

## 금융공학 교육을 위한 R통계 패키지의 활용방안

홍창수\* · 백재승\*\*

### <요 약>

빅데이터에 대한 관심과 함께 오픈소스 언어에 대한 관심도 높아지고 있다. 데이터 분석에 파이썬(Python)과 R 프로그램이 많이 사용되면서, 많은 사람들에 의해 제공되는 기부 패키지를 기본으로 통계적 프로그래밍과 시각화를 위한 작업이 많이 진행되고 있다. R은 다양한 통계분석이 가능하며 그래픽에도 강점을 가지고 있으며, 오픈소스로 운영되기 때문에 모든 것을 무료로 사용할 수 있다. 이는 SPSS, SAS, STATA 등 다른 통계프로그램이 유료이며 상당히 고가라는 점과 비교할 때 매우 큰 장점이다. 이러한 장점에도 불구하고 R은 통계분석과 계량경제학 분야에서 그 사용이 확산되고 있으며, 금융공학 분야 사람들에게 보편적으로 사용되지 않고 있다. 금융시계열 분석을 위한 통계패키지로 많이 활용되는 것은 SAS, STATA, EVIEWS가 널리 활용되고 있고, 금융공학을 위한 실무언어로는 VBA, MATLAB, C++이 주로 사용되고 있다. 실제 금융실무에서 보수적인 시각으로 인해 오래전부터 사용되어왔던 GPU와 오픈소스 언어에 대한 사용 논의가 최근에서야 활발해졌다는 점은 이러한 기법과 언어의 안정성이 높아졌다는 점에 기인할 것으로 생각된다. 본고는 R프로그램 소개와 활용, 금융공학 모델링 절차와 R패키지 소개, 금융공학을 위한 R의 활용에 대해 고찰해 보았다. 향후 데이터를 통한 의사결정이 중요해지고 이를 위한 컴퓨팅 자원과 분석 알고리즘이 필수가 될 것으로 판단된다. R의 장점은 실제 금융공학에 유용한 패키지를 활용하여 교육을 진행할 수 있다는 장점이 있다. 금융모델링 후 계산 값이 맞는지 해당 패키지를 불러 정확한 값을 검증할 수 있다. R을 활용한 금융공학교육 커리큘럼은 시계열 데이터분석, 포트폴리오 최적화, 자산 가격결정 모형, 채권가격계산, 파생상품가격결정, 위험관리에 집중해서 교육할 수 있으며, 관련 패키지를 교육에 모두 활용할 수 있다. R의 유용성은 학교에서 배운 내용을 금융산업 현장에서 바로 활용할 수 있다는 높은 강점이 있다고 판단된다.

주제어 : 금융공학, R 패키지, 포트폴리오 이론, 옵션 모형, 시장위험관리

\* 제1저자, NICE P&I 금융공학연구소 실장(cshong@nicepni.co.kr)

\*\* 교신저자, 한국의국어대학교 국제금융학과 교수(jbaek@hufs.ac.kr)

## I. 서론

빅데이터 분석과 함께 오픈소스 언어에 대한 관심이 높아지고 있다. 현재 데이터 분석 분야에서 파이썬(Python)과 R은 주요 언어로 자리잡게 되었다. 이것은 많은 사람들에 의해 제공되는 기부 패키지(contributed package)를 기본으로 통계적 프로그래밍과 시각적인 표현을 정교하게 구현할 수 있다는 장점이 있기 때문이다. R은 컴퓨터 언어이면서 동시에 통계분석과 그래픽을 위한 다양한 패키지들의 집합이라 할 수 있다. R은 다양한 통계분석이 가능하며 그래픽에도 강점을 가지고 있으며, 오픈소스로 운영되기 때문에 모든 것을 무료로 사용할 수 있다. 이는 SPSS, SAS, STATA 등 다른 통계프로그램이 유료이며 통계패키지별로 가격차이가 있으나, 비교적 고가라는 점에 비해 매우 큰 장점이다.

일반적으로 무료소프트웨어라고 생각하면, 특정한 기능이 빠졌을 것이라 생각할 수 있지만 R은 전세계 모든 분야의 사용자들이 자신의 분야에서 패키지를 개발 및 공개하고 있어 상용프로그램에 뒤지지 않는 분석을 이용할 수 있도록 하고 있다. 또한, R은 인터넷이 연결되어 있는 컴퓨터에서 언제 어디서나 쉽게 다운로드하여 설치하고 사용할 수 있으며 우수한 도움말 기능을 가지고 있어 통계비전공자들도 부담 없이 사용할 수 있는 통계패키지이다. 아울러 R은 여러 개의 그래픽장치를 활용하지 않더라도 많은 정보를 담고 있는 복잡한 그래프 작성이 가능하여 다양하고 우수한 그래프를 그릴 수 있다. 이는 최근 쟁점화되고 있는 저작권 문제와 사용의 연속선상에서 자유로움을 들 수 있다. 또한, 대학에서 학생들이 R을 배운다면 사회의 여러 분야에 진출하여도 R을 계속 사용할 수 있기 때문에 통계패키지 활용의 연속선상과 비용절감의 측면에서 있어 도움이 될 것이다.

R에서는 다양한 함수와 데이터들이 패키지를 통해 제공된다. 기본 패키지에는 R을 작동시키기 위한 기본 함수들과 데이터, 그리고 표준 통계함수와 표준 그래픽함수들이 들어 있다. 표준 패키지 또는 추천패키지로도 불리는 25개 정도의 기본 패키지와 CRAN(CRAN.R-project.org)을 통해 다양한 저자들에 의해 제공되는 수천개의 패키지들이 존재한다. 이러한 장점에도 불구하고 R은 통계분석과 계량경제학 분야에서만 그 사용이 확산되고 있으며, 금융공학 분야 사람들에게 보편적으로 사용되지 않고 있다. 현재 금융시계열 분석을 위한 통계패키지로 많이 활용되는 것은 SAS, STATA, EViews가 널리 활용되고 있고, 금융공학을 위한 실무언어로는 VBA, MATLAB, C++이 주로 사용되고 있다.<sup>1)</sup> 실제 금융실무에서 보수적인 시각으로 인해 오래전부터

1) 다른 오픈소스인 PYTHON의 활용도 또한 점차 증가하고 있는 추세이다. 개발자입장에서의 분석기

사용되어왔던 GPU와 오픈소스언어에 대한 사용 논의가 최근에서야 활발해졌다는 점은 이러한 기법과 언어의 안정성이 높아졌다는 점에 기인할 것으로 생각된다. 본고에서는 R프로그램 소개와 활용, 금융공학 모델링 절차와 R패키지 소개, 금융공학을 위한 R의 활용에 대해 고찰해 보기로 한다.

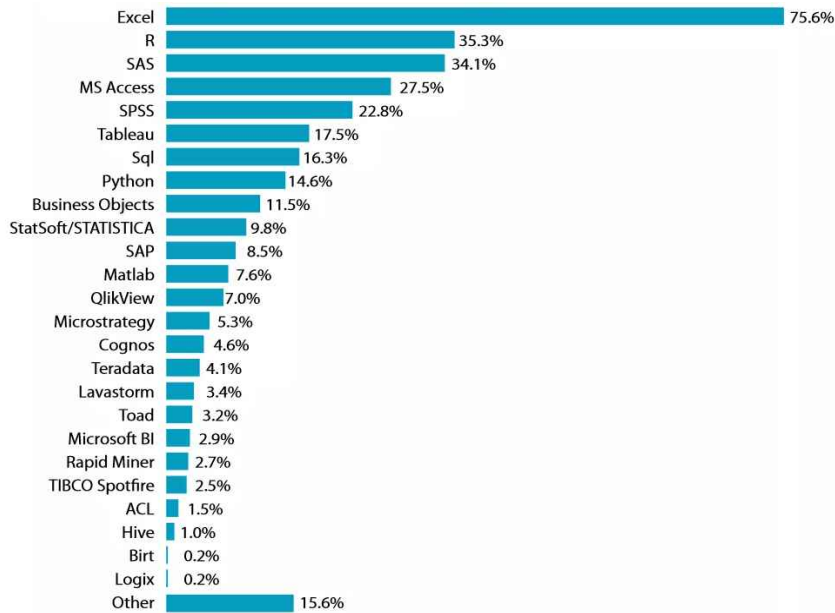
## II. R프로그램의 소개와 활용 입문

### 1. R의 설치와 기본적 사용법

R은 통계학에서 사용되던 언어 S를 벨 연구소의 John Chambers를 포함한 몇몇 개발자들이 오픈소스로 구현한 언어로서 강력하고 유연하다는 특징이 있다. R은 여러 가지 이유로 S와 상용 S-Plus를 제치고 사랑받고 있다. R은 무료이며 최소 6000개 이상의 다양한 패키지를 사용할 수 있는데 역시 R로 마찬가지로 대부분 무료이다. R은 맥이나 윈도우, 리눅스 시스템에서 작동한다. R은 사실 엑셀과 같은 스프레드시트 프로그램이나 전용 통계학 프로그램에 비하면 초반에 배우는 것이 힘든 것이 사실이지만 일단 배우기만 하면 매우 효과적이며 배울만한 가치가 있는 언어임을 알 수 있다. R의 가장 큰 장점은 강력한 패키지(Package) 생태계를 가지고 있다는 점과 시각화를 위한 도표 작성에 뛰어난 이점을 가지고 있다. 통상 실무에서 MATLAB과 비슷하게 R을 편리하게 사용할 수 있는 통합 개발 환경 툴인 RStudio를 사용하여 분석하고 있다. R설치([www.r-project.org](http://www.r-project.org)) 후 RStudio홈페이지에서 RStudio를 쉽게 무료로 프로그램을 다운로드받아 설치할 수 있다. [그림 1]은 Lavastom Inc.에서 2013년에 LinkedIn.커뮤니티 그룹을 대상으로 설문조사한 자료로 분석도구 중 엑셀이 75.6%로 가장 많은 응답을 받았고, 그 뒤를 R이 35.3%, SAS 34.1%, SPSS 22.8%로 그 뒤를 쫓고 있어 데이터분석에서 R의 위상을 알 수 있다.

---

술을 학습한다면 PYTHON프로그래밍을 응용하는 것이 효과적이며, 데이터분석 지향적인 관점의 입문과 활용까지를 생각해 볼 때 R프로그래밍의 장점이 높다고 할 수 있다. 본고에서는 R프로그래밍의 활용관점에 대해서만 논하기로 한다.



[그림 1] 데이터 분석도구 사용에 대한 Lavastom의 조사

통계분석을 할 때 패키지(package)를 사용한다는 점은 R을 SAS나 SPSS와 같은 다른 통계프로그램과 구분해주는 중요한 차이점 중 하나라고 할 수 있다. 이들 통계 프로그램은 특정한 분석을 위해서 프로그램에 제공된 명령어를 사용하면 된다. 그렇지만, R에서는 분석을 위해서는 패키지(package)라는 것을 사용해야 한다. 패키지는 통계분석 등의 특정한 목적을 위해서 사용할 수 있는 다양한 함수와 명령어, 데이터들의 집합이라고 할 수 있다. 즉 R에서 통계분석을 하기 위해서는 자신의 원하는 통계분석을 가능하게 해주는 패키지를 선택하여 설치하고, 그 패키지를 불러온 후에 패키지에 포함되어 있는 명령어를 사용하여 분석해야 한다. R에서는 기본적으로 제공하는 Base에서 제공하는 함수들 외에 특정한 통계분석이 가능하도록 수많은 사람들에게 의해서 만들어져서 제공되는 함수들의 묶음인 수많은 패키지들이 존재한다. R을 설치하면 기본적으로 제공하는 주요 패키지만 설치되고 그 밖의 통계분석 방법들은 다양한 패키지들의 형태로 별도로 제공된다. 이러한 패키지는 CRAN에 게시되고 RStudio에서 쉽게 설치하고 업데이트할 수 있게 된다.

## 2. 금융공학을 위한 모델링 절차와 R 패키지 소개

일반적으로 데이터분석을 하기위한 프로세스로는 ①문제에 대한 확인, ②데이터 요구 사항에 대한 확인, ③데이터 전처리, ④데이터 분석 수행, ⑤데이터 시각화 순으로 나뉠 수 있다. 즉, 비즈니스가 요구하는 데이터분석에 관한 문제를 확인한 후 특정문제를 해결하기 위해 다루어야 할 데이터의 속성을 확인하고 데이터를 확보하게 된다. 이후 데이터 작업을 수행하기 위한 전처리 작업부분이 이루어지면 실제 분석알고리즘을 실행하기 위한 데이터가 사용가능하게 된다. 이러한 데이터분석 절차를 확장하여 금융공학 모델링로 치환하게 되면 다음과 같은 절차로 실제 모델링이 이루어지게 된다.

### 2.1. 금융공학 모델링 절차

#### 2.1.1 금융모델링을 위한 문제 정의

실무적으로 금융모델링의 문제정의는 대부분 시장의 요구에서 출발된다. 실제 금융상품모델링을 통해 실제 상품이 출시되며, 이러한 금융상품이 출시되었을 때 인기를 가질 수 있는 시장환경인지 파악하여야 한다. 고객성향과 자금의 이동상황에 따라 금융상품의 구성이 달라질 수 있으므로 시장상황별 금융상품 모델링을 달리 가져가는 것이 중요하다. 선진 투자은행의 경우 시장상황별로 수 백개의 금융상품을 모델링하여 실제 경기사이클에 따라 시장에 맞는 금융상품을 출시하고 있다.

#### 2.1.2 구현모델의 선정 : 모형 선택 및 개발

금융모형은 실제 계산만을 위한 엔진을 개발하기도 하지만, 실제 금융시스템에서 사용될 수 있게 확장성을 고려하여 엔진을 개발하기도 한다. 일반적으로 가격(price)계산만을 위해서는 몬테카를로 시뮬레이션 방식의 모형을 선택하고 있으나, 최근에는 옵션 그리스(greeks)의 안정성을 위해 유한차분법(FDM)의 방식으로 금융모델링하는 것이 보편화되고 있다. 따라서 유지보수 관점에서 확장성을 고려한 모형선택인지, 운용목적의 그리스(greeks)산출의 안정성면인지를 고려하여 파생상품 모형을 선택하게 된다. 또한, 금융모형과 더불어 고려해야할 것은 다음 3가지 데이터의 상세정의가 필요하다. 첫 번째, 상품의 통화 및 거래상대방과 관련된 국가의 캘린더 혹은 통화별로 지정된 거래규약이 필요하다(static data) 둘째, 기초자산의 시장가격, 수익률곡선, 변동성 커브 등 시장데이터(market data)가 필요하다. 셋째, 거래조건표(term-sheet)에 명시되는 거래정보, 손익구조, 만기 등 거래데이터(deal data)가 필요하다.

### 2.1.3 구현언어의 선정

통상적으로 구현언어는 모형개발을 위한 프로토타입 언어와 시스템 구현을 위한 개발언어로 나눌 수 있다. 프로토타입의 언어는 시제품 개발을 위한 금융모형개발에 필요하며 실제 건축의 경우 설계도나 도면이 이에 해당한다고 할 수 있다. 실무적으로는 MATLAB같은 언어가 많이 사용되나 점차 R과 Python의 활용도가 높아지리라 예측할 수 있다. 이러한 개발로직을 토대로 실제 시스템 구현을 위해 C++로 구현된 엔진이 개발되며 최종 dll과 같은 형식의 파일로 시스템과 유기적으로 활용하게 된다. 이와 관련한 내용을 <표 1> 금융공학을 위한 프로그래밍 언어구분으로 정리하였다.

<표 1> 금융공학을 위한 프로그래밍 언어구분

구분	모형개발을 위한 프로토타입 언어	시스템 구현을 위한 개발언어
언어명	VBA, MATLAB, R	C++, C#, VB , JAVA, Python <sup>2)</sup> (주로 C++로 개발)
설명	시제품 개발을 위한 금융 로직개발 (건축의 경우 설계/도면에 해당)	실제 시스템 개발 (전사, 프론트, 미들 시스템)

### 2.1.4 금융모델링 실행 : 모형개발

앞서 구현모형의 선정에서 설명했듯이 모형개발 시에 고려해야 하는 부분은 기본으로 정적데이터(static data), 시장데이터(market data), 거래데이터(deal data)를 고려하여 확장성 있게 금융엔진을 개발하여야 한다. 실제로 금융모형은 시간이 흐를수록 점차 복잡해지고 있으며, 이러한 복잡성으로 인해 기존에 개발된 엔진을 사용할 수 없는 상황도 발생할 수 있으며, 실제 금융엔진을 개발한 개발자의 경우도 각주처리를 하지 않아 과거에 프로그래밍한 코드를 정확히 알 수 없는 경우가 발생하기도 한다. 또한, 계산속도의 문제를 해결하기 위해 효율적인 프로그래밍이 되어야 하며, 속도개선을 위해 병렬처리를 통한 GPU엔진을 활용하기도 한다. 인텔의 경우 CPU를 통한 병렬프로그래밍을 가능하게 한 제온/제온파이(Xeon/Xeon Phi)제품이 출시되어 좀 더 빠르고 안정적인 금융상품의 가격계산이 가능하게 하고 있다.

2) PYTHON은 해외각국의 시스템 개발에 많이 활용되고 있으며, 국내에서도 금융공학 시스템을 비롯한 금융시스템 구현에 PYTHON활용 빈도가 점차 높아지고 있어 향후 시스템구현을 위한 언어로 널리 활용되리라 판단된다.

2.1.5 금융모델링 검증(validation)

모형검증의 경우 실제 모형로직을 수리적으로 정확하게 풀었는지에 대한 검증과 이를 실제로 정확하게 프로그래밍 했는지가 검증의 중요한 문제가 된다. 실제 시장의 니즈에 맞는 모형을 수리적으로 잘 해결했는지에 대한 부분과 프로그래밍 하여 계산된 가격과 그릭스(greeks)가 시장에서 판매되고 있는 상품과 같은 동일한 계산 값이 나오는지 1차적인 검증의 문제가 된다. 금융모형에서 개발된 계산 값을 검증하는 방법으로는 시물레이션 값과 유한차분법(Finite Difference Method) 값의 비교, 제3자가 개발한 모형과의 비교, 계산 값과 상용프로그램(Numerix, FINCAD 등)과의 비교를 실무적으로 활용할 수 있다.

2.2. 금융공학을 위한 R 패키지(package) 소개

패키지(package)는 R의 라이브러리를 의미한다. 기본적인 패키지는 처음부터 사용할 수 있도록 프로그램 기동 시 로딩된다. 일부는 사용 시 library(package)를 호출해 주어야한다. 만약 패키지 자체가 인스톨되어 있지 않다면, install.packages(package)로 인스톨한 후 위와 같이 로딩하여 사용하면 된다. R패키지는 통계기반의 패키지므로 계량경제학 관련 패키지가 많으며, 금융과 관련한 패키지도 많이 존재하고 있다. 아래는 금융공학을 위한 R패키지를 분류해보았다. 예를 들어, 제프리 라이언(Jeffrey Ryan)이 개발한 재무분석 패키지인 Quantmod는 주식가격을 가져와서 분석할 뿐만 아니라 그래프를 그리는 작업을 쉽게 해준다. Quantmod패키지를 설치하고, 로드해서 어떤 기업의 주식가격을 검색한 다음에 barChart함수를 이용해서 차트를 그리는 작업은 단지 몇줄의 짧은 코드만을 필요로 한다. 명령어를 R 편집기 창이나 애플리케이션 데이터인 경우에는 콘솔에 입력해서 실행할 수 있다.

<표 2> 금융공학을 위한 R 패키지 분류

구 분	패키지(package) 이름
금융데이터	fImport, quantmod, XML, XBRL, pdfetch, fExtremes
포트폴리오	portfolio, PortRisk, MarkowitzR, portfolioSim, FRAPO
금융모델링	quantmod, m4fe, fMultivar, fAsset, SMFI5, moments
트레이딩	fttradong, backtest
옵션가격	RQuantlib, fOptions, fAsianOptions, fExoticOptions, AmericanCallOpt, LSMonteCarlo, opefimor
이자율기간구조	termstrc, YieldCuve
시장미시구조	TAQMNGR, PIN, FinAsym
기타 금융분석	financial, timeSeries, FinancialInstrument

### Ⅲ. 금융공학을 위한 R과 R패키지의 상세활용

#### 1. 포트폴리오 관리를 위한 R 방법론

Markowitz(1952)의 포트폴리오 이론은 현대투자이론의 시작을 알리는 기념비적인 연구로써, Sharpe(1962), Lintner(1965), Mossin(1966) 등 주요 자산가격결정이론들의 토대를 이루고 있을 뿐만 아니라, 투자 의사결정을 위한 유용한 벤치마크 모형으로 지금도 실무에서 널리 활용되고 있다. “평균-분산 모형”으로 알려진 이 이론에서 투자자들은 투자의 평균수익률과 수익률분산(또는 표준편차)에만 관심을 가지게 되는데, 이는 자산의 수익률이 정규분포를 따르거나 혹은 투자자들의 효용함수가 이차함수 형태를 가지고 있다고 전제하기 때문이다. 비록 이러한 가정이 현실적인지에 대해서는 많은 비판이 제기될 수 있지만, 수익률의 비정규성이 강한 일부자산(헤지펀드와 같은 대체투자자산, 옵션과 같은 파생상품)들을 제외하면 적어도 벤치마크모형으로써는 충분히 기능할 수 있다는 사실에 큰 이견이 없다.

---

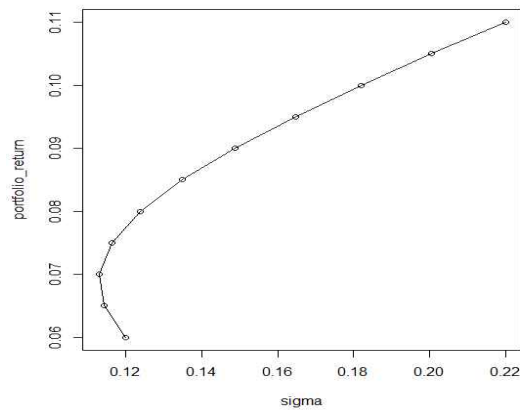
```

#두 자산 포트폴리오 문제
#자산1
r1 =.06
sigma1=.12
#자산2
r2 =.11
sigma2=.22
rho =.19
covar =rho*sigma1*sigma2
p =c(0,.1,.2,.3,.4,.5,.6,.7,.8,.9,1) #자산1의 퍼센트
##자산1의 퍼센트 함수로써 수익률의 포트폴리오##
variance=p^2*sigma1^2+(1-p)^2*sigma2^2+2*p*(1-p)*covar
sigma=sqrt(variance)
portfolio_return=r1*p+r2*(1-p)
cbind(p,sigma,return)
plot(sigma,portfolio_return,type='o')

```

---



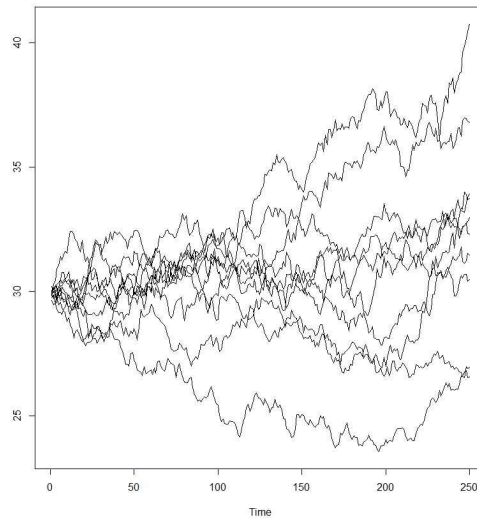


[그림 2] 두 자산의 포트폴리오 문제

## 2. 옵션 가격결정을 위한 R 방법론

옵션가격계산을 위해서는 Rmetrics에서 개발된 패키지를 이용하여 옵션가격계산을 할 수 있다.<sup>3)</sup> Rmetrics에서 제공하고 있는 옵션계산용 패키지는 fOptions, fAsianOptions, fExoticOptions을 통해 유럽형 옵션, 아시안 옵션, 이색 옵션 가격계산이 가능하다. 이외에도 OptionPricing, RQuantLib, AmericanCallOpt을 통해서도 가능하다. 일반적으로 산식이 존재하는 경우에 R코딩을 통해 프로그래밍한 옵션 계산 값과 위에서 설명한 패키지를 통한 계산값을 통해 가격검증이 가능하다. 아래 R스크립트는 로그정규분포의 주가 시뮬레이션과 콜옵션 가격계산을 위한 R스크립트를 보여주고 있다.

3) Rmetrics 전체 패키지(package)를 설치하는 입력문은 다음 2줄의 코드만 있으면 모두 설치된다.  
 > source("http://www.rmetrics.org/Rmetrics.R")  
 > install.Rmetrics( )



[그림3] 로그정규분포 주가 시뮬레이션

---

```

#로그정규분포(lognormal) 주가 시뮬레이션
days <- 250 #영업일
n <- 10 #주가 경로 수
S0=30 # 최초주가
dt<-1/days
r<-0.06
sigma<-0.12
S <- matrix((r-0.5*sigma^2)*dt+sigma*sqrt(dt)*rnorm(days*n),
ncol=days, nrow=n)
S <- S0*exp(apply(S,1,cumsum))
ts.plot(S)
#콜옵션 가격계산
C=6.187974 #콜 가격
X=25 #행사가격
C_t=pmax(S[,200,]-X,0)
exp(-r*200*dt)*mean(C_t)-C

```

---

블랙-숄츠(1973)의 옵션가격결정모형은 기초자산가격이 매순간마다 변동하는 경우를 고려하여 콜옵션의 균형가격을 도출하고자 한 모형이며 그 모형을 통해 R로 구현하면 아래와 같은 콜옵션 가격계산을 가능하게 해준다.

---

```
#블랙-숄츠 콜옵션 가격결정 공식
S =30      #현재주가
X =25      #행사가격
r =.06     #무위험이자율
dt =.8     #잔존만기(연간)
sigma =.12 #주가변동성, 표준편차
d1 =(log(S/X)+(r+0.5*sigma^2)*dt)/(sigma*sqrt(dt))
d2 =d1-sigma*sqrt(dt)
C =S*pnorm(d1)-X*exp(-r*dt)*pnorm(d2)
P1 =C-S*X*exp(-r*dt) #풋-콜 패리티 이용
P2 =X*exp(-r*dt)*pnorm(-d2)-S*pnorm(-d1)
```

---

옵션의 내재변동성(implied volatility)은 옵션의 존속기간 중 평균 변동성에 대한 시장의 믿음을 반영한다. 이는 과거 일정기간의 수익률을 이용하여 계산한 이동표준편차로 정의되는 역사적변동성(historical volatility)과 차이가 있다. 아래 R 스크립트는 현재주가, 행사가격, 무위험이자율, 잔존만기에 대한 정보가 주어졌을 때 현재의 옵션가격이 주어지면 미지수인 변동성 파라미터, 즉 내재된 변동성을 계산하는 방식을 코딩한 것이다.

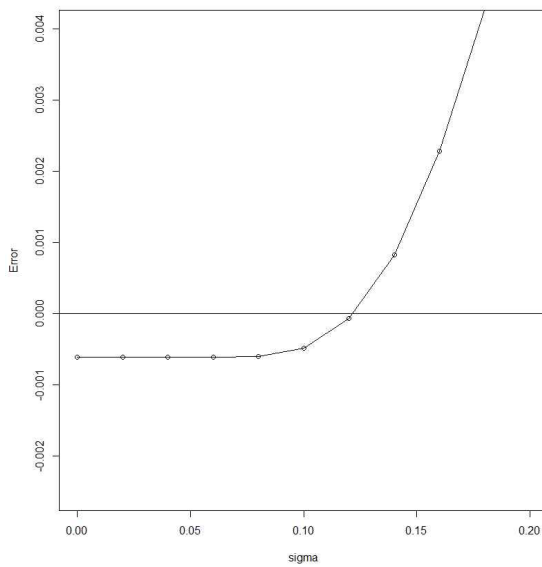
---

```
#블랙-숄츠 내재변동성 계산공식
S =30      #현재 주가
X =25      #행사가격
r =.06     #무위험이자율
dt =.8     #잔존만기(연간)
C =6.19    #콜옵션 가격
sigma =seq(0,.2,.02)
d1 =(log(S/X)+(r+0.5*sigma^2)*dt)/(sigma*sqrt(dt))
d2 =d1-sigma*sqrt(dt)
Error = pnorm(d1) - X/S*exp(-r*dt)*pnorm(d2) - C/S
```

---

```

cbind(sigma,Error)
plot(sigma,Error,type='o',ylim=c(-.0025,.004))
abline(h=0)
    
```



[그림 4] 블랙-숄즈 내재변동성 계산

### 3. 시장위험관리를 위한 R 방법론

통계학에서 불확실성을 나타내는 위험은 분산 또는 표준편차와 같은 변동성 지표로 측정하고 있다. 또한, 주식의 경우 베타, 채권의 경우 듀레이션, 옵션의 경우 델타 등으로 측정하고 있다. 그러나 각각의 상품별로 그 측정도구가 달라서 위험에 노출된 총금액을 파악하기 어려운 단점이 있다. 따라서, 여러 가지 금융상품의 위험을 종합적으로 파악하려면 VaR(Value at Risk)를 이용할 필요가 있다. VaR는 정상적인 금융시장에서 주어진 신뢰수준 하에에서 보유기간 동안에 발생할 수 있는 잠재적인 최대손실금액으로 정의된다. VaR가 1억원이라면 과거의 시장가격 동향으로 볼 때 정상적인 시장에서 99%의 확률로 10일 동안 발생할 수 있는 최대손실규모가 1억원이라는 의미이다. 다시 말하면 10일 동안 보유포지션에서 1억원을 초과하는 손실이 발생할 확률은 1%임을 의미한다. 한편, 금융위기 이후에 널리 사용되는 척도가 Expected Shortfall(이하, ES)이다. 이는 종종 조건부 VaR 또는 극단손실이라고 불린다. VaR는

“얼마나 악화될 수 있는가?”를 묻지만 ES는 “실제로 악화된다면 얼마의 손실이 기대되는가?”를 묻는다. ES는 T기간 동안 실제손실이 손실분포의 X번째 백분위 수를 넘어간 상황에서 기대손실을 나타낸다. 아래는 VaR와 ES에 대한 계산로직을 R의 예제로 표현한 것이다.

---

```

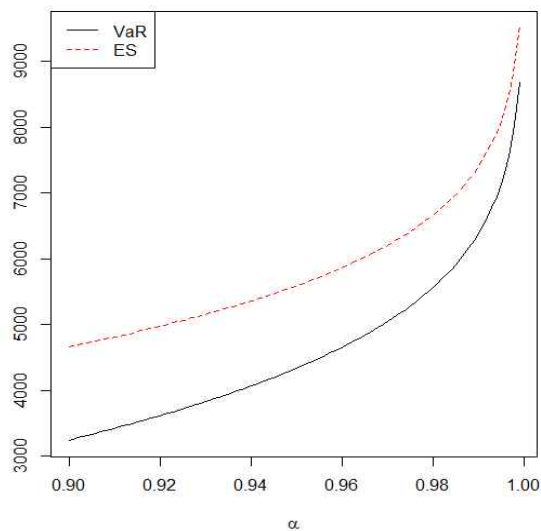
#VaR 계산
alpha <- c(0.95, 0.99)
qnorm(alpha, mean=-0.03, sd=0.15) * 20000

#Expected shortfall 계산
(-0.03 + 0.15 * dnorm(qnorm(alpha)) / (1 - alpha)) * 20000

#상기 계산식에 대한 시각화 그래프
x <- seq(0.9, 0.999, length=100)
yVaR <- (-0.03 + 0.15 * qnorm(x)) * 20000
yES <- (-0.03 + 0.15 * dnorm(qnorm(x)) / (1 - x)) * 20000
plot(x, yVaR, type="l", ylim=range(yVaR, yES),
      xlab=expression(alpha), ylab="")
lines(x, yES, lty=2, col=2)
legend("topleft", legend=c("VaR", "ES"), col=1:2, lty=1:2)

```

---



[그림 5] VaR와 Expected Shortfall 비교

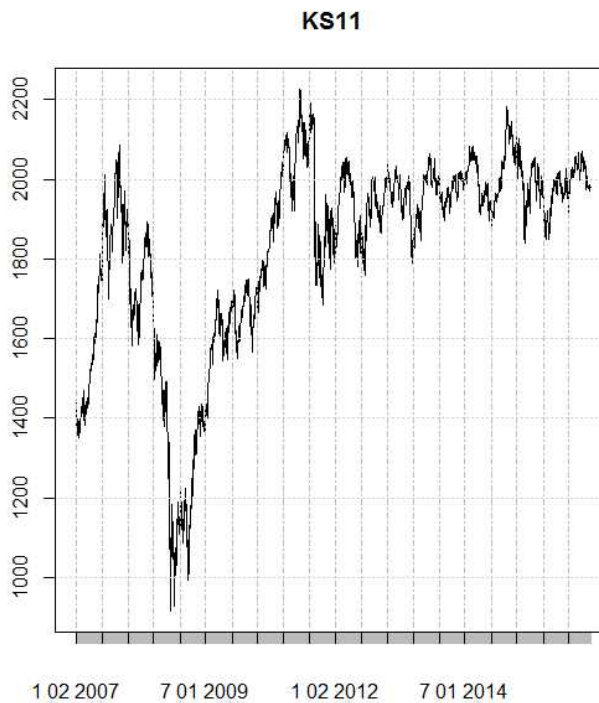
#### 4. 퀀트 트레이딩을 위한 R 방법론

R을 이용한 트레이딩과 기술적 분석은 `quantmod` package를 이용하여 분석을 주로 한다. 앞서 패키지 소개에서도 설명한 `quantmod`는 퀀트 트레이더를 위해 개발된 패키지로 통계적인 거래모형의 실행을 위한 전략개발, 테스트를 활용하기 위해 개발되었다. 아래는 `quantmod` 패키지를 설치하고 종합주가 지수 차트를 생성하는 간단한 R 스크립트를 보여주고 있다.

---

```
install.packages("quantmod") # quantmod 패키지 설치
library(quantmod)           # quantmod 패키지 로딩
getSymbols("^KS11", src="yahoo") # 야후에서 KOSPI지수 데이터 가져오기
plot.xts(KS11) # KOSPI 종합주가지수 일반차트생성
```

---



[그림 6] KOSPI 종합주가지수 차트 생성

본고 분량상 퀀트 트레이딩에 관한 내용을 모두 언급할 수 없으나, Quantmod 패키지를 이용하여 앞서 설명한 R코드에 다양한 기술적 함수를 추가하여 시장을 분석해 볼 수 있다.<sup>4)</sup> 특정 유형 데이터를 다운로드 받기를 원한다면 Quandl과 Datamarket에서 자료를 찾아볼 수 있다. 여기서는 앞서 패키지를 설치할 필요 없이 데이터를 R 포맷으로 다운로드받아 활용할 수 있다.

## 5. 금융공학 분야의 R 활용분야 고찰

전통적인 금융공학 분야인 자산 가격결정(asset pricing)과 위험관리(risk management)는 학계 뿐만 아니라 실무적으로도 퀀트(Quant)의 주요업무로 자리매김하고 있다. 최근 빅데이터 관련 시장의 지속적인 성장과 더불어 대규모 데이터 속에서 새로운 가치를 창출하기 위한 빅데이터 분야의 지식에 대한 수요가 증가하고 있다. 앞으로 모든 분야에서 데이터를 통한 의사결정이 중요해지고 이를 위한 컴퓨팅 자원과 분석 알고리즘이 필수가 될 것으로 판단된다. 앞서 <표 2>에서 설명했듯이 실제 금융공학에 유용한 패키지를 활용하여 교육을 진행할 수 있다는 장점이 있다. 실습교육을 통해 실제 함수를 작성하여 금융모델링을 한 후 그 계산 값이 맞는지 관련된 패키지(package)를 불러 정확한 값을 검증할 수 있다.

R을 활용한 금융공학교육 커리큘럼은 시계열 데이터분석, 포트폴리오 최적화, 자산 가격결정 모형, 채권가격계산, 파생상품가격결정, 위험관리에 집중해서 교육할 수 있으며, 관련 기부 패키지를 모두 활용할 수 있다. 이것은 학교에서 배운 내용을 실무에서 직접 활용할 수 있는 방향으로 확장할 수 있다는 장점이 있다. 또한, 향후 오픈소스 프로그램이라는 장점으로 점차 더 높은 경쟁력을 갖출 수 있으리라 판단되며, 결국 상업용 프로그램의 특징을 모두 장착할 것이라 예상할 수 있다. 현재 상업용 프로그램과 협업이 이루어지며 R과 다른 상업용 프로그램의 호환과 확장이 이루어지고 있으며, 상업용 프로그램과 R분석도구가 통합된 프로그램 상품이 출시될 것으로 보인다.<sup>5)</sup>

4) 관련 내용은 웹사이트 <http://www.quantmod.com/example/intro> 에서 더 자세한 정보를 볼 수 있다.

5) 마이크로소프트(Microsoft)는 오픈소스 분석도구인 레볼루션 R오픈을 인수하고 명칭도 MRO(Microsoft R Open)으로 변경하였다. 향후 마이크로 소프트는 R분석도구를 자사의 서비스와 상품에 통합할 예정이라고 밝히고 있다.

## VI. 결론

빅데이터에 대한 관심과 함께 오픈소스 언어에 대한 관심이 높아지고 있다. 이것은 많은 사람들에 의해 제공되는 기부 패키지(contributed package)를 기본으로 통계적 프로그래밍과 시각적인 표현을 정교하게 구현할 수 있다는 장점이 있기 때문이다. R은 컴퓨터 언어이면서 동시에 통계분석과 그래픽을 위한 다양한 패키지들의 집합이다. R은 다양한 통계분석이 가능하며 그래픽에도 강점을 가지고 있으며, 오픈소스로 운영되기 때문에 모든 것을 무료로 사용할 수 있다. 이는 SPSS, SAS, STATA 등 다른 통계프로그램이 유료이며 상당히 고가라는 점과 비교할 때 매우 큰 장점이다. 이러한 장점에도 불구하고 R은 통계분석과 계량경제학 분야에서 그 사용이 확산되고 있으나, 금융공학 분야의 사람들에게 보편적으로 사용되지 않고 있다. 금융시계열 분석을 위한 통계패키지로 많이 활용되는 것은 SAS, STATA, EVIEWS가 널리 활용되고 있고, 금융공학을 위한 실무언어로는 VBA, MATLAB, C++이 주로 사용되고 있다. 실제 금융실무에서 보수적인 시각으로 인해 오래전부터 사용되어왔던 GPU와 오픈소스언어에 대한 사용 논의가 최근에서야 활발해졌다는 점은 이러한 기법과 언어의 안정성이 높아졌다는 점에 기인할 것으로 생각된다.

본고에서는 R프로그램 소개와 활용, 금융공학 모델링 절차와 R 패키지 소개, 금융공학을 위한 R의 활용에 대해 고찰해 보았다. 최근 빅데이터 관련 시장의 지속적인 성장과 더불어 대규모 데이터 속에서 새로운 가치를 창출하기 위한 수요가 증가하고 있다. 앞으로 모든 분야에서 데이터를 통한 의사결정이 중요해지고 이를 위한 컴퓨팅 자원과 분석 알고리즘이 필수가 될 것으로 판단된다. 앞서 설명했듯이 실제 금융공학에 유용한 패키지를 활용하여 교육을 진행할 수 있다는 장점이 있다. 실제 함수를 작성하여 금융모델링을 한 후 그 계산 값이 맞는지 패키지를 불러 정확한 값을 검증할 수 있다. 시계열분석 모형, 자산가격결정모형, 위험관리 등에 관한 패키지를 모두 학계와 산업현장에 활용 가능하다. R의 유용성은 무엇보다도 학교에서 배운 내용을 금융산업 현장에서 바로 활용할 수 있는 분석용 언어라 판단되며, 오픈소스언어의 특성을 가지고 있어 상업용 통계패키지를 대체하는 수준까지 성장할 것으로 기대된다.



## 참고문헌

- 강병진 공역(2013), “금융기관의 위험관리”, 시그마프레스
- 김도형 역(2016), “파이썬을 활용한 금융분석”, 한빛미디어
- 김성재 역(2014), “R로 배우는 데이터분석 기본기 데이터 시각화”, 한빛미디어
- 박창래, 안창희(2015), “병렬컴퓨팅을 이용한 주가연계증권 가격결정 : 인텔 제온-제온  
파이 환경에서”, 5개 재무금융학회 발표논문
- 이금희, 김종오(2014), “금융데이터의 이해”, KNOU PRESS
- 이운환(2016), “제대로 알고 쓰는 R통계분석”, 한빛아카데미
- 유시용, 김삼용(2008), “국내 금융공학교육 현황과 향후 발전방향”, 응용통계연구  
21(5), pp. 765-774
- 유충현(2010), “예제로 배우는 R프로그래밍”, 자유아카데미
- 장기천, 강병진(2016), “R언어로 짜는 금융프로그래밍”, 서울경제경영
- 조민호(2016), “빅데이터분석을 위한 R프로그래밍”, 정보문화사
- 전희원(2014), “R로 하는 데이터 시각화”, 한빛미디어
- 홍창수(2014), “장외파생상품 실무입문”, 서울경제경영
- Bennett and Hugen(2016), “Financial Analytics with R”, Cambridge.
- Berlinger, E. (2015), “Mastering R for Quantitative Finance”, PACKT Publishing.
- Christopher Conlan(2016), “Automated Trading with R”, Apress
- Clifford S. Ang(2015), “Analyzing Financial Data and Implementing Financial Model  
using R”, Springer
- Daroczi, G. (2013), “Introduction to R for Quantitative Finance”, PACKT Publishing.
- Florian Heiss(2016), “Using R for Introductory Econometrics”, www.urfie.net
- Georgakopoulos, H.(2015), “Quantitative Trading with R- UnderstandingMathematical and  
Computational Tools from a Quant’s Perspective”, PalgraveMacmillan US.
- Hadley Wickham and Garrett Golemund(2017), “R for Data Science”, OREILY
- Jack Xu(2016), “Practical Quantitative Finance with R”, UniCAD Publishing.
- Karthik Ramasubramanian and Abhishek Singh, “Machine Learning using R”,  
Apress
- Tsay, R. S. (2014), “Analysis of Financial Time Series”, John Wiley & Sons
- Yuxing Yan(2016), “Financial Modeling using R”, Tate Publishing & Enterprises

Abstract

## A Study on the Usefulness of Statistical Package R for Financial Engineering Education

*Changsoo Hong\* and Jae-Seung Bae\*\**

R is a powerful open source functional programming language that provides high level graphics and interfaces to other languages. Its strength lies in data analysis, graphics, visualization, and data manipulation. R is becoming a widely used modeling tool in finance, science and engineering.

R program is also supported by a large international team of volunteers who maintain versions of R for multiple platforms. In particular, there are dozens of package for solving problems in finance. This research introduces financial package that is financial engineering education using R statistic calculation and graphics. R provides a wide variety of financial package (timeseries, finance, risk management, data management) and graphical technique, and is highly extensible. In the long run, it seems likely that R and oter open source packages will survive and prosper. Because of their higher growth rate, they will eventually provide almost all of the features of commercial products. When that point will be reached is unknown. In the classroom, where the focus is on learning and adaptability, the free R program is rapidly displacing other alternatives.

Key Words : Financial Engineering, Package, Portfolio, Option model, risk management

---

\* Financial Engineering Research Center, NICE Pricing and Information Inc.

\*\* Department of International Finance, Hankuk University of Foreign Studies