

# PEMODELAN POISSON RIDGE REGRESSION (PRR) PADA BANYAK KEMATIAN BAYI DI JAWA TENGAH\*

Wulandari

Badan Pusat Statistik Kabupaten Wonosobo, Indonesia, wulanda@bps.go.id

**Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)**  
**Vol 4 No 2 (2020), 392 - 400**

Copyright © 2020 Wulandari. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

The decline of infant mortality is one of the targets of the Indonesian government in the health sector, including the Government of Central Java. To achieve this goal, it is necessary to identify factors that affect many infant mortalities in the district/city of Central Java. Infant mortalities are count data, so Poisson regression is commonly used. The data in the study showed the existence of multicollinearity in several predictor variables, so an appropriate model was needed. Poisson Ridge Regression (PRR) is a Poisson modeling that accommodates multicollinearity. In this study, the PRR model was used to model infant mortality in Central Java district/city. The results showed that the parameter estimation of the PRR model was slightly different than the estimated Poisson regression model. Modeling infant mortality with the PRR model, out of five predictor variables, three variables harmed many infant deaths, while the other two variables had a positive effect on many infant deaths.

**Keywords:** infant mortality, multicollinearity, poisson ridge regression.

## 1. Pendahuluan

Kehidupan sehat dan kesejahteraan merupakan salah satu tujuan yang terdapat dalam 17 Tujuan Pembangunan Berkelanjutan atau SDGs. SDGs adalah seperangkat tujuan, sasaran, dan indikator pembangunan yang bersifat universal, yang disepakati bersama oleh berbagai negara di dunia, Indonesia adalah salah satunya. Salah satu tujuan di bidang kesehatan yang ditetapkan oleh Indonesia, termasuk Provinsi Jawa Tengah adalah penurunan angka kematian bayi (AKB). Angka kematian bayi

---

\* Received Aug 2019; Accepted Jul 2020; Published online on Jul 2020

didefinisikan sebagai banyak kematian bayi, yaitu usia 0 – 11 bulan, setiap 1.000 kelahiran dalam kurun waktu satu tahun (BPS (2016)).

Berdasarkan data Badan Pusat Statistik, kenaikan AKB pada tahun 2012 terjadi di semua provinsi di Indonesia. Salah satu provinsi dengan kenaikan AKB tertinggi adalah Provinsi Jawa Tengah. AKB Jawa Tengah mengalami kenaikan tertinggi yaitu 52% jika dibanding tahun 2010. Bahkan AKB di Jawa Tengah adalah yang paling tinggi, sama seperti Provinsi Banten yaitu 32%. Untuk mencapai tujuan penurunan AKB tersebut, perlu diidentifikasi faktor-faktor yang mempengaruhi banyak kematian bayi. Alat analisis yang bisa digunakan adalah analisis regresi.

Banyak kematian bayi adalah data cacah, sehingga membutuhkan analisis regresi yang sesuai untuk data cacah agar estimasi parameter yang dihasilkan tidak bias, yaitu regresi Poisson (Cameron & Trivedi, 2013). Terdapat beberapa asumsi yang harus dipenuhi agar regresi menghasilkan estimasi parameter yang bersifat *unbiased* dan varians minimum, salah satunya tidak terdapatnya kolinieritas antar variabel penjelas. Terlanggarnya asumsi ini akan menyebabkan estimasi parameter menjadi tidak akurat serta *standard error* model menjadi besar (Gujarati, 2009). Beberapa cara dapat digunakan untuk mengatasi adalah multikolinieritas antar variabel penjelas, salah satunya adalah penggunaan *ridge* model. Prinsip *ridge* adalah dengan menambahkan suatu konstanta yang nilainya kecil pada diagonal utama matrik  $X'X$ .

Dalam penelitian ini, untuk memodelkan banyak kematian bayi dengan adanya kasus multikolinieritas, seperti yang akan dijelaskan pada bagian Hasil dan Pembahasan, digunakan analisis *Poisson Ridge Regression* (PRR). PRR adalah analisis yang dikembangkan Månsson & Shukur (2011), yang kemudian mengalami pengembangan model untuk data cacah yang lain selain Poisson oleh peneliti-peneliti lain pada beberapa tahun kemudian.

## 2. Metodologi

### 2.1 Bahan dan Data

Penelitian ini meliputi kabupaten/kota di Jawa Tengah yaitu sebanyak 35 kabupaten/kota. Variabel respon dalam penelitian adalah banyak kematian bayi (Y), sedangkan variabel prediktor terdiri dari 5 variabel yaitu banyak bayi yang memperoleh asi eksklusif ( $X_1$ ), banyak bayi yang memperoleh imunisasi dasar lengkap ( $X_2$ ), cakupan pelayanan kesehatan bayi ( $X_3$ ), jumlah rumah yang memenuhi syarat rumah sehat ( $X_4$ ), dan jumlah bayi berat badan lahir rendah ( $X_5$ ). Data yang digunakan dalam penelitian adalah data sekunder yang diperoleh dari publikasi “Jawa Tengah Dalam Angka 2016” (BPS, 2016) dan “Profil Kesehatan Jawa Tengah Tahun 2015” (Dinkes, 2015). Pengolahan dalam penelitian ini dilakukan dengan menggunakan software R.

### 2.2 Metode Penelitian

Tahapan analisis yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Memodelkan data dengan regresi Poisson, kemudian cek heterogenitas dengan menggunakan uji Breusch-Pagan

Data banyak kematian bayi dimodelkan dengan regresi Poisson, selanjutnya diuji

heterogenitasnya dengan uji Breusch-Pagan. Jika struktur varian-kovarian bersifat homoskedas maka bisa menggunakan analisis Poisson dan lanjut uji yang lain. Jika bersifat heteroskedas maka diperlakukan perlakuan untuk mengatasi hal tersebut.

2. Cek multikolinieritas dengan melihat nilai VIF  
Selanjutnya adalah uji multikolinieritas model Poisson dengan menggunakan nilai VIF. Jika nilai VIF < 10 maka tidak disimpulkan tidak terjadi multikolinieritas sedangkan jika nilai VIF > 10 disimpulkan terjadi multikolinieritas dan dibutuhkan beberapa alternatif penanganan multikolinieritas salah satu dengan metode *ridge*. Cara inilah yang akan digunakan pada penelitian ini.
3. Menghitung nilai estimasi parameter *ridge*  
Sebelum melakukan pemodelan dengan *ridge regression*, dihitung dulu estimasi parameter *ridge*. Selanjutnya, angka estimasi parameter *ridge* tersebut akan digunakan untuk memodelkan data dengan PRR
4. Memodelkan dengan PRR dan menghitung estimasi parameter model PRR  
Memodelkan data dengan model PRR dan selanjutnya kita akan mendapatkan estimasi parameter model PRR dan mengetahui besaran pengaruh variabel-variabel bebas terhadap variabel tak bebas dalam model.
5. Analisis data  
Analisis berupa pengaruh variabel-variabel bebas terhadap variabel tak bebas dalam model.

#### a. Distribusi Poisson

Distribusi Poisson merupakan distribusi kejadian pada interval waktu atau ruang tertentu, dengan hasil pengamatan berupa data diskrit. Misal  $Y$  adalah variabel acak berdistribusi Poisson dengan parameter  $\mu$ , maka fungsi peluangnya adalah sebagai berikut (Cameron & Trivedi, 2013):

$$f_Y(y_i, \mu_i) = \begin{cases} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, & y_i = 0, 1, 2, \dots \\ 0 & , \text{lainnya} \end{cases} \quad (1)$$

#### b. Regresi Poisson

Regresi Poisson adalah regresi standar untuk memodelkan data yang berbentuk cacah (*count*). Model umum regresi Poisson adalah:

$$\ln(\mu_i) = \beta_0 + \sum_{j=1}^k \beta_j x_{ji} \quad (2)$$

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$$

dimana

$$i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, k$$

Penaksiran parameter regresi Poisson dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) dan penyelesaian akhirnya menggunakan iterasi *Newton-Raphson Iteratively Reweighted Least Square* (IRWLS). Bentuk umum rumus estimasi parameter regresi Poisson dengan *intersept*, dalam kondisi konvergen

adalah:

$$\hat{\beta}_{(m+1)} = (\mathbf{X}^T \mathbf{A}_{(m)} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{A}_{(m)} \mathbf{z}_{(m)}) \tag{3}$$

dimana:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix}$$

$\hat{\beta}_{(m+1)}$  adalah matrik koefisien regresi yang diperoleh pada iterasi ke-( $m+1$ )  
 $\mathbf{A}$  adalah matrik penimbang varian, yaitu matrik diagonal ukuran  $n \times n$  yang elemen diagonalnya adalah  $\hat{\mu}_i = \exp(\mathbf{x}_i \hat{\beta}_{(m)})$

$$\mathbf{z}_{(m)} = (z_{1(m)} \ z_{2(m)} \ \dots \ z_{n(m)})^T$$

dengan  $z_{i(m)} = \mathbf{x}_i^T \hat{\beta}_{(m)} + \frac{y_i - \hat{y}_{i(m)}}{\hat{y}_{i(m)}}$

**c. Uji Heterogenitas (Breusch-Pagan Test)**

Dalam analisis regresi, salah satu asumsi yang harus dipenuhi adalah bahwa struktur varian-kovarian dari model harus bersifat homoskedasti. Uji heterogenitas yang dilakukan dalam penelitian ini adalah dengan menggunakan Uji Breusch-Pagan. Hipotesis yang digunakan adalah:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1: \text{minimal terdapat satu } \sigma_i^2 \neq \sigma$$

Statistik uji Breusch-Pagan adalah:

$$BP = \frac{1}{2} \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \sim \chi_k^2 \tag{4}$$

dimana  $i=1,2,\dots, n$  adalah banyak observasi;  $\mathbf{f} = (f_1 \ f_2 \ \dots \ f_n)^T$ ,  $f_i = \left(\frac{y_i - \hat{y}_i}{\sigma^2} - 1\right)$ ;  $\mathbf{Z}$  adalah matrik berukuran  $n \times k$  yang berisi variabel prediktor yang sudah distandarkan. Kriteria uji yang digunakan adalah tolak hipotesis nol jika nilai  $BP > \chi_k^2$ , dengan  $k$  adalah banyaknya variabel prediktor.

**d. Multikolinieritas**

Multikolinieritas adalah terdapatnya hubungan linier yang sempurna atau hampir sempurna antara beberapa atau semua variabel penjelas model. Multikolinieritas sempurna menyebabkan koefisien parameter regresi menjadi tidak dapat ditentukan dan memiliki *standard error* yang tidak terhingga. Sedangkan pada multikolinieritas yang kurang sempurna, walaupun koefisien parameter regresi dapat ditentukan, tetapi memiliki *standard error* yang besar, yang berarti koefisien regresi tersebut tidak dapat diestimasi dengan akurat (Gujarati, 2009).

Terdapat berbagai cara untuk mendeteksi multikolinieritas, salah satunya dengan melihat nilai VIF (*variance inflation factor*). Nilai VIF dari suatu variabel penjelas menunjukkan seberapa besar estimasi koefisien regresi dari variabel tersebut akan

berinflasi (mengalami peningkatan) jika terdapat korelasi antara variabel tersebut dengan minimal satu variabel penjelas lainnya.

Nilai VIF pada koefisien regresi ke- $i$  dirumuskan sebagai berikut:

$$VIF_i = \frac{1}{1-R_i^2} \quad (5)$$

dimana  $R_i^2$  adalah koefisien determinasi yang diperoleh dengan meregresikan variabel  $x_i$  dengan variabel penjelas  $x_{j \neq i}$  lainnya. Semakin tinggi korelasi antar variabel variabel penjelas, nilai VIF akan semakin besar. Sebaliknya jika tidak ada kolinieritas antarvariabel penjelas, nilai VIF akan mendekati 1. Menard (1995) dalam O'brien (2007) menyatakan bahwa nilai VIF yang lebih besar dari 5 menunjukkan multikolinieritas perlu diperhatikan, dan nilai VIF yang lebih dari 10 menunjukkan terjadinya masalah multikolinieritas yang serius.

#### e. Regresi Ridge

Menurut Draper & Smith (1998), multikolinieritas terjadi ketika terdapat hubungan linier yang hampir sempurna maupun sempurna antar dua variabel. Misalkan matriks  $\mathbf{X}$  berisi nilai-nilai variabel independen, maka multikolinieritas yang sempurna akan menyebabkan  $|\mathbf{X}^T\mathbf{X}|=0$ . Hal ini mengakibatkan taksiran untuk parameter regresi menjadi tidak unik. Sedangkan pada multikolinieritas yang kurang sempurna, determinan  $\mathbf{X}^T\mathbf{X}$  mendekati nol, akibatnya varians dari koefisiensi regresi membesar sehingga *standard error* koefisien regresi membesar atau tidak efisien.

Salah satu cara mengatasi multikolinieritas adalah dengan menggunakan regresi *ridge*. Prosedur dalam regresi *ridge* dikembangkan oleh Hoerl & Kennard (1970). Prosedur ini didasarkan pada matriks  $(\mathbf{X}^T\mathbf{X} + k\mathbf{I})$ , dimana  $\mathbf{I}$  adalah matriks identitas dan  $k$  adalah parameter *ridge*, yaitu parameter skalar bernilai positif bernilai  $0 < k < 1$ . Fungsi penambahan  $k\mathbf{I}$  adalah untuk mengatasi  $\mathbf{X}^T\mathbf{X}$  yang mendekati matriks singular. *Ridge* dapat mengurangi varian dengan mengorbankan sedikit bias.

Estimasi parameter model *ridge* adalah sebagai berikut:

$$E(\hat{\beta}_{RR}) = (\mathbf{X}^T\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^T\mathbf{X}\beta \quad (6)$$

$k$  adalah parameter *ridge*, dan nilai dari parameter ini sangat penting dalam pemodelan *ridge*. Terdapat banyak penelitian untuk menentukan nilai parameter *ridge*. Estimasi parameter *ridge* yang digunakan dalam penelitian ini menggunakan salah satu formulasi diperkenalkan oleh Muniz & Kibria (2009), dengan rumusan sebagai berikut:

$$\hat{k} = \text{Median} \left( \frac{1}{m_i} \right) \quad (7)$$

dimana:

$$m_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-p} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p} \text{ adalah residual mean square error}$$

$\hat{\alpha}_{max}^2$  adalah kuadrat nilai maksimum dari elemen  $\hat{\alpha}$  dimana  $\hat{\alpha} = \gamma \hat{\beta}$  dan  $\gamma$  *eigenvector* dari matrik  $X^T X$

f. **Poisson Ridge Regression (PRR)**

Metode *maximum likelihood* (ML), yang merupakan metode yang umumnya digunakan untuk mengestimasi regresi data cacah, sangat sensitif terhadap multikolinieritas (Månsson & Shukur, 2011). Månsson & Shukur (2011) mengadopsi dan memodifikasi metode *ridge regression*, yang diperkenalkan oleh Hoerl & Kennard (1970), pada data cacah untuk mengatasi multikolinieritas. Model ini dinamakan *Poisson Ridge Regression* (PRR).

Penurunan metode PRR dilakukan dengan menggunakan prinsip bahwa metode ML memperkirakan nilai minimum *weighted sum of square error* (WSSE) (Månsson & Shukur, 2011). Dengan menggunakan metode *Lagrange*, estimasi parameter model PRR adalah sebagai berikut:

$$\hat{B} = \frac{(kI + X^T A X)^{-1} X^T A X \hat{\beta}_{ML}}{Z} \tag{8}$$

$$= Z \hat{\beta}_{ML}$$

**3. Hasil dan Pembahasan**

Untuk memodelkan variabel respon dengan variabel prediktor, dibutuhkan analisis regresi yang sesuai dengan kondisi data, agar analisis yang dihasilkan memberikan hasil yang tidak bias dan akurat. Dalam penelitian ini, pertama akan memodelkan data dengan menggunakan regresi Poisson. Kemudian akan dicek asumsi homoskedastisitas dan multikolinieritas dari regresi Poisson. Tabel 1 adalah ringkasan pemodelan banyak kematian bayi dengan menggunakan model Poisson.

Tabel 1: Nilai Estimasi Parameter Model Poisson.

Variabel	Estimasi Parameter	p-value
Banyak bayi yang memperoleh ASI eksklusif (X <sub>1</sub> )	-0,804	<2e-16
Banyak bayi yang memperoleh imunisasi dasar lengkap (X <sub>2</sub> )	5,143	<2e-16
Jumlah cakupan pelayanan kesehatan bayi (X <sub>3</sub> )	-1,275	6,02e-16
Jumlah rumah yang memenuhi syarat rumah sehat (X <sub>4</sub> )	0,268	1,38e-08
Jumlah bayi berat badan lahir rendah (X <sub>5</sub> )	-0,374	<2e-16

Setelah memodelkan dengan regresi Poisson, selanjutnya adalah pengecekan heteroskedastisitas. Cek heteroskedas dilakukan dengan menggunakan *Breusch-Pagan Test*. Dengan menggunakan formulasi seperti pada persamaan (4), nilai statistik uji BP adalah 0,52149, sedangkan nilai  $\chi^2_{(0,05;5)} = 11,070$ . Karena nilai BP lebih

kecil dari  $\chi^2_{(0,05;5)}$  maka pada taraf 5% kita menerima hipotesis nol dan menyimpulkan bahwa struktur varians bersifat homoskedas.

Deteksi multikolinieritas dalam penelitian ini dilakukan dengan melihat nilai *variance inflation factors* (VIF) (Tabel 2). Setelah memodelkan data dengan regresi Poisson, dengan menggunakan formula (5) di atas, berikut adalah nilai VIF regresi Poisson dengan lima variabel prediktor.

Tabel 2: Nilai VIF Lima Variabel Prediktor Regresi Poisson

Variabel	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
VIF	5,85	154,6	121,2	8,63	1.44

Berdasarkan Tabel 2, dapat dilihat bahwa terdapat multikolinieritas antar variabel prediktor model regresi Poisson. Hal ini dapat dilihat dari nilai VIF variabel X<sub>2</sub> dan X<sub>3</sub> yang lebih dari 10. Dari dua pengecekan sebelumnya, ternyata regresi Poisson struktur varians nya bersifat homoskedas dan terdapat multikolinieritas antar variabel prediktor, sehingga kita membutuhkan pemodelan regresi Poisson yang mengakomodir multikolinieritas, salah satunya adalah *Poisson Ridge Regression* (PRR).

Seperti yang telah dijelaskan pada 2.2, bahwa salah satu hal yang sangat penting dalam regresi *ridge*, termasuk PRR, adalah menentukan nilai estimasi parameter *ridge* ( $k$ ). Estimasi parameter *ridge* dalam penelitian ini akan dihitung dengan menggunakan formulasi seperti pada rumus (7). Dengan menggunakan formulasi tersebut dan dengan menggunakan program R, didapat nilai estimasi parameter *ridge* adalah 0.025. Setelah didapat nilai estimasi parameter *ridge*, estimasi parameter model PRR dapat dicari dengan menggunakan formula (8). Dengan menggunakan program R, nilai estimasi parameter model PRR dapat dilihat pada Tabel 3.

Tabel 3: Nilai Estimasi Parameter Model PRR

Variabel	Estimasi Parameter
Banyak bayi yang memperoleh ASI eksklusif (X <sub>1</sub> )	-0.804
Banyak bayi yang memperoleh imunisasi dasar lengkap (X <sub>2</sub> )	5.136
Jumlah cakupan pelayanan kesehatan bayi (X <sub>3</sub> )	-1.269
Jumlah rumah yang memenuhi syarat rumah sehat (X <sub>4</sub> )	0.269
Jumlah bayi berat badan lahir rendah (X <sub>5</sub> )	-0.374

Estimasi parameter yang dihasilkan model PRR sedikit mengalami perubahan jika dibanding dengan model Poisson. Dari tabel di atas, dapat dilihat dari lima variabel prediktor, 3 parameter bernilai negatif sedangkan 2 variabel lain bernilai positif. Banyak bayi yang memperoleh ASI eksklusif memberikan pengaruh negatif terhadap banyak kematian bayi, artinya semakin banyak bayi yang memperoleh ASI eksklusif akan menurunkan jumlah kematian bayi.

Banyak bayi yang memperoleh imunisasi dasar lengkap memberikan pengaruh positif terhadap banyak kematian bayi, artinya semakin banyak bayi yang memperoleh imunisasi dasar lengkap di suatu kabupaten, semakin banyak pula banyak kematian

bayi. Hal ini bertentangan dengan hal yang diharapkan, yaitu semakin banyak bayi yang memperoleh imunisasi dasar lengkap, semakin sedikit banyak kematian bayi.

Jumlah cakupan pelayanan kesehatan bayi memberikan pengaruh negatif terhadap banyak kematian bayi, artinya semakin banyak bayi yang memperoleh cakupan pelayanan kesehatan, maka semakin sedikit banyak kematian bayi di suatu wilayah. Hal ini sesuai dengan yang diharapkan oleh pemerintah.

Jumlah rumah yang memenuhi syarat rumah sehat ternyata juga memberikan pengaruh positif terhadap banyak kematian bayi. Artinya semakin banyak rumah yang memenuhi syarat rumah sehat di suatu wilayah, semakin banyak pula banyak kematian bayi. Hal ini bertentangan dengan hal yang diharapkan pemerintah, sama seperti variabel banyak bayi yang memperoleh imunisasi dasar lengkap.

Jumlah bayi berat badan lahir rendah memberikan pengaruh negatif terhadap banyak kematian bayi. Artinya semakin banyak bayi yang lahir dengan berat badan rendah, maka semakin sedikit banyak kematian bayi.

#### 4. Simpulan dan Saran

Dari hasil penelitian di atas, dapat disimpulkan bahwa penggunaan *ridge* pada regresi Poisson atau model *Poisson Ridge Regression* menyebabkan perubahan nilai estimasi parameter pada beberapa variabel prediktor. Dari lima variabel prediktor, tiga variabel prediktor yaitu jumlah bayi yang memperoleh ASI eksklusif, jumlah cakupan pelayanan kesehatan bayi, dan bayi berat badan lahir rendah. memberikan pengaruh negatif terhadap banyak kematian bayi, yang artinya semakin banyak variabel prediktor tersebut maka semakin sedikit banyak kematian bayi. Sedangkan dua variabel prediktor yang lain, yaitu banyak bayi yang memperoleh imunisasi dasar lengkap, jumlah rumah yang memenuhi syarat rumah sehat, memberikan pengaruh positif, artinya semakin banyak variabel prediktor tersebut di suatu wilayah maka semakin banyak pula banyak kematian bayi.

#### Daftar Pustaka

- [BPS] Badan Pusat Statistik. (2016). *Provinsi Jawa Tengah Dalam Angka 2016*. Semarang (ID): BPS Jawa Tengah.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge university press.
- [Dinkes] Dinas Kesehatan Jawa Tengah. (2015). *Profil Kesehatan Provinsi Jawa Tengah Tahun 2015*. Semarang: Dinkes Jawa Tengah di akses dari <http>. Semarang (ID): Dinas Kesehatan Jawa Tengah.
- Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). John Wiley & Sons.
- Gujarati, D. N. (2009). *Basic econometrics*. Tata McGraw-Hill Education.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1): 55–67.

- Månsson, K., & Shukur, G. (2011). A Poisson ridge regression estimator. *Economic Modelling*, 28(4): 1475–1481.
- Muniz, G., & Kibria, B. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics—Simulation and Computation*®, 38(3): 621–630.
- O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5): 673–690.