Hindawi

Research Article

Testing Heteroscedasticity in Nonparametric Regression Based on Trend Analysis

Si-Lian Shen,[1] Jian-Ling Cui,[2] and Chun-Wei Wang[1]

[1] *School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang 471003, China*
[2] *Electronic Equipment Test Center, Luoyang 471003, China*

Correspondence should be addressed to Si-Lian Shen; meipersuit@126.com

Received 27 January 2014; Revised 26 May 2014; Accepted 26 May 2014; Published 11 June 2014

Academic Editor: Zhihua Zhang

Трендийн анализд үндэслэсэн параметрийн бус регрессийн гетероскедастикийг илрүүлэх тест

Орчуулсан:

МБУС-ийн Математикийн тэнхмийн багш Б.Алтансувд

1. Удиртгал

Сүүлийн үед тоон шинжилгээний янз бүрийн салбарт параметрийн бус регрессийн загвар өргөн хэрэглэгдэж байна. Регрессийн функцийн үнэлэлт болон параметрийн бус загварын статистик таамаглалууд нь алдаанууд гомоскедастик байх таамаглалд үндэслэдэг. Гэвч бодит тоо өгөгдөлд тулгардаг таамаглалуудыг баталж чадах эсэхээ бид өмнө нь мэддэг. Иймээс эхлээд дүгнэлт хийх замаар дэвшүүлсэн асуудлууд болон тохирсон загварт шалгаж үзэн алдаануудын гетероскедастикийг илрүүлэх загварыг хөгжүүлэх шаардлагатай байна.

Параметрийн бус регрессийн тухай номнуудад гетероскедастик тестийн тухай бичсэн байдаг[1-8]. Энэ өгүүлэлд Dette, Munk [2] нарын хөгжүүлсэн аргад үндэслэн дисперсийн функцийг тогтмолоор, ойролцоолон тооцсон L^2 хамгийн сайн дөхөлтийг үнэлсэн ба хосын шугаман регрессийн загварыг You, Chen [5] нарын аргаар өргөтгөсөн. Dette [1] параметрийн бус гетероскедастик тестийг санал болгосон. Eubank, Thomas [3] нар үлдэгдэлд үндэслэсэн параметрийн бус загварын алдааны гетероскедастикийг илрүүлэхийг санал болгосон. Түүнээс гадна Zhang, Mei [7] нар үлдэгдлийн анализд үндэслэн алдааны загварын дисперс тогтмол байх тестийг гаргасан.

Дээр дурьдагдсан аргуудын хамгийн сайн нь параметрийн таамаглал шалгах тестийн аргад хамаарна. Алдааны загварууд нь өмнө нь таамаглаж байсан тархалттай давхцаж байвал эдгээр аргуудтай ажиллахад сайн үр дүнтэй байдаг. Харин тархалтыг нь баталж чадахгүй байвал эдгээр аргуудын үр дүн нь буурна. Иймээс алдааны тархалтын robust /бат бөх сайн/ тестийг хөгжүүлэх шаардлагатай.

Энэ өгүүлэлд бид параметрийн бус регрессийн алдаануудын гетероскедастикийг илрүүлэх параметрийн бус таамаглал шалгах тестийн аргыг танилцуулж байна. Энэ статистик тест нь локал шугаман үнэлэлттэй регрессийн загвартай тохирсоны дараа үлдэгдлүүдийн тохиромжтой хувиргалтыг байгуулдаг. Дэвшүүлсэн загварын үр дүнг үнэлэхийн тулд бид Zhang, Mei [7]-ий арга болон дэвшүүлсэн аргыг ашиглан бодит тоон өгөгдөлд хийсэн шинжилгээний үр дүнд харьцуулалтыг хийсэн.

Өгүүллийн бусад хэсэг нь дараах байдлаар зохион байгуулагдсан.

Хоёрдугаар хэсэгт локал шугаман үнэлэлтийн аргыг товч тайлбарласан.

Гуравдугаар хэсэгт локал шугаман үнэлэлттэй регрессийн загварт тохирсоны дараах үлдэгдлийг ашиглах болон параметрийн бус статистикийн трендийн анализын ойлголтын хэрэглээг тайлбарласан.

Дөрөвдүгээр хэсэгт тестийн үр дүнгийн үнэлэлтэд зарим симуляци хийсэн.

Тавдугаар хэсэгт бодит тоон өгөгдөлд анализ хийж, дэвшүүлсэн аргын хэрэглээг оруулсан.

Өгүүллийн төгсгөл хэсэгт зарим санамжийг оруулсан.

2. Локал шугаман үнэлэлтийн товч тайлбар

Параметрийн бус univariate регрессийн загварыг авч үзсэн.

$$Y_i = m\left(X_i\right) + \sigma\left(X_i\right) e_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

Энд Y, X-ээр харгалзан хамаарах ба тайлбарлагч хувьсагчийг харин $(Y_i, X_i)(i = 1,2,\dots,n)$ нь (1) загвараас санамсаргүй сонгосон түүвэр болно. $m(\cdot), \sigma(\cdot)$ нь үл мэдэгдэх регресс ба дисперсийн функц. $e_1, e_2, \dots, e_n$ нь хоорондоо үл хамаарах ба 0 дундаж, 1 дисперстэй нэгэн ижил тархалттай санамсаргүй хувьсагчууд. Мөн X ба e нь үл хамаарах.

Хэд хэдэн математик харьцааны тусламжтайгаар локал шугаман үнэлэлтийн зарчимыг (1) загварт туршиж үзсэн. Ялангуяа (1) загварын $m(x)$ регрессийн функцийн II эрэмбийн уламжлал нь X хувьсагчийн D мужид тасралтгүй ба x_0 нь D мужид өгөгдсөн цэг. Тейлорын өргөтгөл ёсоор x_0 цэгийн орчинд

$$m\left(x\right) \approx m\left(x_0\right) + m'\left(x_0\right)\left(x - x_0\right), \tag{2}$$

байна. Энд $m'(x_0)$ нь x_0 цэг дээрх $m(x)$-ийн I эрэмбийн уламжлалыг тэмдэглэв.

$m(x)$-ийн (2) шугаман ойролцооллыг (1) загвар дахь $m(x)$-д орлуулж ХБКА-ыг хэрэглэвэл x_0 цэг дээрх $m(x)$ регрессийн функцийн локал шугаман үнэлэлт дараах жинлэгдсэн ХБКА-аар тооцоогдоно.

$$\text{minimize} \quad \sum_{i=1}^{n}\left[Y_i - m(x_0) - m'(x_0)(X_i - x_0)\right]^2 K_h\left(X_i - x_0\right) \tag{3}$$

Энэ нь $m(x_0), m'(x_0)$ –оос хамаарна. Энд $K_h(\cdot) = K(\cdot/h)/h$ ба $K(\cdot)$ нь тэгш хэмт магадлалын нягтын функцтэй Кернелийн функцаар тодорхойлогдоно. Мөн h cross-validation (солбицуулсан сөргүүлэн батлах), өргөтгөсөн cross-validation, засварлагдсан AIC аргууд болон тоон өгөгдөлд суурилсан аргаар тодорхойлогдсон бүлгийн өргөн. Ялангуяа сөргүүлэн батлах арга нь h -ийг

$$\mathrm{CV}\,(h) = \sum_{i=1}^{n} \left[Y_i - \widehat{Y}_{(i)}(h) \right]^2, \tag{4}$$

илэрхийллийн хамгийн бага утгаар сонгодог.

Бид тохиромжтойг бодож матрицан тэмдэглэгээгээр бичье.

$$\mathbf{X}\,(0) = \begin{pmatrix} 1 & X_1 - x_0 \\ 1 & X_2 - x_0 \\ \vdots & \vdots \\ 1 & X_n - x_0 \end{pmatrix}, \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix},$$

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \sigma\,(X_1)\, e_1 \\ \sigma\,(X_2)\, e_2 \\ \vdots \\ \sigma\,(X_n)\, e_n \end{pmatrix}, \tag{5}$$

$$\mathbf{W}\,(0) = \mathrm{Diag}\left(K_h\,(X_1 - x_0)\,, \; K_h\,(X_2 - x_0)\,, \ldots, K_h\,(X_n - x_0)\right).$$

(3) бодлогыг жинлэгдсэн ХБКА-аар тооцвол $x = x_0$ цэг дээрх $m(x)$ −ийн локал шугаман үнэлэлт

$$\widehat{m}\,(x_0) = \mathbf{e}_1^{\mathrm{T}} \left[\mathbf{X}^{\mathrm{T}}(0)\mathbf{W}(0)\mathbf{X}(0) \right]^{-1} \mathbf{X}^{\mathrm{T}}\,(0)\,\mathbf{W}\,(0)\,\mathbf{Y}, \tag{6}$$

болно. Энд e_1 эхний элемент нь 1, бусад элемент нь 0 байх хоёр хэмжээст вектор.

(6) дахь x_0-г $X_1, X_2, \ldots, X_n$ тус бүр авах ба $Y = (Y_1, Y_2, \ldots, Y_n)^T$-ийн тохиромжтой утга болно. Түүнийг $Y = (Y_1, Y_2, \ldots, Y_n)^T$ гэж тэмдэглэсэн.

$$\widehat{\mathbf{Y}} = \begin{pmatrix} \widehat{m}(X_1) \\ \widehat{m}(X_2) \\ \vdots \\ \widehat{m}(X_n) \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{e}_1^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}(1)\mathbf{W}(1)\mathbf{X}(1)\right]^{-1}\mathbf{X}^{\mathrm{T}}(1)\,\mathbf{W}(1)\,\mathbf{Y} \\ \mathbf{e}_1^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}(2)\mathbf{W}(2)\mathbf{X}(2)\right]^{-1}\mathbf{X}^{\mathrm{T}}(2)\,\mathbf{W}(2)\,\mathbf{Y} \\ \vdots \\ \mathbf{e}_1^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}(n)\mathbf{W}(n)\mathbf{X}(n)\right]^{-1}\mathbf{X}^{\mathrm{T}}(n)\,\mathbf{W}(n)\,\mathbf{Y} \end{pmatrix} = \mathbf{LY}, \tag{7}$$

Энд

$$\mathbf{L} = \begin{pmatrix} \mathbf{e}_1^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}(1)\mathbf{W}(1)\mathbf{X}(1)\right]^{-1}\mathbf{X}^{\mathrm{T}}(1)\,\mathbf{W}(1) \\ \mathbf{e}_1^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}(2)\mathbf{W}(2)\mathbf{X}(2)\right]^{-1}\mathbf{X}^{\mathrm{T}}(2)\,\mathbf{W}(2) \\ \vdots \\ \mathbf{e}_1^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}(n)\mathbf{W}(n)\mathbf{X}(n)\right]^{-1}\mathbf{X}^{\mathrm{T}}(n)\,\mathbf{W}(n) \end{pmatrix} \tag{8}$$

гэж тодорхойлогдох ба “hat” матриц гэж нэрлэдэг.
Түүнчлэн үлдэгдлийн вектор

$$\widehat{\boldsymbol{\varepsilon}} = (\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \ldots, \widehat{\varepsilon}_n)^{\mathrm{T}} = \mathbf{Y} - \widehat{\mathbf{Y}} = (\mathbf{I} - \mathbf{L})\,\mathbf{Y}, \tag{9}$$

гэж тодорхойлогдох ба үүнийг дараагийн хэсэгт ашиглах болно.

3. Параметрийн бус регрессийн гетероскедастикийг илрүүлэх арга

Бодит тоон шинжилгээнд алдаанууд гомоскедастик шинжтэй эсэхийг урьдчилан мэдэх нь ховор болохыг удиртгал хэсэгт бид дурьдсан.
Одоо

$$\mathrm{H}_0: \ \sigma^2(X_i) = \sigma^2 \longleftrightarrow \mathrm{H}_1: \sigma^2(X_i) \neq \sigma^2, \tag{10}$$

таамаглалыг шалгая. Энд $\sigma^2 > 0$ тогтмол.
$\widehat{\boldsymbol{\varepsilon}} = (\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \ldots, \widehat{\varepsilon}_n)^{\mathrm{T}} = \mathbf{Y} - \widehat{\mathbf{Y}} = (\mathbf{I} - \mathbf{L})\,\mathbf{Y},$ гэж тодорхойлогдох үлдэгдлийн вектор байг.
Параметрийн бус регрессийн алдааны гетероскедастикийг тооцох тохиромжтой тестийн статистикийг зохиохын тулд бид

$$r_i = \frac{\widehat{\varepsilon}_i}{\sqrt{\widehat{\sigma}_0^2 h_{ii}}}, \quad i = 1, 2, \ldots, n, \tag{11}$$

хувиргасан үлдэгдлүүдийг ашигласан. Энд

$$\widehat{\sigma}_0^2 = \frac{\mathbf{Y}^{\mathrm{T}}(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\,\mathbf{Y}}{\mathrm{tr}\left[(\mathbf{I}-\mathbf{L})^{\mathrm{T}}(\mathbf{I}-\mathbf{L})\right]} \tag{12}$$

ба "tr" нь $H = (I-L)^T(I-L)$ матрицын trace утга, h_{ii} нь энэ матрицын диагоналийн i-р элемент.

Хэрэв (10) тэг таамаглал үнэн бол (1) загварын алдааны дисперс тогтмол, ба хэрэв гетероскедастик шинжтэй байвал $r_i^2,\ i = 1,2,\dots,n$ –ийн утгууд трендтэй байж чадахгүй. Ийм шинжтэй байхад $r_1^2, r_2^2, \dots, r_n^2$-үүд нь өөрчлөлттэй байна. Иймээс бид r_i^2-ийн трендийн анализаар алдааны гетероскедастикийг шалгаж чадна. Эндээс (10) таамаглал нь дараах таамаглалтай нэгэн утгатай.

$$\begin{aligned} &\mathrm{H}_0 : r_1^2, r_2^2, \dots, r_n^2 \text{ have no trend} \\ &\quad\longleftrightarrow \mathrm{H}_1 : r_1^2, r_2^2, \dots, r_n^2 \text{ have certain trend.} \end{aligned} \tag{13}$$

Diblasi, Bowman [15] нарын номд өгүүлсэнээр $r_1, r_2, \dots, r_n$ санамсаргүй хувьсагчууд нь үл хамаарах ба нэгэн ижил тархалттай.

$$c = \begin{cases} \dfrac{n}{2}, & n \text{ is even;} \\ \dfrac{n+1}{2}, & n \text{ is odd,} \end{cases} \qquad n' = \begin{cases} c, & n \text{ is even;} \\ c-1, & n \text{ is odd,} \end{cases}$$

$$D_i = r_i^2 - r_{i+c}^2, \quad i = 1,2,\dots,n'. \tag{14}$$

$D_1, D_2, \dots, D_n$ нь H_0 таамаглал үнэн үед үл хамаарах ба $P_{H_0}(D_i > 0) = P_{H_0}(D_i < 0) = 1/2$. Иймээс тестийн статистик

$$T = \sum_{i=1}^{n} I\left(D_i > 0\right), \tag{15}$$

тодорхойлогдоно. Энд $I(\cdot)$ заагдсан функц.

Хэрэв (10) H_0 (эсвэл (13)) тэг таамаглал үнэн бол алдааны загвар гомоскедастик ба

$$T \sim B\left(n', \frac{1}{2}\right), \tag{16}$$

байна. Энд $B\left(n', \frac{1}{2}\right)$-ээр n' түүврийн хэмжээтэй $\frac{1}{2}$ параметртэй бином тархалтыг тэмдэглэв. Тэмдэглэснээр тестийн статистик T нь $\frac{n'}{2}$-той симметр эсвэл ойролцоо симметр байна. Хэрэв алдаа гетероскедастик шинжтэй бол тестийн статистик $\left|T - \frac{n'}{2}\right|$ утга ихсэх хандлагатай. Иймээс H_1-тэй харьцуулахад H_0 тестийн P утга T статистик дээр үндэслэсэн ба

$$
\begin{aligned}
p &= P_{\mathrm{H}_0}\left(\left|T-\frac{n'}{2}\right| \ge \left|t-\frac{n'}{2}\right|\right)\\
&= P_{\mathrm{H}_0}\left(T \le \frac{n'}{2}-\left|t-\frac{n'}{2}\right|\right)+P_{\mathrm{H}_0}\left(T \ge \frac{n'}{2}+\left|t-\frac{n'}{2}\right|\right)\\
&= 2P_{\mathrm{H}_0}\left(T \le \frac{n'}{2}-\left|t-\frac{n'}{2}\right|\right)\\
&= 2\times\frac{1}{2^{n'}}\sum_{k=0}^{(n'/2)-|t-n'/2|} C_{n'}^{k}\\
&\approx 2\left[1-\Phi\left(\frac{|T-n'/2|}{\sqrt{n'}/2}\right)\right],
\end{aligned} \tag{17}
$$

тодорхойлогдоно. Энд t (15)-аар тодорхойлогдсон T-ийн утга. α ач холбогдолын түвшин өгөгдөх ба хэрэв $p < \alpha$ бол H_0 таамаглалыг няцаана. Харин эсрэг тохиолдолд H_0 таамаглалыг хүлээн авна.

4. Симуляци тооцоолол

Параметрийн бус загварын гетероскедастикийг илрүүлэх Zhang, Mei [7] нарын дэвшүүлсэн тестийн аргыг удиртгал хэсэгт дурьдсан. Энэ аргад тэд t тестийг (9)-д харуулсан үлдэгдлүүдийн квадратад хэрэглэсэн.

Энэ хэсэгт дэвшүүлсэн тестийн аргын үндэслэлтэй эсэхэд үнэлгээ өгөхөөр Zhang, Mei-ны аргатай харьцуулалтыг хийсэн.

Регресс ба дисперсийн дараах 3 төрлийн функцийг авч үзсэн.

$$
\begin{aligned}
&(1)\ m(x) = 1 + x, \sigma(x) = \sigma(1 + a\sin(8x))^2;\\
&(2)\ m(x) = 1 + \sin x, \sigma(x) = \sigma(4 + 4\ a\cos(4x));\\
&(3)\ m(x) = 1 + \sin x, \sigma(x) = \sigma\exp(ax),
\end{aligned}
$$

Энд $\sigma = 0.5$ ба a-тогтмол тоо.

Регресс, дисперсийн функцийн дээрх 3 загварыг харгалзан Model1, Model2, Model3 гэж тэмдэглэе. Загвар тус бүрд X тайлбарлагч хувьсагчийн $X_1, X_2, \dots, X_n$ туршилтын утгуудыг $[0; 1]$ завсарт ижил зайтай ө.х, $X_i = \frac{i}{n}$ байхаар авна. Дисперсийн функцийн хувьд a тогтмолыг 0, 0,5, 1-ээр авсан. $a = 0$ үед алдаа гомоскедастик шинжтэй загварт хамаарах ба a-ын утга өсөхөд дисперсийн функц гомоскедастикаас хазайна. Түүврийн хэмжээг n=100, n=200 гэж авсан.

Model	α	$N(0,1)$		$U(-\sqrt{3},\sqrt{3})$		$\chi^2(4)$	
		Proposed method	Zhang and Mei [7]	Proposed method	Zhang and Mei [7]	Proposed method	Zhang and Mei [7]
				$n = 100$			
Model 1	0	0.022	0.038	0.032	0.022	0.058	0.024
	0.50	0.346	0.014	0.402	0.002	0.348	0.026
	1.00	0.540	0.020	0.570	0.006	0.506	0.012
Model 2	0	0.020	0.038	0.038	0.024	0.048	0.030
	0.50	0.294	0.020	0.490	0.048	0.352	0.064
	1.00	0.996	0.032	1.000	0.056	0.996	0.080
Model 3	0	0.024	0.038	0.038	0.026	0.044	0.032
	0.50	0.162	0.510	0.258	0.796	0.202	0.248
	1.00	0.544	0.954	0.736	1.000	0.574	0.714
				$n = 200$			
Model 1	0	0.034	0.034	0.034	0.018	0.062	0.032
	0.50	0.596	0.012	0.678	0.002	0.620	0.014
	1.00	0.870	0.014	0.884	0.004	0.952	0.016
Model 2	0	0.032	0.040	0.038	0.018	0.058	0.030
	0.50	0.436	0.022	0.690	0.052	0.514	0.068
	1.00	1.000	0.034	1.000	0.058	1.000	0.072
Model 3	0	0.038	0.042	0.038	0.018	0.058	0.030
	0.50	0.236	0.692	0.358	0.946	0.318	0.362
	1.00	0.714	0.996	0.912	1.000	0.750	0.862

Түүнчлэн алдааны тархалтын хувьд тестийн аргуудын робуст \бат бөх-сайн арга\ чанарыг үнэлэхийн тулд $e_1, e_2, \dots, e_n$ санамсаргүй хувьсагчид нь $N(0;1), U(-\sqrt{3}, \sqrt{3})$ ба 4 чөлөөний зэрэгтэй стандартчилагдсан Хи квадрат тархалтуудаас хамааралгүй сонгогдсон.

Өгөгдсөн 1,2,3-р загварын хувьд тогтмол α-ын утга бүрд N=500 удаа симуляци хийсэн. Zhang, Mei-ий арга болон дэвшүүлсэн аргын хувьд тестийн үр дүнг шалгахын тулд симуляцийг 500 удаа хийсэн. Гаусс Кернеллийн функц $K(x) = exp(-x^2/2)/\sqrt{2\pi}$ хэлбэртэй ба h сөргүүлэн-батлах аргаар сонгогдсон. Энэ 500 удаагийн давталтын үр дүнд $\alpha = 0.05$ ач холбогдлын түвшинд тэг таамаглалыг няцаах давталтын тоог бүртгэж авсан бөгөөд харгалзах үр дүнг хүснэгт 1-д харуулсан.

Хүснэгтээс харахад алдааны хувьсагч нормаль тархалттай байхад H_0 таамаглалын хувьд 2 аргыг няцаах давтамж нь 2 түүврийн хэмжээний харгалзах ач холбогдлын түвшинд маш ойрхон байна. Нөгөө талаас 2 тестийн арга нь альтернатив таамаглалын хувьд дисперсийн функцийн ялгаатай хэлбэрүүдийн хувьд нилээд өөр өөр үр дүнг өгч байна. Хэдийгээр бидний дэвшүүлсэн аргаар тооцсон няцаах давтамж нь монотон дисперсийн функцийн хувьд бага хэмжээтэй байх хандлагатай байгаа боловч өндөр давтамжийн дисперсийн функцийн хувьд Zhang, Mei-ий аргаар олж авснаас маш их байна. Энэ нь бидний арга нь гетероскедастикийг илрүүлэхэд хангалттай чадалтай байна гэдэг нь харагдаж байна. Тухайлбал, дисперсийн функц нь олон янзын хувилбартай байхад ч чадалтай байна. Хэрэв загварын хувьд үлдэгдэл хувьсагч нь нормаль биш байвал хүснэгт 1-ээс тэг таамаглалын хувьд бидний аргаар тооцсон магадлал нь алдааны тархалтын өөр өөр түвшинд Zhang, Mei аргаар тооцсон үнэлэгдсэн утга нь илүү тогтвортой ба харгалзах ач холбогдолын түвшинтэй илүү ойрхон байна. Цаашилбал, $\alpha \neq 0$ үед 2 тестийн аргын хувьд няцаах давтамжын утга нь ижилхэн байгаа бөгөөд энэ нь бидний тестийн аргаас өндөр давтамжтай дисперсийн функцийг илрүүлэхэд илүү чадвартай байгааг харуулж байна.

5. Дэвшүүлсэн аргыг хэрэглэсэн жишээ

Энэ хэсэгт дэвшүүлсэн аргыг хэрэглэж бодит тоон өгөгдөлд шинжилгээ хийсэн. Тухайлбал, 1951 оны 1 сарын 1-нээс 2000 оны 12 сарын 31-ны хоорондох Хятад улсын өдөр бүрийн дундаж температурын өгөгдөлийг авсан. 50 жилийн ижил өдрүүдийн дундаж температуруудын багцын дундажийг дундаж температурын утгаар авсан. Харин 50 жилийн турш дахь 2 сарын 29-ны өдрийн өгөгдөлийг хасаж тооцоход буруудах зүйлгүй. Мөн түүнчлэн регрессийн функцийн тайлбарлагч хувьсагч X-ийн туршилтын утгуудыг 1 сарын 1-нээс 12 сарын 31 хүртэлх хугацааны дараалллаар авсан. Зураг 1-д дүрслэгдсэн $(AT_i, X_i), i = 1,2,\ldots,365$ туршилтын утгуудад үндэслэн параметрийн бус регрессийн загварыг

$$\mathrm{AT}_i = m(X_i) + \sigma(X_i)\, e_i, \quad i = 1, 2, \ldots, 365, \qquad (18)$$

хэлбэртэй авсан. Энд e_i нь $E = (e_i/X_i) = 0$ ба $Var = (e_i/X_i) = 1$ нөхцөл хангана гэж үзсэн. Бид загварын алдаа гомоскедастик шинжтэй эсэхийг шалгана. Гаусс Кернелийн функцтэй дэвшүүлсэн тестийн аргыг ашиглан сэргүүлсэн баталгаа зарчимаар тооцсон h бүлгийн өргөний оновчтой утга h=2 ба харгалзах p-ийн утга 4.713×10^{-27} байна.

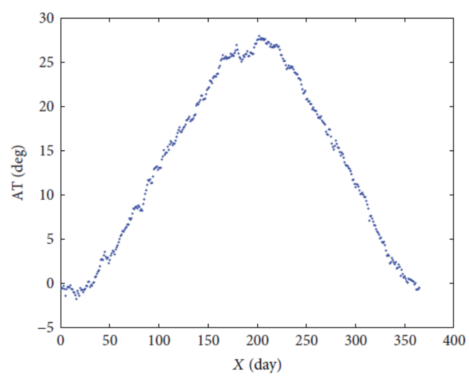

FIGURE 1: The original data of the response AT in model (18).

6. Санамж

Энэ өгүүлэлд алдааны тархалтуудын төрлүүдийг чөлөөтэй сонгосон тестийг параметрийн бус регрессийн загварын гетероскедастикийг илрүүлэхэд өргөтгөсөн. Тухайлбал, Статистик нь регрессийн загварыг локал шугаман үнэлгээтэй тохируулсны дараах болон параметрийн бус статистикийн трендийн анализын санаан дээр үндэслэн гаргасан.

Дэвшүүлсэн аргын гүйцэтгэлд үнэлгээ өгөхийн тулд бид бусад зарчимуудтай харьцуулсан симуляци хийсэн ба өндөр давтамжтай дисперсийн функцийн хувьд хангалттай сайн үр дүн гарсан.

Zhang, Mei-ий аргатай харьцуулахад гетероскедастик шинжтэй үед дэвшүүлсэн аргын чадал нь монотон дисперсийн функцийн хувьд бага байх хандлагатай байсан. Тэр нь эхлээд үлдэгдлийн квадратын монотон хандлага дээр үндэслэгдэж томъёологдсон ба хоёрдугаарт тэмдэгийн тестийн аргад үндэслэгдэж томъёолодсоноос хамаарч байна. Ер нь бидний дэвшүүлсэн арга агуулгын хувьд энгийн, хэрэглэхэд хялбар байгаагаараа параметрийн бус регрессийн гетероскедастик шинжийг шалгахад ач холбогдолтой байна. Ялангуяа дисперсийн функц олон хувилбартай үед илүү үр дүнтэй байна.