IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (13) المجلد (1): P-P 179-192

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN 2222-2995
University of Kirkuk Journal For Administrative and Economic Science

babakr Sawen Othman. & Badal Mohammed AbdulMajeed. Using Multilevel Poisson Regression with Wavelets Filters. *University of Kirkuk Journal For Administrative and Economic Science* (2023) 13 (4): 179-192.

Using Multilevel Poisson Regression with Wavelets Filters

[1]Sawen Othman babakr, [2]Mohammed AbdulMajeed Badal

[(1,2)] Salahaddin University, Erbil, Iraq

[1]sawen.babakr@su.edu.krd

[2]Mohammed.badal@ su.edu.krd

Abstract. The proposed method in this study is used to identify the influence of the drug on the thyroid disease as well as the ordinary least squares method (OLS) and the generalized linear regression method (GLM) are used to prepare the estimate of the linear model. The study concluded that wavelet filters (Haar, Symlets, and Daubechies) produce the best results for estimating the Poisson Regression model when compared to the Generalized Linear Regression model based on (MSE, R^2, and P-value). The data in this study has been collected from Smart Health Tower in Sulaimani, which contains 128 cases of thyroid disease, as well as the simulation data used in this study. The paper's finding showed that the STOP-TSH (sym72) wavelet with the (fixed from thresholding) threshold technique and soft threshold rule was the most efficient when compared to all other proposed methods and the classical approach for both real and simulation data. The STOP_TSH (sym52) wavelet with the (fixed from thresholding) threshold method and hard threshold rule was the most efficient compared with all other proposed methods and the classical method for both real and simulation data. All the proposed methods have better efficiency than the classical method in estimating the Poisson Regression model, depending on the average of Akaike Information Criteria (AIC) and Bayesian information criteria (BIC).

Keywords: Poisson Regression, GLM, Wavelet Shrinkage, thresholding rules.

IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (13) المجلد (1): P-P 179-192

استخدام انحدار بواسون متعدد المستويات مع مرشحات المويجات

[1]ساوين عثمان بابكر، أ.م.د. [2]محمد عبد المجيد بدل
(١،٢) جامعة صلاح الدين-أربيل، العراق
[1]sawen.babakr@su.edu.krd
[2]Mohammed.badal@ su.edu.krd

المستخلص. تهدف الدارسة إلى التعرف على تأثير الدواء على مرضى الغدة الدرقية عن طريق تحليل الانحدار بطريقة المربعات الصغرى (OLS) وانحدار بواسون للنموذج العام وتقدير معالم نموذج الانحدار الخطي الانحداربواسون(GLM) لخصت نتائج الدراسة أن مرشح المويجة (Harr,Symlets and (Daubechies أعطى افضل النتائج لتقدير معالم نموذج الانحدار الخطي العام من خلال المقاييس الإحصائية-MSE, R2 and p) (value.اخذت بيانات الدراسة من مركز الصحة المتقدم في مدينة السليمانية لعينة (١٢٨) من المرضى المصابون بأمراض الغدة الدرقية. مع تطبيق اسلوب المحاكاة وتوصلت الدراسة أن استخدام طريقة قطع العتبة الناعم والثابت (fixed from thresholding)للمويجة أعطى نتائج كفوءة عند التوقف (STOP-TSH(sym72عن إعطاء الدواء للمريض المصاب بالغدة الدرقية مقارنة مع بقية الطرق للانحدار الكلاسيكي وانحدار بواسون وكذلك كانت ذات كفاءة مع اسلوب المحاكاة وخلال المقاييس الإحصائية . كانت المويجات (STOP_TSH (sym52 مع طريقة العتبة (الثابتة من العتبة) وقاعدة العتبة الصلبة هي الأكثر كفاءة مقارنة بجميع الطرق المقترحة الأخرى والطريقة الكلاسيكية لكل من البيانات الحقيقية والمحاكاة. ووجد أن جميع الطرق المقترحة ذات كفاءة أفضل من الطريقة الكلاسيكية في تقدير نموذج انحدار بواسون ، اعتمادا على متوسط معايير معلومات (Akaike (AICومعايير معلومات (Bayesian (BIC .

الكلمات المفتاحية: انحدار بواسون ، GLM ، انكماش المويجات ، قواعد العتبة.

1. Introduction:

There is one major regression models used for censored data: poisson regression model. On the basis of one or more independent or explanatory variables, regression analysis creates an explanatory model of a dependent variable (typically designated Y) (denots X1, X2, etc). A countback happens when the dependent variable is a count, such as the number of times an event occurs or the quantity of persons in a specific group, Poisson regression is applicable. It might be especially beneficial if any of the observations have extremely low values. In geography, as well as the social and environmental sciences in general, the technique is still mostly unknown. In the journal literature, however, there are an increasing number of uses. The number of migrants from one city to another plant species on islands (Vietnam) are some of the dependent variables that have been studied (Gatrell1986), spina bifida (Lovett and Gatrell 1988), robbed New York City businesses, and international newspapers (Flowerdew and Lovett 1984). Lovett (1984) has written a comprehensive reference to the approach, which is included in the GLIM statistical program. Eddy (2011) suggest technique for estimating the function appropriate compactly supported wavelets such as the Daubechies, Symlets, or Coiflets family of wavelets, the smoothness and time-frequency features of these wavelets allow us to derive asymptotically estimators of the slope coefficient of the linear model. Rogério (2016) suggest extraction of an observation in the presence of random noise by wavelet shrinkage has been studied under assumptions that the containate is independent and identically distributed and that the samples are evenly spaced with time. Xing et al. (2017), discuss the estimation a models with censor data by using wavelet method when the survival function and the censors times has a stationary α-mixing sequence, and of the wavelet estimators for varying functions. Christophe et al. (2019) suggest a deal with the estimation of a non-parametric regression with both additive and contaminate, for uniform multiplicative contaminate is considered, and develop a projection estimator by using a several wavelets. Jinru et al. (2020), the point wise asymptotic convergence rates of wavelet estimators of censored mixture density were discussed.

2. Poisson distribution:

The benchmark model for count data is the Poisson distribution. It's a good idea to go through some of the Poisson distribution's essential features and characterization results first (for derivation seen Taylor and Karlin, 1994). Y has density if the discrete r.v's Y is Poisson-distributed with an intensity or rate parameter $\mu, \mu > 0$, and t is the exposure, define as the period of time during which the events are recorded.

$$\Pr[Y = y] = \frac{e^{-\mu t}(\mu t)^{y}}{y!} \qquad y = 0,1,2,\dots \qquad (1)$$

The Poisson distribution can be described in a variety of ways. We'll look at four of them here. The first is the law of unusual occurrences, which is a typical Poisson motive. The second, the Poisson count processing, is frequently seen in stochastic process introductions. The third is merely a mirror image of the second, with the count replaced by the duration between events.

3. Poisson Regression:

To investigate the relationship between several factors of interest and count result data, regression modeling approaches can be utilized. However, due to a number of modeling assumptions breaches, traditional linear regression is typically not feasible for count data. The generalized linear model (GLM) is the foundation of traditional statistics teaching, and it comprises standard statistics procedure such as the t-test, analysis of variance, and linear regression that assume a continuous and normally distributed dependent variable. Count data on uncommon events or occurrences with a mean of 10 is usually always skewed and non-normal, therefore traditional methodologies based on normalcy are insufficient. Count data, on the other hand, may be represented using a broader set of statistical approaches that do not need the dependent variable is regularly distributed.

The method used to analyze count data, particularly the regression framework chosen, is sometimes determined by how the counts are expected to occur. There are two types of formulations that are often used. They originate from a firsthand observed of a point process in the first case. Counts emerge from the discretization ("ordinalizated") of continuous latent data in the second. Other less-common formulations make use of the law of uncommon evented or the Poisson binomial stops.

3.1. Methods Poisson regression:

A traditional Poisson regression model assumes that the number of occurrences in a given unit of time follows a Poisson distribution with a mean n, and that the rate li for observation i is correlated to a vector of explanatory variables, Xi, by a linear relationship.

$$Log(\mu_i) = \log(n_i) + X_i b \qquad (2)$$

Where ni indicates the period at risk and is comparable to the denominator of the rate, and b is a vector of unknown parameters to be approximated. The "offset" of the model is typically referred to as the quantity log (ni). A cohort is often cross-classified by degrees of exposure and other predictor factors, Xi, in this model, and the period at risk is determined for each of the resultant combinations of X.

In a Poisson regression model, the maximum likelihood estimate Xi of the average of a Poisson-distributed variable Yi is the projected value of the dependent variable for qase i. This estimate's natural logarithm is equivalent to a linear combination of the explanatory variables' corresponding values; if there is just one explanatory variable, the collated takes the form

$In(\mu_i) = \beta_0 + \beta_1 x_i$ (3)

There is the same OLS regression in Poisson Regression. The regression model has two properties that ought to be discussed further. The Poisson model, unlike OLS regression, does not presuppose homoscedasticity in the data. Each instance's variance is, in fact, equal to the expected value for that case. As a result, the variances associated with Yi values cannot be the equivalent and the distribution cann't be normal. Second, Yi is a count of independent occurrences produced by a Poisson distribution with parameter Xi. When an analyst has to estimatting the value of a dependent variable Y based on a linear predictor, he or she can utilize generalized linear models (a linear combination of independent variables X1, X2, etc.). A random (or error) component i is added to an observed value Ɛi:

$Y_i = \beta_i + Ɛ_i$ (4)

Where Y_i is the Y value seen in item i, β_i is the Y value predicted in item i and $Ɛ_i$ is error term for random distribution.

3.2 Poisson Regression in GLIM:

Generalized linear models (GLMs) are the name for these approaches (GLM). In the same way that linear regression incorporates predictions based on the dependent variable, the GLM does as well. Unlike linear regression, which models the dependent variable's actual observed value, the GLM models the dependent variable's function. The link function is what it's called. Think about it. The log–odds scale is used to analyze the regression coefficients (Betas) in this equation. To put it another way, the value of e 1 is a probability ratio. Poisson regression employs a different link function, the log-link function, to represent the natural log of count outcome data (Y), which is written as:

$Log(Y) = \beta_0 + \beta_1 X_1$ (5)

The regression model with the highest log-likelihood value outperforms the one with the lowest log-likelihood value. When α = 0, the GPR model is reduced to the PR model. We test the hypothesis to see if the GPR model is better than the PR model.

$H_0: \alpha = 0$ against $H_1: \alpha = 0$ (6)

The importance of the dispersion parameter is tested by H0 in (6). Whenever H0 is rejected, the GPR model should be used instead of the PR model. The asymptotical normal Wald type "t" statistics, define as the ratio of the estimate of to its standard error, may be used to perform the test in (6). When the null hypothesis is true, an alternate test for the null hypothesis in (6) is using the likelihood ratio test statistics, which is nearly chi-square distribution with one degree of freedom (d.f.).

And $\mu_i = \mu_i(x_i) = \exp(x_i)$, where xi is a (k-1) dimensional vector of Explanatory factors, which includes personal qualities of both husband and wife in a household as well as certain demographic information, and is a k-dimensional vector of regression parameters. Yi's mean and variance are calculated as follows:

$$E\ (Y_i|x_i)\ =\ \mu_i \qquad (7)$$

and

$$V\ (Y_i|x_i)\ =\ \mu_{\ i}(1+\ \alpha\ \mu_i)^2 \qquad (8)$$

The nonlinear variance and average structure (8) is same to that described by Winkelmann and Zimmermann in their generalized event count (GECk) model (1994). It is stated that Var (Yi|xi) = (σ2-1) µiq+1 + µi with σ2 ∈ R+and q ∈ R. It is easy to see that σ2 =1 implies equidispersion, σ2 > 1 over dispersion, and 0 < σ2 < 1 and µiq ≤ (1- σ2) under-dispersion.

$$\text{Ln L } (\alpha;\ \beta;\ y_i) = \sum_{i=1}^{n}\left\{y_i \log\left(\frac{\mu_i}{1+\alpha\mu_i}\right)+(y_i-1)\log(1+ay_i)-\frac{\mu_{i(1+\alpha y_i)}}{1+\alpha\mu_i}-\log(y_i\,!)\right\}. \qquad (9)$$

The greatest likelihood equations for estimating and are obtained by taking the partial derivatives of (9) and equating them to zero. As a result of this,

$$\frac{\partial \ln L}{\partial\alpha}=\sum_{i=1}^{n}\left\{\frac{-y_i\mu_i}{1+\alpha\mu_i}+\frac{y_i(y_i-1)}{1+\alpha y_i}-\frac{\mu_i(y_i-\mu_i)}{(1+\alpha\mu_i)^2}\right\}=0\ , \qquad (10)$$

and

$$\frac{\partial \ln L}{\partial\beta_r}=\sum_{i=1}^{n}\frac{y_i-\mu_i}{\mu_i(1+\alpha\mu_i)^2}\frac{\partial\mu_i}{\partial\beta_r}=0\ ,\ \ r=1,2,\dots,k\ . \qquad (11)$$

An iterative approach is used to solve equations (10) at the same time. The finally estimated from fitting a Poisson regression model to the data is the beginning estimate of b for the iteration. The primary estimated of can be set to zero or calculated by multiplying the chi-square statistic by the (d.f.). This is provided by

$$\sum_{i=1}^{n}\frac{(y_i-\mu_i)^2}{\text{V (Yi|xi)}}=n-k, \qquad (12)$$

Where $V(Y_i|x_i)$ is given by (3). When $\alpha<0$(The value of (in the case of under-dispersion) is such that $1+\ \alpha\mu_i>0$ and $1\ +\ \alpha\, y_i\ >\ 0$, i.e., $\alpha\ >\ \min\,(-1/\max\,(\mu i),-1/\max\,(yi))$ as specified in Eq. (1). A FORTRAN Eqs. Are solved with the help of a software. (10) Simultaneously.

4. Wavelet:

The study of wavelets becomes more intriguing and valuable as a result of this. Noise contaminates many scientific data sets, either as a result of the data collection process or as a result of external causes. A basic step in studying such datasets is de-noising, or predicting the unknown signal of interest from the existing noisy data. Signals and images can be de-noised in a number of ways. De-noising, de-blurring, smoothing, and restoration have minor differences while having similar aesthetic results.

Smoothing frequently eliminates high frequency while preserving low frequency (with blurring). De-noising seeks to remove all noise present in a noisy signal regardless of its spectral content, De-blurring, on the other hand, increases the clarity of signal characteristics by enhancing high frequencies. Restoration is a type of de-noising that attempts to recover the original signal as accurately as possible by balancing de-blurring and smoothing. Wavelet transformations can be used to represent signals that have a lot of sparsity. Wavelet de-noising, a non-linear wavelet-based signal estimation approach, is based on this idea. Wavelet de-noising aims to eliminate noise from a signal while keeping its properties, independent of the signal's frequency content. Smoothing should not be confused with wavelet de-noising. Smoothing just removes the high frequencies while keeping the lower ones, as previously stated.

4.1. Wavelet transform:

There are a variety of wavelet basis functions available, including Harr, Duabechies, Symlet, Meyer, biorthogonal wavelet, and others. The climbing function, also well-known as the father wavelet, and the wavelet purpose, sometimes known as the mother wavelet, are used to define the wavelet (Daubechies, 1992; Van Fleet, 2007). Assume, following Karim et al. (2008), that a function ϕ (t) ∈ L2(R) exists in such a way that the functional family

$$\phi(\mathrm{t}) = 2^{j/2}(2^j t - k), j, k \in Z. \quad (13)$$

Is a foundation that is orthonormal?

The wavelet series can be defined as follows:

$$f(x) = \sum_k \alpha_k \, \varphi_{0k}(x) + \sum_{j=0}^{\infty} \sum_k \beta_{jk} \, \psi_{jk}(x), \quad (14)$$

Where αk, βjk, Eq (14) defines the coefficients, and $\{\psi_{jk}\}$,k ∈ Z is a basis for Wj . A multiresolution extension of f is the relation in (13). The following equation is used to convert (13) into wavelet expansion.

$$\psi_{jk}(x) = 2^{\frac{j}{2}}\psi(2^j x - k), j, k \in Z\,. \quad (15)$$

In essence, the functionφ_{jk}(x) and ψ_{jk}(x) the mother wavelet and the scaled function (father wavelet) are referred to as the scaled function and mother wavelet, respectively. Meanwhile

$$\alpha_k = \int f(x)\overline{\varphi_{0k}(x)}\, dx\,, \beta_{jk} = \int f(x)\, \overline{\psi_{jk}(x)} dx\,, \quad (16)$$

The approximation/coarser coefficients are α_k, while the detaicoefficients are β_{jk}.

4.2. The discrete wavelet transform:

The wavelet transformation is a method for assessing in what way fit a set of wavelet functions accurately represents the information under consideration. The wavelet coefficients represent the function's quality of fit to the signal. The resulting concluded a set of constants for two independent variables: translate and dilate. The term "translate" usually refers to the passage of time, whereas "scaling" refers to the way in which material is seen in terms of frequency. Lower frequencies equate to larger scales.

The wavelet coefficient, αi,j , of the j wavelets at each level i are determined via the forward wavelet transform. For the signal f (n), the DWT is

$$\alpha_{i,j} = \alpha^i_{2+j} = \sum_n f(n) \psi_{i.j}(n) \quad (17)$$

IWT is the inverse wavelet transform that corresponds to it.

$$f(n) = \sum_i \sum_j \alpha_{i,j} \psi_{i.j}(n) \quad (18)$$

Where $\psi_{i.j}(n) = 2^{i/2}\psi(2^i n - j)$

4.3. Beyond the Haar Wavelet:

The Haar MODWT and related ODWT have been the focus of our discussion thus far, although there are additional variations of these transforms. These transformations may be expressed for a given maximum level J0 in terms of wavelet filtering of levels $j = 1, \dots, J_0$ and a scaled filter of level J0 Comparing analysis based on the Harr wavelet against analyses based on other wavelets is a good way to see if the Haar wavelet is acceptable or not. There is no need to use anything other than the Haar wavelet if the analyses are essentially the same; if they aren't, a different wavelet analysis than the Haar wavelet may be necessary. When non-Haar MODWTs and ODWTs are used, new boundary coefficients are produced. Hence discarding a Haar-based analysis comes at a cost.

$$\Psi(u) = \begin{cases} 1, & 0 \leq u < \frac{1}{2} \\ -1, & \frac{1}{2} \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases} \qquad (19)$$

The measurement function is stated as followed:

$$\phi(u) = \begin{cases} 1, & 0 \leq u \leq 1 \\ 0, & \text{Otherwise} \end{cases} \qquad (20)$$

4.4. Wavelets Daubechies:

All of the wavelets we've looked at thus far, such as Haar and Shannon wavelets, have significant limitations. Haar wavelets have a dense but discontinuous support. Shannon wavelets are extremely smooth, yet they span the whole real line. Before Ingrid Daubechies discovered the hierarchy of wavelets that includes the Haar wavelet, which is the only discontinuous one, these wavelets and a few others with comparable qualities were the only ones accessible. The hierarchy's other wavelets are compactly supported and continuous. The features of a wavelet with compact support are numerous. They can be built to have a specific number of derivatives and vanishing moments.

$$L1 = N/2 \qquad (٢1)$$

In general, dbN represents the family of the (N) moijat, although Moija Har is a member of this family because db1 is the same as the Harr wavelet.

4.5. The Continuous wavelet transform:

The Continuous wavelet transform (CWT) is a technique for displaying and assessing time and scale-dependent signal properties. The CWT, like the Fourier transform, is based on a single scaled function. Unlike the Fourier transform, however, we additionally shift the function, making the CWT an operator that accepts a signal and outputs a function with two variables: time and scale. Which may be thought of as a surface or picture.

This is accomplished by employing scales that are closer together than the 2i relationship, and it provides the foundation for the continuous wavelet transform (CWT). The CWT takes the following shape:

$$\alpha_{i,j} = \int_{\infty}^{-\infty} f(t)\psi(i,j,t)dt \qquad (22)$$

Where i denotes dilation and j denotes translation. The integral will be replaced by a summation for a finite digitally sampled signal, and the time t will be replaced by the discrete n.

The discrete wavelet transform (DWT) is a signal process algorithm that is commonly used in the biomedical area to de-noise and compress data [M. Unser and A. Al-droubi]. The DWT decomposition procedure is as follows:

$$\alpha_{j+1}(k) = \sum_{n} h(n-2k)\alpha_j\,(k) \qquad (23)$$

$$d_{j+1}(k) = \sum_{n} g_0\,(n-2k)\alpha_j(k) \qquad (24)$$

In the DWT, a half-band low-pass filter h0 and α half-band high-pass filter g0 yield coarse approximation results αj and detail results dj, respectively.

5. Threshold Estimation:

The primary principle of wavelet de-noising is to extract the ideal components of a signal from a noisy signal, which necessitates noise level measurement. There are a variety of ways to estimating the noise level, as well as a thorough examination of their effectiveness [8, 9]. Four alternative threshold settings were tested in this study to see how successful they were. Rigresure is the process of selecting an adoption threshold using Stein's unbiased risk estimate criterion.

$$\text{ThValue} = \sigma\sqrt{2\log(N\log_2 N)} \qquad (25)$$

Where the noise standard deviation and N is the signal length. (Sqtwolog is the universal threshold; the latter is determined from the detail coefficient at the first level of signal decomposition; Dxmedian 674.0)

$$\text{ThValue} = \sigma\sqrt{2\ \log_e(N)} \qquad (26)$$

Heursure is a heuristic that employs a combination of the prior principles. The minmax concept is used to select a threshold in Minimaxi. When compared to an ideal approach, a defined threshold is chosen to attain the minimum of the highest MSE generated by the best function in a assumed set.

We look at a unique type of nonlinear estimator called a truncated threshold wavelet estimator. As in (2), define the empirical coefficients $\hat{\alpha}_{jk}, \hat{\beta}_{jk}$ and use hard thresholding:

$$\hat{\beta}_{jk} = \begin{cases} \hat{\beta}_{jk} & if\ |\hat{\beta}_{jk}| > KC(J)n^{-1/2} \\ 0 & if\ |\hat{\beta}_{jk}| \le KC(J)n^{-1/2} \end{cases} \qquad (27)$$

The estimator TW is then applied to the functions$j_0(n), j_1(n), C(j)$, and K.

$$TW(x) = \hat{f}_{n,j_1} + \hat{D}_{j_1,j_0} = \sum_{k\in Z} \hat{\alpha}_{j_1k}\ \phi_{j_1k}\,(x) + \sum_{j_1}^{j_0}\sum_{k\in Z} \hat{\beta}_{jk}\,\psi_{jk}\,(x) \qquad (28)$$

Before we look at the attributes of this estimator, let's talk about why we're using it. The linear wavelet estimator, LW (corresponding to, $j_0 < j_1$, and hence no "detail" term $\hat{D}_{j_1,j_0}$) cannot be optimum if p < p', as we saw in the previous sections. The dissection of the error into bias and variance components may explain this. When LW employs level $j(n)$, it has a bias of order $2^{-j(n)\acute{\sigma}\acute{p}}$, whereas the stochastic term is of order$(2^{j(n)}/n)^{\acute{p}/2}$. This leads to the notion of starting with a low-frequency estimator $LW(j_1(n)))$, with $j_1(n)$set low enough that the stochastic term has the correct rate, and then adding in specific "details" up to the higher order $j_0(n)$in such a manner that the bias term has the correct order as well. (It is clear that if p' = p, selecting $j_0 < j_1$suffices, however if p' > p, selecting $j_0 > j_1$ is required.)

5.1. Thresholding Rules:

There are many rules for the thresholding. The two types used in this research will be discussed.

A. Soft Thresholding:

Soft thresholding of the wavelet coefficient is another typical approach for wavelet de-noising. Also proposed by Donoho and Johnostone, which is defined as follows (Jeena, 2013).

B. Hard Thresholding:

Donoho and Johnstone proposed hard thresholding, it is a simplest scheme thresholding interprets the declaration of (kill or keep). For wavelet implementation, hard thresholding employed a simple approach. De-noising (Katsuyuki, 2021).

6. Evaluation criteria:

The AIC (Akaike information criterion) and BIC (Bayesian information criterion) are two different types of information criteria. Log-likelihood (LL) will be used as selection criteria for the models. The model with the lowest value of AIC and BIC term appears the best model to Poisson regression (Rinku and Manash 2016)

$$AIC = -2(LL) + \frac{2K}{P} \quad (29)$$

$$BIC = -2(LL) + K.Ln(n) \quad (30)$$

7. Experimental and Application:

To compare between the classical and the proposed method in terms of efficiency and accuracy of the estimated model, an experimental part was done by simulating the Poisson regression and an applied of the real data about Thyroid disease for the patients in Smart Health Tower Sulaimani in Kurdistan-Iraq, based on AIC and BIC. And by designing a program in MATLAB (version 2018b) dedicated to this purpose (Appendix).

7.1. Experimental Part:

For case the time-independent and dependent covariates, Real data about Thyroid disease for the patients in Smart Health Tower Sulaimani in Kurdistan-Iraq. The case was selected for the sample size (128). This data content (three dependent and six independent) variables with give effected two times on the patient. Noises with added to the Poisson Regression model, dependent variable de-noise and thresholding de-noise shown figure (1-3).

Table 1: Vector Auto regression (Variable) for variables before, stop and after TSH Wavelet for symlets-stop-hard 52.

	Log likelihood	SBIC	AIC	Number of obs.
Before	-4.261	9.908	12.522	128
Stop	-233.495	505.807	482.991	128
After	-2.208	5.802	8.416	128

From the table (1) the Variable stop model is the best because AIC (482.991) value is less than SBIC values.

Table 2: Test of regression model (Variable) for variables Before, Stop and after.

Equation	No. Parms.	RMSE	R-sq.	chi2	P-value
Before	7	97.349	0.034	7.4796E-30	0.636
Stop	7	2.379	0.721	287.872378	3.1206E-31
After	7	178.293	0.0590	1.2161E-12	0.279

Tables (1-2) show that all the proposed methods have better efficiency than the classical method in estimating the Poisson Regression model depending on both average of criteria (AIC and BIC) for Thyroid disease selected samples, the case (n = 128), for the (Sym52) wavelet, (fixed from thresholding) threshold method, hard rule. Hard threshold rule was the best efficient compared with all other proposed methods and with the classical method because it has the lowest average of both criterions and for Thyroid disease selected samples (AIC = 482.991, BIC = 505.807), respectively. For most real data experiments, (sym52) Wavelet was better than, Also hard rule better than soft for all cases.

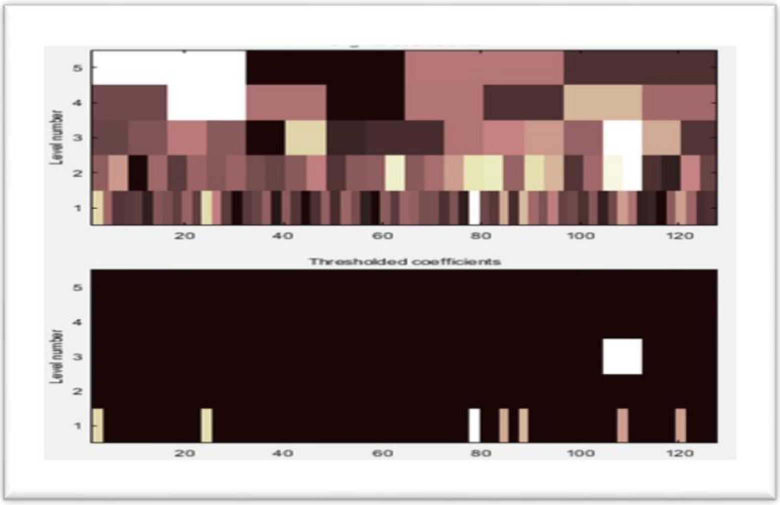

Figure 1: Dependent variable (Observation before TSH) de-noising and thresholding de-noising.

For the purpose of the comparison between the proposed and classical method in estimating the Poisson Regression, also the experiment for real data about Thyroid disease for the patients and the average criteria for AIC and BIC was calculated. One wavelet (Sym52) was used with same methods in estimating the threshold level (fixed from thresholding), for two threshold rule (Soft and Hard), and for sample size (128). The results are summarized in Tables (1-2).

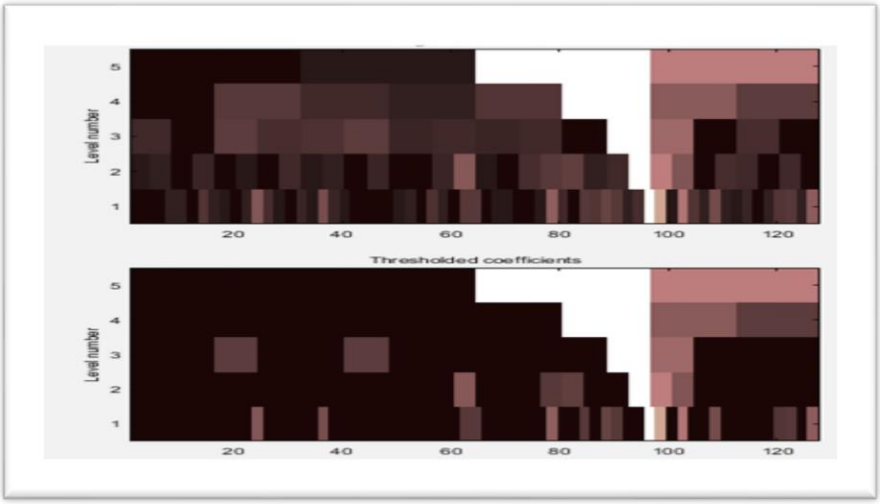

Figure 2: Dependent variable (Observation stop TSH) de-noising and thresholding de-noising.

For the purpose of the comparison between the proposed and classical method in estimating the Poisson Regression, also the experiment for real data about Thyroid disease for the patients and the average criteria for AIC and BIC was calculated. One wavelet (Sym52) was used with same methods in estimating the threshold level (fixed from thresholding), for two threshold rule (Soft and Hard), and for sample size (128). The results are summarized in Tables (1-2).

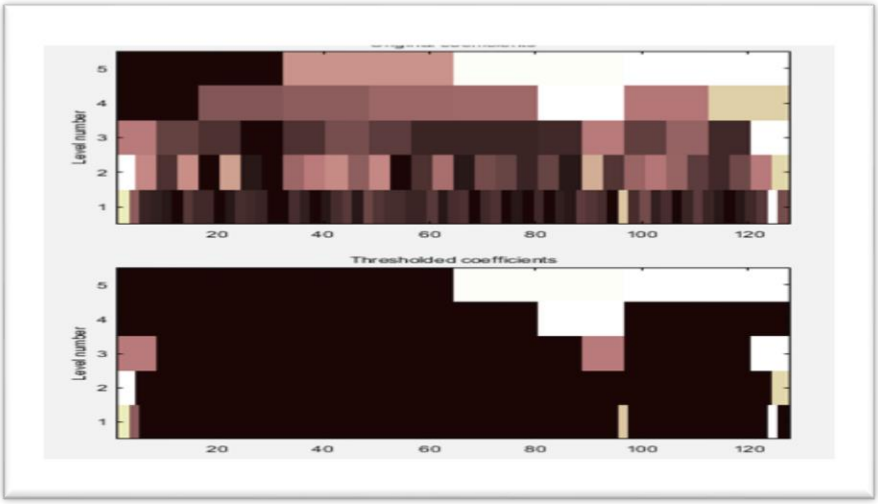

Figure 3: Dependent variable (Observation after TSH) de-noising and thresholding de-noising.

For the purpose of the comparison between the proposed and classical method in estimating the Poisson Regression, also the experiment for real data about Thyroid disease for the patients and the average criteria for AIC and BIC was calculated. One wavelet (Sym52) was used with same methods in estimating the threshold level (fixed from thresholding), for two threshold rule (Soft and Hard), and for sample size (128). The results are summarized in Tables (1-2).

7.2. Experimental part for Simulation data:

For case the time-independent and dependent covariates, simulation data visualizations (Appendix -program-1). One case was selected for the sample size (200). Noises with added to the Poisson

Regression model, dependent variable de-noising and thresholding de-noise for the simulation with n = 200 shown figure (4).

Table 3: Average of criteria AIC and SBIC when n = 200.

	Wavelet	Log likelihood	SBIC	CIA	Threshold Rule
Simulation data	Harr7	-39.821	123.310	97.641	Soft
	Harr7	-178.464	400.596	374.927	Hard
	Sym72	-61.010	165.687	140.020	Soft
	Sym72	-188.414	420.496	394.828	Hard
	db72	-61.010	165.688	140.020	Soft
	db72	-188.414	420.496	394.828	Hard

From the table (3) the Variable model is the best because AIC (140.020) value is less than SBIC values.

Table 4: Test of regression model (Variable) for simulation data n=200.

Equation	No. Parms.	RMSE	R-sq.	chi2	P-value
Harr7 soft	7	0.116	0.092	13.962	0.106
Harr7 hard	7	1.015	0.102	121.833	0.069
Sym72 soft	7	0.162	0.621	19.442	1.5138E-22
Sym72 hard	7	1.186	0.180	142.326	0.001
db72 soft	7	0.162	0.621	19.442	1.5138E-22
db72 hard	7	1.186	0.180	142.326	0.001

Tables (3-4) show that all the proposed methods have better efficiency than the classical method in estimating the Poisson Regression model depending on both average of criteria (AIC and BIC) for surgery thyroid selected samples, the case (n = 200), for the (Sym72) wavelet, (fixed from thresholding) threshold method, soft rule. Also, (db72) wavelet with (fixed from thresholding) threshold method and Soft threshold rule was the best efficient compared with all other proposed methods and with the classical method because it has the lowest average of both criterions and for surgery thyroid selected samples (AIC = 140.020, BIC = 165.687), respectively. For most simulation experiments, (db72) Wavelet was better than (Sym72), Also Soft rule better than Hard for all cases.

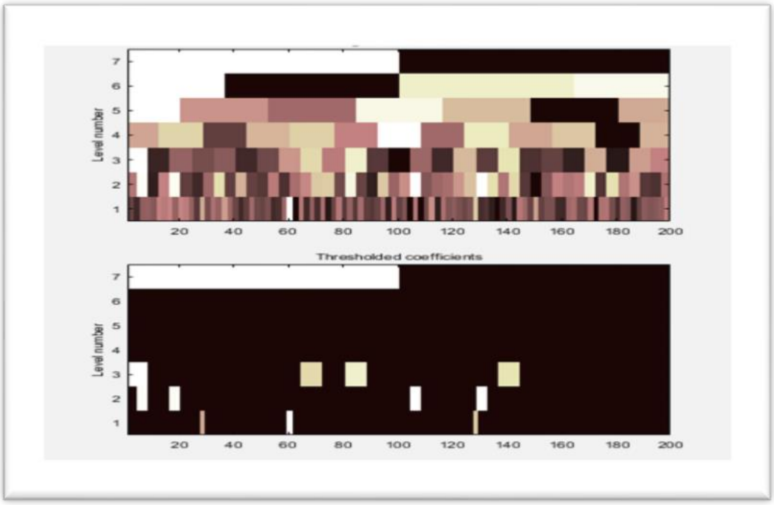

Figure 4: Dependent variable (Simulation data) de-noising and thresholding de-noising.

For the purpose of the comparison between the proposed and classical method in estimating the Poisson Regression, also the experiment was repeated to (1000) times and the average criteria for AIC and BIC was calculated. Two wavelets (Sym72) and (db72) were used with same methods in estimating the threshold level (fixed from thresholding), for two threshold rule (Soft and Hard), and for sample size (200). The results are summarized in Tables (3-4).

8. Conclusions:

1. The aims methods (thresholding) have the primary principle of wavelet de-noising is to extract the main components of a signal from a noisy signal, which necessitates noise level measurement. In estimating the Poisson regression model depending on both average of criteria (AIC and BIC) for selected samples Thyroid disease for the patients, simulation data and application).
2. STOP_TSH (sym72) wavelet with (fixed from thresholding) threshold method and soft threshold rule was the best efficient compared with all other proposed methods and with the classical method for Thyroid disease for the patients selected samples (for simulation and applications).
3. STOP_TSH (sym52) wavelet with (fixed from thresholding) threshold method and hard threshold rule was the best efficient compared with all other proposed methods and with the classical method for Thyroid disease for the patients selected samples (for Real data and applications).

References:

1. Chang, S.G., Yu, B. and Vetterli, M., 2000. Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing*, *9*(9), pp.1532-1546.
2. Christophe Chesneau, Junke Kou and Fabien Navarro, (2019), linear wavelet estimation in regression with additive and multiplicative noise, HAL Id: hal-01877543 https://hal.archives-ouvertes.fr/hal-01877543v2
3. Consul, P. and Famoye, F., 1992. Generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, *21*(1), pp.89-109.
4. Donoho D. L. and Johnstone I. M., (1994), Ideal Spatial Adaptation by Wavelet Shrinkage, Biometrika , 81 , 425-550
5. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D., 1996. Density estimation by wavelet thresholding. *The Annals of statistics*, pp.508-539.
6. Ergen, B., 2013. Comparison of wavelet types and thresholding methods on wavelet based denoising of heart sounds. *Journal of Signal and Information Processing*, *4*(3B), p.164.
7. Hayat, M.J. and Higgins, M., 2014. Understanding poisson regression. *Journal of Nursing Education*, *53*(4), pp.207-215.
8. Hutchinson, M.K. and Holtman, M.C., 2005. Analysis of count data using poisson regression. *Research in nursing & health*, *28*(5), pp.408-418.
9. Iolanda Lo Cascio, (2007), Wavelet Analysis and De-noising: new Tool for Economists, University of London.
10. Jeena Joy1, Salice Peter2, Neetha John3, (2013), de-noising using soft thresholding, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 3.
11. Jinru Wang, Zhenming Zhang, Xue Zhang, and Xiaochen Zeng (2020), Wavelet pointwise estimations under multiplicative censoring, International Journal of Wavelets, Multiresolution and Information Processing.
12. Karim, S.A.A., Karim, B.A., Ismail, M.T., Hasan, M.K. and Sulaiman, J., 2010, November. Compression of temperature data by using daubechies wavelets. In *Proceedings of International Conference on Mathematical Sciences (ICMS2 2010)* (Vol. 30, pp. 726-734).
13. Katsuyuki Hagiwara, (2021), Bridging between soft and hard thresholding by scaling, Faculty of Education Mie University (Mie Univ.)

14. Percival, D.B., 2008. Analysis of geophysical time series using discrete wavelet transforms: An overview. *Nonlinear Time Series Analysis in the Geosciences*, pp.61-79.
15. Ramazan Gengay, Faruk Selguk and Brandon Whitcher, (2002), an Introduction to Wavelet and other Filtering Methods in Finance and Economics.
16. Weiss, G. and Wilson, E.N., 2001. The mathematical theory of wavelets. In *Twentieth century harmonic analysis—a celebration* (pp. 329-366). Springer, Dordrecht.
17. Xing-cai Zhou, Ying-zhi Xu, and Jin-guan Lin, (2017), Wavelet estimation in varying coefficient models for censored dependent data.