

## ANALISIS IPM DI PULAU JAWA MENGGUNAKAN ANALISIS REGRESI KUANTIL

<sup>1</sup>Vendy Eka Wahyudi, <sup>2</sup>Ismaini Zain

<sup>1,2</sup> Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Noverber, Surabaya  
Alamat e-mail : <sup>1</sup>ve.wahyudi@yahoo.com

### ABSTRAK

Analisis regresi merupakan metode statistika untuk mengetahui hubungan antara variabel respon dengan variabel prediktor. Dalam analisis regresi, salah satu metode estimasi yang sering digunakan adalah OLS (*Ordinary Least Square*). Namun metode ini sangat rentan dipengaruhi oleh adanya pencilan. Pencilan dapat menyebabkan hasil estimasi parameter menjadi tidak stabil. Metode OLS dianggap kurang tepat untuk menganalisis sejumlah data yang tidak simetris karena nilai *mean* sebagai penduga bagi nilai tengah data menjadi sangat peka dengan adanya data pencilan. Selanjutnya dikembangkan metode regresi kuantil (*Quantile Regression*). Penduga parameter regresi kuantil diperoleh dengan meminimumkan jumlah nilai mutlak dari error. Secara umum ada beberapa kelebihan dari regresi kuantil, salah satunya adalah dapat meminimumkan pengaruh dari pencilan. Penelitian dan pengaplikasian tentang regresi kuantil banyak dilakukan diberbagai bidang oleh negara-negara di dunia. Dari ide itulah maka penelitian ini membahas tentang model regresi kuantil pada kasus IPM (Indek Pembangunan Manusia).

**Kata Kunci** : IPM, OLS, Regresi, Regresi kuantil

### PENDAHULUAN

Analisis regresi merupakan suatu metode statistik yang mempelajari tentang pola hubungan secara matematis antara variabel respon (Y) dengan salah satu atau lebih prediktor (X). Dalam regresi linear terdapat beberapa metode estimasi parameter, salah satunya adalah metode OLS (*Ordinary Least Square*). Prinsip dari metode ini adalah meminimumkan jumlah kuadrat residu (*error*). Namun demikian, metode ini sangat rentan dipengaruhi adanya pencilan (*outlier*). Pencilan dapat menyebabkan hasil estimasi parameter menjadi tidak stabil. Selain itu, analisis regresi dengan metode OLS didasarkan pada fungsi distribusi *mean*. Nilai *mean* menunjukkan ukuran pemusatan dari suatu distribusi sehingga hanya sedikit informasi yang diketahui dari keseluruhan distribusi.

Oleh sebab itu pendekatan dengan metode ini hanya mampu menduga model dari fungsi bersyarat *mean* dan tidak mewakili keseluruhan data dari distribusi. Pendekatan *mean* menjadi kurang tepat digunakan sebagai penduga bagi nilai tengah data<sup>[17]</sup>. Akhirnya dikembangkan metode regresi kuantil (*Quantile Regression*).

Regresi kuantil ini merupakan metode yang berguna dalam mengestimasi parameter, metode ini tidak mudah terpengaruh oleh kehadiran pencilan sehingga pencilan menjauh dan tidak mengganggu kestabilan data yang diperoleh. Selain itu, metode ini dapat memberikan hasil yang tepat dan stabil pada kehadiran pencilan serta dapat membatasi pengaruh dari pencilan<sup>[12]</sup>. Kelebihan dari regresi kuantil, salah satunya adalah dapat meminimumkan

pengaruh dari pencilan<sup>[19]</sup>. Penaksir parameter yang digunakan dalam regresi kuantil sama dengan metode OLS yaitu meminimumkan jumlah kuadrat sisaan<sup>[17]</sup>.

Penelitian sebelumnya yang berkaitan dengan regresi kuantil pernah dilakukan oleh<sup>[7], [10], [11], [18], [21], [22]</sup>. Analisis regresi kuantil juga pernah diaplikasikan ke bidang lain, seperti sosiologi<sup>[14]; [15]; [16]</sup>, ekologi dan ilmu lingkungan<sup>[9]; [25]</sup>, bidang kesehatan<sup>[2]; [26]</sup>, serta bidang sosial dan kependudukan<sup>[1]</sup>.

IPM (Indek Pembangunan Manusia) merupakan salah satu contoh kasus pada bidang sosial dan kependudukan. IPM adalah suatu tolak ukur angka kesejahteraan suatu daerah atau negara yang dilihat berdasarkan tiga dimensi yaitu: (1) Angka harapan hidup pada waktu lahir (*life expectancy at birth*); (2) Angka melek huruf (*literacy rate*) dan rata-rata lama sekolah (*mean years of schooling*); (3) Kemampuan daya beli (*purchasing power parity*). Ketiga indikator tersebut saling mempengaruhi satu sama lain, selain itu dapat dipengaruhi oleh faktor-faktor lain (*United Nation Development Programme, UNDP, 1990*).

Metode-metode yang pernah dilakukan dalam penelitian tentang faktor-faktor yang mempengaruhi IPM adalah regresi probit spasial<sup>[24]</sup>. Kajian mengenai indikator IPM juga telah banyak dilakukan antara lain, variabel-variabel yang berpengaruh terhadap IPM Sulawesi Selatan adalah pertumbuhan ekonomi, persentase penduduk miskin, pengeluaran pemerintah di bidang pendidikan, pengeluaran pemerintah di bidang kesehatan dan ketimpangan distribusi pendapatan<sup>[23]</sup>. Menguji bagaimana pengaruh pengeluaran pemerintah daerah khususnya bidang pendidikan dan kesehatan, investasi swasta dan distribusi pendapatan proksi indeks Gini terhadap IPM dalam konteks regional (antar provinsi) di Indonesia<sup>[6]</sup>. Semakin tinggi angka keluhan kesehatan maka akan mengurangi angka harapan hidup sehingga

mengakibatkan penurunan IPM suatu daerah<sup>[4]</sup>. Selain itu, rata-rata umur kawin pertama wanita mempunyai pengaruh yang positif terhadap nilai IPM<sup>[3]</sup>.

Indikator-indikator yang mempengaruhi IPM Kabupaten/Kota di Pulau Jawa memiliki data yang menyebar. Selain itu dari diagram pencar antara variabel Y dengan masing-masing variabel X menunjukkan data yang *outlier*<sup>[24]</sup>, maka regresi kuantil cocok diterapkan untuk menganalisis data ini dan sejauh ini penelitian tentang indikator IPM dengan regresi kuantil belum pernah dilakukan.

Model regresi kuantil (*Quantile Regression*) pertama kali diperkenalkan oleh<sup>[19]</sup>, dapat dianggap sebagai perluasan dari model OLS. Secara khusus, regresi OLS hanya memperkirakan bagaimana variabel prediktor terkait dengan nilai rata-rata variabel respon, sedangkan regresi kuantil memungkinkan untuk model prediktor terhadap berbagai lokasi/pengukuran variabel respon. Karena sifatnya yang robust terhadap pencilan maka regresi kuantil cocok untuk menganalisis sejumlah data yang bentuknya tidak simetris serta distribusi datanya tidak homogen.

Secara statistik, fungsi distribusi probabilitas dari variabel random  $X$  dapat dinyatakan sebagai berikut<sup>[8]; [20]</sup>:

$$F(x) = Prob(X \leq x)$$

dan fungsi distribusi probabilitas pada kuantil ke- $\tau$  dari  $X$  dapat ditulis sebagai berikut.

$$F^{-1}(\tau) = inf\{x: F(x) \geq \tau\}$$

di mana  $0 < \tau < 1$ .

Seperti halnya dengan metode OLS yang meminimumkan jumlah kuadrat sisaan untuk mencari nilai dugaan bagi  $\beta$ , maka dalam regresi kuantil, kuantil ke- $\tau$  dari  $F$  dapat diperoleh dengan meminimumkan fungsi (1) terhadap  $\hat{x}$ .

$$E\rho_{\tau}(X - \hat{x}) = (\tau - 1) \int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + \tau \int_{\hat{x}}^{\infty} (x - \hat{x})dF(x) \quad (1)$$

Dengan meminimumkan persamaan (1), maka diperoleh persamaan sebagai berikut.

$$\begin{aligned}
 0 &= (1 - \tau) \int_{-\infty}^{\hat{x}} dF(x) - \tau \int_{\hat{x}}^{\infty} dF(x) \\
 &= (1 - \tau)F(\hat{x}) - \tau[1 - F(\hat{x})] \\
 &= F(\hat{x}) - \tau
 \end{aligned}$$

Sehingga kuantil ke- $\tau$  merupakan solusi dari  $F$ .

Jika  $X$  merupakan fungsi dari  $Y$  yang telah diketahui, memiliki peluang  $F(x)$ , kuantil ke- $\tau$  dari fungsi tersebut dapat ditulis sebagai  $\hat{x} = Q(\tau) = F^{-1}(\tau)$ .  $Q(\tau)$  merupakan fungsi dari  $Y$  yang dapat diselesaikan dengan persamaan sebagai berikut.

$$\begin{aligned}
 &\min(\tau - 1) \int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + \\
 &\tau \int_{\hat{x}}^{\infty} (x - \hat{x})dF(x) \quad (2)
 \end{aligned}$$

$Q(0.5)$  adalah median  $X$  (sebagai fungsi dari  $Y$ ) yang menunjukkan titik simetri dari  $F$ ; untuk  $\tau$  mendekati 0 atau 1,  $Q(\tau)$  menunjukkan ekor kiri atau kanan dari  $F$ . Dalam notasi matrik, jika  $Q(\tau)$  adalah fungsi linear  $X'\beta$  yang dinotasikan dengan  $\xi(\tau)$ , maka persamaan (2) dapat ditulis menjadi.

$$\begin{aligned}
 &\min(\tau - 1) \int_{-\infty}^{X'\beta} (x - X'\beta)dF(x) + \\
 &\tau \int_{X'\beta}^{\infty} (x - X'\beta)dF(x) \quad (3)
 \end{aligned}$$

Solusi dari persamaan (3) ini dinotasikan sebagai  $\beta_0$  dan kuantil  $X$  (sebagai fungsi dari  $Y$ ) ke- $\tau$  adalah  $Q(\tau) = X'\beta_0$ .

Misal diberikan data  $(y_i, x_i)$  untuk  $i = 1, 2, \dots, n$ , di mana  $x_i$  berukuran  $k \times 1$ , maka model linear dari persamaan regresi kuantil dapat dituliskan sebagai berikut.

$$y_i = X_i'\beta + e_i$$

Dengan  $Q_\tau(y_i|x_i) = X_i'\beta$  merupakan kuantil ke- $\tau$  ( $0 < \tau < 1$ ) dari  $y$  dengan suatu nilai  $x_i$  tertentu. Penduga bagi  $\beta$  dari regresi kuantil ke- $\tau$  diperoleh dengan meminimumkan jumlah nilai mutlak dari error dengan pembobot  $\tau$  untuk error positif dan pembobot  $(1 - \tau)$  untuk error negatif yaitu sebagai berikut.

$$\begin{aligned}
 \hat{\beta}(\tau) &= \min_{\beta} \{ \tau \sum_{i: y_i \geq x_i} |y_i - X_i'\beta| + \\
 &(1 - \tau) \sum_{i: y_i < x_i} |y_i - X_i'\beta| \} \quad (4)
 \end{aligned}$$

$$\min_{\xi \in R} \sum_{i=1}^n \rho_{\tau}(y_i - \xi) \quad (5)$$

di mana  $\rho_{\tau}(u) = u(\tau - I(u < 0))$

Estimasi koefisien variabel penjelas kemudian dapat diperoleh dengan memecahkan fungsi sebagai berikut.

$$\hat{\beta}(\tau) = \operatorname{argmin}_{\xi \in R} \sum_{i=1}^n \rho_{\tau}(y_i - X_i'\beta)$$

Solusi dari persamaan (3) tidak dapat diperoleh secara analitik, melainkan secara numerik seperti dengan metode *simplex*, metode *interior point* atau metode *smoothing*.

Interpretasi pada regresi kuantil sedikit berbeda dengan regresi OLS. Dalam regresi OLS, koefisien prediktor tertentu  $X$ , merupakan perubahan yang diharapkan dalam variabel dependen/respon yang terkait dengan suatu unit perubahan dalam  $X$ , sedangkan pada regresi kuantil, koefisien prediktor  $X$  pada kuantil ke- $\tau$  dapat diinterpretasikan secara marginal, di mana relative sesuai dengan nilai kuantil ke- $\tau$  dengan suatu unit perubahan dalam  $X$ . Nilai  $\tau$  berkisar antara 0 dan 1, koefisien yang diestimasi dalam pemodelan dapat lebih dari satu atau banyak, dan umumnya nilai kuartil yang digunakan adalah 0,5.

Dalam analisis regresi kuantil terdapat tiga metode untuk menghitung selang kepercayaan bagi  $\beta$ , yaitu metode *sparsity*, metode *rank*, dan metode *resampling* [19].

## METODE PENELITIAN

### Sumber Data dan Variabel Penelitian

Penelitian ini akan membahas indikator yang mempengaruhi IPM di Pulau Jawa dengan menggunakan regresi kuantil. Data yang digunakan dalam penelitian ini adalah data sekunder hasil Survei Sosial Ekonomi Nasional (SUSENAS) di Pulau Jawa oleh BPS (Badan Pusat Statistik) tahun 2010. Unit pengamatan pada penelitian ini terdiri dari 118 kabupaten/kota di Pulau Jawa.

Variabel yang digunakan pada penelitian ini sebanyak 7 variabel. Variabel tersebut terbagi atas satu variabel

respon dan enam variabel prediktor. Variabel yang berperan sebagai variabel respon (Y) adalah Indeks Pembangunan Manusia (IPM). Variabel respon ini bersifat continue.

Sedangkan variabel prediktor yang dipakai dalam penelitian ini meliputi

1. Persentase penduduk yang tinggal di perkotaan ( $X_1$ ) adalah jumlah penduduk yang tinggal di daerah perkotaan dalam jangka waktu tertentu.
2. Persentase penduduk yang berpendidikan di atas SLTP ( $X_2$ ) adalah penduduk yang telah menamatkan pendidikan setingkat SLTP atau jenjang pendidikan yang lebih tinggi.
3. Rasio ketergantungan penduduk ( $X_3$ ) menunjukkan beban tanggungan penduduk usia produktif (15-64 tahun) terhadap penduduk usia muda (kurang dari 15 tahun) dan penduduk usia tua (65 tahun atau lebih) <sup>[3]</sup>.
4. Peranan sektor industri dalam PDRB ( $X_4$ ). PDRB merupakan jumlah nilai tambah atas barang dan jasa yang dihasilkan oleh berbagai unit produksi di suatu wilayah dalam jangka waktu tertentu.
5. Persentase penduduk yang mengalami keluhan kesehatan ( $X_5$ ). Semakin tinggi angka keluhan kesehatan maka akan mengurangi angka harapan hidup sehingga mengakibatkan penurunan IPM suatu daerah <sup>[4]</sup>.
6. Rata-rata umur kawin pertama wanita ( $X_8$ ). Semakin tinggi rata-rata umur kawin pertama wanita di suatu provinsi menyebabkan nilai IPM di provinsi tersebut semakin tinggi <sup>[3]</sup>.

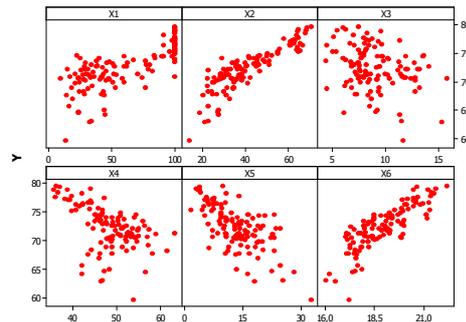
**Metode Analisis**

Berdasarkan studi awal menunjukkan adanya pencilan pada data, langkah pertama dalam analisis regresi kuantil adalah melakukan estimasi pada nilai kuantil, yaitu  $\tau = 0,50$ . Parameter regresi

kuantil diduga dengan metode simpleks, selang kepercayaan koefisien regresi kuantil diperoleh dengan metode resampling.

**HASIL PENELITIAN**

*Scatterplot* antara nilai IPM dan indikator yang mempengaruhi dapat dilihat pada Gambar 1.



**Gambar 1** *Scatterplot* antara nilai IPM dan faktor-faktor yang mempengaruhinya

*Scatterplot* antara variabel Y dengan variabel X terlihat bahwa terdapat pencilan data yaitu pada kabupaten Sampang, Bondowoso, Probolinggo, Situbondo, Pamekasan dan Jember. Langkah selanjutnya adalah melakukan estimasi nilai kuantil  $\tau = 0,50$  untuk masing-masing koefisien regresi. Hasil estimasi dengan regresi kuantil ditunjukkan pada Tabel 1.

**Tabel 1** Estimasi Parameter Regresi Kuantil

Variabel	Estimasi	p_value
Konstanta	43.745330	0.000
X1	0.005925	0.540
X2	0.129669	0.001
X3	0.138378	0.048
X4	0.115428	0.005
X5	-0.168921	0.000
X6	0.963710	0.000

Berdasarkan hasil yang diperoleh pada Tabel 1, pemodelan indikator yang mempengaruhi IPM Kabupaten/Kota di Pulau Jawa dengan pendekatan regresi kuantil adalah sebagai berikut.

$$y = 43.75 + 0.0059X_1 + 0.1297X_2 + 0.1384X_3 + 0.1154X_4 - 0.1689X_5 + 0.9637X_6$$

Pada metode regresi kuantil, variabel-variabel yang berpengaruh signifikan pada taraf  $\alpha = 5\%$  adalah persentase penduduk yang berpendidikan di atas SLTP ( $X_2$ ), rasio ketergantungan penduduk ( $X_3$ ), peranan sektor industri dalam PDRB ( $X_4$ ), persentase penduduk yang mengalami keluhan kesehatan ( $X_5$ ), dan rata-rata umur kawin pertama wanita ( $X_6$ ). Nilai dari koefisien determinasi atau R square adalah 0.5726 hal ini menunjukkan bahwa 57% indikator IPM mampu dijelaskan oleh variabel-variabel  $X_1, X_2, X_3, X_4, X_5, X_6$ , sedangkan 43% sisanya dijelaskan oleh variabel lainnya.

Dari model  $y = 43.75 + 0.0059X_1 + 0.1297X_2 + 0.1384X_3 + 0.1154X_4 - 0.1689X_5 + 0.9637X_6$  dapat dijelaskan bahwa apabila nilai dari semua variabel  $X$  sama dengan nol maka nilai IPM adalah 43.75. Sebaliknya, dengan kenaikan 1 satuan persentase penduduk yang tinggal di perkotaan dan nol buat variabel yang lainnya maka akan meningkatkan IPM sekitar 0.0059. Kenaikan 1 satuan persentase penduduk yang mengalami keluhan kesehatan dan nol buat variabel yang lainnya maka akan mengurangi IPM sekitar 0.1689.

Selang kepercayaan bagi penduga parameter  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$  dengan nilai  $\alpha = 5\%$  adalah sebagai berikut.

**Tabel 2** Selang Kepercayaan

Variabel	Selang Kepercayaan	
	batas bawah	batas atas
Konstanta	36.445270	51.04540
X1	-0.013194	0.02504
X2	0.055973	0.20336
X3	0.000975	0.27578
X4	0.035489	0.19537
X5	-0.251409	-0.08643
X6	0.494532	1.43289

## KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa dengan menggunakan regresi kuantil indikator yang mempengaruhi IPM Kabupaten/Kota di pulau Jawa adalah persentase penduduk yang berpendidikan di atas SLTP ( $X_2$ ), rasio ketergantungan penduduk ( $X_3$ ), peranan sektor industri dalam PDRB ( $X_4$ ), persentase penduduk yang mengalami keluhan kesehatan ( $X_5$ ), dan rata-rata umur kawin pertama wanita ( $X_6$ ). Nilai dari koefisien determinasi atau R square adalah 0.5726 hal ini menunjukkan bahwa 57% indikator IPM mampu dijelaskan oleh variabel-variabel  $X_1, X_2, X_3, X_4, X_5, X_6$ , sedangkan 43% sisanya dijelaskan oleh variabel lainnya.

## DAFTAR PUSTAKA

- [1] Abreveya, J. (2001). The effects of Demographics and Maternal Behavior on the Distribution of Birth Outcomes. *Empirical Economics*, 26, 247-257.
- [2] Austin, P., Tu, J., Daly, P., & Alter, D. (2005). The use of quantile regression in health care research: A case study examining gender differences in the timeliness of thrombolytic therapy. *Statistics in Medicine*, 24, 791-816.
- [3] BPS. (2008). *Indeks Pembangunan Manusia 2009-2010*. BPS, Jakarta.
- [4] BPS. (2009). *Indeks Pembangunan Manusia 2009-2010*. BPS, Jakarta.
- [5] BPS., (2011). *Indeks Pembangunan Manusia 2009-2010*. BPS, Jakarta.
- [6] Brata, A.G. (2005). *Investasi Sektor Publik Lokal, Pembangunan Manusia, dan Kemiskinan*. Yogyakarta: Lembaga Penelitian Universitas Atma Jaya.
- [7] Buchinsky, M. (1994). Changes in the U.S. wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62, 405-458.

- 
- 
- [8] Buhai, S. (2005). Quantile regression: overview and selected applications. *Ad-Astra The Young Romanian Scientists' Journal*, 5(1), 1e17
- [9] Cade, B. S., Terrell, J. W., & Schroeder, R. L. (1999). Estimating effects of limiting factors with regression quantiles. *Ecology*, 80, 311–323.
- [10] Chamberlain, G. (1994). *Quantile regression, censoring and the structure of wages*. In C. Skins (Ed.), *Advances in Econometrics* (pp. 171–209). Cambridge, UK: Cambridge University Press.
- [11] Eide, E. R., & Showalter, M. H. (1999). Factors affecting the transmission of earnings across generations: A quantile regression approach. *The Journal of Human Resources*, 34, 253–267.
- [12] Furno, M. (2007). Parameter Instability in Kuantil Regressions. *Statistical Modelling*. 7(4) : 345-362.
- [13] Ginanjar, K. (1996). *Pembangunan Untuk Rakyat, Memadukan Pertumbuhan dan Pemerataan*. Jakarta, Pusat Cidessindo.
- [14] Hao, L. (2005). Immigration and wealth inequality: A distributional approach. *Invited seminar at The Center for the Study of Wealth and Inequality*, Columbia University.
- [15] Hao, L. (2006a). Sources of wealth inequality: Analyzing conditional distribution. *Invited seminar at The Center for Advanced Social Science Research*, New York University.
- [16] Hao, L. (2006b). Sources of wealth inequality: Analyzing conditional location and shape shifts. *Paper presented at the Research Committee on Social Stratification and Mobility (RC28) of the International Sociological Association (ISA) Spring meeting 2006 in Nijmegen, the Netherlands*.
- [17] Hao, L., & Naiman, D. Q. (2007). *Quantile Regression*. Sage Publications, Inc.
- [18] Kim, T., & White, H. (2003). Estimation, inference and specification testing for possibly misspecified quantile regression. In Fomby T dan Hills RC eds, *Maximum likelihood estimation of misspecified models: twenty years later*. New York: Elsevier, 107-132.
- [19] Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica: Journal of the Econometric Society*, 46(1), 33e50.
- [20] Koenker, R., & Hallock, K. F. (2001). Quantile regression: an introduction. *The Journal of Economic Perspectives*, 15(4), 143 e156.
- [21] Machado, J., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20,445–465.
- [22] Melly, B. (2005). *Decomposition of differences in distribution using quantile regression*. Labour Economics, 12, 577–590.
- [23] Patta, D. (2012). Analisis Faktor-faktor yang Mempengaruhi Indeks Pembangunan Manusia di Sulawesi Selatan Periode 2010-2011. *Tesis Master*. Universitas Hasanudin.
- [24] Puspit, F. I. (2013). Model Probit Spasial pada Faktor-Faktor yang Mempengaruhi Klasifikasi IPM di Pulau Jawa. *Tesis Master*. ITS. Surabaya.
- [25] Scharf, F. S., Juanes, F., & Sutherland, M. (1989). Inferring ecological relationships from the edges of scatter diagrams: Comparison of regression techniques. *Ecology*, 79, 448–460.
- [26] Wei, Y., Pere, A., Koenker, R., & He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, 25, 1369–1382.