 이유는 두 함수에서 이용하는 평활기법의 차이에 있다. pcd.fd 함수는 스플라인 함수, 퓨리에 함수를 이용하여 평활하고 FPCA 함수는 국소 가중 선형 평활 (locally weighted linear smoother)을 이용한다. 결론적으로, 소수의 주성분으로 전체 변동을 설명하려는 주성분분석의 목적을 생각해 볼 때 고빈도 시계열 변동성 분석에서 FPCA 함수를 이용하는 것이 설명비율이 높은 함수형 주성분을 얻을 수 있을 것이라 생각된다. 본 연구에서는 사례분석 예제를 통해 FPCA 유용성을 알아보았으며 향후 연구로서 연관된 이론 연구 수행을 제안하고자 한다. 특히 정칙적으로 얻어진 촘촘한 자료에 PACE 알고리즘을 적용하는 경우 발생할 수 있는 문제점과 이를 보완하기 위한 방법에 대한 연구가 필요하다고 생각된다.

References

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

Conway, J. B. (1994). *A Course in Functional Analysis*, Springer, New York.

Hadjipantelis, Dai, Ji, Han, Muller, and Wang (2017). *Functional PCA in R*, R manual.

Hormann, S., Horvath, L., and Reeder, R. (2013). A functional version of the ARCH model, *Econometric Theory*, **29**, 267–288.

Horvath, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, Springer.

Kim, J.-M. and Hwang, S. Y. (2020). Functional ARCH directional dependence via copula for intraday volatility based on high-frequency financial time series, *Applied Economics*, online published.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*, Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*, Springer, New York.

Ramsay, J. O., Hooker, G. and Silverman, B.W. (2009). *Functional Data Analysis with R and MATLAB*, Springer, New York.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves, *Biometrics*, **14**, 1–17.

Rao, C. R. (1987). Prediction in growth curve models (with discussion), *Statistical Science*, **2**, 434–471.

Tucker, L. R. (1958) Determination of parameters of a functional relationship by factor analysis, *Psychometrika*, **23**, 19–23.

Wang, J. L., Chiou, J. M., and Muller, H. G. (2016). Functional data analysis, *Annual Review of Statistics and Its Application*, **3**, 257–295.

Yao, F., Muller, H. G., and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data, *Journal of the American Statistical Association*, **100**, 577–590.

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.

FPCA를 통한 고빈도 시계열 변동성 분석: R함수 소개와 응용

윤재은[a] · 김종민[b] · 황선영[a,1]

[a]숙명여자대학교 통계학과, [b]미네소타대학교

(2020년 9월 17일 접수, 2020년 10월 17일 수정, 2020년 10월 22일 채택)

요 약

본 논문은 최근 금융시계열 분야에서 자주 등장하는 고빈도 시계열 변동성 분석을 다루고 있다. 고빈도 시계열 변동성 분석을 위해 차원 축소를 목적으로 하는 함수형 주성분분석을 적용하였으며 이를 수행하는 R의 두 함수를 비교하고 있다. 응용으로서, KOSPI 고빈도 자료에 적용해 보았다.

주요용어: 함수형 주성분분석, 고빈도 시계열, R-함수

본 연구는 한국연구재단의 지원을 받았습니다 (NRF-2018R1A2B2004157).

[1] 교신저자: (04310) 서울특별시 용산구 청파로47길100, 숙명여자대학교 통계학과. E-mail: shwang@sookmyung.ac.kr