*The Korean Journal of Applied Statistics* (2017) DOI: http://dx.doi.org/10.5351/KJAS.2017.30.6.877
**30**(6), 877–888

# Bayesian estimation for frequency using resampling methods

Ro Jin Pak[a,1]

[a]Department of Applied Statistics, Dankook University

(Received September 5, 2017; Revised October 6, 2017; Accepted October 25, 2017)

**Abstract**

Spectral analysis is used to determine the frequency of time series data. We first determine the frequency of the series through the power spectrum or the periodogram and then calculate the period of a cycle that may exist in a time series. Estimating the frequency using a Bayesian technique has been developed and proven to be useful; however, the Bayesian estimator for the frequency cannot be analytically solved through mathematical equations and may be handled numerically or computationally. In this paper, we make an inference on the Bayesian frequency through both resampling a parameter by Markov chain Monte Carlo (MCMC) methods and resampling data by bootstrap methods for a time series. We take the Korean real estate price index as an example for Bayesian frequency estimation. We have found a difference in the periods between the sale price index and the long term rental price index, but the difference is not statistically significant.

Keywords: filtering, fourier transform, Markov chain Monte Carlo, *R*-package, signal processing, spectral analysis, spectrum, time series

## 1. 서론

정수 순열 $\{n; \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$에 대하여 관측값 $\{x_n\}$을 갖는 순열을 이산시간신호(discrete time signal)라 하고 흔히 시계열(time series)이라고 부른다. 시계열 자료의 주기를 파악하는 방법으로 베이지안 추정을 소개하고 시계열 재표본을 활용한 유용성을 부동산 가격 지수 데이터를 예로 들어 시연하여 보았다.

자연 현상 혹은 사회 현상에서 많은 종류의 시계열을 접하게 되는데 그 시계열이 어떤 주기를 갖고 있는지가 매우 중요한 관심사 중에 하나이다. 예를 들어 부동산 관련 데이터 속에 어떤 순환 주기가 존재하는 가를 알면 부동산 가격의 예측에 크게 도움이 될 것이다. 주기를 파악하기 위해 주기의 역수인 주파수에 관심을 갖게 되었고 주파수의 전체적인 특징을 찾고자 하는 노력들이 이어졌다.

주어진 시계열이 갖고 있는 주파수(frequency; $f$) 특징을 나타내는 전력스펙트럼밀도(power spectral density; PSD) 함수를 아래와 같이 정의한다.

$$P_{xx}(f) = \lim_{M\to\infty} E\left[\frac{1}{2M+1}\left|\sum_{n=-M}^{M} x_n \exp(-j2\pi f n)\right|^2\right].$$

[1]Department of Applied Statistics, Dankook University, 152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 16890, Korea. E-mail: rjpak@dankook.ac.kr

그런데 실제로 주어지는 시계열은 관측치가 예를 들어 자연수 $N$개로 이루어진 유한 시계열인 경우가 많고 이런 경우 위에서 정의한 PSD의 추정량인

$$\hat{P}_{xx}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x_n \exp(-j2\pi f n)\right|^2$$

을 사용하고 이를 피리오도그램(periodogram)이라 부른다. 위에서 정의된 피리오도그램이 데이터의 개수가 많아지면 성능이 향상될 수 있으나 PSD에 대한 불편추정량(unbiased) 또한 일치추정량(consistent)이 아니라는 약점을 갖고 있다. Bartlett (1948)과 Welch (1967) 등에 의해 여러 가지 비모수적 방법을 통해서도 성능이 향상됨을 보였으나 여전히 주파수를 명확하게 파악하는데 어려움이 있다.

한편에서는 PSD 자체를 추정하기 보다는 모수로서의 주파수의 추정에 집중하는 연구들이 진행되었다. 그 중에서 본 연구에서는 베이지안 방법론을 소개하고자 한다. 그런데 주파수에 대한 베이지안 추정량의 경우 그 추정량이 확률변수들의 함수 혹은 어떤 공식으로 표현되지 않아 대수적으로 구할 수 없고 심도 있는 추정이 용이하지 않다. 다행히 이러한 문제는 수치 해석적 방법을 통해 계산적으로 해결될 수 있다. 일반적으로 베이지안 추정의 경우에는 Markov chain Monte Carlo (MCMC)와 같은 방법을 사용할 수 있겠으나 본 논문에서 구해진 사후확률함수가 특수한 형태를 갖고 있어 안정된 결과를 보장할 수 없는 상황이다. 본 논문에서는 베이지안 추정의 전통적 방법과 다소 차이가 있으나 하나의 현실적 대안으로서 마치 부트스트랩 추정을 하듯이 시계열 자료를 재표본하는 방법을 통해서 주파수에 대한 통계적 추정을 시도하여 보았다. 즉, 모수를 재표본하는 방법과 데이터를 재표본하는 방법을 사용하여 보았다. 예제로서 부동산 매매/전세 가격 지수 데이터를 사용하였고 매매와 전세 가격 지수 간에 3.7개월 정도의 주기 차이가 존재하나 그 차이가 통계적으로는 유의한 차이가 아니라고 볼 수도 있음을 보였다.

## 2. 베이지안 주파수 추정

주파수에 대한 베이지안 추정은 Jaynes (1987)에 의해 처음 소개되었다. 본 연구에서는 Gregory (2005)와 Bretthorst (2013)의 책을 참조하였다. 두 책은 본래 신호처리를 위해 고안된 방법을 제시하고 있으나 시계열에 맞도록 용어나 방식을 어느 정도 추가 혹은 생략하는 과정을 통해 재구성하여 아래에 간단하게 설명하려 한다.

시간 $\{n = 0, 1, 2, \ldots, N-1\}$에 따른 관측치 $x_n$가 모형 $m(n)$과 오차 혹은 소음 $\epsilon_n$의 결합으로

$$x_n = m(n) + \epsilon_n$$

로 정의된다고 하자. 특별히 모형은

$$m(n) = A_1 \cos(\omega n) + A_2 \sin(\omega n), \quad \omega = 2\pi f,\ f = \text{frequency},$$

또한 $\epsilon_n$의 확률함수는 평균이 0이고 분산이 $\sigma^2$인 가우시안 혹은 정규분포로 가정한다.

관측치에 대한 확률함수는

$$P(x_n|A_1, A_2, \omega, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[x_n - m(n)]^2}{2\sigma^2}\right)$$

가 된다. 만일 $N$개의 관측치가 주어지면 우도함수는 $X = \{x_1, \ldots, x_N\}$에 대하여

$$P(X|A_1, A_2, \omega, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}[x_n - m(n)]^2\right) \tag{2.1}$$

라고 쓸 수 있다.

먼저, 우리의 목표는

$$P(\omega,\sigma|X)=\frac{P(\omega,\sigma,X)}{P(X)}=\frac{\iint P(\omega,\sigma,A_1,A_2,X)P(A_1)P(A_2)dA_1dA_2}{P(X)} \tag{2.2}$$

를 찾는 것이 되겠다. 계산상의 어려움을 감소하기 위해 $\omega$, $\sigma$, $A_1$, $A_2$에 대해 균등 사전분포(uniform prior)를 가정하기로 하자. 식 (2.2)의 베이즈 공식 속 분모 $P(X)$는 $X$의 함수이므로 사후분포는 사실상 베이즈 공식의 분자의 비례식으로 구하여도 주파수를 추정하는데 충분하다고 하겠다.

식 (2.1)의 지수함수에 포함된 제곱합을 $D^2$라 쓰고

$$\begin{aligned}D^2&=\sum_{n=0}^{N-1}[x_n-m(n)]^2\\&=\sum_{n=0}^{N-1}x_n^2+\sum_{n=0}^{N-1}m^2(n)-2\sum_{n=0}^{N-1}x_nm(n)\\&=\sum_{n=0}^{N-1}x_n^2+\sum_{n=0}^{N-1}m^2(n)-2[A_1R(\omega)+A_2I(\omega)],\\R(\omega)&=\sum_{n=0}^{N-1}x_n\cos(\omega n),\ I(\omega)=\sum_{n=0}^{N-1}x_n\sin(\omega n)\end{aligned} \tag{2.3}$$

와 같이 전개할 수 있다. 이제, 식 (2.3)의 두 번째 항을 전개하면

$$\sum_{n=0}^{N-1}m^2(n)=A_1^2\sum_{n=0}^{N-1}\cos^2(\omega n)+A_2^2\sum_{n=0}^{N-1}\sin^2(\omega n)+2A_1A_2\sum_{n=0}^{N-1}\cos(\omega n)\sin(\omega n)$$

가 된다. 여기서, $x^N=1$의 해를 $e^{j\omega}$라 하면

$$\begin{aligned}\sum_{n=0}^{N-1}\cos^2(\omega n)&=\frac{N}{2}+\frac{1}{2}\sum_{n=0}^{N-1}\cos(2\omega n)=\frac{N}{2}+\frac{1}{2}\Re\left(\sum_{n=0}^{N-1}e^{j\omega n}\right)\\&=\frac{N}{2}+\frac{1}{2}\Re\left(\frac{1-e^{j\omega N}}{1-e^{j\omega}}\right)=\frac{N}{2}\end{aligned}$$

이 되고 비슷한 방법으로

$$\begin{aligned}\sum_{n=0}^{N-1}\sin^2(\omega n)&=\frac{N}{2}-\frac{1}{2}\sum_{n=0}^{N-1}\cos(2\omega n)=\frac{N}{2},\\\sum_{n=0}^{N-1}\cos(\omega n)\sin(\omega n)&=\frac{1}{2}\sum_{n=0}^{N-1}\sin(\omega n)=0\end{aligned}$$

이 된다. 결국, $\bar{x}^2=\sum x^2/N$로 표시하면

$$D^2=\sum_{n=0}^{N-1}[x_n-m(n)]^2=N\bar{x}^2+\frac{N}{2}\left(A_1^2+A_2^2\right)-2[A_1R(\omega)+A_2I(\omega)]$$

가 된다.

이제 식 (2.1)의 우도함수의 지수 부분은

$$\exp\left(\frac{-D^2}{2\sigma^2}\right) = \exp\left(\frac{-N\bar{x}^2}{2\sigma^2}\right)\exp\left(-\frac{NA_1^2}{4\sigma^2} + \frac{R(\omega)A_1}{\sigma^2} - \frac{NA_2^2}{4\sigma^2} + \frac{R(\omega)A_2}{\sigma^2}\right) \tag{2.4}$$

가 된다. 식 (2.4)는 아래 적분식을 활용하여

$$\int_{-\infty}^{\infty}\exp(-ax^2 - bx)dx = \sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2}{4a}\right) \quad (a > 0)$$

$A_1$과 $A_2$에 대하여 적분을 수행하면

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(\frac{-D^2}{2\sigma^2}\right)dA_1dA_2 = 4\sigma^2\frac{\pi}{N}\exp\left(\frac{-N\bar{x}^2}{2\sigma^2}\right)\exp\left(\frac{R^2(\omega)+I^2(\omega)}{N\sigma^2}\right)$$

가 되고 $\omega$와 $\sigma$에 대한 사후확률함수를 비례식의 형태로 아래와 같이 구할 수 있다.

$$P(\omega,\sigma|X) \propto \sigma^{2-N}\exp\left(\frac{-N\bar{x}^2}{2\sigma^2} + \frac{R^2(\omega)+I^2(\omega)}{N\sigma^2}\right). \tag{2.5}$$

이제, $\omega$에 대한 사후확률함수를 구하기 위해 식 (2.5)에서 (i) $\sigma$에 대해 적당한 추정량 $\hat{\sigma}$를 넣거나 (ii) $\sigma$를 장애모수(nuisance parameter)로 여겨 $\sigma$에 대하여 적분을 수행하게 되면 (i) $\sigma$를 알거나 추정을 하는 경우

$$\begin{aligned} P(\omega|X) &\propto \exp\left(\frac{R^2(\omega)+I^2(\omega)}{N\hat{\sigma}^2}\right) \\ &= \exp\left(\frac{\left|\sum_{n=0}^{N-1}x_ne^{-j\omega n}\right|^2}{N\hat{\sigma}^2}\right) = \exp\left(\frac{\hat{P}_{xx}(\omega)}{\hat{\sigma}^2}\right), \end{aligned} \tag{2.6}$$

(ii) $\sigma$를 장애모수로 처리하는 경우

$$\begin{aligned} P(\omega|X) &= \int P(\omega,\sigma|X)d\sigma \\ &= \sigma^1\int_0^{\infty}\sigma^{1-N}\exp\left(\frac{-N\bar{x}^2}{2\sigma^2} + \frac{R^2(\omega)+I^2(\omega)}{N\sigma^2}\right)d\sigma \\ &\propto \left(\frac{N\bar{z}^2}{2} - \frac{\hat{P}_{xx}(\omega)}{\hat{\sigma}^2}\right)^{\frac{2-N}{2}} \end{aligned} \tag{2.7}$$

를 얻게 된다. 식 (2.7)은 아래와 같은 변환과 적분을 통해 구하여진다. 즉,

$$a = \frac{N\bar{z}^2}{2} - \frac{R^2(\omega)+I^2(\omega)}{N}; \quad x = \frac{1}{\sigma}$$

로 대입하면

$$\int_0^{\infty}x^{N-3}\exp\left(-ax^2\right)dx = \frac{1}{2}\Gamma\left(\frac{N-2}{2}\right)a^{\frac{2-N}{2}} \quad (a > 0, N > 2).$$

따라서 (i)의 식 (2.6) 혹은 (ii)의 식 (2.7)에서 구한 $P(\omega|X)$의 기대값 $E[\omega|X]$를 $\omega$의 추정량으로 하겠다. 공학적 관점에서는 $P(\omega|X)$를 극대화 시키는 $\omega$로써 추정량을 삼을 수도 있을 것이다. $\omega$에 대한 추정량이 구해지면 주파수 ($f = \omega/2\pi$)에 대한 베이지안 추정량을 계산할 수 있다.

그런데 실제적인 문제는 베이지안 추정량이 수식으로 나타내기 어려운 음함수(implicit function) 형태로 묶여 있어 심도 있는 추정은 수치 해석적 계산으로 수행하여야 되는 상황이다. 예컨대 식 (2.6)과 식 (2.7)에 있는 사후확률함수의 누적확률함수를 구하기 위한 적분조차도 사실상 매우 어렵다는 점이다. 이러한 경우 MCMC를 활용한 경험적 사후확률함수 추정이 널리 사용되고 있다. 다만 주파수에 대한 사후확률함수가 마치 크로네커(Kronecker) 혹은 디렉(Dirac) 델타 함수처럼 중심부의 폭이 좁고 뾰족한 모양을 갖는 경우에 표본을 생성하는 과정에서 확률값이 매우 작은 부분에서 데이터가 선택되면 그 부분을 벗어날 수 없는 경우가 생길 수 있다. 따라서 MCMC를 이용해 추정된 사후확률함수는 본래의 사후확률함수보다 폭이 넓고 부드러운 모양을 가질 수도 있고 그에 따른 추정 오류가 생길 여지가 있다고 하겠다. 따라서 본 연구에서는 특별히 메트로폴리스-헤이스팅스 알고리즘 (Hastings, 1970)을 활용한 경험적 사후확률함수 추정을 통한 방법과 더불어 마치 부트스트랩을 통해 추정을 수행하듯이 시계열의 재표본을 생성하여 경험적으로 주파수의 신뢰구간을 구하는 방법을 시도해보려 한다.

## 3. 시계열 재표본 방법론

재표본에 관한 아이디어를 소개한 Efron (1979)은 데이터들의 독립성을 기본 가정으로 하고 있다. 관측치들 사이에 독립성이 확실치 않은 경우 Efron (1979)의 방법을 직접적으로 사용할 수는 없다. 시계열 자료처럼 근본적으로 서로 종속적인 관측치들에 대한 특별한 재표본 방법론이 요구된다. 시계열 자료 같은 종속적 데이터의 재표본 방법은 모수적 방법과 비모수적 방법이 있다.

먼저 모수적 방법은 모델을 상정하고 데이터에 모델을 적합하여 잔차를 구한 다음 잔차의 독립성을 이용하여 일반적인 재표본 방법을 적용하는 것이다. 한편 비모수적 방법은 시계열이 만일 정상적(stationary)이라면 관측 기간에서 동일한 통계적 (확률적) 구조를 가질 것이라는 가정 하에 전체 데이터를 몇 개의 블록으로 나누고 블록을 재표본하고 블록 표본들을 이어 붙여서 새로운 재표본 시계열을 얻는 것이다. 자세한 설명은 Kunsch (1989), Politis와 Romano (1994), Davison과 Hinkley (1997), Lahiri (1999), 그리고 Kreiss와 Lahiri (2012)에 잘 정리되어 있다. 아래에 두 가지 방법에 대해 추가적으로 설명을 하겠다.

### 3.1. 모수적 방법

Efron (1979)의 아이디어는 독립표본을 가정하고 있기 때문에 시계열 자료의 경우 모델을 설정할 때 오차항을 독립으로 가정한다는 점에서 모델 적합 후 얻은 전차들로부터 재표본을 얻어 시계열 재표본을 얻어내자는 것이다. 그 과정은 다음과 같다. 먼저 시계열 자료 가 주어졌다고 하자.

- 적당한 모델을 적합한다. 계산상의 편리를 위해 주로 자기회귀모형(autoregressive model; AR)이 선호된다.
- 모형에 대한 추정을 수행하고 잔차를 구한다.
- 잔차에 대한 표준화를 수행하여 표준화 잔차를 구하고 그들로부터 재표본을 구한다.
- 잔차의 재표본을 통해 추정된 모형을 역으로 이용하여 재표본을 하나씩 축차적으로(sequentially) 시계열 재표본을 구한다.

### 3.2. 비모수적 방법

주어진 시계열이 정상적이라고 하자.

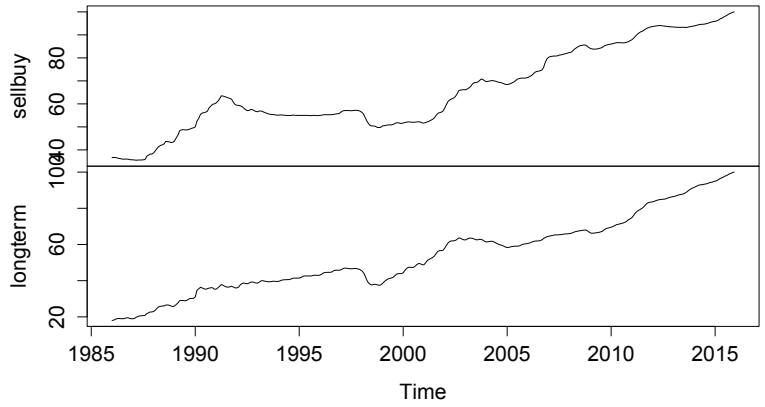


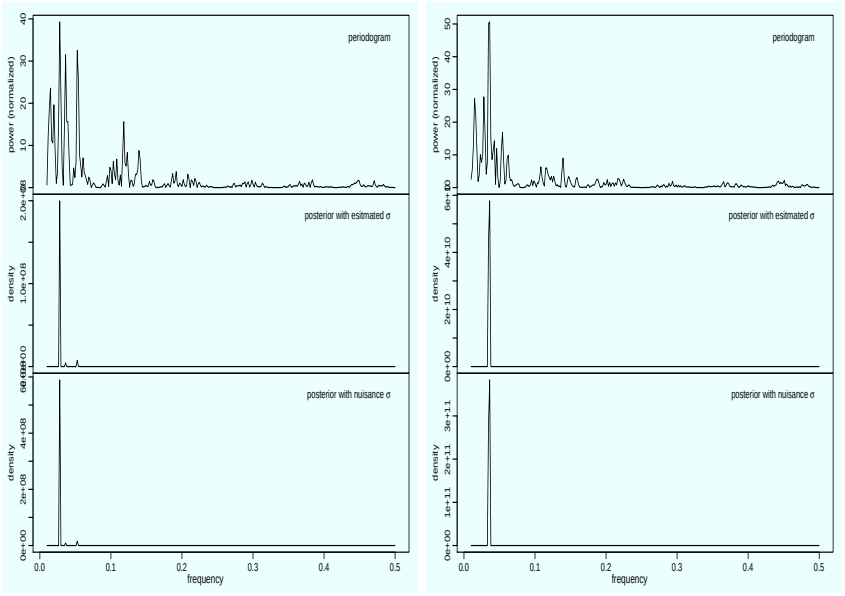

**Figure 4.1.** Real estate price index.

- 적당한 자연수 $n$을 택하여 길이가 $n$인 데이터 블록을 만들되 (1) 블록들이 겹치지 않도록 만들거나, (2) 겹침을 허락하여 만들거나, (3) 처음부터 $n$을 랜덤하게 선택하여 블록을 만든다.
- 만들어진 블록들 중에서 중복이 허락되도록 블록들을 랜덤 추출하여 뽑힌 블록들을 시간 상 흐름에 맞도록 이어 붙이되 원 시계열 자료 개수와 같도록 블럭 길이와 블록 수를 선택하여 재표본을 구성하게 된다.

블록을 이용하는 비모수적 방법의 경우에 블록의 길이를 정하는 것이 가장 중요한 일로서, Hall 등 (1995)이 먼저 최소 평균제곱오차법을 제안하였고 Dudek 등 (2014)은 보다 정교하고 일반화 가능한 방법을 제안하였다. 비록 지금까지 블록 크기에 대한 다양한 방법들이 제시되어 있지만 뚜렷하게 탁월하다고 여겨지는 방법은 없어 보인다. 무난한 방법은 이론적 길이들이 주기의 함수로 표현 되는 경우가 많은 바, 계절성 등에 의한 주기가 뚜렷하게 존재하는 경우 간단하게 주기를 블록의 최소 단위로 사용할 수 있다고 사려 된다. 본 논문에서는 Hayfield와 Racine (2017)이 앞선 여러 방법들과 Patton 등 (2009)을 근거로 만든 R패키지 중 NP패키지의 B.STAR함수를 이용해 블록의 최적 길이를 구하려 한다.

## 4. 데이터 분석

1. 데이터 설명: 본 연구에서 예제로 사용한 데이터는 국민은행에서 보유하고 있는 부동산 관련 자료 중 1986년 1월에서부터 2015년 12월까지의 월별 매매 가격 종합 지수와 전세 가격 종합 지수이다. 이번에 사용한 국민은행 부동산 데이터는 전국 아파트 및 주택을 모집단으로 하여 층화 2단 집락 확률비례추출법을 통해 추출된 표본들의 가격 지수를 매달 일회씩 조사한 데이터이다. 가격 지수는 선택된 표본들의 부동산 상태에 따른 가중치를 이용하여 가격 가중평균을 구하고 2015년 12월을 100으로 하여 상대적으로 계산된다. 2015년 12월의 매매(가격)지수와 전세(가격)지수는 모두 100으로 계산되는데 그것이 매매가와 전세가가 같다는 의미는 아니다. 매매(sell & buy)와 전세(long term)지수의 월별 가격 지수들의 시계열 그림을 Figure 4.1에 그려 넣었다. 전체 구간에서 보면 상승 추세를 보이며 매매와 전세 지수가 변화하는 모습이 아주 유사함을 알 수 있다.
2. 주파수 추정: 부동산 데이터를 일반차분과 12개월 계절차분을 1회씩 실시하여 정상화시킨 후 피리오도그램과 베이지안 사후확률밀도함수 추정법을 수행하였다. 분석 결과를 정리하면 피리오도그램에 비해 베이지안 추정의 경우 뾰족한 모양이 하나만 존재하여 명확하게 유력 주파수를 확인할 수 있을




(a) Sell & buy (b) Long term rent

**Figure 4.2.** Periodogram, Bayesian posterior density.

**Table 4.1.** Bayesian posterior estimate for the frequency

| | | Frequency points | | | | |
|---|---|---|---|---|---|---|
| | | 300 | 500 | 700 | 900 | 1000 |
| Sell & buy | $\sigma$-estimated | 0.0290 | 0.0326 | 0.0309 | 0.0311 | 0.0311 |
| | $\sigma$-nuisance | 0.0287 | 0.0323 | 0.0303 | 0.0305 | 0.0305 |
| Long term | $\sigma$-estimated | 0.0355 | 0.0355 | 0.0352 | 0.0354 | 0.0354 |
| | $\sigma$-nuisance | 0.0355 | 0.0355 | 0.0352 | 0.0354 | 0.0354 |

만큼 해상도가 높다는 것을 확인할 수 있다 (Figure 4.2).

베이지안 추정치는 $E[\omega|X]=\int \omega p(\omega|X)d\omega$로서 사후확률밀도함수가 지수함수에 근거한 연속형 함수이지만 대수적으로 적분이 용이하지 않아 수치 해석적 적분 계산을 수행하였다. 이를 위해 주파수축을 짧은 구간들로 나누는 점의 개수를 300, 500, 700, 900, 1,000으로 정하여 적분을 수행하였다. 실제로 눈금 수에 따라 추정치의 차이가 조금씩 있지만 매매는 0.031 그리고 전세는 0.035에 접근한다고 하겠다. 즉, 매매는 유력 주파수가 약 0.031 그리고 전세는 약 0.035라고 하겠다 (Table 4.1). 이는 주기로 보면 매매는 32.2개월 그리고 전세는 28.5개월로 전세 주기가 3.7개월 정도 짧다고 하겠다. 부동산 경제학적 측면에서 3.7개월 차이가 어떤 의미가 있을 수 있다면 그 차이가 통계적으로 유의한 차이인지 검증할 필요가 있겠다. MCMC에 의한 방법과 시계열 재표본을 이용하여 경험적으로 차이를 검증해 보고자 한다.

3. MCMC에 의한 추정치: 매매/전세 시계열 데이터로 부터 구한 사후확률함수로부터 $\omega$에 대한 MCMC 표본을 추출하고 $\omega$의 추정치를 구하였다. 앞서 언급했듯이 사후확률함수는 막대기처럼 뾰족하나 추정함수들은 폭이 넓고 경우에 따라 오른쪽에 형성되는 경향이 있고 추정함수들의 평균이 본래의 사후확률함수의 평균보다 큰 값을 가질 수 있다. 예로서 전세지수에 대해 식 (2.6)에 의한 사

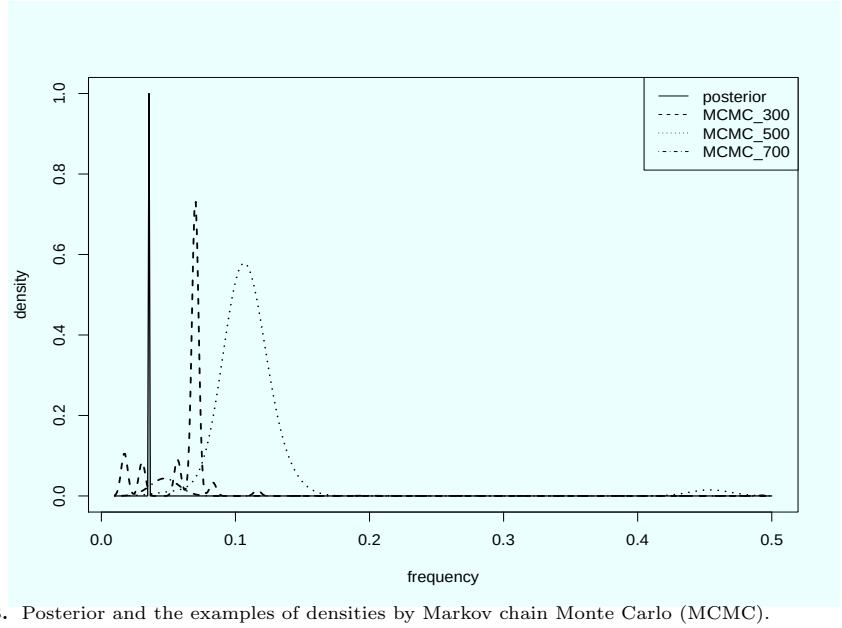



**Figure 4.3.** Posterior and the examples of densities by Markov chain Monte Carlo (MCMC).

**Table 4.2.** Means of the Bayesian posterior estimates for the frequency by MCMC

| | | Sampling repetition = 1000 | | | Sampling repetition = 2000 | | |
|---|---|---|---|---|---|---|---|
| | | Sample size | | | Sample size | | |
| | | 300 | 500 | 700 | 300 | 500 | 700 |
| Sell & buy | $\sigma$-estimated | 0.0409 | 0.0382 | 0.0363 | 0.0413 | 0.0384 | 0.0367 |
| | $\sigma$-nuisance | 0.0416 | 0.0374 | 0.0365 | 0.0415 | 0.0377 | 0.0357 |
| Long term | $\sigma$-estimated | 0.0382 | 0.0364 | 0.0362 | 0.0380 | 0.0364 | 0.0359 |
| | $\sigma$-nuisance | 0.0380 | 0.0366 | 0.0361 | 0.0376 | 0.0366 | 0.0360 |

MCMC = Markov chain Monte Carlo.

후확률함수와 표본크기가 300, 500, 그리고 700인 MCMC에 의한 추정함수를 각각 한 개씩 Figure 4.3에 그려보았다. 전세지수 데이터로 부터 구한 사후확률함수보다 오른쪽에 분포하는 경우가 있음을 볼 수 있다. 실제로 MCMC를 통해 얻은 추정함수들로부터 계산된 평균들은 모의 실험의 반복수 (1000회, 2000회)나 표본크기에 상관없이 참 값인 0.031 (매매)와 0.035 (전세)를 상회하고 있다 (Table 4.2). 표본크기가 300, 500, 그리고 700인 MCMC에 의한 표본을 추출하는 과정을 1,000과 2,000번 수행하고 추정치들의 평균에 대한 95%-신뢰구간을 구하고 평균의 상자그림과 더불어 신뢰구간 그림을 그려보았다 (Table 4.3, Figure 4.4). 표본크기가 클수록 신뢰구간이 좁아짐을 볼 수 있고 대부분의 상자가 0.031을 상회함을 볼 수 있다.

4. 시계열 재표본을 통한 추정치: (1) 매매/전세 시계열 데이터로부터 동수인 360개 데이터를 모델을 이용한 방법과 블록을 이용한 방법을 통해 추출하고 (2) 사후확률함수를 구한 후 (3) 기대값을 적분을 통해 구하여 $\omega$의 추정치를 계산하였다. 블록의 크기는 R의 b.star에의해 매매의 경우 12개월, 전세의 경우 55개월로 하였다. 블록 (비모수적 방법) 그리고 모델 (모수적 방법)을 기반으로 하는 재표본에 따른 $\omega$의 신뢰구간을 수치해석 적분을 위한 주파수축 구분점의 개수 (300, 500, 700)에 따른 결과와 재표본 횟수 (1000, 2000)에 따른 결과로 구분하여 Table 4.4에 수록하였다. 시각적 표현을 위해 상자 도표와 95% 신뢰구간 도표를 Figure 4.5에 그려 보았다.

**Table 4.3.** MCMC confidence interval of the posterior estimates

| | | | Sampling repetition = 1000 | | | Sampling repetition = 2000 | | |
|---|---|---|---|---|---|---|---|---|
| | | | Sample size | | | Sample size | | |
| | | | 300 | 500 | 700 | 300 | 500 | 700 |
| Sell & buy | $\sigma$-estimated | $l$ | 0.0204 | 0.0219 | 0.0229 | 0.0209 | 0.0225 | 0.0226 |
| | | $u$ | 0.0615 | 0.0545 | 0.0498 | 0.0617 | 0.0544 | 0.0509 |
| | $\sigma$-nuisance | $l$ | 0.0202 | 0.0223 | 0.0227 | 0.0205 | 0.0220 | 0.0225 |
| | | $u$ | 0.0632 | 0.0525 | 0.0496 | 0.0627 | 0.0534 | 0.0490 |
| Long term | $\sigma$-estimated | $l$ | 0.0205 | 0.0255 | 0.0280 | 0.0205 | 0.0253 | 0.0275 |
| | | $u$ | 0.0560 | 0.0474 | 0.0445 | 0.0557 | 0.0475 | 0.0445 |
| | $\sigma$-nuisance | $l$ | 0.0210 | 0.0257 | 0.0283 | 0.0210 | 0.0254 | 0.0280 |
| | | $u$ | 0.0552 | 0.0475 | 0.0441 | 0.0542 | 0.0478 | 0.0442 |

$l$ = 95%-lower limit; $u$ = 95%-upper limit. MCMC = Markov chain Monte Carlo.

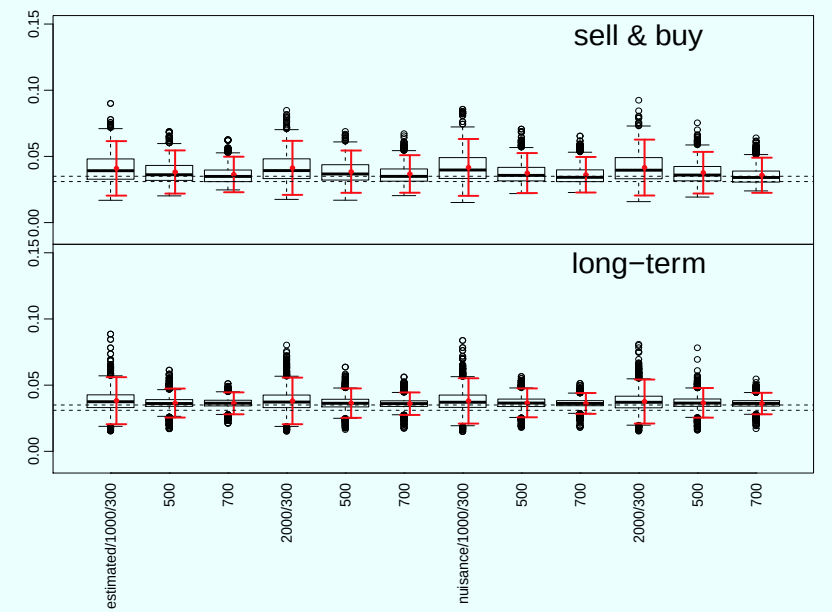



**Figure 4.4.** Box and 95%confidence interval plots by Markov chain Monte Carlo; the dotted lines are at 0.031 and 0.035.

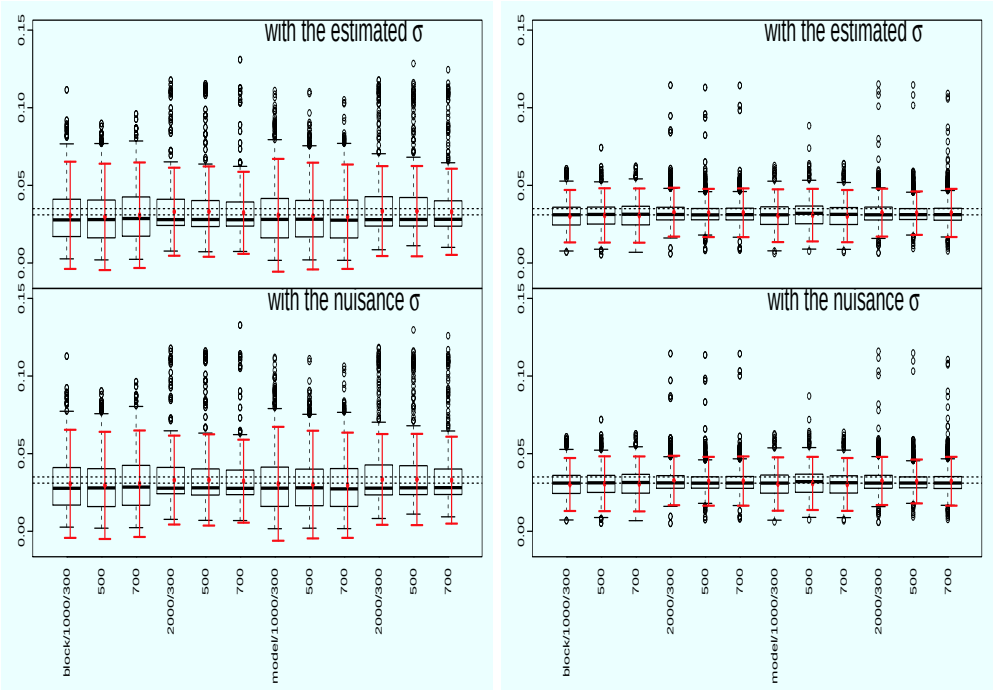
재표본 방법에 따른 평균은 매매의 경우 작게는 0.0296 (블록, 700 구분점, 2000회 반복) 부터 크게는 0.0334 (모델, 500 구분점, 2,000회 반복) 그리고 전세의 경우 작게는 0.0301 (블록, 300 구분점, 1000회 반복) 부터 크게는 0.0322 (모델, 700 구분점, 2,000회 반복)까지 형성되는데 참 값과는 비슷하고 MCMC에 의한 평균보다는 낮은 수치들이다.

전체적으로 재표본 횟수, 구분점 개수가 클수록 신뢰구간의 길이가 짧아지는 경향을 보인다. 표준편차를 추정한 경우와 장애모수로 처리한 경우는 모두 전체적으로 신뢰구간의 길이가 유사하게 계산되었다. 한편, 블록방법과 모델방법에 따른 신뢰구간의 길이에 대하여는 특정한 패턴을 찾을 수 없었다. 신뢰구간의 장단 여부는 재표본 반복수나 구분점의 수 같은 컴퓨터 계산을 위한 조건에 보다 영향을 받는 것으로 보인다.

**Table 4.4.** Confidence interval of the posterior estimates

| | Methods | Sampling repetition | | Frequency point | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 300 | | 500 | | 700 | |
| | | | | $l$ | $u$ | $l$ | $u$ | $l$ | $u$ |
| Sell & buy | Block | 1000 | $\sigma$-estimated | −0.0038 | 0.0652 | −0.0046 | 0.0640 | −0.0032 | 0.0647 |
| | | | $\sigma$-nuisance | −0.0042 | 0.0654 | −0.0050 | 0.0641 | −0.0037 | 0.0649 |
| | | 2000 | $\sigma$-estimated | 0.0047 | 0.0614 | 0.0040 | 0.0622 | 0.0059 | 0.0587 |
| | | | $\sigma$-nuisance | 0.0044 | 0.0616 | 0.0037 | 0.0625 | 0.0055 | 0.0590 |
| | Model | 1000 | $\sigma$-estimated | −0.0056 | 0.0670 | −0.0041 | 0.0646 | −0.0038 | 0.0634 |
| | | | $\sigma$-nuisance | −0.0061 | 0.0673 | −0.0046 | 0.0648 | −0.0042 | 0.0635 |
| | | 2000 | $\sigma$-estimated | 0.0045 | 0.0624 | 0.0043 | 0.0625 | 0.0052 | 0.0607 |
| | | | $\sigma$-nuisance | 0.0042 | 0.0626 | 0.0040 | 0.0628 | 0.0050 | 0.0609 |
| Long term | Block | 1000 | $\sigma$-estimated | 0.0133 | 0.0470 | 0.0132 | 0.0481 | 0.0131 | 0.0480 |
| | | | $\sigma$-nuisance | 0.0131 | 0.0472 | 0.0130 | 0.0483 | 0.0128 | 0.0482 |
| | | 2000 | $\sigma$-estimated | 0.0170 | 0.0485 | 0.0167 | 0.0477 | 0.0167 | 0.0480 |
| | | | $\sigma$-nuisance | 0.0168 | 0.0486 | 0.0165 | 0.0479 | 0.0165 | 0.0482 |
| | Model | 1000 | $\sigma$-estimated | 0.0135 | 0.0475 | 0.0139 | 0.0477 | 0.0134 | 0.0470 |
| | | | $\sigma$-nuisance | 0.0133 | 0.0476 | 0.0137 | 0.0479 | 0.0132 | 0.0472 |
| | | 2000 | $\sigma$-estimated | 0.0170 | 0.0478 | 0.0181 | 0.0461 | 0.0167 | 0.0477 |
| | | | $\sigma$-nuisance | 0.0169 | 0.0479 | 0.0180 | 0.0462 | 0.0165 | 0.0479 |

$l$ = 95%-lower limit; $u$ = 95%-upper limit.




(a) Sell & buy (b) Long term rent

**Figure 4.5.** Box and 95%confidence interval plots by Bootstrap; the dotted lines are at 0.031 and 0.035.

5. 경험적 검정 결과: 분석을 수행한 모든 경우에서 신뢰구간이 0.031과 0.035를 포함함을 볼 때, 매매지수와 전세 지수의 주파수의 차이 0.004 는 통계적으로 유의한 것으로 보이지 않는다. 주기로 보면 매매 지수와 전세 지수 사이에 3.7개월 정도의 차이가 있으나 그 차이가 통계적으로 유의한 차이는 아니라고 하겠다. 본 논문의 경우에서는 MCMC에 의한 신뢰구간이 시계열 재표본에 의한 신뢰구간보다 다소 짧은 경향이 있음을 관측할 수 있었다.

## 5. 맺음말

시계열 자료 속에 존재할지 모르는 주기를 파악하기 위한 베이지안 주파수 추정에 관하여 간단히 설명하고 부동산 관련 자료를 예제로 분석을 수행하였다. 주어진 데이터의 경우 주기의 차이는 통계적으로 유의하지 않음을 두 가지 계산적 방법을 통해서 보였다. 본 연구는 수리적 측면에 초점이 있으면 결과에 대한 부동산 경제학적 언급이 없음은 연구의 한계점이라고 하겠다.

## References


Bartlett, M. S. (1948). Smoothing periodograms from time-series with continuous spectra, *Nature*, **161**, 686–687.

Bretthorst, G. L. (2013). *Bayesian Spectrum Analysis and Parameter Estimation (Lecture Notes in Statistics)*, **48**, Springer-Verlag, New York.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge.

Dudek, A. E., Leśkow, J., Paparoditis, E., and Politis, D. N. (2014). A generalized block bootstrap for seasonal time series, *Journal of Time Series Analysis*, **35**, 89–114.

Efron, B. (1979). Bootstrap methods: another look at the jacknife, *Annals of Statistics*, **7**, 1–26.

Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica$\mu$l$k$a Support*, Cambridge University Press, Cambridge.

Hall, P., Horowitz, J., and Jing, B. (1995). On blocking rules for the bootstrap with dependent data, *Biometrika*, **82**, 561–574.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97–109.

Hayfield, T. and Racine, J. S. (2017). Nonparametric Econometrics: The np Package, R-package version 0.60-3, URL http://www.jstatsoft.org/v27/i05/.

Jaynes, E. T. (1987). Bayesian spectrum and chirp analysis, *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, 1–37.

Kreiss, J. P. and Lahiri, S. N. (2012). Bootstrap methods for time series, *Time Series Analysis: Methods and Applications*, **30**, 3–25.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations, *Annals of Statistics*, **17**, 1217–1241.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods, *Annals of Statistics*, **27**, 386–404.

Patton, A., Politis, D. N., and White, H. (2009). CORRECTION TO Automatic block-length selection for the dependent bootstrap by D. Politis and H. White, *Econometric Reviews*, **28**, 372–375.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap, *Journal of the American Statistical Association*, **89**, 1303–1313.

Welch, P. D. (1967). The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Transactions on Audio and Electroacoustics*, **15**, 70–73.

# 재표본 방법론을 활용한 베이지안 주파수 추정

박노진[a,1]

[a]단국대학교 응용통계학과

(2017년 9월 5일 접수, 2017년 10월 6일 수정, 2017년 10월 25일 채택)

**요 약**

시계열 자료의 주기를 파악하기 위해 스펙트럴 분석이 널리 이용되고 있다. 전력 스펙트럼이나 피리오도그램을 통해서 주파수를 추정하고 그로부터 순환 주기를 계산한다. 한편에서는 통계학의 한 축인 베이지안 기법을 활용한 주파수 추정법이 연구되어 사용되고 있다. 그런데 베이지안 주파수 추정량이 수학 공식을 통해 분석적으로 표현이 가능하지 않음으로 인해 신뢰구간 추정 같은 심도 깊은 통계학적 분석이 용이하지 않은 상화에서 컴퓨터를 이용한 수치해석적인 방법으로 신뢰구간을 추정하였다. 본 논문에서는 베이지안 주파수에 대한 보다 심도 있는 분석을 위해 모수를 재표본하는 Markov chain Monte Carlo (MCMC)을 이용한 추정과 데이터를 재표본하는 시계열 재표본을 통한 추정을 시도해 보았다. 예제로서 부동산 매매/전세 가격 지수 데이터을 사용하였고 매매와 전세 가격 지수간에 3.7개월 정도의 주기 차이가 존재하나 통계학적으로는 유의미한 차이라고 할 수 없음을 알았다.

Keywords: 베이즈 추정, 재표본, 피리오도그램, 스펙트럴 분석, 스펙트럼, 시계열

[1](16890) 경기도 용인시 수지구 죽전로 152, 단국대학교 응용통계학과. E-mail: rjpak@dankook.ac.kr