



Kharazmi University

Wavelet Threshold Estimator of Semiparametric Regression Function with Correlated Errors

Hamid Karamikabir¹ , Mahmoud Afshari² ✉, Abouzar Bazyari³ 

1. Department of Statistics, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr Iran.

E-mail: h_karamikabir@yahoo.com

2. Department of Statistics, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr Iran.

✉E-mail: afshar@pgu.ac.ir

3. Department of Statistics, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr Iran.

E-mail: ab_bazyari@yahoo.com

Article Info

ABSTRACT

Article type:

Research Article

Article history:

Received:

18 February 2019

Revised form:

10 October 2020

Accepted:

21 November 2020

Published online:

22 November 2022

Keywords:

Convergence ratio;

Correlated

errors;

Semiparametric

regression;

Wavelet threshold.

Introduction

One of the most important topics in statistics is regression analysis, which examines the relationship between independent and dependent variables.

According to the type of function that expresses the relationship between two independent variables and the response, regression is divided into two categories: parametric regression and non-parametric regression. In non-parametric regression, the relationship between the independent variables and the response is expressed through an unknown function and is of particular importance due to the initial preconditions.

One of the important tools in estimating the regression function is the wavelet.

Wavelet is considered as an analytical tool in signal processing and mathematical functions, and it has the ability to solve related problems in applied mathematics. For the first time, Donoho and Johnstone (1992) used wavelets in statistical research. Much of the advanced mathematical and statistical material of wavelet theory is given in Chui's book (1992).

Material and methods

In this scheme, the wavelet threshold estimation of semiparametric regression model with correlated errors with having Gaussian distribution is determined

and the convergence ratio of estimator computed. To evaluate the wavelet threshold estimation, the block function and sinusoidal function are used as objective functions and using the simulation method the average of mean square error and standard deviation. Also, the wavelet semiparametric regression model has been fitted to data on the growth rate of the teeth.

Results and discussion

The purpose of this article is to estimate the wavelet threshold of the semiparametric regression function with dependent errors and Gaussian distribution. The wavelet estimator always has a lower mean square error than the wavelet kernel estimator. Also, by increasing the number of observations of the sampling function, the accuracy of the wavelet estimator increases compared to the wavelet kernel estimator, and the error rate of the wavelet estimator tends to zero faster. In addition, the wavelet estimator always has a lower standard error than the wavelet kernel estimator.

Conclusion

The following conclusions were drawn from this research.

- The wavelet threshold estimator of the semiparametric regression function was investigated and its convergence rate was calculated.
- In order to evaluate the proposed estimator, sinusoidal and block functions were used as objective functions and by using simulation, it was shown that the average mean square of the errors obtained by the wavelet threshold method is lower than the average mean square of the errors obtained by the wavelet kernel method.
- According to the property of wavelet and scale bases, the wavelet threshold method estimates sharp and jump points better than other methods and has an acceptable convergence rate.

How to cite: Karamikabir, H., Afshari, M., Bazyari, A., (2022) Wavelet Threshold Estimator of Semiparametric Regression Function with Correlated Errors. *Mathematical Researches*, 8 (3), 180-205



© The Author(s).

Publisher: Kharazmi University

برآوردگر آستانه موجک تابع رگرسیون نیمه پارامتری با خطاهای همبسته

حمید کرمی کبیر^۱، محمود افشاری^۲✉، ابوذر بازیاری^۳

۱. گروه آمار، دانشگاه خلیج فارس، بوشهر، ایران. پست الکترونیکی: h_karamikabir@yahoo.com

۲. نویسنده مسئول، گروه آمار، دانشگاه خلیج فارس، بوشهر، ایران. پست الکترونیکی: afshar@pgu.ac.ir

۳. گروه آمار، دانشگاه خلیج فارس، بوشهر، ایران. پست الکترونیکی: ab_bazyari@yahoo.com

چکیده

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۳۹۷/۱۱/۲۹

تاریخ بازنگری: ۱۳۹۹/۰۷/۱۹

تاریخ پذیرش: ۱۳۹۹/۰۹/۰۱

تاریخ انتشار: ۱۴۰۱/۰۹/۰۱

واژه‌های کلیدی:

آستانه موجک،

خطاهای همبسته،

رگرسیون نیمه پارامتری،

نرخ همگرایی.

تحلیل موجکی یکی از تکنیک‌های مفید در ریاضی است که در سال‌های اخیر در علم آمار بسیار مورد استفاده قرار گرفته شده است. در این مقاله، ضمن معرفی تبدیل موجک، برآورد آستانه موجک تابع رگرسیون نیمه پارامتری با خطاهای همبسته تعیین شده و سپس نرخ همگرایی برآوردگر محاسبه می‌شود. برای ارزیابی برآوردگر آستانه موجک، از دو تابع سینوسی و تابع بلوکی به عنوان توابع هدف استفاده خواهد شد و با استفاده از شبیه‌سازی، متوسط میانگین مربع خطا و مقدار انحراف معیار این برآوردگر با متوسط میانگین مربعات خطاها و مقادیر انحراف معیارهای به دست آمده با روش هسته موجک مقایسه می‌شوند. همچنین برای ارزیابی روش ارائه شده، تابع رگرسیون نیمه پارامتری موجکی بر روی داده‌های مربوط به میزان رشد دندان برازش داده شده است.

استناد: کرمی کبیر، حمید؛ افشاری، محمود؛ بازیاری، ابوذر؛ (۱۴۰۱). برآوردگر آستانه موجک تابع رگرسیون نیمه پارامتری با خطاهای همبسته. پژوهش‌های ریاضی، ۸ (۳)، ۱۸۰-۲۰۵.



© نویسندگان.

ناشر: دانشگاه خوارزمی

۱. مقدمه

یکی از مهمترین مباحث علم آمار تحلیل رگرسیون^۱ است که به بررسی ارتباط بین متغیرهای مستقل و وابسته می‌پردازد. رگرسیون با توجه به نوع تابعی که ارتباط بین دو متغیر مستقل و پاسخ را بیان می‌کند به دو دسته رگرسیون پارامتری^۲ و رگرسیون ناپارامتری^۳ تقسیم می‌شود. در رگرسیون ناپارامتری ارتباط بین متغیرهای مستقل و پاسخ از طریق تابعی مجهول مانند f بیان می‌شود و به دلیل پیش شرط‌های اولیه از اهمیت خاصی برخوردار است. هدف از این مقاله برآورد آستانه موجک^۴ تابع رگرسیون نیمه پارامتری با خطاهای وابسته^۵ و دارای توزیع گوسین است.

یکی از ابزارهای مهم در برآورد تابع رگرسیونی، موجک^۶ است. موجک ابزاری تحلیلی در پردازش سیگنال‌ها و تابع‌بندی ریاضی محسوب می‌شود و در ریاضیات کاربردی توانایی حل مسایل مربوطه را دارا است. اگر سیگنال، نمایشی از تابع زمانی باشد، آنگاه موجک موقعیت بندی مناسب و کارایی را در هر دو زمینه مقیاسی و مکانی ارائه می‌دهد، چیزی که پیش از این برای تبدیل سری فوریه امکان‌پذیر نبود. آنالیز فوریه نقش اساسی در برآورد توابع دارد ولی یک محدودیت در استفاده از تبدیل فوریه این است که برای به دست آوردن برآورد تابع در دامنه فرکانس، نیاز به تمام مقادیر تابع بر روی اعداد حقیقی است. برای این منظور باید مقدار تابع در یک محدوده را داشته باشیم. لذا همزمان نمی‌تواند خاصیت کران موضعی را در دامنه زمان و فرکانس داشته باشد. به عبارت دیگر در روش تبدیل فوریه، قادر خواهیم بود تعیین کنیم که یک فرکانس در چه مکان یا زمان از تابع اصلی به وقوع پیوسته است، ولی اطلاعات ناچیزی راجع به مقدار فرکانس سیگنال اصلی خواهیم داشت و بالعکس.

در تبدیل موجک به دلیل ویژگی‌های پایه‌های متعامد که اساس کار موجک‌ها است، توابع موجک می‌توانند دارای محمل متناهی در دامنه زمان باشند و همزمان خاصیت کران موضعی را در دامنه فرکانس داشته باشند و یا در مقابل، دارای محمل متناهی در دامنه فرکانس بوده و خاصیت کران موضعی در دامنه زمان داشته باشند. این خصوصیت ممتاز تبدیل موجک، نتایج خوبی در تشخیص نقاط پرش، تحلیل دوره‌های متناوب پنهان، فشرده‌سازی اطلاعات دارد. بنابراین تبدیل موجک هم در حوزه زمان و هم در حوزه فرکانس دارای رزولوشن بالایی است به عبارت دیگر این تبدیل نه تنها مقدار فرکانس‌های موجود در سیگنال را مشخص می‌کند، بلکه تعیین می‌کند که آن فرکانس‌ها در چه زمانی از سیگنال به وقوع می‌پیوندند. بطور کلی هر تجزیه و تحلیل مبتنی بر تبدیل فوریه را می‌توان با استفاده از موجک‌ها نیز تابع‌بندی کرد و اطلاعات زمانی و مکانی بیشتری به دست آورد.

¹ Regression analysis

² Parametric regression

³ Nonparametric regression

⁴ Wavelet threshold estimation

⁵ Corroleted errors

⁶ Wavelet

در سال‌های اخیر، نظریه موجک‌ها کاربرد فراوانی در بسیاری از زمینه‌های علوم مختلف از جمله آمار داشته است. اولین بار داناو و جانستون (۱۹۹۲) موجک را در تحقیقات آماری به کار بردند. بسیاری از مطالب پیشرفته ریاضی و آماری نظریه موجک در کتاب چوی (۱۹۹۲) آمده است. با افزایش توجه به موجک، کاربردهای بیشتری از این ابزار در علم آمار ارائه شده که به عنوان برخی از کارهای کلی در این زمینه می‌توان به آنتونیودیس (۲۰۰۷) و ویداکویک (۲۰۰۹) اشاره کرد.

بسیاری از روش‌های موجکی توسط نسون (۲۰۱۱) با استفاده از نرم افزار R انجام شده است. در زمینه برآورد تابع رگرسیون بر اساس موجک‌ها تحقیقات فراوانی انجام گرفته که می‌توان به هال و پاتیل (۱۹۹۶)، واسرمن و گنوس (۲۰۰۵)، دوستی و همکاران (۲۰۰۸) و افشاری و همکاران (۲۰۱۷) اشاره کرد.

داناو و جانستون (۱۹۹۵) برآوردگرهای موجک غیرخطی^۷ تابع رگرسیون ناپارامتری را معرفی کردند. آبرامویچ و بنجامین (۱۹۹۶) آستانه ضرایب موجک^۸ را به عنوان روش آزمون فرضیه‌های چندگانه در نظر گرفتند و محدودیت‌های آستانه‌سازی غیرسازگار را با استفاده از شبیه‌سازی^۹ نشان دادند. اوجدن و پارزن (۱۹۹۶) برآوردگرهای آستانه‌ای موجک را با داده‌های وابسته مورد مطالعه قرار دادند. جانسون و سیلورمن (۱۹۹۷) برآوردگرهایی را بر اساس روش اعتبارسنجی متقابل^{۱۰} برای انتخاب پارامتر آستانه به دست آوردند. کای و سیلورمن (۲۰۰۱) برآوردگر آستانه‌ای بلوک موجک را در تابع رگرسیون ناپارامتری محاسبه و نشان دادند که برآوردگر پیشنهادی ویژگی یک برآوردگر خوب شامل نارایی و سازگاری را دارا است. آنتونیادیس و همکاران (۲۰۰۱) برآوردگر موجکی تابع رگرسیون ناپارامتری را با وجود داده‌های با نوفه‌های گوسین^{۱۱} محاسبه و برای این داده‌ها، برآوردگر موجک را با برآوردگر انقباضی مقایسه کردند.

همچنین برخی از نویسندگان به برآورد موجک در تابع رگرسیون نیمه‌پارامتری^{۱۲} توجه کردند. برای مثال، گاناژ (۲۰۰۷) برآورد آستانه موجک را در تابع رگرسیون نیمه‌پارامتری جزئی با ضرایب رگرسیونی مجهول محاسبه و نیز نتایج کلی را برای اندازه نمونه بزرگ، به دست آورد. همچنین با استفاده از مطالعه شبیه‌سازی به مقایسه بین برآوردگر به دست آمده با برآوردگرهای دیگر پرداخت. همچنین هو (۲۰۰۵) با استفاده از برآورد ستیغی^{۱۳} در رگرسیون ناپارامتری، روزبه و کفاکی (۲۰۱۵) با استفاده از رگرسیون استوار از طریق روش کمترین توان‌های دوم پیراسته^{۱۴} و هی و سورینی (۲۰۱۶) با استفاده از راهکار به کار بردن نوعی درست‌نمایی جمع‌بسته^{۱۵} در رگرسیون نیمه‌پارامتری به بررسی این مسأله پرداختند.

⁷ Nonlinear wavelet estimators

⁸ Thresholding of wavelet coefficients

⁹ Simulation

¹⁰ Cross-validation

¹¹ Gaussian noise

¹² Semiparametric regression models

¹³ Ridge estimation

¹⁴ Least trimmed squares estimator

¹⁵ Integrated likelihood

لی و وو (۲۰۰۷) برآورد موجک در تابع رگرسیون نیمه پارامتری را تحت دنباله‌ای از خطاهای همگرا با وابستگی ضعیف بین آن‌ها محاسبه کردند. زو و لین (۲۰۱۰) برای یافتن برآوردهای موجک پارامترها در تابع رگرسیون نیمه پارامتری از روش بوت‌استرپ^{۱۶} استفاده کردند. آماره‌های بوت‌استرپ را با روش بازنمونه‌گیری افرون محاسبه و سپس همگرایی آماره‌های به دست آمده را مورد بررسی قرار دادند. هوو و وو (۲۰۱۲) نرخ همگرایی^{۱۷} برآوردهای موجک در تابع رگرسیون نیمه پارامتری با خطاهای همبسته را مورد بررسی قرار دادند. نیو و لیو (۲۰۱۲) قضیه حد مرکزی را برای برآوردهای موجک در تابع رگرسیون نیمه پارامتری با خطاهای همبسته مورد مطالعه قرار دادند. ژو و لین (۲۰۱۳) به بررسی ویژگی‌های مجانبی برآوردهای موجک در تابع‌های رگرسیون نیمه پارامتری با خطاهای همبسته پرداختند. یالاز (۲۰۱۸) برآورد موجکی خطاها در تابع رگرسیون نیمه پارامتری را با فرض معلوم بودن توزیع خطاها محاسبه و نتایج را با شبیه‌سازی مورد بررسی قرار داد. در بخش دوم مقاله، به معرفی آستانه موجک، مقدار آستانه و نحوه استفاده از این روش در برآوردیابی و نوفه‌زدایی پرداخته شده و برآوردگر هسته موجک مدل رگرسیون ارائه شده است. در بخش سوم، تابع رگرسیون نیمه پارامتری با فرض این‌که خطاها همبسته و دارای توزیع گوسین باشند، برآورد آستانه موجک محاسبه و نرخ همگرایی برآوردگر در قالب چند لم و قضیه بیان می‌شود. در بخش چهارم، برای ارزیابی روش پیشنهادی، از دو تابع سینوسی و تابع بلوکی به عنوان توابع هدف استفاده خواهد شد و با استفاده از شبیه‌سازی متوسط میانگین مربع خطا^{۱۸} و مقدار انحراف معیار به دست آمده با روش آستانه موجک با متوسط میانگین مربعات خطاها و مقادیر انحراف معیارهای به دست آمده با روش هسته موجک مقایسه می‌شوند. در بخش پنجم، به عنوان یک مثال کاربردی، تابع رگرسیون نیمه پارامتری موجکی بر روی داده‌های مربوط به میزان رشد دندان برازش داده می‌شود. در نهایت بحث و نتیجه‌گیری ارائه شده است. لازم به ذکر است که تمام محاسبات انجام شده شامل مقدار میانگین متوسط خطا و انحراف استاندارد میانگین مربع خطا برای برآوردهای موجکی و هسته‌ای و همچنین رسم نمودارها با استفاده از نرم‌افزار **R** انجام شده است.

۲. آستانه موجک و روش‌های آستانه‌گذاری

موجک‌ها ابزاری قابل انعطاف با محتوای غنی ریاضی هستند که در تجزیه و تحلیل یک تابع نوسانی از زمان و مکان به کار می‌روند. در تجزیه موجکی، موجک اصلی (تابع یا سیگنال) بر روی دسته‌ای از توابع پایه که موجک نامیده می‌شوند تصویر شده و با استفاده از حاصل ضرب موج اصلی و تابع موجک، تجزیه موجکی انجام می‌گیرد. موجک تابعی با میانگین صفر است که بسط بر اساس انتقال و اتساع آن انجام می‌شود. اساس موجک از موجک پدر $\phi(\cdot)$ و موجک مادر $\psi(\cdot)$ تشکیل شده که

¹⁶ Bootstrap approach

¹⁷ Convergence ratio

¹⁸ Average of mean square error

به ترتیب روند کلی داده‌ها و انحرافات را نشان می‌دهند و به فرم‌های

$$\psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{سایر جابجایی‌ها} \end{cases} \quad \phi(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{سایر جابجایی‌ها} \end{cases}$$

می‌باشند. برای هر $j \geq 0$ و $k = 0, 1, \dots, 2^j - 1$ ، انتقال‌های مقیاسی توابع ψ و ϕ به صورت

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \quad (1)$$

تعریف می‌شوند. توابع $\phi(x)$ ، $\psi(x)$ ، $\phi_{j,k}(x)$ و $\psi_{j,k}(x)$ که قطعه‌وارهای ثابت هستند، می‌توانند در برآورد توابع استفاده شوند و برای هر j ثابت $\psi_{j,k}$ را موجک‌های نسل j ام می‌گویند. اگر برای هر مقدار $j \in \mathbb{Z}$ ، $V_j = \text{Span}\{\phi_{j,k}, k \in \mathbb{Z}\}$ باشد، که منظور از Span فضای تولید شده است، آنگاه V_j ها دنباله‌ای از فضاهای تودرتو به صورت $V_j \subset V_{j+1}$ خواهند بود.

تعریف ۱. اگر بردارهایی از فضای برداری V باشند، به طوری که هر بردار دلخواهی در V به صورت ترکیب خطی از بردارهای v_1, \dots, v_n قابل بیان باشد، آنگاه گفته می‌شود که این بردارها فضای برداری V را تولید می‌کنند.

تعریف ۲. دو بردار u و v را متعامد^{۱۹} گویند، هرگاه بر هم عمود باشند یا به عبارت دیگر دو بردار متعامدند، اگر و تنها اگر ضرب داخلی آن‌ها برابر با صفر باشد. دو بردار متعامد u و v را با نماد $u \perp v$ نشان می‌دهند. همچنین دسته‌ای از بردارهای دو به دو متعامد که طول آن‌ها برابر یک است را بردارهای متعامد یکه گویند.

تعریف ۳. بردار u را بر زیر فضای V_1 متعامد گویند، اگر بردار u بر هر بردار در زیر فضای V_1 متعامد باشد و به مجموعه‌ای که شامل تمامی بردارهای متعامد بر زیر فضای V_1 باشد، متمم متعامد^{۲۰} زیر فضای V_1 گویند.

تصویر تابع f روی فضای V_j عبارت است از:

$$P_j f = \sum_{k \in \mathbb{Z}} \alpha_{j,k} \phi_{j,k}(x),$$

که در آن $\alpha_{j,k} = \int_{s_x} t \phi_{j,k}(x) f(x) dx$ است. (دابچیز ۱۹۹۲ و چوی ۱۹۹۲)

اگر فضای تولید شده از متمم متعامد V_j در V_{j+1} در نظر گرفته شود، آنگاه:

$$W_j = \text{Span}\{\psi_{j,k}, k \in \mathbb{Z}\}_{j \in \mathbb{Z}}.$$

¹⁹ Orthogonal

²⁰ Orthogonal complement

همچنین با فرض این که $f \in L^2(R)$ باشد، تقریب تابع f بر روی فضای V_{j_0} بصورت

$$\tilde{f} = \sum_{k \in \mathbb{Z}} \alpha_{j_0, k} \phi_{j_0, k}(x) + \sum_{j \leq j_0} \sum_{k \in \mathbb{Z}} \beta_{j, k} \psi_{j, k} = P_{j_0} f + \sum_{j \leq j_0} D_j f, \quad (2)$$

خواهد بود، که در آن $\beta_{j, k} = \int_{s_x} \psi_{j, k}(x) f(x) dx$ و تابع D_j تصویر تابع f روی فضای W_j است. تبدیل موجک ابزاری برای تجزیه کردن توابع، عملگرها و داده‌ها به اجزای با فرکانس‌های مختلف است که اجازه مطالعه و تجزیه و تحلیل هر کدام از اجزا را بطور جداگانه می‌دهد. در تبدیل موجک گسسته، سیگنال از یک سری فیلترهای بالاگذر برای فرکانس‌های بالا و یک سری فیلترهای پایین‌گذار برای تحلیل فرکانس‌های پایین، عبور داده شده و سیگنال به دو بخش تقسیم می‌شود. بخش حاصل از عبور سیگنال بالاگذر که شامل اطلاعات فرکانس بالا و جزئیات نام دارد، بخش حاصل از عبور سیگنال از فیلتر پایین‌گذر که شامل اطلاعات فرکانس پایین و در بر گیرنده مشخصات هویتی سیگنال بوده و کلیات نامیده می‌شود. در نهایت، گروهی از سیگنال‌ها را خواهیم داشت که همان سیگنال اولیه را نشان می‌دهند اما هر گروه سیگنال به باند فرکانسی متفاوتی مربوط است.

تعریف ۴. (نوفه) هرگونه آلودگی صوتی و یا سیگنالی ناخواسته که شکل سیگنال‌ها را تغییر دهد و باعث بروز اختلال شود را نوفه گویند. نوفه اغلب دارای اجزاء فرکانسی نیز بوده و تجزیه و تحلیل فرکانسی شبیه به سیگنال دارد.

تعریف ۵. (آستانه) مقداری است که بر اساس آن ضرایب موجکی به دو دسته کم اهمیت و پر اهمیت تقسیم می‌شوند و آن را با نماد λ نشان می‌دهند.

تعریف ۶. (تابع حذف): تابع حذف^{۲۱} به دو صورت سخت و نرم قابل استفاده است و به ترتیب به فرم‌های

$$\eta_{hard}(x) = x I_{\{|x| < \lambda\}},$$

و

$$\eta_{soft}(x) = (x + \lambda) I_{\{x < \lambda\}} + (x - \lambda) I_{\{x > \lambda\}},$$

تعریف می‌شوند، که در آن I تابع نشانگر است. (برای اطلاعات بیشتر به دایچیز ۱۹۹۲، رجوع شود).
برای نوفه‌زدایی از سیگنال‌هایی که به همراه نوفه مشاهده می‌شوند، روش‌های مختلفی وجود دارد. برای مثال، داناو و جانستون (۱۹۹۵) روش آستانه‌ای را برای حذف نوفه‌ها ارائه دادند که به روش آستانه موجک مشهور شد. دلایل زیادی برای استفاده از این روش وجود دارد که از جمله می‌توان به سادگی استفاده از آن در مسایل مختلف و کاربردی و سریع بودن آن اشاره کرد. روش آستانه موجک مبتنی بر دو پارامتر اساسی مقدار آستانه و تابع حذف است. با مشخص شدن مقدار آستانه و تابع حذف، مراحل الگوریتم آستانه موجک عبارت است از:

²¹ Deletion function

۱. ابتدا از مشاهدات دارای نوفه، تبدیل موجک گسسته گرفته می‌شود. به عبارت دیگر، با فرض این که ماتریس تبدیل موجکی

گسسته W و تابع مشاهدات به صورت $y = f + \varepsilon$ باشد، ضرایب موجکی از تبدیل موجکی

$$Wy = Wf + W\varepsilon \Rightarrow d^* = d + \varepsilon,$$

به دست می‌آید.

۲. با استفاده از مقدار آستانه، ضرایب موجکی به دو دسته ضرایب پر اهمیت و ضرایب کم اهمیت تقسیم می‌شوند، به این صورت که اگر ضریب موجکی بیشتر از مقدار آستانه باشد، متعلق به دسته ضرایب پر اهمیت و در غیر این صورت جزو دسته ضرایب کم اهمیت خواهد بود. سپس ضرایب کم اهمیت حذف شده و ضرایب پر اهمیت با توجه به تابع حذف، به این صورت اصلاح می‌شوند که در تابع حذف سخت، ضرایب کمتر از مقدار آستانه برابر صفر و باقی ضرایب بدون تغییر باقی می‌مانند. در تابع حذف نرم، ضرایب کمتر از مقدار آستانه برابر صفر و باقی ضرایب به اندازه مقدار آستانه به صفر نزدیک می‌شوند.

۳. سیگنال حاصل با استفاده از تبدیل موجک معکوس بازسازی می‌شود. (برای اطلاعات بیشتر به چوی ۱۹۹۲ و داناهو و جانستون ۱۹۹۵، رجوع شود.)

روش‌های مختلفی برای تعیین آستانه وجود دارد که از آن جمله می‌توان به آستانه جهانی (داناهو و جانستون ۱۹۹۵)، اعتبار متقابل (نیسون ۱۹۹۶) و برآوردگر مخاطره ناریب استاین^{۲۲} (داناهو و جانستون ۱۹۹۵) اشاره نمود. در این مقاله، برای برآورد تابع از مقدار آستانه جهانی $\lambda = \sigma\sqrt{2\log n}$ استفاده شده که در آن σ نشان دهنده انحراف معیار نوفه‌ها و n تعداد مشاهدات است. برای مطالعه بیشتر در این زمینه به کرمی و افشاری (۲۰۱۹ و ۲۰۲۰) و کرمی کبیر و همکاران (۲۰۲۰) مراجعه کنید.

۱.۲. برآوردگر هسته موجک

فرض کنید $\{X_n, n \geq 1\}$ دنباله‌ای از متغیرهای تصادفی دارای تابع توزیع $F(x)$ و تابع چگالی $f(x)$ باشد، تابع $f(x)$ کراندار و دارای محمل فشرده روی فضای احتمال (Ω, F, P) باشد. یک برآورد طبیعی تابع توزیع $F(x)$ ، تابع

توزیع تجربی

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x),$$

است. چون طبق تعریف، $f(x)$ مشتق تابع $F(x)$ می‌باشد، بنابراین می‌توان برای یک h مناسب، برآوردگر $f(x)$ را به

صورت

²² Stein's unbiased risk estimator

$$\hat{f} = \frac{1}{2h} (\hat{F}(x+h) - \hat{F}(x-h)) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) = \frac{1}{n} \sum K_h(x, X_i),$$

نوشت، که در آن

$$K(x) = \begin{cases} \frac{1}{2} & x \in (-1, 1) \\ 0 & o.w, \end{cases}$$

و

$$K_h(x, X_i) = \frac{1}{h} K\left(\frac{x-X_i}{h}\right),$$

است. h را پارامتر هموارساز و تابع K را تابع هسته‌ای می‌نامند. برای جفت نمونه‌های (X_i, Y_i) ، $i=1, \dots, n$ ، می‌خواهیم ارتباط بین دو متغیر را بررسی کنیم. در حالت خاص برای مدل رگرسیون تصادفی

$$Y_i = f(X_i) + e_i, \quad i = 1, 2, \dots, n$$

فرض می‌شود (x_1, \dots, x_n) مقادیر عددی متغیرهای تصادفی (X_1, \dots, X_n) بوده و هدف برآورد تابع رگرسیون از داده‌ها است. اگرچه رگرسیون پارامتری بطور گسترده در موضوعات مختلف استفاده شده، اما در بسیاری از مسایل ممکن است علاقمند باشیم که برای داده‌هایی مشخص، یک مدل به خصوص پیش‌بینی شود. رگرسیون ناپارامتری یک روش بسیار خوب برای این منظور است بدون این‌که فرضیات زیادی روی تابع برقرار باشد. مثلاً ممکن است فرض شود که f تابعی غیر هموار بوده و یا f حداقل یک مشتق پیوسته داشته باشد.

حال برای مدل رگرسیونی، فرض کنید (X_i, Y_i) ، $i=1, \dots, n$ ، متغیرهای تصادفی و هم‌توزیع با میانگین شرطی $E(Y_i | X_i = x_i)$ باشند. برای این مدل آنتونیادیس (۲۰۰۷) برآوردگر

$$\tilde{f}(x_i) = \frac{\sum_{i=1}^n y_i k_i(x, X_i)}{\sum_{i=1}^n k_i(x, X)}, \quad (3)$$

را پیشنهاد کرد که در واقع برآوردگرهای هسته موجک است. همچنین نشان داد که اگر $j \rightarrow \infty$ و $n2^j \rightarrow \infty$ ، آنگاه برآوردگر فوق، به ازای تمام x ‌ها سازگار بوده و $E(Y | X = x^*)$ برای x^* هایی که در همسایگی x است، کراندار است.

۳. برآورد آستانه موجک تابع رگرسیون نیمه پارامتری

در بیشتر روش‌های رگرسیون ناپارامتری، تابع مورد نظر پیوسته و همواره مشتق پذیر در نظر گرفته می‌شود، در حالی که در برآورد موجکی نیازی به مشتق پذیری تابع رگرسیون نیست. در این بخش، ابتدا تابع رگرسیون نیمه پارامتری را معرفی نموده و برآورد آستانه موجک این تابع را محاسبه و سپس نرخ همگرایی برآوردگر به دست آمده بررسی می‌شود. تابع رگرسیون نیمه پارامتری، ترکیبی از تابع‌های پارامتری و ناپارامتری رگرسیون است و اغلب زمانی به کار برده می‌شود که رگرسیون ناپارامتری خوب عمل نکند و یا پژوهشگر بخواهد از تابع پارامتری استفاده کند که شکل تابعی آن بستگی به زیر تابع رگرسیونی داشته باشد که توزیع آن معلوم نباشد.

تعریف ۷. (رابینسون، ۱۹۸۸) فرض کنید

$$Y_i = X_i' \beta + g(Z_i) + e_i, \quad i = 1, \dots, n, \quad (4)$$

که در آن Y_i متغیر وابسته، X_i ها بردار مشاهدات $p \times 1$ بعدی، β بردار پارامتر $p \times 1$ بعدی بوده و $Z_i \in R^p$ دسته‌ای از متغیرهای پیوسته و نیز e_i خطاهای تصادفی دارای توزیع نرمال $N(0, \sigma^2)$ است. چنین مدلی را مدل نیمه پارامتری گویند، که قسمت پارامتری آن $X_i' \beta$ و قسمت ناپارامتری آن $g(Z_i)$ است.

اکنون فرض کنید برای متغیرهای تصادفی پیوسته X_i ، Z_i و Y_i ، $i = 1, \dots, n$ ، در مدل (۴) شرایط

$$E[e_i | X_i, Z_i] = 0, \quad \text{Var}[e_i | X_i, Z_i] = \sigma_{(x_i, z_i)}^2,$$

برقرار باشند. ابتدا برآوردگر سازگار بردار پارامتر β معرفی می‌شود. برای این منظور با گرفتن امید ریاضی شرطی از طرفین رابطه (۴)، داریم:

$$E[Y_i | Z_i] = E[X_i' | Z_i] \beta + g(Z_i). \quad (5)$$

با کم کردن رابطه (۵) از رابطه (۴) خواهیم داشت:

$$[Y_i - E(Y_i | Z_i)] = [X_i' - E(X_i' | Z_i)] \beta + e_i. \quad (6)$$

گائو و همکاران (۲۰۱۵)، برآوردگر سازگار بردار پارامتر β ی داده شده در رابطه (۶) را به صورت

$$\hat{\beta} = \left(\sum_{i=1}^n [X_i - E(X_i | Z_i)][X_i - E(X_i | Z_i)]' \right)^{-1}$$

$$\times \left(\sum_{i=1}^n [X_i - E(X_i | Z_i)][X_i - E(X_i | Z_i)] \right) = \beta_0,$$

محاسبه کردند. بنابراین عبارت $X_i' \beta_0$ مقداری معلوم بوده و تنها قسمت مجهول در رابطه (۴) عبارت $g(Z_i)$ است. به دلیل همخوانی با بخش دوم، مدل رابطه (۴) به فرم

$$y_i = f(x_i) + e_i, \quad e_i \sim N(0, \sigma^2), \quad (7)$$

بازنویسی می‌شود، که تنها قسمت مجهول آن تابع f است. با توجه به تعریف ۲،۱ داده شده در ایچیمورا (۱۹۹۳)، مدل رابطه (۷) یک مدل نیمه پارامتری است و هدف برآورد کردن تابع مجهول f به روش انقباض موجک است. حال فرض کنید $((x_1, y_{(x_1)}), \dots, (x_n, y_{(x_n)}))$ یک نمونه تصادفی از مدل رابطه (۷) باشند، که $x_i, i=1, 2, \dots, n$ متغیرهای تصادفی، هم توزیع و مستقل دارای توزیع یکنواخت $U[0,1]$ و $0 \leq x_{(1)} \leq \dots \leq x_{(n)} \leq 1$ نشان دهنده آماره‌های ترتیبی x_1, \dots, x_n هستند و در شرایط

$$V_{ar}(x_{(i)}) \leq \frac{1}{n}, \quad |E(x_{(i)} - \frac{i}{n})| \leq \frac{1}{\sqrt{n}}, \quad i=1, \dots, n, \quad (8)$$

$$Cov(e(x_{(i)}), e(x_{(j)})) = \gamma(|i-j|), \quad \lim_{n \rightarrow \infty} \sum_{u=-(n-1)}^{n-1} |\gamma(u)| < \infty, \quad (9)$$

صدق می‌کنند. با استفاده از رابطه (۲)، برآوردگر آستانه موجکی تابع f عبارت است از:

$$\hat{f}(x) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} \hat{d}_{j,k} \psi_{j,k}(x), \quad (10)$$

که در آن $\hat{d}_{j,k}$ ها ضرایب موجکی، $\hat{c}_{j_0,k}$ ها ضرایب مقیاسی به دست آمده از تبدیل موجک گسسته و J' بزرگترین عدد صحیحی است که برای مقدار ثابت $K > 0$ در نامساوی $2^{J'} \leq K \sqrt{n \log n}$ صدق کند. در عمل معمولاً ثابت k طوری انتخاب می‌شود که $J' \geq J$ باشد و ضرایب آستانه روی همه سطوح ابتدایی در سطح J_0 انجام شود. تعریف ۸ (ویدا کوویک، ۲۰۰۹) اگر تابع $f^{(m)}$ نشان دهنده m امین مشتق تابع f و $[\alpha]$ بزرگترین عدد صحیح کوچکتر مساوی α باشد. آنگاه برای هر مقدار حقیقی α ، تابع f متعلق به کلاس هولدر $\Lambda^\alpha(M)$ روی بازه $[0,1]$ است هرگاه:

۱. برای $0 \leq \alpha \leq 1$ ، همواره نامساوی $|f(x) - f(y)| \leq M |x - y|^\alpha$ برقرار باشد.

۲. برای $\alpha > 1$ ، همواره نامساوی‌های $|f^{(1)}(x)| \leq M$ و $|f^{([\alpha])}(x) - f^{([\alpha])}(y)| \leq M |x - y|^{\alpha - [\alpha]}$ برقرار

باشند.

۱.۳. لم‌ها و قضایا

در این زیربخش، دو قضیه تحت عنوان قضیه ۱ و قضیه ۲ و نیز لم‌هایی که در اثبات قضیه ۲ مورد استفاده قرار می‌گیرند، ارائه شده است. برای سادگی در نوشتن، در اثبات لم‌ها و قضیه اصلی از نمادهای C و C_i ، $i=1,2,\dots$ به عنوان ضرایب ثابت استفاده می‌شود.

قضیه ۱. فرض کنید برای هر $i, r, s=1, \dots, n$ ، $y_i = f(t_{(i)}) + e_i$ که e_1, \dots, e_n یک فرآیند مانای گوسین با $E(e_i) = 0$ و $Var(e_{(i)}) = \sigma^2$ و $Cov(e_r, e_s) = \gamma(|r-s|)$ ، باشد. اگر $\sum_{u=-\infty}^{\infty} |\gamma(u)| < \infty$ و رابطه (۸) برقرار باشد، آن‌گاه با فرض نامساوی $t_{(1)} < \dots < t_{(n)}$ ، برای هر $\alpha > 0$ که $f \in \Lambda^\alpha(M)$ باشد، تقریب خطا عبارت است از:

$$\frac{1}{n} \sum_{i=1}^n E \left(\left(f(t_{(i)}) - f\left(\frac{i}{n}\right) \right)^2 \right) \leq O(n^{-s(\alpha)}).$$

برهان: با توجه به تعریف (۸) خواهیم داشت:

$$|f(x) - f(y)| \leq M |x - y|^{s(\alpha)},$$

که در آن $s(\alpha) = \min(\alpha, 1)$ است. با استفاده از نامساوی جنسن^{۲۳} و رابطه (۸) تقریب خطا به صورت زیر محاسبه می‌شود:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n E \left(\left(f(t_{(i)}) - f\left(\frac{i}{n}\right) \right)^2 \right) &\leq \frac{M}{n} \sum_{i=1}^n E \left(\left(t_{(i)} - \frac{i}{n} \right)^{2s(\alpha)} \right) \\ &\leq \frac{M}{n} \sum_{i=1}^n \left[E \left(\left(t_{(i)} - \frac{i}{n} \right)^2 \right) \right]^{s(\alpha)} \\ &= \frac{M}{n} \sum_{i=1}^n \left\{ \text{Var} \left(t_{(i)} - \frac{i}{n} \right) + \left[E \left(t_{(i)} - \frac{i}{n} \right) \right]^2 \right\}^{s(\alpha)} \end{aligned}$$

²³ Jensen's inequality

$$\leq \frac{M}{n} \sum_{i=1}^n \left[\frac{1}{n} + \left(\frac{1}{\sqrt{n}} \right)^2 \right]^{s(\alpha)}$$

$$= O(n^{-s(\alpha)}).$$

لم ۱. فرض کنید برای هر $\alpha > 0$ ، $f \in \Lambda^\alpha(M)$ و موجک مادر ψ دارای r گشتاور صفر با فرض $r \geq \alpha$ باشد. آنگاه برای هر $n \geq 3$ نامساوی

$$\sum_{j=J'}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k}^2 \leq C_2 \left(\frac{\log n}{n} \right)^{\frac{\alpha}{1+\alpha}},$$

برقرار است.

برهان: با توجه به قضیه ۲،۹،۱ در دوبیچز (۱۹۹۲) و لم ۱ در چوی (۱۹۹۲)، به ازای هر $\alpha > 0$ و $n \geq 3$ می‌توان

نوشت:

$$\begin{aligned} \sum_{j=J'}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k}^2 &\leq C_1 \sum_{j=J'}^{\infty} 2^j 2^{2(-j(1/2+\alpha))} = C_1 \sum_{j=J'}^{\infty} 2^{-2j\alpha} \\ &= C_1 \left(\sum_{j=0}^{\infty} 2^{-2j\alpha} - \sum_{j=0}^{J'-1} 2^{-2j\alpha} \right) \\ &= C_1 \frac{(2^{-2\alpha})^{J'}}{1-2^{-2\alpha}} \leq C_1 (2^{J'})^{-2\alpha} \\ &= C_1 (2^{J'+1-1})^{-2\alpha} = 2^{2\alpha} C_1 (2^{J'+1})^{-2\alpha} \\ &\leq 2^{2\alpha} C_1 \left(K^2 \frac{n}{\log n} \right)^{-2\alpha/2} = C_2 \left(\frac{\log n}{n} \right)^{\alpha} \\ &\leq C_2 \left(\frac{\log n}{n} \right)^{\frac{\alpha}{1+\alpha}}. \end{aligned}$$

لم ۲. فرض کنید برای هر $\alpha > 0$ ، $f \in \Lambda^\alpha(M)$ و موجک مادر ψ دارای r گشتاور صفر با فرض $r \geq \alpha$ باشد. همچنین اگر برای هر متغیر تصادفی Y ، $E_1(Y) = E(Y | t_1, t_2, \dots, t_n)$ باشد. آنگاه نامساوی

$$\sum_{k=0}^{2^{j_0}-1} E \left(\left(\hat{c}_{j_0,k} - c_{j_0,k} \right)^2 \right) \leq C_3 2^{2j_0} n^{-1} + 2 \sum_{k=0}^{2^{j_0}-1} \left(\tilde{a}_{j_0,k}^2 + E(\tilde{b}_{j_0,k}^2) \right),$$

برقرار است.

$$\begin{aligned}
 \text{برهان: مقدار عبارت } E\left(\left(\hat{c}_{j_0,k} - c_{j_0,k}\right)^2\right) \text{ بصورت} \\
 E\left(\left(\hat{c}_{j_0,k} - c_{j_0,k}\right)^2\right) &= E(E_1((\hat{c}_{j_0,k} - c_{j_0,k})^2)) \\
 &= E(E_1(\hat{f}_{j_0,k} + \tilde{a}_{j_0,k} + \tilde{b}_{j_0,k}^2)) \\
 &= E(E_1(\hat{f}_{j_0,k} + \tilde{a}_{j_0,k} + \tilde{b}_{j_0,k})) \\
 &= E(\hat{f}_{j_0,k}^2) + E(E_1((\tilde{a}_{j_0,k} + \tilde{b}_{j_0,k})^2)),
 \end{aligned}$$

نوشته می‌شود. بنابراین نامساوی

$$\begin{aligned}
 E\left(\left(\hat{c}_{j_0,k} - c_{j_0,k}\right)^2\right) &\leq E(\hat{f}_{j_0,k}^2) + E(E_1((2\tilde{a}_{j_0,k}^2 + 2\tilde{b}_{j_0,k}^2))) \\
 &= E(\hat{f}_{j_0,k}^2) + 2\tilde{a}_{j_0,k}^2 + 2E(\tilde{b}_{j_0,k}^2) \\
 &\leq C_3 2^{j_0} n^{-1} + 2\tilde{a}_{j_0,k}^2 + 2E(\tilde{b}_{j_0,k}^2),
 \end{aligned}$$

برقرار است. با مجموع گیری از طرفین نامساوی فوق و استفاده از نامساوی

$$(a+b)^2 \leq 2a^2 + 2b^2, \quad a, b \in R$$

اثبات کامل می‌شود.

لم ۳. تحت شرایط لم‌های ۱ و ۲ نامساوی

$$\begin{aligned}
 E\left(\left(\hat{d}_{j,k} - d_{j,k}\right)^2\right) &\leq 8d_{j,k}^2 \wedge 6 \frac{2^j C_4}{n} \log n \\
 &\quad + 20a_{j,k}^2 + 20E(b_{j,k}^2) + \frac{2^{j+1} C_4}{n^2},
 \end{aligned}$$

برقرار است، که در آن $x \wedge y = \min(x, y)$ است.

برهان: ابتدا توجه کنید که بنا به پورتو و همکاران (۲۰۰۹) نامساوی

$$E\left(\left(\hat{d}_{j,k} - d_{j,k}\right)^2\right) = E\left(\left(\hat{d}_{j,k} - d'_{j,k} + a_{j,k} + b_{j,k}\right)^2\right) \quad (11)$$

$$\begin{aligned} &\leq E\left(2(\hat{d}_{j,k} - d'_{j,k})^2 + 2(a_{j,k} + b_{j,k})^2\right) \\ &= E\left(E_1\left(2(\hat{d}_{j,k} - d'_{j,k})^2\right) + 2(a_{j,k} + b_{j,k})^2\right), \end{aligned}$$

برقرار است. حال فرض کنید $\frac{\sigma_{j,k;1}^2}{n} = E_1(\hat{f}_{j,k}^2)$ باشد. با استفاده از لم ۴ داده شده در چوی (۱۹۹۲)، می‌توان نوشت:

$$\begin{aligned} E_1\left(\left(\hat{d}_{j,k} - d'_{j,k}\right)^2\right) &\leq \left(2(d'_{j,k})^2 + \frac{\sigma_{j,k;1}^2}{n^2}\right) \wedge (2\log n + 1) \frac{\sigma_{j,k;1}^2}{n} \\ &\leq \left(2(d'_{j,k})^2 + \frac{\sigma_{j,k;1}^2}{n^2}\right) \wedge \left(3\log n + \frac{1}{n}\right) \frac{\sigma_{j,k;1}^2}{n} \\ &\leq \left(2(d'_{j,k})^2 + \frac{\sigma_{j,k;1}^2}{n^2}\right) \wedge \left(3\frac{\sigma_{j,k;1}^2}{n} \log n + \frac{\sigma_{j,k;1}^2}{n^2}\right) \\ &= 2(d'_{j,k})^2 \wedge 3\frac{\sigma_{j,k;1}^2}{n} \log n + \frac{\sigma_{j,k;1}^2}{n^2}. \end{aligned}$$

با استفاده از عبارت فوق و رابطه (۱۱) به سادگی خواهیم داشت:

$$\begin{aligned} &E_1\left(2(\hat{d}_{j,k} - d'_{j,k})^2\right) + 2(a_{j,k} + b_{j,k})^2 \\ &\leq E_1\left(2(\hat{d}_{j,k} - d'_{j,k})^2\right) + 4a_{j,k}^2 + 4b_{j,k}^2 \\ &\leq 8d_{j,k}^2 \wedge 6\frac{\sigma_{j,k;1}^2}{n} \log n + 20\left(a_{j,k}^2 + b_{j,k}^2 + \frac{\sigma_{j,k;1}^2}{10n^2}\right). \end{aligned}$$

بنابراین نامساوی زیر برقرار است:

$$+20\left(a_{j,k}^2 + E(b_{j,k}^2) + E\left(\frac{\sigma_{j,k;1}^2}{10n^2}\right)\right) E\left(\left(\hat{d}_{j,k} - d'_{j,k}\right)^2\right) \leq 8d_{j,k}^2 \wedge 6E\left(\frac{\sigma_{j,k;1}^2}{n}\right) \log n$$

$$+20a_{j,k}^2 + 20E(b_{j,k}^2) + \frac{2^{j+1}C_4}{n^2} \leq 8d_{j,k}^2 \wedge 6\frac{2^j C_4}{n} \log n$$

لم ۴. تحت شرایط لم ۳، نامساوی زیر برقرار است:

$$\sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^{j-1}} E\left(\left(\hat{d}_{j,k} - d_{j,k}\right)^2\right) \leq C_5 \left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2+2\alpha}} + 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^{j-1}} a_{j,k}^2 + E(b_{j,k}^2). \quad (12)$$

برهان: با توجه به لم ۳ می‌توان نوشت:

$$\begin{aligned} \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^{j-1}} E\left(\left(\hat{d}_{j,k} - d_{j,k}\right)^2\right) &\leq \frac{6C_4 \log n}{n} \sum_{j=j_0}^{J_1-1} \sum_{k=0}^{2^{j-1}} 2^j + 8 \sum_{j=J_1}^{J'-1} \sum_{k=0}^{2^{j-1}} d_{j,k}^2 \\ &+ 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^{j-1}} a_{j,k}^2 + E(b_{j,k}^2) + \frac{2C_4}{n^2} \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^{j-1}} 2^j \\ &\leq \frac{6C_4 \log n}{n} \sum_{j=j_0}^{J_1-1} 2^{2j} + 8C_1 \sum_{j=J_1}^{J'-1} 2^j 2^{-j(1+2\alpha)} + 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^{j-1}} a_{j,k}^2 + E(b_{j,k}^2) + \frac{2C_4}{n^2} \sum_{j=j_0}^{J'-1} 2^{2j} \\ &\leq \frac{6C_4 \log n}{n} 2^{2J_1} + 8C_1 \left(\frac{2^{-2\alpha J'} - 2^{-2\alpha J_1}}{2^{-2\alpha} - 1}\right) + 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^{j-1}} a_{j,k}^2 + E(b_{j,k}^2) + \frac{2C_4}{n^2} 2^{2J'} \\ &= \frac{6C_4 \log n}{n} 2^{2J_1} + C_6 2^{-2\alpha J_1} - C_6 2^{-2\alpha J'} \\ &+ 20 \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^{j-1}} a_{j,k}^2 + E(b_{j,k}^2) + \frac{2C_4}{n^2} 2^{2J'}. \end{aligned} \quad (13)$$

همچنین در آخرین عبارت رابطه (۱۳)، بنا به قضیه ۲،۹،۱ دوبیچز (۱۹۹۲) و لم ۱ چوی (۱۹۹۲)، نامساوی‌های

$$\frac{6C_4 \log n}{n} 2^{2J_1} \leq \frac{24C_4 \log n}{nC_6^2} \left(\frac{n}{\log n}\right)^{\frac{1}{1+\alpha}} = C_7 \left(\frac{\log n}{n}\right)^{\frac{\alpha}{1+\alpha}},$$

$$C_6 2^{-2\alpha J_1} \leq C_6 \left[\frac{1}{C_5} \left(\frac{n}{\log n}\right)^{\frac{1}{2+2\alpha}} \right]^{-2\alpha} = C_8 \left(\frac{\log n}{n}\right)^{\frac{\alpha}{1+\alpha}},$$

$$\frac{2C_4}{n^2} 2^{2j'} \leq \frac{2C_4 K}{n} \leq \frac{2C_4 K}{n^{\frac{\alpha}{1+\alpha}}} \leq 2C_4 K \left(\frac{\log n}{n} \right)^{\frac{\alpha}{1+\alpha}},$$

برقرار هستند. با جایگذاری روابط فوق در رابطه (۱۳)، اثبات کامل می‌شود.

قضیه ۲. فرض کنید شرایط (۸) و (۹) در تابع (۷) برقرار باشد. همچنین فرض کنید موجک مادر ψ ، دارای r گشتاور متمایل به صفر و محمل فشرده است. آنگاه برآوردگر \hat{f} تعریف شده در رابطه (۱۰) تحت یک عامل لگاریتمی دارای نرخ همگرایی بهینه در فضای هولدر $\Lambda^\alpha(M)$ است. به عبارت دیگر برای هر $M \in (0, \infty)$ و $\alpha > 0$ خواهیم داشت:

$$\sup_{f \in \Lambda^\alpha(M)} E \|\hat{f} - f\|_2^2 \leq C \left(\frac{\log n}{n} \right)^{\frac{\alpha}{1+\alpha}}.$$

برهان: عبارت $E(\|\hat{f} - f\|_2^2)$ به صورت

$$\begin{aligned} E(\|\hat{f} - f\|_2^2) &= E\left(\int_0^1 [\hat{f}(x) - f(x)]^2 dx\right) \\ &= E\left(\int_0^1 \left[\sum_{k=0}^{2^{j_0-1}} \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{j'-1} \sum_{k=0}^{2^j-1} \hat{d}_{j,k} \psi_{j,k}(x) \right. \right. \\ &\quad \left. \left. - \sum_{k=0}^{2^{j_0-1}} c_{j_0,k} \phi_{j_0,k}(x) - \sum_{j=j_0}^{j'-1} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x) \right. \right. \\ &\quad \left. \left. - \sum_{j=j'}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x) \right]^2 dx\right) \\ &= E\left(\int_0^1 \left[\sum_{k=0}^{2^{j_0-1}} (\hat{c}_{j_0,k} - c_{j_0,k}) \phi_{j_0,k}(x) \right. \right. \\ &\quad \left. \left. + \sum_{j=j_0}^{j'-1} \sum_{k=0}^{2^j-1} (\hat{d}_{j,k} - d_{j,k}) \psi_{j,k}(x) - \sum_{j=j'}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x) \right]^2 dx\right), \end{aligned}$$

نوشته می‌شود. از طرفی به دلیل متعامد یکه بودن پایه موجک، عبارت فوق برابر با

$$E(\|\hat{f} - f\|_2^2) = \sum_{k=0}^{2^{j_0}-1} E\left((\hat{c}_{j_0,k} - c_{j_0,k})^2\right) + \sum_{j=j_0}^{J'-1} \sum_{k=0}^{2^j-1} E(\hat{d}_{j,k} - d_{j,k})^2 + \sum_{j=J'}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k}^2,$$

خواهد بود. در نهایت با استفاده از قضیه ۱ و لم‌های ۱، ۲ و ۴ و بنا به پورتو و همکاران (۲۰۰۹) چون نامساوی

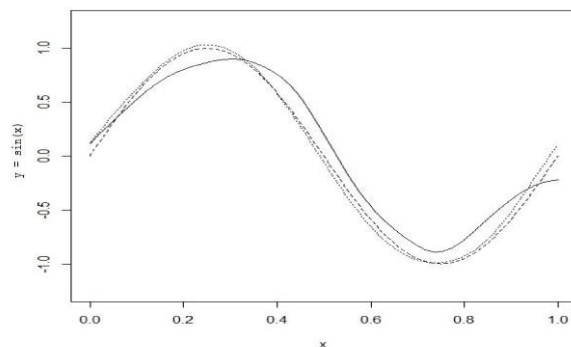
$$s(\alpha) = \min(\alpha, 1) > \frac{\alpha}{1+\alpha}, \quad \alpha > 0$$

برقرار است، بنابراین اثبات به صورت زیر کامل می‌شود:

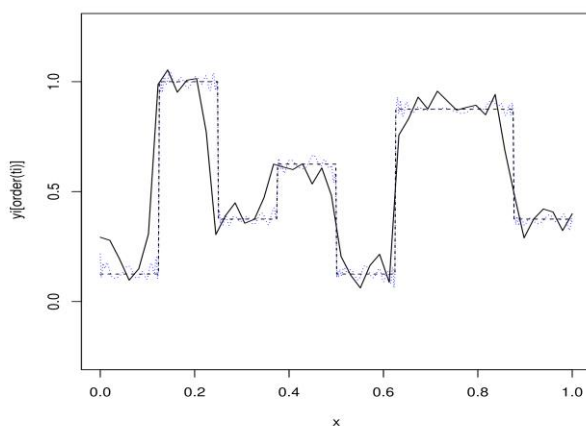
$$E(\|\hat{f} - f\|_2^2) \leq \frac{C_3 2^{2j_0}}{n} + \frac{C_9}{n^{2(1/2+s(\alpha))}} + \frac{C_{10}}{n^{s(\alpha)}} + C_5 \left(\frac{\log n}{n}\right)^{\frac{\alpha}{1+\alpha}} + C_2 \left(\frac{\log n}{n}\right)^{\frac{\alpha}{1+\alpha}} \leq C(\log n/n)^{\frac{\alpha}{1+\alpha}}.$$

۴. مطالعه شبیه‌سازی

در این بخش، نتایج به دست آمده در بخش سوم، با استفاده از شبیه‌سازی مورد بررسی قرار گرفته و نتایج به دست آمده با نتایج حاصل از روش برآورد رگرسیون نیمه‌پارامتری هسته موجک با استفاده از معیارهای متوسط میانگین مربعات خطا و انحراف معیار مقایسه می‌شوند. همچنین از تابع سینوسی و تابع بلوکی به عنوان توابع هدف استفاده شده است. هر آزمایش شامل ۲۵۶، ۵۱۲ و ۱۰۲۴ مشاهده از تابع هدف بر اساس ۱۰۰ بار تکرار با شرایط تابع رگرسیون نیمه‌پارامتری است. در هر آزمایش متوسط میانگین مربعات خطا و مقدار انحراف معیار برآوردگرها محاسبه و نتایج در جدول ۱ آورده شده است. همان‌طور که گفته شد، شبیه‌سازی با استفاده از نرم‌افزار **R** و مقدار آستانه جهانی و تابع حذف سخت انجام گرفته است. برآورد تابع سینوسی و تابع بلوکی با استفاده از دو روش آستانه موجک و هسته موجک برای یک نمونه تصادفی به اندازه ۵۱۲ رسم و در شکل‌های ۱ و ۲ نشان داده شده‌اند. همان‌طور که در شکل‌های ۱ و ۲ ملاحظه می‌شود برآوردگر موجکی نسبت به برآوردگر هسته موجک به تابع هدف نزدیک‌تر است و همچنین برآوردگر موجکی، نقاط نوک‌تیز با پرش را نسبت به برآوردگر هسته موجک بهتر برآورد می‌کند.



شکل ۱: برآورد تابع سینوس با استفاده از دو روش آستانه موجک و هسته موجک. خط نقطه چین و خط پیوسته به ترتیب نشان دهنده برآوردگر آستانه و هسته موجک هستند. تابع هدف با خط بریده نشان داده شده است.



شکل ۲: برآورد تابع بلوکی با استفاده از دو روش آستانه موجک و هسته موجک. خط نقطه چین و خط پیوسته به ترتیب نشان دهنده برآوردگر آستانه و هسته موجک هستند. تابع هدف با خط بریده نشان داده شده است.

در جدول ۱ ملاحظه می شود که برآوردگر موجکی همواره مقدار متوسط میانگین مربعات خطای کمتری نسبت به برآوردگر هسته موجک دارد. همچنین با افزایش تعداد مشاهدات تابع نمونه گیری، دقت برآوردگر موجکی نسبت به برآوردگر هسته موجک بیشتر شده و میزان خطای برآوردگر موجکی با سرعت بیشتری به سمت صفر میل می کند. علاوه بر این، برآوردگر موجکی همواره خطای استاندارد کمتری نسبت به برآوردگر هسته موجک دارد.

جدول ۱: مقدار متوسط میانگین مربع خطا و انحراف استاندارد میانگین مربع خطاها برای توابع هدف بر اساس

$$AMSE = \frac{1}{100} \sum_{j=1}^{100} \sum_{i=1}^n \frac{(f(x_i) - \hat{f}(x_i^j))^2}{n} \cdot 100 \text{ بار تکرار.}$$

۱۰۲۴		۵۱۲		۲۵۶		n	تابع هدف
SD	AMSD	SD	AMSD	SD	AMSD	برآوردگر	
0/041	0/074	0/085	0/976	0/126	0/974	موجکی	تابع سینوس
0/058	0/971	0/093	0/985	0/128	1/03	هسته موجک	
0/018	0/176	0/026	0/182	0/043	0/188	موجکی	تابع بلوکی
0/021	0/190	0/029	0/190	0/046	0/217	هسته موجک	

۵. مثال کاربردی

به عنوان یک مثال کاربردی برای برازش دادن تابع رگرسیون نیمه پارامتری موجکی، داده‌های مربوط به رشد دندان که بر روی بسته *datasets* نرم‌افزار *R* قرار دارد، مورد بررسی قرار گرفته است. داده‌ها اثر ویتامین *C* را بر روی میزان رشد طول دندان‌ها نشان می‌دهند و شامل ۶۰ داده از این دسته هستند. اکنون با استفاده از تبدیل موجک گسسته روی بردار مشاهدات می‌خواهیم اطلاعات چند مقیاسی را از داده‌ها استخراج کنیم. این اطلاعات شامل جزییات (ضرایب موجک مادر) و کلیات (ضرایب موجک پدر) است. چون ضرایب موجک مادر و پدر از تفریق و مجموع زوج مشاهدات مجاور هم به دست می‌آیند، بنابراین در روش موجکی تعداد داده‌ها باید ضریبی از ۲ باشد. برای برطرف کردن این مشکل، یک نمونه تصادفی و بدون جایگذاری به اندازه ۴ از بین داده‌های موجود انتخاب و به داده‌های اولیه اضافه شده است. هدف برآورد آستانه موجک تابع رگرسیون نیمه پارامتری به ازای موجک‌های دوبیچز^{۲۴} و کویفلت^{۲۵} با در نظر گرفتن تعداد پالایش‌های^{۲۶} متفاوت است. در این راستا با استفاده از مقادیر میانگین مربعات خطا، تعداد پالایش مناسب برای هر کدام از موجک‌ها یافته می‌شود. مقادیر میانگین مربعات خطا به ازای تعداد پالایش‌های مختلف برای موجک‌های دوبیچز و کویفلت محاسبه و در جدول ۲ آورده شده‌اند. همان‌طور که در جدول مشخص است، در موجک کویفلت به ازای تعداد پالایش ۵ و در موجک دوبیچز به ازای تعداد پالایش ۱۰، مقدار کمترین میانگین مربعات خطا مشاهده شده است. لازم به ذکر است که در این مطالعه، در موجک دوبیچز به ازای مقادیر تعداد پالایش بیش از ۱۰ و در موجک کویفلت به ازای تعداد پالایش بیش از ۵، ضرایب موجکی بسیار کوچک و نزدیک به صفر است که در این حالت برآوردی صورت نخواهد گرفت. برای پی بردن به دلیل این موضوع به بخش (۲،۴) کتاب نسون (۲۰۱۱) مراجعه شود.

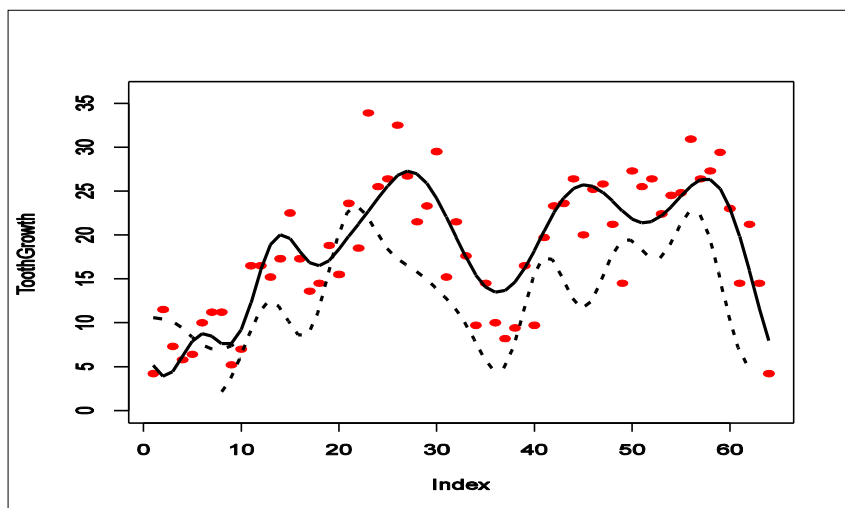
در ادامه برآورد آستانه موجک تابع رگرسیون نیمه پارامتری به ازای موجک‌های دوبیچز و کویفلت در حالتی که تعداد پالایش‌ها

²⁴ Daubechies

²⁵ Coiflets

²⁶ Filter number

بهینه است، رسم و در شکل ۳ نشان داده شده است. همان طور که در شکل ۳ ملاحظه می شود، تابع رگرسیون نیمه پارامتری موجکی مدلی مناسب در برازش دادن به داده های موجود است.



شکل ۳: برآورد آستانه موجک تابع رگرسیون نیمه پارامتری (خط ممتد برای موجک دوبیچز با تعداد پالایش ۱۰ و نقطه چین برای موجک کویفلت با تعداد پالایش ۵).

جدول ۲: میانگین مربعات خطای رگرسیون نیمه پارامتری برای موجک های دوبیچز و کویفلت

موجک		تعداد پالایش ها
دوبیچز	کویفلت	
24/70682	19/83758	۱
24/66351	21/85813	۲
26/07223	15/90004	۳
16/28669	19/36442	۴
16/06318	15/31501	۵
19/28729	—	۶
19/97472	—	۷
18/39431	—	۸
22/35943	—	۹
14/95963	—	۱۰

۶. بحث و نتیجه‌گیری

در این مقاله، برآوردگر آستانه موجک تابع رگرسیون نیمه‌پارامتری مورد بررسی قرار گرفت و نرخ همگرایی آن محاسبه شد. جهت ارزیابی برآوردگر پیشنهادی، از توابع سینوسی و بلوکی به عنوان توابع هدف استفاده و با بکارگیری شبیه‌سازی، نشان داده شد که متوسط میانگین مربعات خطاهای به دست آمده با روش آستانه موجک کمتر از متوسط میانگین مربعات خطاهای به دست آمده با روش هسته موجک است. همچنین بنا به خاصیت پایه‌های موجکی و مقیاسی، روش آستانه موجک نقاط تیز و جهشی را بهتر از سایر روش‌ها برآورد کرده و دارای نرخ همگرایی قابل قبول است. به عنوان یک مثال کاربردی، داده‌های مربوط به میزان رشد دندان در نظر گرفته شد و تابع رگرسیون نیمه‌پارامتری موجکی بر روی داده‌ها برازش داده شد.

قدردانی

نویسندگان مقاله از سردبیر محترم مجله پژوهش‌های ریاضی و داوران محترم بخاطر پیشنهادات ارزشمندشان که موجب تغییرات در راستای هر چه بهتر شدن مقاله شد، نهایت تشکر و قدردانی را دارند.

References

1. Abramovich, F. and Benjamini, Y., "Adaptive thresholding of wavelet coefficients", Computational Statistics and Data Analysis, 22 (1996) 351-361.
2. Afshari, M., Lak, F., and Gholizadeh, B., "A new Bayesian wavelet thresholding estimator of nonparametric regression", Journal of Applied Statistics, 44 (2017) 649-666.
3. Antoniadis, A., "Wavelet methods in statistics, some recent developments and their applications", Statistical Surveys, 1 (2007) 16-55.
4. Antoniadis, A., Bigot, A. J. and Sapatinas, T., "Wavelet estimators in nonparametric regression: a comparative simulation study", Journal of Statistical Software, 6 (2001) 1-83.
5. Cai, T. T. and Silverman, B. W., "Incorporating information on neighboring coefficients into wavelet estimation", Sankhya, Series A, 63 (2001) 127-148.
6. Chui, C. K., (1992). "An Introduction to Wavelets", Academic Press, Boston, MA.

7. Daubechies, I., Ten Lectures in Wevelets, CBMS-NSF regional conferences series in applied mathematics, SFAM, Philadelphia, (1992).
8. Donoho, D. L. and Johnstone, I. M., "Ideal spatial adaptation by wavelet shrinkage", *Biometrika*, 81 (1992) 425-455.
9. Donoho, D. L. and Johnstone, I. M., "Adapting to unknown smoothness via wavelet shrinkage", *Journal of the American Statistical Association*, 90 (1995) 1200-1224.
10. Doosti, H., Afshari, M. and Niroomand, H. A., "Wavelets for nonparametric stochastic regression with mixing stochastic process", *Communication in Statistics -Theory and Methods*, 37 (2008) 373-385.
11. Gannaz, I., "Robust estimation and wavelet thresholding in partially linear models", *Statistics and Computing*, 17 (2007) 293-310.
12. Gao, Q., Liu, L. and Racine, J. S., "A partially linear kernel estimator for categorical data", *Econometric Reviews*, 34(6-10) (2015) 959-978.
13. Hall, P. and Patil, P., "On the choice of smoothing parameter, threshold and truncation in nonparametric regression by non-linear wavelet methods", *Journal of the Royal Statistical Society, Series B*, (1996) 361-377.
14. He., H. and Severini, T. A., "A flexible approach to inference in semiparametric regression models with correlated errors using Gaussian processes ", *Computational Statistics & Data Analysis*, 103 (2016) 316-329.
15. Hu, H. and Wu, L., "Convergence rates of wavelet estimators in semiparametric regression models under NA samples", *Chinese Annals of Mathematics, Series B*, 33 (2012) 609-624.
16. Hu, H., "Ridge estimation of a semiparametric regression model", *Journal of Computational and Applied Mathematics*, 176(1) (2005) 215-222.

17. Ichimura, H., "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models", *Journal of Econometrics*, 58 (1993) 71-120 .
18. Johnstone, I. M. and Silverman, B. W., "Wavelet threshold estimators for data with correlated noise", *Journal of the Royal Statistical Society, Series B*, 59 (1997) 319-351.
19. Karamikabir, H. and Afshari, M., "Wavelet Shrinkage Generalized Bayes Estimation for Elliptical Distribution Parameter's under LINEX Loss", *International Journal of Wavelets, Multiresolution and Information Processing*, 17(3) (2019) 1950009.
20. Karamikabir, H. and Afshari, M., "Generalized Bayesian Shrinkage and Wavelet Estimation of Location Parameter for Spherically Distribution under Balance-type Loss: Minimavity and Admissibility", *Journal of Multivariate Analysis*, 177 (2020) 104583.
21. Karamikabir, H. and Afshari, M., "Wavelet threshold based on Stein's Unbiased Risk Estimators of Restricted Location Parameter in Multivariate Normal", *Journal of Applied Statistics*, (2020) *Published online*.
22. Li, Q. and Wu, Y., "On wavelet estimation in semiparametric regression models under convergence system error sequences", *International Conference on Wavelet Analysis and Pattern Recognition*, Beijing, China (2007).
23. Nason, G., (1996). Wavelet shrinkage using cross validation, *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(2) (1996) 463-479.
24. Nason, G., *Wavelet Methods in Statistic with R*, Springer-Verlag, New York (2011).
25. Niu, S. and Liu, Y., "CLT of wavelet estimator in semiparametric model with correlated errors", *Journal of Systems Science and Complexity*, 25 (2012) 567-581.
26. Ogden, R. T. and Parzen, E., "Data dependent wavelet thresholding in nonparametric regression with change-point applications", *Computational Statistics and Data Analysis*, 22 (1996) 53-70.

27. Porto, R., Morettin, P. and Aubin, E., "Wavelet regression with correlated errors on a piecewise Hölder class", *Statistics and Probability Letters*, 78 (2009) 2739-2743.
28. Robinson, P. M., "Root-n consistent semiparametric regression", *Econometrica*, 56 (1988) 931-954.
29. Roozbeh, M. and Babaie-Kafaki, S., "Extended least trimmed squares estimator in semiparametric regression models with correlated errors", *Journal of Statistical Computation and Simulation*, 86(2) (2015) 357-372.
30. Vidakovic, B., "Statistical Modeling By Wavelets", John Wiley and Sons, New York (2009).
31. Wasserman, L. and Genovese, R., "Confidence sets for nonparametric wavelet regression", *Institute of Mathematical Statistics*, 3 (2005) 698-729.
32. Xue, L. G. and Lin, Q., "Bootstrap approximation of wavelet estimates in a semiparametric regression model", *Acta Mathematica Sinica, English Series*, 26 (2010) 763-778.
33. Yalaz, S., "Wavelet estimation of semiparametric errors in variables model", *Communications Faculty of Sciences, University of Ankara, Mathematics and Statistics*, 68 (2018) 595-601.
34. Zhou, X. C. and Lin, J. G., "Asymptotic properties of wavelet estimators in semiparametric regression models under dependent errors", *Journal of Multivariate Analysis*, 122 (2013) 251-270.