ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2019-64979

Semiparametric Mixed Models for Longitudinal Data: Wavelets Analysis as Smoothing Approach

Longitudinal Veriler İçin Semiparametrik Karışık Modeller: Smoothing Yaklaşımı Olarak Dalgacık Analizi

Marwa BEN GHOUL[a],
Berna YAZICI[a],
Ahmet SEZER[a]

[a]Department of Statistics,
Eskişehir Technical University,
Faculty of Science,
Eskişehir, TURKEY

Received: 22.01.2019
Received in revised form: 18.02.2019
Accepted: 19.02.2019
Available Online: 19.04.2019

Correspondence:
Marwa BEN GHOUL
Eskişehir Technical University,
Faculty of Science,
Department of Statistics, Eskişehir,
TURKEY/TÜRKİYE
benghoulmarwa@gmail.com

ABSTRACT Objective: This paper consists on examining a longitudinal data, output of generated data, via constructing a semiparametric model, the wavelets analysis will be applied as smoothing approach for the nonparametric part of the model. **Material and Methods:** Mixed effects models have been largely examined due to their flexibility in handling data without constraints. Mixed models could be characterized with their parametric and nonparametric features. Indeed, semiparametric mixed models which combine parametric and nonparametric features, started to receive more attention notably for examining longitudinal data. Regarding the nonparametric features, smoothing approaches should be applied. Recently, the wavelets analysis has been considered as a powerful mathematic tool to decompose a series due to its multiresolution features (frequential and temporal) and some researchers mentioned it as a smoothing approach for large data, but the wavelets features for smoothing still not commonly applied on longitudinal data. **Results:** A data is generated referring to a previous published hypertension study by National Institute Health. The results show that the wavelets analysis has a strong capacity as smoothing approach compared to well-known other smoothing methods; Root Mean Square Errors are calculated, and via the constructed semiparametric model, it has been confirmed that the incident hypertension is related to the high Systolic blood pressure, high diastolic blood pressure and low BMI.

Keywords: Longitudinal data; smoothing approaches; wavelets decomposition; semiparametric model; mixed models

ÖZET Amaç: Bu makale, türetilmiş bir longitudinal veri seti için, bir yarı parametrik model kurularak incelenmesi üzerine olup, modelin parametrik olmayan kısmı için pürüzsüzleştirme yaklaşımı dikkate alınarak dalga analizi uygulanacaktır. **Gereç ve Yöntemler:** Karışık etki modelleri, verilerin kısıtlama olmadan kullanılmasındaki esneklikleri nedeniyle yaygın biçimde çalışılmaktadır. Karışık modeller parametrik ve parametrik olmayan özellikleri dikkate alınarak incelenebilir. Nitekim, parametrik ve parametrik olmayan özellikleri birleştiren yarı parametrik karışık modeller, özellikle longitudinal verilerin analizi için daha fazla çalışılmaya başlanmıştır. Parametrik olmayan özellikler ile ilgili olarak, pürüzsüzleştirme yaklaşımları uygulanmalıdır. Son zamanlarda, dalga analizi, çoklu çözünürlük özellikleri (sıklık ve zamansal) nedeniyle bir dizi ayrıştırmak için güçlü bir matematiksel araç olarak kabul edilmektedir, ancak dalga analizi hala longitudinal veri seti için yaygın bir biçimde kullanılmamaktadır. **Bulgular:** National Institute Health tarafından daha önce yayınlanmış bir hipertansiyon çalışmasına atıfta bulunan bir veri üretilmiştir. Sonuçlar, bilinen diğer pürüzsüzleştirme yöntemlerine kıyasla dalga analizinin, pürüzsüzleştirme yaklaşımı olarak güçlü bir yaklaşım olduğunu göstermektedir; Ortalama Kare Kök Hatalar hesaplanmış ve oluşturulan semiparametrik model aracılığıyla, hipertansiyonunun yüksek Sistolik kan basıncı, yüksek diyastolik kan basıncı ve düşük BMI ile ilişkili olduğu doğrulanmıştır.

Anahtar Kelimeler: Longitudinal veri; pürüzsüzleştirme yaklaşımları; dalga ayrışması; yarı parametrik model; karma modeller

In the last decades, mixed models have gained a lot of attention in the statistical researches over the traditional analyses due to their flexibility to handle multilevel or clustered data without constraints (as equal number of observations or non-missing observations). Also, mixed models allow a variety of choices for modeling correlations in the data. Mixed models are current in the design which combines random and fixed effects and widely used in several fields such as pharmaceutical industry and economics.[1] Mixed effects models could be characterized as parametric and nonparametric; the parametric mixed models may take different covariates into account, but they require parametric assumptions, the nonparametric mixed models are flexible to fit longitudinal data and robust against model misspecification, but they involve few numbers of covariates and may be computationally intensive. Recently, semiparametric mixed models which combine parametric and nonparametric features, started to receive more attention notably for longitudinal data but still under development.[2-7]

For nonparametric features, smoothing approaches have been applied, the most famous ones are kernel approaches, regression polynomial splines, smoothing splines, penalized splines or p-splines and the wavelets technique.[5-11]

Longitudinal data is the data which tracks the same type of information for the same subjects at different timepoints. It is very popular in the pharmaceutical industry. The most common issue for this type of data is the attrition (loss of follow-up, subject discontinuation etc.).[12-15]

Historically, since the 18th century, Fourier transform has been known as an important mathematical tool used in the spectral analysis. Nevertheless, a major flaw was identified in this tool; time information is missing. Indeed, the Fourier transform gives the information about the number of frequencies contained in the signal but hides the times of the diffusion of these frequencies. The wavelet was the alternative approach that breaks down a signal both in time and in frequency. For a musical note, wavelet is the equivalent of a score signal which provide which frequencies (notes) should be played and what time should those notes be played.[16]

Thus, wavelets have increasingly become a popular tool in various fields such as image processing (digital borrowing, medical X-rays, seismic waves etc.), audio processing (voices, musical notes, etc.) and recently in the economic and financial areas.[17] However, even though the wavelets have been considered as a very powerful mathematical tool, they are not commonly used in the longitudinal data.

To sum up, the objective of this research is to construct a semiparametric model after considering the wavelets decomposition as smoothing technic for longitudinal data.

The remainder of the paper is organized as follows. Second section is devoted to the material and methods and it was subdivided into two sections: the semiparametric model and the theoretical framework exploring the smoothing methods notably the wavelets analysis. Then the dataset used in this paper will be defined. Section three outlines the findings and the results of this research. The final section summarizes this research.

MATERIAL AND METHODS

This section is devoted to the theoretical framework providing a general overview about the semiparametric models, the smoothing approaches and the wavelets' analysis.

SEMIPARAMETRIC MODEL

Some researchers showed, a semiparametric mixed-effects model, as a part of the variations in the response variable can be explained by given parametric models of some covariates in the fixed effect component and/or the random-effect component, while the remaining is explained by a nonparametric function of time.[18] In a time, varying coefficient mixed-effects model, the coefficients of the fixed-effects and random-effects covariates are smooth functions of time. These two kinds of models are very important and useful in practical longitudinal data analysis. Semiparametric regression is concerned with the flexible incorporation of nonlinear functional relationships in regression analyses and any application area that uses regression analysis can benefit from semiparametric regression.[19]

Therefore, a SPMEF (Semi Parametric Mixed Effects Model) for longitudinal data will be constructed, after the data generation. The following equation presents the general format of the semi-parametric model.[18]

$$\mathbf{y_{ij}} = \mathbf{x_{ij}^T}\boldsymbol{\beta}(\mathbf{t_{ij}}) + \boldsymbol{\eta}(\mathbf{t_{ij}}) + \mathbf{z_{ij}^T b_i} + \mathbf{v_i}(\mathbf{t_{ij}}) + \boldsymbol{\varepsilon}_{\mathbf{ij}}(\mathbf{t_{ij}}) \quad \textbf{Equation 1}$$

β is the parametric fixed-effects (population parameter), b_i is the parametric random-effects (individual parameters), $v_i(t)$ is the nonparametric random-effect function, $\eta(t)$ is the nonparametric fixed-effect function. $\varepsilon_{ij}(t)$ are the measurement errors, x_{ij}^T and z_{ij}^T are the associated fixed-effects and the random-effects design matrices, n presents the number of subjects, ni presents the number of observations for the i-th subject.

$v_i(t) \sim GP(0, \gamma)$; $b_i \sim N(0, D_b)$ and $\varepsilon_{ij} \sim GP(0, \gamma_0)$ (GP : Gaussian process).

Equation 1. will be followed to construct the semiparametric model for the generated data in the next sections.

WAVELETS AS SMOOTHING APPROACH

Parametric modeling has restrictions which may not be satisfied by the non-linear models. To overcome this issue, smoothing methods have become largely applied in data analysis and in inferential analysis. The next paragraph describes the basic idea behind the smoothing technics.

The basic idea of these nonparametric approaches is to let the data determine the most suitable form of the functions. There are one or two so-called smoothing parameters in each of these methods for controlling the model complexity and the trade-off between the bias and variance of the estimator. For example, the bandwidth h in local kernel smoothing determines the smoothness of the regression function and the goodness-of-fit of the model to the data so that when $h = \infty$, the local nonparametric model becomes a global parametric model; whereas when $h = 0$, the resulting estimate essentially interpolates the data points. Thus, the boundary between parametric and nonparametric modelling may not be clear-cut if one takes the smoothing parameter into account.[4]

Indeed, kernel smoothing, local polynomial fitting, polynomial regression splines, smoothing splines, penalized splines, locally weighted scatter plot smoothing (LOWESS), wavelet-based methods and other orthogonal series-based approaches could be considered as the frequently used smoothing approaches.[4]

Four major smoothing approaches to nonparametric modeling could be determined: smoothing splines; series-based smoothers, including wavelets; kernel methods, including local regression; and regression splines.[19-30] The type of data is the reference to choose among those approaches.[19] In addition, the penalized splines, labelled also as P-splines, pseudosplines, and low-rank spline smoothers have been largely discussed in the literature.[19]

The research related to nonparametric methods applied on longitudinal data had not been largely investigated until 1988 with Müller.[15] Indeed, Müller's approach consisted on estimating distinctly each individual curve, consequently the within-subject correlation of the longitudinal data was not examined in the modelling which make those methodologies essentially like the nonparametric regression methods for cross-sectional data.[4]

Recently, the development of nonparametric regression methods for longitudinal data analysis including utilization of kernel-type smoothing methods have been largely exposed, smoothing spline methods and regression (polynomial) spline methods.[31-40]

Wavelets analysis have become more used for the nonparametric regression and as a smoothing approach. Wavelets analysis was applied with S to carry out nonlinear regression and image compression, this application emphasized the use of thresholding of wavelet coefficients as method for attempting to extract signal from noise.[41] A new threshold algorithm was created basically on the wavelet analysis to smooth noise for nonlinear time series, by detailing the signals decomposed onto different scales, the details were smoothed by using the updated thresholds to different characters of a noisy nonlinear signal. This method is an improvement of Donoho's wavelet methods to nonlinear signals.[42]

Also, a wavelet shrinkage was proposed for signal smooth by using a generated data.[41,43-45]

However, the application of the wavelets in the pharmaceutical industry for the nonparametric models as smoothing method still not highly covered.

WAVELETS AND MULTIRESOLUTION DECOMPOSITION

Fourier transform was used in the extraction of the frequential information of a variable of interest (for example the length of an economic cycle) without the identification of the temporal information (which is fixed). To overcome the issue of the fixed temporal resolution, the concept of wavelets was developed.[46]

Wavelets have been known as an extension of Fourier analysis, developed and extended mainly by Grossman, Morlet, Mallat, Daubechies and Meyer.[16,47-49]

A wavelet is a small wave, the term small mainly explains in this context that the wave increases and decreases in a limited time support. The main feature of a wavelet is the compact support i.e. the wavelet function is limited in time and frequency domains.

Theoretically, a wavelet is simply a time function that follows a basic rule, known as the wavelet eligibility condition: $\boldsymbol{C_\Psi = \int_0^\infty \frac{|\Psi(f)|}{f}\, df} < \infty$ with $\boldsymbol{\Psi(f) = \int_{-\infty}^{+\infty} \psi(t) e^{-2\pi i f t}\, dt}$, Fourier transform.

f is a frequency function for $\Psi(f)$, ψ is called mother wavelet or analyzing wavelet.

This condition ensure that the mother wavelet $\Psi(f)$ quickly tends to 0 if $f \to \infty$.[50]

To ensure that $C_\Psi < \infty$, the following conditions, related to the mother wavelet, must be imposed:

a. $\psi(0) = 0$, or $\int_{-\infty}^{\infty} \psi(t)dt = 0$

b. $\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$, representing the energy unit.

The Discrete wavelet transforms (DWT) use low pass and high pass digital filters in cascade. Indeed, at each level of decomposition, the output signal of the low pass is again filtered and separated into two frequency bands which is sub-sampled, retaining only one sample out of two. Thus, the number of coefficients is constant from one stage to another. The bands denoted by d_i are said to be detailed since they comprise the elements with higher frequency content and the bands f_i are called the approximations.

The equations summarized in Table 1. show the decomposition from the discrete signal f_i at the level i into a high frequency part presented by d_{i+1} (details) and a low frequency part explained by f_{i+1} (approximation), via the numeric filters low pass and high pass.

DWT is based on two so-called discrete wavelets; mother wavelet $\psi_l = (\psi_0, \cdots, \psi_{l-1})$ and father wavelet $\Phi_l = (\Phi_0, \dots, \Phi_{l-1})$. The mother wavelet is characterized by three properties:

First the condition showing that the mother wavelet also called differentiation function is a high pass filter $\sum_{l=0}^{L-1} \psi_l = 0$. Then, $\sum_{l=0}^{L-1} {\psi_l}^2 = 1$ and the last propriety $\sum_{l=0}^{L-1} \psi_l \, \psi_{l+2n} = 0$ for each integer $n \neq 0$.

The three properties ensure that the mother wavelet preserves the variance of the original data and a multi-resolution analysis can take place. The scale wavelet or the father wavelet aims to capture the long scale i.e. the low frequencies, it must respect this property: $\sum_{l=0}^{L-1} \Phi_l = 0$.

The application of the mother and father wavelet allows the separation between the low and high frequencies. In addition, a bandpass can be constructed from a successive recursive application low pass and high pass filters. Mathematically, the table summarizes the above section:[51-53]

TABLE 1: Signal constructed by successive refinement.

Signal constructed by successive refinement Approximation + detail	
Father $\Phi(t)$	Mother $\psi(t)$
$\phi_{j,k}(t) = 2^{-\frac{j}{2}}\phi(2^{-j}t - k)$ Approximation coefficients at scale j $a(j,k) = \int_{-\infty}^{\infty} x(t)\,\phi_{j,k}(t)dt$	$\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t - k)$ Detail coefficient at scale j $W(a,b) \cong \int x(t)\psi_{j,k}^{*}(t)dt = d_{j,k}$
Approximation +	Detail
$f_j(t) = x_j(t) = \sum_{k=-\infty} a(j,k)\phi_{j,k}(t)$	$g_j(t) = \sum_{-\infty} d_{j,k}\,\psi_{j,k}(t)$

The approximation coefficients presented in the table 1. will be used as smoothing coefficients in this paper.

DATASET

The applied data in the empirical section of this research is a generated data. The parameters used to generate the longitudinal data are based on a previous study published by (NIH, 2018). Data generation, referring previous clinical studies, was based on previous researches.[54,55]

The Framingham Heart Study was published by the National Institutes of Health (NIH). It is a long-term cohort study of the etiology of cardiovascular disease among subjects from the community of Framingham, Massachusetts.

For this paper, the Incident Hypertension was selected as outcome of interest. To get the predictive variables of the model, most of the defined variables in the study were selected. After a mixed model construction, the following list of variables had significant relationship with the incident hypertension; age, Systolic Blood Pressure (SYSBP), Diastolic Blood Pressure (DIABP) and Body Mass Index (BMI). The following Table 2 resumes the variables definition:

TABLE 2: Variables definitions.

Variable	Abbreviation	Description
Incident Hypertension	HYPERTEN	Hypertensive. Defined as the first exam treated for high blood pressure or second exam in which either Systolic is $\geq$ 140 mmHg or Diastolic $\geq$ 90mmHg.
Age	Age	Age (years) at examination.
Systolic BP	SYSBP	Systolic Blood Pressure (mean of last two of three measurements) (mmHg). Range of count: 83.5-295.
Diastolic BP	DIABP	Diastolic Blood Pressure (mean of last two of three measurements) (mmHg) Range of count: 30-150
Body Mass Index	BMI	Weight in kilograms/height meters squared

Source: NIH institute

To generate the data, a previous proposed Statistical Analysis Software (SAS) macro for sample size calculation based on a previous clinical trial was updated to feat the study case of this research, 100 subjects were considered, each subject was generated 10 times (each subject has 10 timepoints).[54]

First step of SAS macro consists on defining the coefficients of the variables.

```
/*Define the coefficients*/
proc glimmix data=hypert; /*generalized mixed model*/
model HYPERTEN = BPMEDS PERIOD SEX TOTCHOL AGE SYSBP DIABP CURSMOKE
CIGPDAY BMI DIABETES HDLC LDLC BPMEDS*PERIOD/s solution;
random PERIOD;
/* The OUTPUT statement saves the predictions of the mean of each observations to the data set gmxout.*/
output out=gmxout pred(blup)=pred;
nloptions tech=newrap;
run;
```

Second step of SAS macro is to use the output coefficients from the above glimmix in data generation. The following program is an extraction of the full SAS macro.

```
/*Estimation coefficients*/
%let beta0=-1.1; %let beta1=0.0037; %let beta2=0.0081; %let beta3=0.0041; %let beta4=0.007;
%let n = 100;
%let age_av= 55;
%let age_max = 81;
%let age_min = 32;
%let age_sd = 9.564;
%macro sim(iterations);
%do i = 1 %to &iterations;
Data hypert;
do id = 1 to &n;
b0    = &sigma0.*rannor(6427);
b2    = &sigma2.*rannor(1234);
r0    = &delta0.*rannor(4875);
Age   = &age_av. + &age_sd.*rannor(4875);
SYSBP = &SBP_av. + &SBP_sd.*rannor(1245);
DIAPB = &dbp_av.+ &dbp_sd.*rannor(5484);
BMI   = &BMI_av.+ &BMI_sd.*rannor(67848);
do CYCLE = 0 to 10;
y = (&beta0. + b0) + &beta1.*age + &beta2.*SYSBP+ &beta3.*SYSBP +&beta4.*DIAPB + &beta5.*BMI ;
output;
end;
end;
run;
%end;
%mend;
Options macrogen symbolgen mlogic mprint mfile;
%sim(10000);
```

After running the SIM SAS macro, the generated data will be used to construct the model. The next section will provide the results.

RESULTS

This section presents the empirical findings of this research.

Before constructing the semiparametric model, the scatter plots of the variables were graphed to examine which variable will be considered in the nonparametric part of the model.

Two variables will be used in the nonparametric part of the model, Systolic BP and Diastolic BP from the generated data of incident hypertension.

Figure 1 and Figure 2 show the wavelet decomposition of the variables Systolic BP and Diastolic BP. The wavelets analysis allows the decomposition of the series into details (to extract the detailed information of the variable) and approximation (to see the tendency of the variable). Through the wavelet's decomposition, the coefficients of the details and the approximation could be extracted.

The approximation coefficients after the wavelet's decomposition will be used as smoothed variable.

Table 3 compares the Root Mean Square Error (RMSE) associated to the smoothing approaches applied for this paper. RMSE was calculated as following: RMSE = sqrt (mean ((Raw variable value – smoothed variable value) ^2)). MATLAB software was used in this section.

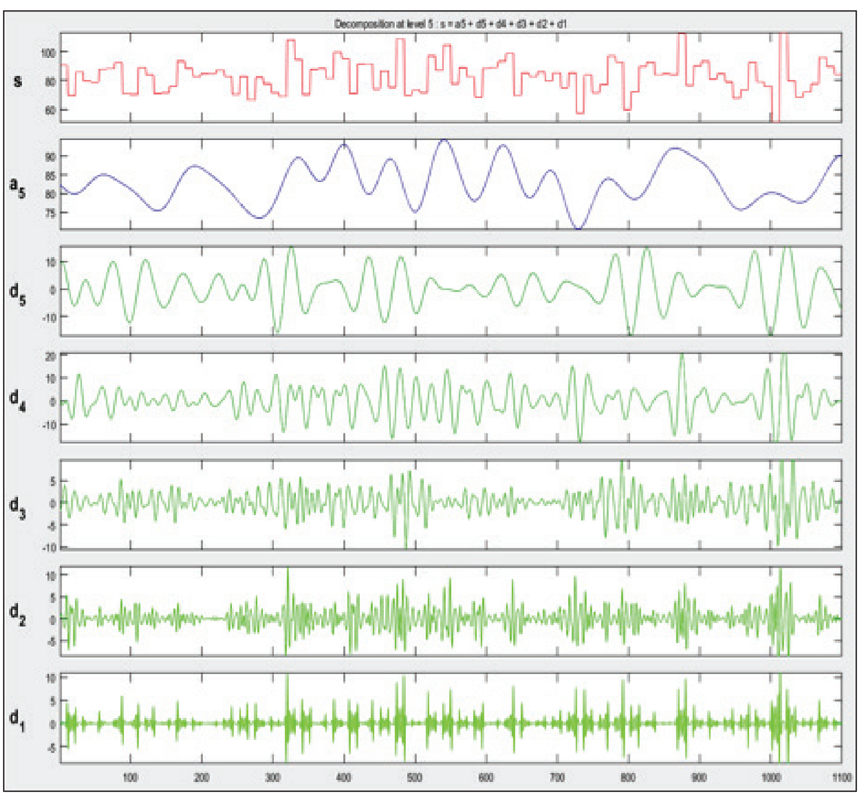

FIGURE 1: Wavelets decomposition of Diastolic BP.

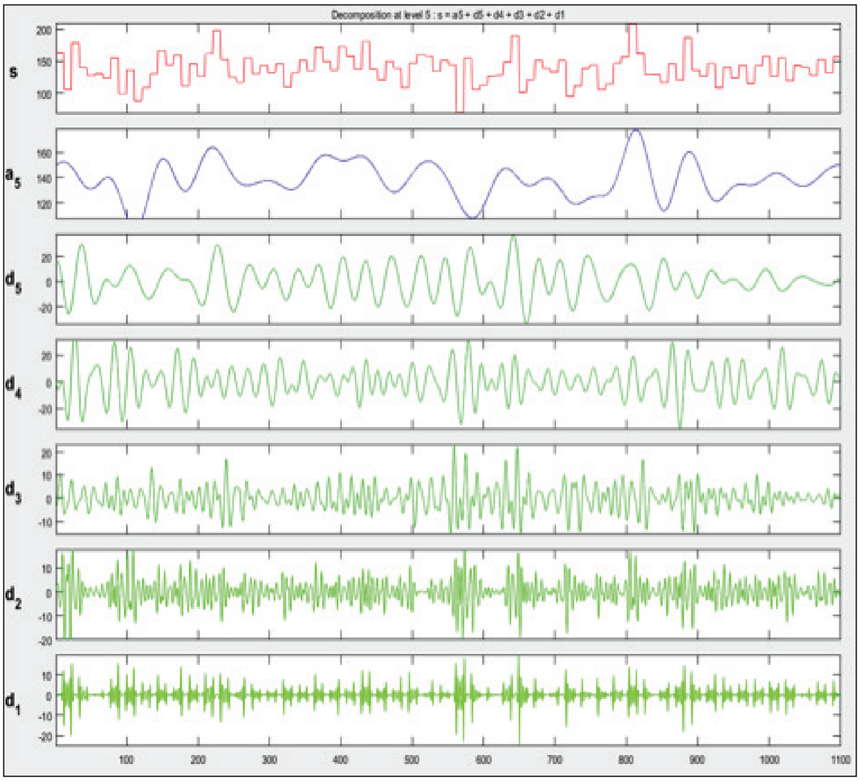

FIGURE 2: Wavelets decomposition of Systolic BP.
Source: Own elaboration

TABLE 3: Comparison of RMSE between smoothing approaches.

Smoothing approach*	RMSE (DIABP)	RMSE (SYSBP)
Wavelets	1.9583	3.9148
movmedian	2.4942	0
movmean	4.5991	9.4958
loess	2.5075	5.0061
lowess	3.4718	6.9313
rloess	1.8979	4.2089
rlowess	3.2293	6.7460
sgolay	3.1262	6.4511
gaussian	3.2955	6.9424
Smoothing spline	3.049	5.2122

*'moving': Moving average (default). A lowpass filter with filter coefficients equal to the reciprocal of the span.
'lowess': Local regression using weighted linear least squares and a 1st degree polynomial model
'loess': Local regression using weighted linear least squares and a 2nd degree polynomial model
'sgolay': Savitzky-Golay filter. A generalized moving average with filter coefficients determined by an unweighted linear least-squares regression and a polynomial model of specified degree (default is 2). The method can accept nonuniform predictor data.
'rlowess': A robust version of 'lowess' that assigns lower weight to outliers in the regression. The method assigns zero weight to data outside six mean absolute deviations.
'rloess': A robust version of 'loess' that assigns lower weight to outliers in the regression. The method assigns zero weight to data outside six mean absolute deviations. (MATLAB website)

Source: Own elaboration

As shown in Table 3, the wavelets decomposition has the least RMSE for both variables Systolic BP and Diastolic BP except the movmedian for the case of SYSBP and rloess for the case of DIABP (no notable difference). Indeed, the wavelets approach showed interesting performance for both variables, therefore the coefficients of the approximation results of the wavelets decomposition will be used in the modelling.

A semiparametric model will be performed in the modeling section. For the nonparametric part, the wavelets analysis was applied as smoothing approach. The cycle (timepoints) is included in the parametric part as random effect variable and the variables age and BMI are included in the parametric part as variables with fixed effect.

A generalized mixed model was used to estimate the coefficients, the estimation method is pseudo-likelihood technic.

The results of the semiparametric model are presented in the following Table 4:

TABLE 4: Semiparametric modelling.

Incident Hypertension as dependent variable		
Term	**Coefficient**	**P-Value**
Intercept	-2.2237	<.0001
Age	0.004583	0.0888
BMI	-0.05498	<.0001
Diastolic BP	0.01804	<.0001
Systolic BP	0.01306	<.0001

Source: Own elaboration

The model results confirm that high systolic blood pressure, high diastolic blood pressure and low BMI cause the incident hypertension. However, the age has not a significant impact on the incident hypertension.

The normality of residuals and the homoscedasticity of the residuals were rechecked and proved after the model construction (Figure 3).

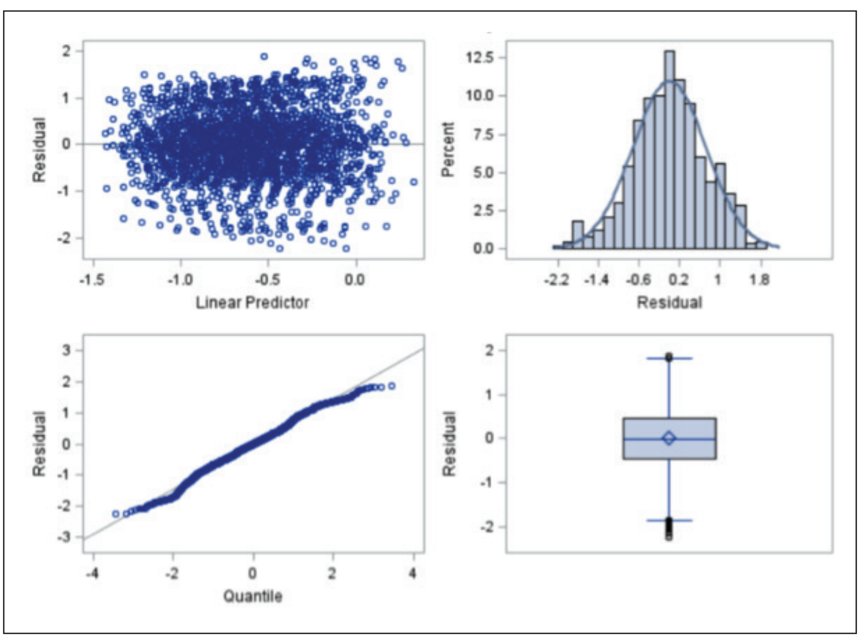

FIGURE 3: Residuals of Y.
Source: Own elaboration

DISCUSSION

Longitudinal data have been largely discussed and used in several researches notably in the pharmaceutical industry. Modelling this type of data still challenging due to the attrition issue and the violation of some parametric hypothesis. Many researches constructed semiparametric models to analyze longitudinal data, as this type of model combine the parametric and the nonparametric features. Smoothing methods are required in the nonparametric part of the model, different smoothing approaches have been found as performant such as splines, Pseudo spline and kernel. In the literature, one method was noted as smoothing approach but not highly examined, which is the wavelets. This research paper consists on figuring out the performance of the wavelets analysis as smoothing method. Indeed, this mathematical tool was largely used in imaging analysis, and signal treatments and started to gain more attention in the financial and economic fields, due to its ability to denoise a signal (or a series), to decompose it and reconstruct it again without loosing information. Longitudinal data provided by the NIH was used to get the parameters of a model to explore the relationship between incident hypertension and variables such as age, BMI, diastolic blood pressure and systolic pressure. Those parameters were used to generate a data which will be analyzed via a semiparametric model. After the data generating, systolic blood pressure and diastolic blood pressure were found to be in the nonparametric part of the model due to the normality assumption violation. Matlab was used to apply a smoothing approach for both variables. Here, different smoothing approaches were applied and RMSE was calculated to choose the most performant approach. Although, wavelets were ranked as the second approach having RMSE, but the results were quite interesting for both of variables. So, wavelets method was picked in this analysis. SAS software was used to generate longitudinal data and then to construct the semiparametric model. The modelling results showed that high systolic blood pressure, high diastolic blood pressure and low BMI cause the incident hypertension.

CONCLUSION

Recently, Wavelets analysis has been considered as a powerful mathematical tool to decompose time series and to avoid the non-stationarity problem. This research consists on applying this tool as smoothing approach for longitudinal data. A generated data, based on a previous clinical study, was used for the application section.

The results of wavelets decomposition demonstrated a strong capacity of this method, for this application case, vis a vis the other smoothing approaches notably the splines method. Therefore, the wavelets analysis as smoothing method was used in the nonparametric part of the semi parametric model. The advantage of the semiparametric model is its capacity to include nonparametric effects and parametric effects, however due to the modeling complexity this type of model still not highly covered. The results confirmed that the incident hypertension is highly related to the systolic blood pressure, diastolic blood pressure and the BMI.

This research attempted to cover two complex pillars (wavelets analysis and semiparametric models) and highlight their capacity which can be more applied to analyze longitudinal data. This paper is the start of further applications as the wavelets' analysis can be applied to handle outliers before the modeling and this is will be the objective of future work.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

***Idea/Concept:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer; **Design:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer; **Control/Supervision:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer; **Data Collection and/or Processing:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer; **Analysis and/or Interpretation:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer; **Literature Review:** Marwa BEN GHOUL, Berna YAZICIa, Ahmet Sezer; **Writing The Article:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer; **Critical Review:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer; **References and Fundings:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer; **Materials:** Marwa BEN GHOUL, Berna YAZICI, Ahmet Sezer*

REFERENCES

1. Maxwell SE, Delaney HD, Kelley K. Designing Experiments and Analyzing Data A Model Comparison Perspective. 3rd ed. New York: Routledge; 2018. p.879-81.
2. Ruppert D, Wand MP, Carroll RJ. Semiparametric Regression. 1st ed. New York: Cambridge University Press; 2003. [Crossref]
3. Szczesniak RD, Li D, Raouf SA. Semiparametric mixed models for medical monitoring data: an overview. J Biom Biostat. 2015;6(2). https://doi.org/10.4172/2155-6180.1000234. [Crossref] [PubMed] [PMC]
4. Hulin W, Jin-Ting Z. Semiparametric models. Nonparametric Regression Methods for Longitudinal Data Analysis. Mixed-effects Modeling Approaches. 1st ed. New Jersey: WILEY Series in Probability and Statistics; 2007. p.229-70.
5. Ngo L, Wand MP. Smoothing with mixed model software. J Statist Softw. 2004;9(1):1-54. https://doi.org/10.18637/jss.v009.i01. [Crossref]
6. Szczesniak RD, McPhail GL, Duan LL, Macaluso M, Amin RS, Clancy JP. A semiparametric approach to estimate rapid lung function decline in cystic fibrosis. Ann Epidemiol. 2013;23(12):771-7. https://doi.org/10.1016/j.annepidem.2013.08.009. [Crossref]
7. Zhang D, Lin X. Hypothesis testing in semiparametric additive mixed models. Biostatistics. 2003;4(1):57-74. https://doi.org/10.1093/biostatistics/4.1.57. [Crossref]
8. Cleveland W. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979;74(368):829-36. [Crossref]
9. Wu CO, Chiang CT. Kernel smoothing on varying coefficient models with longitudinal dependent variable. Stat Sin. 2000;10:433-56.
10. Verbyla AP, Cullis BR, Kenward MG, Welham SJ. The analysis of designed experiments and longitudinal data by using smoothing splines. Journal of the Royal Statistical Society Series C-Applied Statistics. 1999;48:269-300. https://doi.org/10.1111/1467-9876.00154. [Crossref]
11. Wand MP, Jones MC. Univariate kernel density estimation. Kernel Smoothing. Chapter 2. 1st ed. New York: Chapman & Hall/CRC; 1995. p.10-4.
12. Diggle PJ, Heagerty PJ, Liang K, Zeger SL. Analysis of Longitudinal Data. 2nd ed. United Kingdom: Oxford University Press; 2013.
13. Fitzmaurice GM, Laird N, Ware HJ. Applied Longitudinal Analysis. 1st ed. New Jersey: John Wiley & Sons; 2004.
14. Hedeker D, Gibbons R. Longitudinal Data Analysis. 1st ed. New Jersey: Wiley-Blackwell; 2006.
15. Müller HG. Nonparametric Regression Analysis of Longitudinal Data. 1st ed. Berlin Heidelberg New York, London, Paris, Tokyo: Springer-Verlag; 1988. [Crossref]
16. Quarta L. Une Introduction (Elémentaire) à La Théorie Des Ondelettes. 2nd ed. Belgium: Mons-Hainaut University; 2001. p.4-18.
17. Gençay R, Selçuk F, Whitcher B. An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. Chapters 3, 4 & 5. 1st ed. California; Elsevier, Academic Press; 2002. p.96-234. [Crossref]
18. Jin-ting Z. Nonparametric Mixed-Effects Models for Longitudinal Data. South Korea: University of Seoul; 2007.
19. Ruppert D, Wand MP, Carroll RJ. Semiparametric Regression. 1st ed. New York: Cambridge University Press; 2003. https://doi.org/10.1017/CBO9780511755453. [Crossref]
20. Lee TCM. Smoothing parameter selection for smoothing splines: a simulation study. Computational Statistics & Data Analysis. 2002;42:139-48. https://doi.org/10.1016/S0167-9473(02)00159-7. [Crossref]
21. Eubank RL. Nonparametric Regression and Spline Smoothing. 2nd ed. New York: Marcel Dekker Inc.; 1999.
22. Wahba G. Estimating the smoothing parameter. Spline Models for Observational Data. 1st ed. Philadelphia: CBMS-NSF Regional Conference Series in Applied Mathematics; 1990. p.45-65. https://doi.org/10.1137/1.9781611970128. [Crossref]
23. Green PJ, Silverman BW. Nonparametric Regression and Generalized Linear Models. 1st ed. London: Chapman & Hall/CRC; 1994. [Crossref]
24. Tarter ME, Lock MD. Model-free Curve Estimation. 1st ed. New York, Chapman & Hall/CRC; 1993.
25. Ogden RT. Essential Wavelets for Statistical Applications and Data Analysis. 1st ed. Boston, Birkhäuser; 1997. [Crossref]
26. Ramsay JO. Kernel smoothing approaches to nonparametric item characteristic curve estimation. Psychometrika. 1991;56(4):611-30. http://doi.org/10.1007/BF02294494. [Crossref]
27. Fan J, Gijbels I. Local Polynomial Modelling and Its Applications. 1st ed. London: Chapman & Hall/CRC; 1996.
28. Friedman JH. Multivariate adaptive regression splines (with discussion). Ann Statist. 1991;19(1):1-67. http://doi.org/10.1214/aos/1176347969. [Crossref]
29. Stone CJ, Hansen MH, Kooperberg C, Truong YK. Polynomial splines and their tensor products in extended linear modeling. Ann Statist. 1997;25:1371-425. [Crossref]
30. Hansen MH, Kooperberg C. Spline adaptation in extended linear models (with discussion). Statist Sci. 2002;17(1):2-51. [Crossref]

31. Fan J, Zhang JT. Two-step estimation of functional linear models with applications to longitudinal data. Journal of Royal Statistical Society, Series B 2000;62:303-22. 10. http://doi.org/1111/1467-9868.00233. [Crossref]
32. Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika. 1998;85:809-22. http://doi.org/10.1093/biomet/85.4.809. [Crossref]
33. Wu H, Zhang JT. The study of long-term HIV dynamics using semiparametric nonlinear mixed-effects models. Statistics in Medicine. 2002a;21:3655-75. http://doi.org/10.1002/sim.1317. [Crossref]
34. Wang N, Carroll RJ, Lin X. Efficient semiparametric marginal estimation for longitudinal /clustered data. Journal of American Statistical Association; 2005; 100:147-57. https://doi.org/10.1198/016214504000000629. [Crossref]
35. Brumback B, Rice J. Smoothing spline models for the analysis of nested and crossed samples of curves. Journal of American Statistical Association. 1998;93:961-94. https://doi.org/10.1080/01621459.1998.10473755. [Crossref]
36. Wang Y. Mixed-effects smoothing spline ANOVA. Journal of Royal Statistical Society Series B. 1998;93:341-8.
37. Wang Y. Smoothing spline models with correlated random errors. Journal of American Statistical Association. 1998;93:341-8. [Crossref]
38. Liang H, Wu H, Carroll RJ. The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient semiparametric models with measurement error. Biostatistics. 2003;4:297-312. 10. https://doi.org/1093/biostatistics/4.2.297. [Crossref]
39. Rice JA, Wu CO. Nonparametric mixed effects models for unequally sampled noisy curves. Biometrika. 2001;57:253-9. https://doi.org/10.1111/j.0006-341X.2001.00253.x. [Crossref]
40. Wu H, Zhang JT. Local polynomial mixed-effects models for longitudinal data. Journal of American Statistical Association. 2002;97(459):883-97. https://doi.org/10.1198/016214502388618672. [Crossref]
41. Nason GP, Silverman BW. The discrete wavelet transform in S. J Comput Graph Stat. 1994;3(2):163-91. https://doi.org/10.1198/1061860031671. [Crossref]
42. Han M, Liu Y, Xi J. Guo W. Noise smoothing for nonlinear time series using wavelet soft threshold. IEEE Signal Process Lett. 2007;14(1):62-5. https://doi.org/10.1109/LSP.2006.881518. [Crossref]
43. Horgan GW. Using wavelets for data smoothing: a simulation study. J Appl Stat. 1999;26(8):923-32. https://doi.org/10.1080/02664769921936. [Crossref]
44. Schimek M. Smoothing and Regression Approaches, Computation, and Application. 1st ed. USA: John Wiley & Sons, Inc.; 2000. https://doi.org/10.1002/9781118150658. [Crossref] [PMC]
45. Nason GP, Silverman BW. The stationary wavelet transform and some statistical applications. Wavelets and Statistics. 1995;281-300. [Crossref]
46. Masset P. Analysis of Financial Time-Series Using Fourier and Wavelet Methods. SSRN. 2008 ;1-36. https://doi.org/10.2139/ssrn.1289420. [Crossref]
47. Polikar R. The wavelet tutoriel. Index to Series of Tutoriels to Wavelet Transform. 2nd ed. New Jersey: Rowan University; 2006. p.2-15.
48. Conway P, Frame D. A spectral analysis of New Zealand output gaps using Fourier and wavelet techniques. Discussion Paper Series. 2000;1-20.
49. Debdas S, Quereshi MF, Reddy A, Chandrakar D, Pansari D. A Wavelet based multiresolution analysis for real time condition monitoring of AC machine using vibration analysis. International Journal of Scientific & Engineering Research 2011;2(10).
50. Grossmann A, Morlet J. Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM Journal of Mathematical Analysis and Applications. 1984;15(4):723-36. https://doi.org/10.1137/0515056. [Crossref]
51. Andreopoulos Y, Schaar MVD. Incremental Refinement of Computation for the Discrete Wavelet Transform IEEE Transactions on Signal Processing. 2008;56(1):140-57. https://doi.org/10.1109/TSP.2007.906727. [Crossref]
52. Equitz WHR, Cover TM. Successive Refinement of Information. IEEE Transactions on information theory. 1991;37(2). https://doi.org/10.1109/ITW.1989.761420. [Crossref]
53. Graps A. An Introduction to Wavelets. IEEE Computational Science and Engineering. 1995;2(2):50-61. https://doi.org/10.1109/99.388960. [Crossref]
54. Chen J, Stock S, Deng C. Sample size estimation through simulation of a random coefficient model by using SAS abstract. PharmaSUG. 2008;3.
55. Psioda M. Random Effects Simulation for Sample Size Calculations Using SAS. 1st ed. Carolina: University of North Carolina, Chapel Hill NC; 2012. p.1-11.