

# Robust spectral estimator from M-estimation point of view: application to the Korean housing price index

Ro Jin Pak<sup>a,1</sup>

<sup>a</sup>Dankook University, Department of Applied Statistics

(Received February 12, 2016; Revised April 11, 2016; Accepted April 11, 2016)

---

## Abstract

In analysing a time series on the frequency domain, the spectral estimator (or periodogram) is a very useful statistic to identify the periods of a time series. However, the spectral estimator is very sensitive in nature to outliers, so that the spectral estimator in terms of M-estimation has been studied by some researchers. Pak (2001) proposed an empirical method to choose a tuning parameter for the Huber's M-estimating function. In this article, we try to implement Pak's estimation proposal in the spectral estimator. We use the Korean housing price index as an example data set for comparing various M-estimating results.

Keywords: housing price index, m-estimation, periodogram, spectral estimator, tuning parameter

---

## 1. 서론

시계열 데이터를 주파수 영역에서 분석함에 있어 핵심은 스펙트럴 밀도 함수(spectral density function)를 구하는 것이다. 스펙트럴 밀도 함수는 일반적인 추정에서 확률 밀도 함수와 같이 데이터로부터 추정을 해야 할 필요가 있다. 그 추정량으로서 스펙트럴 추정량 혹은, 주기도(periodogram)가 사용된다. 그런데, 스펙트럴 추정량은 부드럽지 않고 굴곡의 변화가 너무 심하여 때때로 해석에 어려움이 있을 수 있다. ‘변화가 심하다’함은 다른 말로 추정량의 분산이 크다고 할 수 있는데, 따라서 그 분산을 줄이기 위해 Bartlett (1948, 1950)은 관측기간을 몇 개의 구간으로 나누어 스펙트럴 추정량을 각각 추정한 후, 추정된 스펙트럴 추정량들의 평균으로 최종적인 스펙트럴 추정량으로 사용할 것을 제안하였다. 이렇게 함으로 비록 해상도는 떨어지나 분산은 작게 되는 효과를 얻을 수 있음을 보였다. 몇 해 후, Welch (1967)는 관측 구간을 나눌 때 겹치는 것을 허락하여 Bartlett의 방법을 사용하면 잡음이 많이 제거된 스펙트럴 추정량을 구할 수 있다고 하였다. 한 걸음 더 나아가 최근에 Spangl과 Dutter (2005)는 Welch (1967)의 방법을 따르되 마지막 단계에서 구간별 스펙트럴 추정량들의 평균 대신에 M-추정량을 사용하면 보다 의미 있는 형태의 스펙트럴 추정량을 구할 수 있다고 하였다.

한편, Pak (2001)은 보편적인 M-추정을 수행할 때 밀도 함수 추정에서 사용하는 띠너비(bandwidth)를 활용하면 로버스트하며 효율적인 추정이 가능함을 보였다. 본 연구에서는 Pak (2001)의 제안을 주기도를 구하는 방법에 활용하려 한다. 모의실험을 통해 제안한 방법이 기존의 방법 보다 주파수 추정에서 정확도가 있음을 보이고 우리나라 주택가격 지수를 이용하여 제안된 방법이 유용한 결과를 도출함을 보이고자 한다.

---

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

## 2. 방법론

시계열  $x(t)$ ,  $t = 0, \dots, N - 1$ 의 스펙트럴 추정량은

$$\hat{\phi}_p(\omega) = \frac{1}{N} \left| \sum_{t=0}^{N-1} x(t) e^{-i\omega t} \right|^2$$

로 정의된다 (Proakis와 Manolakis, 2006). 그런데, 이 스펙트럼 추정량 혹은 주기도는 해상도는 높으나 분산이 큰 문제가 있어 Welch (1967)가 아래의 방법을 제안하였다.

### 2.1. Welch (1967)의 Overlapped Segment Averaging(WOSA)

Welch의 Overlapped Segment Averaging(WOSA)의 과정은 아래와 같다.

- 전체 관측 구간을 길이가  $N_S$ 인  $N_B$ 개의 블록으로 나눈다. 이때, 블록끼리 겹쳐도 되며 50% 겹침이 적당하다고 한다.
- 각 블록에 대하여 스펙트럴 추정량  $\hat{S}_l(\omega)$ 를 계산한다.

$$\hat{S}_l(\omega) = \frac{1}{N_S} \left| \sum_{t=0}^{N_S-1} h_t x_{t+l} e^{-i\omega t} \right|^2, \quad l = 0, \dots, N_B - 1,$$

여기서  $h_0, \dots, h_{N_S-1}$ 는 ‘data taper’라고 부르는데 윈도우(window)를 취해 얻은 값을 의미한다. 윈도우란 Bartlett, Hamming 등으로 대표되는 대개 0에 대하여 대칭인 형태를 갖으며 입력 값의 영향을 제어하는데 사용하는 함수이다.

- WOSA 스펙트럴 추정량은

$$\hat{S}^W(\omega) = \frac{1}{N_B} \sum_{l=0}^{N_B-1} \hat{S}_l(\omega)$$

가 된다.  $\omega = 2\pi f$ 로 대체하여  $\hat{S}^W(f)$ 로 표시할 수도 있다.

### 2.2. M-추정법의 활용

Spangl과 Dutter (2005)는 앞에서의 언급한 WOSA의 마지막 단계에서  $\hat{S}_l(\omega)$ 들의 산술 평균 대신 M-추정량을 구하여  $\hat{S}^W(\omega)$ 로 하길 제안하였다.

M-추정은 Huber (1964)에 의해 깊게 연구되었는데, 어떤 확률 함수  $f(x; \theta)$ 로 부터 확률 표본  $\{X_1, \dots, X_n\}$ 이 주어질 때,  $\theta$ 에 대한 추정량을

$$\min_{\theta} \sum \rho \left( \frac{X_i - \theta}{S} \right)$$

를 만족하는  $\hat{\theta}$  혹은 동일하게

$$\sum \psi \left( \frac{X_i - \theta}{S} \right) = 0, \quad \psi = \rho' \quad (2.1)$$

의 해인  $\hat{\theta}$ 로 하고  $\theta$ 의 M-추정량이라 한다. 여기서,  $S$ 는 표준 편차의 추정량이다.

가장 대표적인 Huber의 M-추정함수는

$$\psi_k(t) = \min\{k, \max\{t, -k\}\} = t \cdot \min\left\{1, \frac{k}{|t|}\right\}, \quad 0 < k < \infty$$

로 정의되는데 식 (2.1)의  $\psi$  대신에 Huber의  $\psi_k$ 를 사용하고  $t_l = (\hat{S}_l(\omega) - \theta)/s$ 로 놓고 M-추정을 시행하여 구한  $\hat{\theta}$ 으로  $\hat{S}^W(\omega)$ 로 삼자는 것이 Spangl과 Dutter (2005)의 제안이다.

이 때, 필요한 것은 띠너비(bandwidth)라는  $k$ 의 값인데, 일반적으로 표준정규분포의 임계치인  $k = 1.28$  또는  $1.65$ 로 고정하여 사용한다. Pak (2001)은  $k$ 에 대하여  $h_{sj}/\text{mad}$ ,  $h_{cv}/\text{mad}$  등을 사용할 것을 제안하였다. mad는 median absolute deviance이고  $h_{sj}$ 는 Sheather와 Jones (1991) 그리고  $h_{cv}$ 는 cross-validation에 의한 띠너비 (Silverman, 1986)를 의미한다. Pak (2001)은 제안한 방법이 기존의 방법들 보다 더 데이터 친화적(data friendly)이라고 주장하였다. 일반적인 모수 추정에서 효과가 있다는 Pak (2001)의 방법을 주기도 추정에도 사용하려한다. 다음 장에서 Spangl과 Dutter (2005)의 방법과 Pak (2001)의 방법에 따른 모의실험을 수행하고 실제 자료에 적용하여 보겠다.

### 3. 데이터 분석

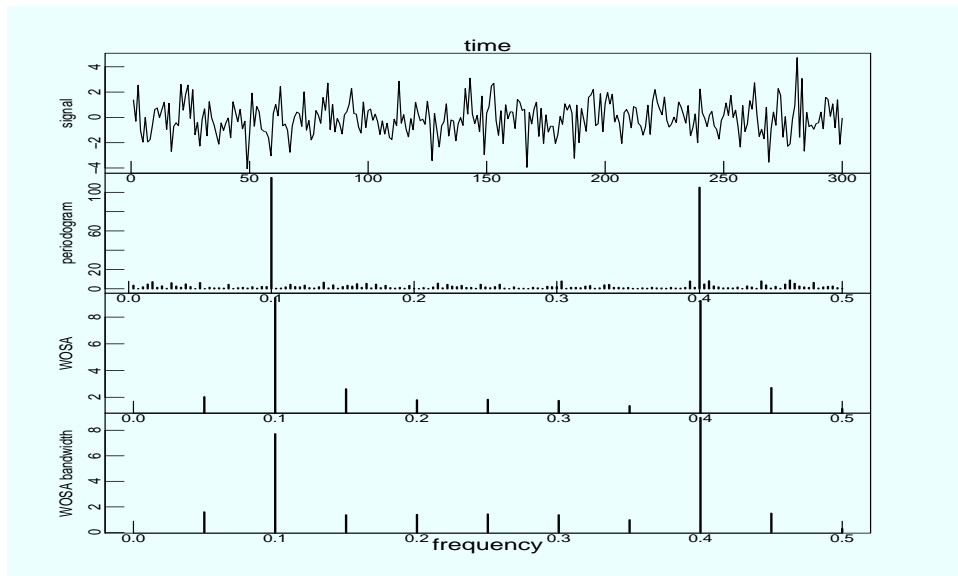
#### 3.1. 모의실험

위에 언급한 추정법들을 비교하기 위해 간단한 모의실험을 아래와 같이 실시하였다. 계산은 'sapa'라는 R-package를 기본으로 이루어졌다.

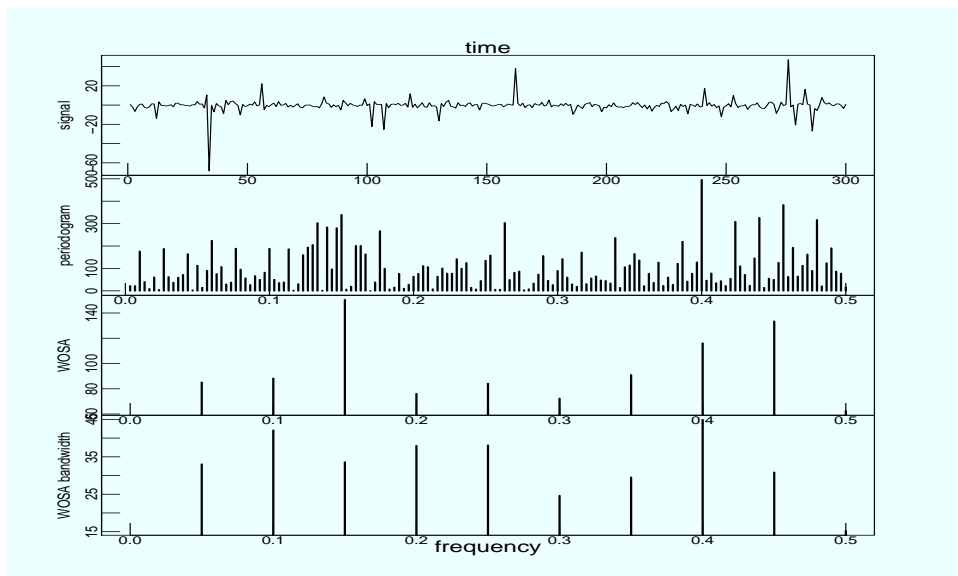
- (1) 주파수가 0.1과 0.4인 모델,  $m(t) = \sin(2\pi(0.1)t) + \sin(2\pi(0.4)t)$ 을 가정하고 시간  $t = 0, \dots, 299$ 에 따라 구한 300개의 자료에  $N(0, 1)$ 과 Cauchy(0, 1)에서 생성한 오차( $e(t)$ )를 각각 더한 신호,  $s(t) = m(t) + e(t)$ 를 생성하여 (i) 일반적인 주기도, (ii) 평균을 이용한 WOSA, (iii) 띠너비를 이용한 WOSA를 추정하여 보았다. 이 과정을 1,000회 실시하여 매번 주파수 적어도 0.1 혹은 0.4 중 하나를 제대로 특정해 내는지를 보았다. Figure 3.1에 추정 결과의 예를 그려 넣었다. 정규분포를 오차로 한 경우 (Figure 3.1(a))는 주파수를 특정 하는데 큰 문제가 없어 보인다. 그런데 꼬리가 긴 코쉬분포 (Figure 3.1(b))를 사용하는 경우 일반적인 주기도나 WOSA는 0.1과 0.4를 잘 특정할 수 없으나 띠너비를 사용한 WOSA를 이용한 경우 시각적으로 더 확실한 모양을 나타내고 있다. 1,000회 모의 실험을 수행한 결과 정규분포를 오차로 가정한 경우 두 개의 주파수 중 적어도 하나를 제대로 추정해 내는 경우는 모든 추정법에서 100%였으나 코쉬분포를 오차로 가정한 경우 일반적 주기도는 15.6%, WOSA는 50.3% 그러나 Pak (2001)이 제안한 띠너비를 이용한 WOSA는 75.1%로 가장 높은 추정률을 보였다. 그런데 Sheather와 Jones의 띠너비는 계산이 매우 안정적이었으나 cross-validation에 의한 띠너비는 계산이 안 되는 경우가 있었다. 따라서 본 논문에서는 Sheather와 Jones의 띠너비만 사용하였다.
- (2)  $z(t) = 0.6z(t-1) - 0.9z(t-2) + e(t)$ ,  $e(t) \sim N(1, 0)$ 를 가정하고 시간  $t = 0, \dots, 299$ 에 따라 생성된 300개의 자료에 대하여 (1)과 같이 주기도, WOSA 그리고 띠너비를 이용한 WOSA를 추정하여 보았다. 주어진 AR(2)모형의 스펙트럼 밀도 함수는

$$S(\omega) = \frac{1}{|1 - 0.6e^{-2\pi i\omega} + 0.9e^{-4\pi i\omega}|^2}$$

이되고  $0.3333 \pm i$ 를 극점(pole)으로 갖으며 단위원에서의 점은  $0.3162 \pm 0.9486i$ 로 이 경우에  $\omega = 0.1987$ 로 주파수가 약 0.2일 때  $S(\omega)$ 가 최대화된다고 할 수 있겠다. Figure 3.2에 추정 결과의 예를



(a) Normal error



(b) Cauchy error

Figure 3.1. Simulation examples.

그러 넣었는데, 일반적인 주기도는 0.2를 특정하지 못하고 있다. 1,000회의 모의실험을 실시하였는데 WOSA와 띠틈너비를 사용한 WOSA는 주파수를 0.2로 특정하는 비율이 거의 100%에 가까우나 일반적인 주기도는 30%를 넘지 못했다. 물론 아주 제한된 상황에서의 실험이지만 WOSA를 기반으로 하는 추정이 매우 효과적이라고 주장하고 싶다.

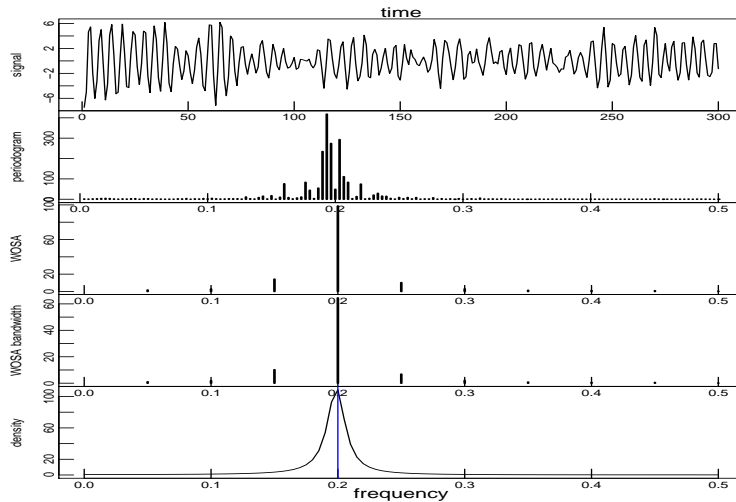


Figure 3.2. Periodograms of an example and the spectral density by AR(2).

### 3.2. 주택가격자료

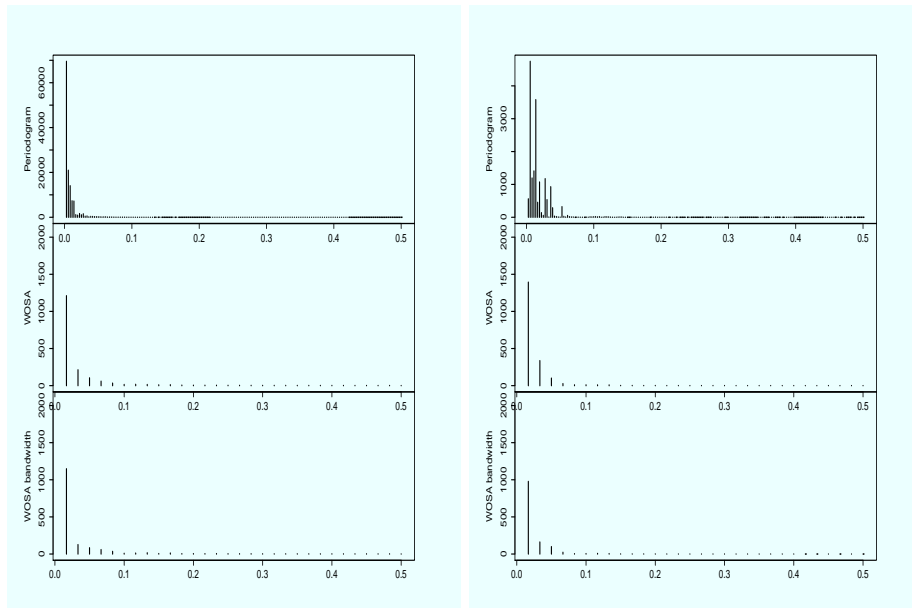
스펙트럼 추정량을 구함에 있어 Pak (2001)이 제안한 방법의 효과를 확인하기 위해 실제 데이터를 이용하여 보려고 한다. Figure 3.3에 1986년 1월부터 2015년 9월까지의 우리나라 주택가격지수와 전년 동기 대비 변화율을 그려 보았다. 자료는 국민은행 부동산 사이트에서 내려 받았다. 주택 가격지수 데이터의 경우에 일반적 주기도의 가장 강한 크기의 주파수는 0.00277 그러나 WOSA는 0.016으로 파악된다. 변화율의 경우는 일반적 주기도는 0.0055와 0.0138에서 크기가 가장 크고 WOSA는 0.016에서 강하게 나타난다. 주기로 따지면 일반적 주기도는 가격지수는 361개월 그리고 변화율은 181개월과 72개월로 판단되고 WOSA는 주택가격지수와 변화율에 대한 모두 62.5개월로 추정한다. WOSA에 의하면 주택 가격 관련 지표의 주기가 대통령 임기인 60개월과 거의 일치한다고 하겠다. Figure 4.2에서 점선으로 표시된 수직선들은 대통령들이 선출된 연도를 표시한다. 대통령들의 재임기간동안 가격지수의 경우 알파벳 U처럼 하강→유지→상승으로 이어지는 패턴들이 보인다. 지수 변화율의 경우는 코사인 형태의 하강→상승→하강의 패턴이 관찰된다.

우리나라 주택가격지수는 전체적으로는 상승 기조를 보인다. 주택가격지수는 U자 형태가 크기를 달리 하면서 반복되는 것으로 보이는데, 1990년대 초반에서 2000년대 초반까지 큰 U자 형태 뚜렷한 모양을 보이나 그 이후 U자의 크기는 점점 작아지며 반복되는 것으로 보인다. 분석 결과 U자의 긴 주기(361개월)와 짧은 주기(72, 60개월)가 관측된다. 변화율은 국소적으로 대통령 임기와 비슷한 62.5개월 동안 상승과 하강의 부침이 있음을 확인할 수 있다. 요약하면 주택 가격은 단기적으로는 그 변화가 롤러코스터 처럼 코사인 곡선을 따라 움직이지만 그 가격은 하강→유지→상승을 반복하며 종국적으로는 상승한다고 하겠다.

실제 예제의 경우 띠너비를 활용한 WOSA나 기존의 WOSA 모두 같은 주파수를 특정해 주고 있다. 굳이 따지면 띠너비를 활용한 WOSA경우가 기존의 WOSA 보다 미미 하지만 첫 번째 막대기가 두 번째 이후의 막대기들과 비교해 상대적으로 확실하다고 하겠다. 기존의 방법에 새로이 제안된 방법을 추가 사용하면 보다 확실한 의사 결정이 가능하다고 하겠다.



(a) Korean housing price index; the vertical lines indicate when the presidency began.



(b) Periodograms for the price index

(c) Periodograms for the rate of change

Figure 3.3. Housing index, periodograms.

#### 4. 결론

시계열 데이터의 주기성을 파악하는데 주기도가 널리 사용되는데 해상도가 지나치게 높아 과대적합의 문제가 존재한다. WOSA는 해상도가 낮지만 나름 유용한 결과를 보여주기도 한다. WOSA에 로버스트 기법을 적용함에 있어 Pak (2001)의 제안을 사용하여 보았고 모의실험과 실제 자료에 적용할 때 의미 있는 결과를 얻을 수 있었다. 본 논문에서 추정량들의 적합도에 대한 수학적 고찰보다 시각적 비교에 그친 것은 다소 아쉬운 점이 있었다.

#### References

- Bartlett, M. S. (1948). Smoothing periodograms from time-series with continuous spectra, *Nature*, **161**, 686–687.
- Bartlett, M. S. (1950). Periodogram analysis and continuous spectra, *Biometrika*, **37**, 1–16.
- Huber, P. J. (1964). Robust estimation of location parameters, *Annals of Mathematical Statistics*, **35**, 73–101.
- Pak, R. J. (2001). The bending constant in Huber's function in terms of a bandwidth in density estimator, *The Korean Journal of Applied Statistics*, **14**, 357–367.
- Proakis, J. G. and Manolakis, D. K. (2006). *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice Hall, New York.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.
- Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society: Series B*, **53**, 683–690.
- Spangl, B. and Dutter, R. (2005). On robust estimation of power spectra, *Austrian Journal of Statistics*, **34**, 199–210.
- 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 periodogram, *IEEE Transactions on Audio Electroacoustics*, **AU-15**, 70–73.

# M-추정에 기반을 둔 로버스트 스펙트럴 추정량: 주택 가격 지수에 대한 응용

박노진<sup>a,1</sup>

<sup>a</sup>단국대학교 응용통계학과

(2016년 2월 12일 접수, 2016년 4월 11일 수정, 2016년 4월 11일 채택)

---

## 요약

주파수 영역에서 시계열 자료를 분석함에 있어 스펙트럴 추정량은 매우 유용한 도구이다. 기존의 스펙트럴 추정량은 이상치에 영향을 받을 수밖에 없는 구조로 되어있어서 M-추정법을 활용하여 로버스트 스펙트럴 추정량이 제안되었다. M-추정을 위해서는 조율모수를 적절하게 선택해 주어야 하는데 Pak (2001)이 제안한 방법을 사용할 때의 효과를 연구하였다. 모의실험과 주택가격지수에의 적용을 통하여 효과가 있음을 확인하였다.

주요용어: M-추정법, 학률 함수 추정, 스펙트럴 추정량, 조율 모수, 띠너비

---

<sup>1</sup>(16890) 경기도 용인시 수지구 죽전로 152, 단국대학교 응용통계학과. E-mail: rjpak@dankook.ac.kr