

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 21 日現在

機関番号：21101

研究種目：研究活動スタート支援

研究期間：2015～2016

課題番号：15H06525

研究課題名(和文)欠損値がある長期記憶過程におけるウェーブレット推定量の統計的性質の研究

研究課題名(英文)The Statistical Property of the Wavelet-Based Estimator for Long-Memory Processes with Missing Data

研究代表者

七宮 圭(NANAMIYA, Kei)

青森公立大学・経営経済学部・講師

研究者番号：20755714

交付決定額(研究期間全体)：(直接経費) 1,800,000円

研究成果の概要(和文)：長期記憶過程に従うモデルから発生した観測値に欠損(欠測)がある場合のパラメータの推定問題について、ウェーブレット解析を利用した近似最尤推定量の導出とその統計的性質について研究を行った。具体的には、欠損値を含む観測系列をウェーブレット変換することで周波数帯ごとに分散の大きさが異なるホワイトノイズ系列で近似し、さらに変換の際に欠損値を含むような係数を取り除くことにより、近似最尤推定量を提案した。また、その推定量が欠損値の影響を受けていないことを示した。さらに、その統計的性質は観測した系列に欠損がない場合と同様のものになることを示した。

研究成果の概要(英文)：We introduce the wavelet-based approximated maximum likelihood estimators for the parameters of the long-memory processes with missing data, and explore the statistical property of these estimators. In our study, by approximating the wavelet coefficients of the processes with missing data as white noise processes in each level and by removing the coefficients including missing data, we propose the wavelet-based approximated maximum likelihood estimators. We show that these estimators are not affected by missing data and have the same statistical property as the wavelet-based approximated maximum likelihood estimators for long-memory processes which are complete data.

研究分野：経済時系列解析

キーワード：時系列解析 長期記憶過程 欠測値 ウェーブレット解析

1. 研究開始当初の背景

(1) 長期記憶過程は時系列分析の主要な研究テーマの一つであり、自然科学から社会科学まで多岐にわたる分野において研究されている。長期記憶過程の特徴は、過去の値への依存度の減衰が遅いため、自己共分散の絶対和が発散することであり、それは周波数領域で考えた場合、原点において発散するようなスペクトル関数を持つことである。また、パラメトリックなモデルにおいては、長期記憶性を表すパラメータの値によって過程が定常から非定常に変化するという特徴がある(詳細は田中(2006)、Beran(1994)、Giritis et al.(2012)、Palma(2007)参照)。

このような特徴があることから長期記憶過程に属する時系列データの分析の方法としては大きく分けて

- 時間領域：時間における変化に着目した時間領域の観点から分析する方法。
- 周波数領域：フーリエ解析によるデータに内包されている周期的な動きのそれぞれ大きさに着目した周波数領域の観点から分析する方法。
- 時間-周波数領域：ウェーブレット解析などの周期的な動きが時間とともにどのように変化するのかに着目する時間-周波数分析による方法。

という3つの方法が存在し、様々な研究が行われている。

(2) 特にウェーブレット解析を利用した推定法では、McCoy and Walden(1996)やPercival and Walden(2000)にあるように、非定常な場合においても長期記憶過程を変換後のウェーブレット係数をスケールのレベルごとに分散の大きさが異なるホワイトノイズの系列として近似することで、パラメータの推定が他の手法よりも簡単にできるということから、その推定量の漸近理論などの研究が多く行われてきた(Moulines et al.(2007, 2008)、Roueff and Taqqu(2009a, b)など)。また、Faÿ et al.(2009)のように他の推定手法との比較も行われている。ただし、上記の方法は変換後の系列がホワイトノイズであると近似して推定する手法であり、変換後の系列の正確な状態はNanamiya(2014)などにあるように複雑なものとなるため、Craigmale et al.(2005)では1階の自己回帰モデルをスケールごとに当てはめた近似最尤推定法を提案している。

(3) 欠損(欠測)値の問題は社会科学や医学などの幅広い分野において取り扱われている。入手したデータの一部が欠損しているということは実証分析ではよくあることであり、これらの欠損データの統計分析手法が近年活発に研究されている。それらの研究の中では欠損メカニズムを、

- MCAR (missing completely at random): データとは無条件に完全にランダムで

欠損。

- MAR (missing at random): 観測されたデータに依存してランダムに欠損。
- MNAR (missing not at random): データの欠損確率が観測したデータだけではなく、欠損したデータの値にも依存。

という3つに分類し、その欠損メカニズムに応じて分析手法を考慮する必要性が指摘されている(詳細は、阿部(2016)、岩崎(2002)、高井他(2016)、Little and Rubin(2002)参照)。

(4) 時系列データにモデルを当てはめて考察する場合には、一般的にはデータには欠損値がなく観測間隔が一定であること、いわゆるデータが完全(complete)であることが前提とされている。仮に観測間隔が不規則であったり、欠損値があったりしても、間隔の違いを調整するためにデータに対して線形補間などの何らかの操作を加えたり、時間の代わりに観測した順番を使用することで、あたかも一定の間隔で観測された完全なデータとみなして取り扱うことが多い。

一方で、不完全なデータを不完全なものとして分析する方法としては、時間領域に関してはカルマン・フィルターを用いた手法(Kim and Shao(2014)参照)や、周波数領域に関してはParzen(1963)以降、Dunsmuir and Robinson(1981)などで研究が行われてきた。

特に長期記憶過程においては、(1)で述べたように過去の値への依存度の減衰が遅いため、欠損値が存在する場合にはモデルの推定に困難が伴う。定常な長期記憶過程に対しては、カルマン・フィルターを用いた状態空間モデルによるパラメトリック・モデルのパラメータの推定問題についての研究(Palma(2007)、Palma and Del Pino(1999)、Palma and Chan(1997))や、自己共分散関数に関する研究(Yajima and Nishino(1999))が行われている。また、Wilson et al.(2003)のように欠損したデータを補完する方法を用いる実証研究もある。

2. 研究の目的

長期記憶過程のパラメトリック・モデルの一つである autoregressive fractionally integrated moving average (ARFIMA) モデルのパラメータの推定問題について、欠損値を含まない区間から計算されたウェーブレット係数だけを用いて長期記憶過程のパラメータの近似最尤推定法が可能であると考えられるため、この着想に基づいた推定手法の確立と理論的性質の解明が本研究の目的である。

3. 研究の方法

(1) 本研究に関連する欠損値および長期記憶過程の推定問題についての先行研究の調査と整理を行い、本研究で用いる欠損のメカニズムを MCAR に決定した。また、その欠損

メカニズムの下で欠損値がある ARFIMA モデルのパラメータについてのウェーブレット解析を利用した近似最尤推定量の統計的な性質の理論的な解明を行った。

(2) コンピュータによるシミュレーションによって、推定量の挙動の検証を行った。

4. 研究成果

(1) 本研究の理論的な成果としては、MCAR の場合に、欠損値を含む観測系列をウェーブレット変換することで周波数帯ごとに分散の大きさが異なるホワイトノイズ系列で近似し、さらに変換の際に欠損値を含むような係数を取り除くことにより、近似最尤推定量を構築した。また、その推定量が欠損値の影響を受けていないことを示し、その統計的な性質は観測した系列に欠損がない場合と同様のものになるということを示した。ただし、欠損の影響により、推定に使用できる変換後のウェーブレット係数の数が減少する。これは、通常の推定問題において標本数が少なくなることを意味する。また、スケールのレベルが高い部分(低周波数帯)については、ウェーブレット変換に使用するフィルターのサポートが長くなるため欠損値を含む可能性が高くなり、ウェーブレット係数が計算できず、推定の際に除外される可能性が高くなる。このため、スケールのレベルが低い部分(高周波数帯)の係数を中心とした推定となる。一方で、ウェーブレット解析を利用しているので、先行研究で用いられた時間領域や周波数領域の手法では分析できなかった非定常な長期記憶過程の分析もできるという優位性がある。

(2) 使用するウェーブレット・フィルターの長さや欠損の確率が推定結果に与える影響を確認するために、欠損がない場合も含めて、コンピュータによるシミュレーション実験を行った。

(3) 今後の課題としては、他の推定手法との比較や、実証分析への応用がある。また、今回の研究では欠損のメカニズムとしては MACR のみを取り扱ったため、他のメカニズムに対する推定手法の確立も課題となっている。

<引用文献>

(和文)

阿部貴行(2016)『欠測データの統計解析』、朝倉書店

岩崎学(2002)『不完全データの統計解析』、エコノミスト社

高井啓二、星野崇宏、野間久史(2016)『欠測データの統計科学 医学と社会科学への応用』、岩波書店

田中勝人(2006)『現代時系列分析』岩波書店。

(英文)

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall, New York.

Craigmile, P. F., Guttorp, P. and Percival, D. B. (2005), Wavelet-Based Parameter Estimation for Polynomial Contaminated Fractionally Differenced Processes, *IEEE Transactions on Signal Processing*, 53, pp.3151-3161.

Dunsmuir, W. and Robinson, P. M. (1981). Parametric estimators for stationary time series with missing observations. *Adv. Appl. Prob.* 13, pp.129-146.

Fay, G., Moulines, E., Roueff, F. and Taqqu, M. S. (2009). Estimators of long-memory: Fourier versus wavelets. *Journal of Econometrics*, vol.151, pp.159-177.

Giraitis, L., Koul, H. L. and Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.

Iglesis, P., Jorquera, H. and Palma, W. (2006). Data analysis using regression models with missing observations and long-memory: an application study. *Computational statistics & Data analysis*, vol.50, pp.2028-2043.

Kim, J. K. and Shao, J. (2014). *Statistical methods for handling incomplete data*. CRC press, Boca Raton.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. Wiley, Hoboken.

McCoy, E. J. and Walden, A. T. (1996). Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics*, 5, pp.26-56.

Moulines, E., Roueff, F. and Taqqu, M. S. (2007). On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter. *Journal of Time Series Analysis*, vol.28, pp.155-187.

Moulines, E., Roueff, F. and Taqqu, M. S. (2008). A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series. *The Annals of Statistics*, 36, 1925-1956.

Nanamiya, K. (2014). Modelling for the wavelet coefficients of ARFIMA processes. *Journal of Time Series Analysis*, vol.35, pp.341-356.

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. Wiley, Hoboken.

Palma, W. and Del Pino, G. (1999). Statistical analysis of incomplete long-range dependent data. *Biometrika*, 86, pp.965-972.

Palma, W. and Chan N. H. (1997). Estimation and forecasting of long-memory processes with missing values. *Journal of Forecasting*, vol.16, pp.395-410.

Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation. *Sankhyā A* 25, pp.383-392.

Percival, D. B. and Walden, A. T. (2000). *Wavelet methods for time series analysis*. Cambridge University press, Cambridge.

Roueff, F. and Taqqu, M. S. (2009a). Asymptotic normality of wavelet estimators of the memory parameter for linear processes. *Journal of Time Series Analysis*, vol.30, pp.534-558.

Roueff, F. and Taqqu, M. S. (2009b). Central limit theorems for arrays of decimated linear processes. *Stochastic Processes and their Applications*, vol.119, pp.3006-3041.

Wilson, P. S., Tomsett, A. C. and Toumi, R. (2003). Long-memory analysis of time series with missing values. *Physical Review E* 68, 017103.

Yajima, Y. and Nishino, H. (1999). Estimation of the autocorrelation function of a stationary time series with missing observations. *Sankhyā A* 61, pp.189-207.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計0件)

〔学会発表〕(計1件)

七宮 圭、Wavelet-based estimation of long-memory processes with missing values, 2016年度統計関連学会連合大会、2016年9月5日、金沢大学(石川県金沢市)

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1) 研究代表者

七宮 圭(NANAMIYA, Kei)
青森公立大学・経営経済学部・講師
研究者番号：20755714

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：

(4) 研究協力者

()