



Black Sea Journal of Statistics

Volume 2 | Issue 1



ISSN: 3108-4265


KAR-YAY



BLACK SEA JOURNAL OF STATISTICS
(BSJ STAT)



Black Sea Journal of Statistics (BSJ Stat) is an international a peer-reviewed journal published electronically twice a year in June and December. BSJ Stat is an international e-journal that publishes open access. The general aim of BSJ Stat is to provide a platform for statistical scientists and related researchers to publish their research findings on a global scale, to encourage their professional development and to provide new perspectives for further research. BSJ Stat publishes original research articles, reviews and technical notes produced from studies carried out in the fields of theoretical and applied statistics. BSJ Stat is published electronically and all articles are provided to the readers free of charge.

e-ISSN: 3108-4265

Phone: +90 362 408 25 15

Fax: +90 362 408 25 15

Email: bsjstat@blackseapublishers.com

Web site: <http://blackseapublishers.online/index.php/stat>

Sort of publication: Periodically 2 times (June and December) in a year

Publication date and place: June 15, 2025- Samsun, TÜRKİYE

Publishing kind: Electronically

OWNER

KARYAY Karadeniz Yayımcılık ve Organizasyon Ticaret Limited Company

DIRECTOR IN CHARGE

Prof. Dr. Hasan ÖNDER

EDITOR BOARDS

EDITOR IN CHIEF

Assoc. Prof. Dr. Cem TIRINK, Iğdır University, Türkiye

SECTION EDITORS*

Prof. Dr. Dariusz PIWCZYŃSKI, Bydgoszcz University of Science and Technology, POLAND

Prof. Dr. Özgür Hakan AYDOĞMUŞ, Social Sciences University of Ankara, TÜRKİYE

Prof. Dr. Taner TUNÇ, Ondokuz Mayıs University, TÜRKİYE

Prof. Dr. Adnan ÜNALAN, Nigde Ömer Halisdemir University, TÜRKİYE

Assoc. Prof. Dr. İlker ÜNAL, Çukurova University, TÜRKİYE

* The ranking is arranged alphabetically within the academic title

STATISTIC EDITOR

Prof. Dr. Mehmet TOPAL, Kastamonu University, TÜRKİYE

ENGLISH EDITOR

Assist. Prof. Dr. Betül ÖZCAN DOST, Ondokuz Mayıs University, TÜRKİYE

TURKISH EDITOR

Prof. Dr. Serkan ŞEN, Ondokuz Mayıs University, TÜRKİYE

REVIEWERS OF THE ISSUE*

Prof. Dr. Çiğdem TAKMA, Ege University, Department of Animal Science, Biometry, TÜRKİYE

Prof. Dr. Mustafa ŞAHİN, Kahramanmaraş Sutcu Imam University, Department of Agricultural Biotechnology, Biometry, TÜRKİYE

Assoc. Prof. Dr. Aycan Mutlu YAĞANOĞLU, Atatürk University, Department of Animal Science, Biometry, TÜRKİYE

Assoc. Prof. Dr. Daniel Zaborski, Szczecin University of Technology, Department of Ruminant Science, POLAND

Assoc. Prof. Dr. Serdar GENÇ, Kırşehir Ahi Evran University, Department of Animal Science, Biometry, TÜRKİYE

Assoc. Prof. Dr. Yasin ALTAY, Eskişehir Osmangazi University, Department of Animal Science, Biometry, TÜRKİYE

Assist. Prof. Dr. Esra YAVUZ, Şırnak University, Department of Accounting and Tax, Biometry, TÜRKİYE

Dr. Ebru ERGÜNEŞ BERKİN, Hassa District Directorate of Agriculture and Forestry, Biometry, TÜRKİYE

Dr. Hasan Alp ŞAHİN, Ondokuz Mayıs University, Hemp Research Institute, Biometry, TÜRKİYE

* The ranking is arranged alphabetically within the academic title

Table of Contents

Research Articles

1. A META-ANALYSIS APPLICATION OF PERSONALITY TRAIT RELIABILITY AND ANIMAL WELFARE
Şenol ÇELİK.....1-8
2. SENSORY CHARACTERISTICS ANALYSIS OF HONEY USING GENERALIZED PROCRUSTES ANALYSIS
Hasan ÖNDER, Uğur ŞEN, Cem TIRINK.....9-11
3. UZAKLIK TEMELLİ REGRESYON MODELLERİNİN PEKİN ÖRDEKLERİNE AİT VÜCUT AĞIRLIKLARI ÜZERİNE İNCELENMESİ
Burcu KURNAZ.....12-14
4. COMPARATIVE ANALYSIS OF REGRESSION METHODS IN SMALL-SAMPLE AGRICULTURAL TIME SERIES: THE CASE OF TURKISH BEEKEEPING DATA
Mehmet BEYAZ, Lütfi BAYYURT.....15-27
5. BIBLIOMETRIC MAPPING OF ZERO-INFLATED STATISTICAL MODELS
Memduh AK, Cem TIRINK.....28-43



A META-ANALYSIS APPLICATION OF PERSONALITY TRAIT RELIABILITY AND ANIMAL WELFARE

Şenol ÇELİK^{1*}

¹Bingöl University, Faculty of Agriculture, Department of Animal Science, Unit of Biometry and Genetics, 12000, Bingöl, Türkiye

Abstract: This study aims to investigate the reasons of variation in the reliability coefficients reported in each study and to determine the mean value for the reliability coefficients reported in the studies involving animal welfare and various personal characteristics. In this regard, 23 articles published between 2018 and 2024 looked at a total of 25574 objects. The Cronbach alpha values of 23 studies that satisfied the inclusion criteria were combined to perform a reliability generalization meta-analysis research. Jamovi 2.6.19 packet routines were useful for data analysis. The studies' Cronbach alpha values were transformed into transformed coefficient values, and the random effects model was employed for the analysis. The 23 studies had a mean Cronbach alpha value of 0.82 (95% CI: 0.79-0.86), which was statistically significant ($P < 0.01$). The findings indicate that this meta-analysis study does not exhibit publication bias. To explore potential causes of variability between individual researches, moderator analyses were carried out. The Cronbach alpha coefficient changed considerably based on the item, number of alternatives, sample content, and structure of the data collection instrument, according to these analysis results. Other variables that are thought to affect reliability can be controlled, and their effects on the reliability coefficient can be examined, in addition to the variables included in this study.

Keywords: Meta-analysis, Heterogeneity, Bias, Reliability, Cronbach's alpha

*Corresponding author: Bingöl University, Faculty of Agriculture, Department of Animal Science, Unit of Biometry and Genetics, 12000, Bingöl, Türkiye

E mail: senolcelik@bingol.edu.tr (Ş. ÇELİK)

Şenol ÇELİK  <https://orcid.org/0000-0001-5894-8986>

Received: November 30, 2025

Accepted: February 11, 2026

Published: June 15, 2026

Cite as: Çelik, Ş. (2026). A meta-analysis application of personality trait reliability and animal welfare. *Black Sea Journal of Statistics*, 2(1), 1-8.

1. Introduction

Reliability is generally defined as the degree of clarity or precision in measurement results that are free from random errors. Numerous reliability coefficients have been proposed to estimate the reliability of measurement instruments. The main reason for the existence of so many coefficients is that they yield different values across different item structures (Osburn, 2000). Among these coefficients, Cronbach's alpha, developed by Cronbach (1951), is most commonly used for obtaining internal consistency reliability, particularly in composite measurements. This widely used coefficient is a test reliability index (Miller, 1995) calculated within the framework of classical test theory (Zumbo, 1999).

The reliability and validity of scores obtained from measurement instruments must be examined (Crocker and Algina, 2008). Among earlier coefficients, Cronbach's Alpha is the most frequently used measure for analyzing and interpreting internal consistency reliability. However, like all coefficients, Cronbach's Alpha may vary from one study to another even when the same scale is used, largely because it is sample-dependent (Özdemir et al., 2020). Reliability coefficients change according to sample characteristics, such as sample size, testing conditions, and timing. These variations across studies have highlighted the need for reliability generalization.

Reliability Generalization (RG), based on meta-analysis,

was first introduced by Vacha-Haase (1998). According to Vacha-Haase, RG examines the magnitude and sources of variability in reliability coefficients reported across different measurements and studies. In other words, an RG study investigates whether reliability coefficients differ significantly across studies. A considerable number of RG studies can be found in the literature (Barnes et al., 2002; Beretvas et al., 2002; Bornmann et al., 2010; Li and Bagger, 2007; Vacha-Haase and Thompson, 2011; Vicent et al., 2019).

As in many scientific fields, reliability analyses are also conducted in studies related to animal science. Animal personality traits are behavioral characteristics that remain relatively stable over time and across contexts (Stamps and Groothuis, 2010; Wolf and Weissing, 2012; Carere and Maestripietri, 2013). In humans and captive animals, the concept of personality refers to a combination of traits that define an individual and distinguish them from others (Caspi et al., 2003; Bergner, 2020). For this reason, the behavioral pattern of different animal species has increasingly attracted research interest.

Identifying and preventing pain in farm animals is one of the primary steps toward improving animal welfare. One of the freedoms that underpin animal welfare is being "free from pain, injury, and disease through prevention or rapid diagnosis and treatment" (Underwood, 2022).



However, production animals such as goats often suffer pain due to routine handling procedures—including castration, hot-iron or freeze branding, and dehorning—which may compromise their welfare (Huxley and Whay, 2006; Alvarez et al., 2009). Therefore, the reliability of pain assessment tools and behavioral indicators becomes especially important.

The purpose of this study is to evaluate the reliability of measurement tools used in research on animal welfare and behavioral traits, and to determine the bias, heterogeneity, and effect sizes of reliability estimates based on reliability generalization.

2. Materials and Methods

2.1. Materials

Within the scope of this study, a total of 25,574 items reported in 23 articles published between 2018 and 2024 were examined through a systematic search conducted in the Web of Science database. Among the 23 included studies whose inclusion criteria were clearly specified 1 study was published in 2018; 3 studies each were published in 2019, 2020, and 2024 (a total of 9); 4 studies each were published in 2021 and 2022 (a total of 8); and 5 studies were published in 2023. The studies included in the analysis evaluated measurement scores related to behavioral and behavioral traits, as well as welfare indicators, in different animal species. The search strategy included combinations of keywords related to animal behavior and welfare, such as animal behavior, behavioral traits, animal welfare, welfare indicators, and measurement scores, using Boolean operators (AND/OR).

A portion of this research was presented orally at the VI. International Applied Statistics Congress (UYIK-2025) held between 14 and 16 May 2025, and an abstract was published in the conference proceedings book.

2.2. Methods

The Cronbach's Alpha coefficient is a weighted standard measure of internal consistency that is obtained by calculating the ratio of the sum of the variances of k items—which contain continuous, interval-scaled, or ordinal responses with 4 or 5 options—to the total variance of the scale (Özdamar, 2013). The Cronbach's Alpha Reliability Coefficient is calculated as given in equation 1 (İnal et al., 2017):

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_p^2} \right) \quad (1)$$

where k is the number of items in the scale, σ_i^2 represents the variance of each individual item, and σ_p^2 denotes the total variance of the scale.

2.2.1. Heterogeneity analysis

To examine whether the reliability coefficients were homogeneous, the Q-test (Hedges and Olkin, 1985) was employed, and this test was supported by the I^2 statistic (Higgins and Thompson, 2002) following the recommendations of Huedo-Medina et al. (2006).

Potential heterogeneity among studies can be identified by calculating Cochran's Q-test and the I^2 value (Higgins and Thompson, 2002). A statistically significant Q-test result or an I^2 value exceeding 75% indicates substantial heterogeneity (Higgins et al., 2003).

The heterogeneity test is performed as follows (Higgins and Thompson, 2002):

The Q statistic is calculated as given in equation 2:

$$Q = \sum_{i=1}^k W * (EF_i - EF_{fixed})^2 \quad (2)$$

and follows a chi-square distribution with $k - 1$ degrees of freedom.

Where k denotes the number of studies included in the analysis, W represents the weight assigned to the i th study, EF_i is the effect size of the i th study, and EF_{fixed} refers to the pooled effect size under the fixed-effects model. The Q statistic follows a chi-square distribution with $k-1$ degrees of freedom.

The I^2 heterogeneity index (equation 3) is obtained by comparing the Q value with its expected value under the assumption of homogeneity (i.e., degrees of freedom, $df = k - 1$) (Higgins and Thompson, 2002):

$$I^2(\%) = \frac{Q - df}{Q} * 100 \quad (3)$$

where $df = k - 1$, and the formula is applicable for $Q > k - 1$.

2.2.2. Effect size estimation and transformation procedures

In reliability generalization studies, transformation and weighting procedures must be performed with great care (Şen, 2022). In the present analysis, Bonett's transformation formula (Bonett, 2002) was used to normalize the sampling distributions and stabilize the variance of reliability coefficients. The transformation is given in equation 4:

$$y = \ln(1 - \alpha) \quad (4)$$

where α denotes Cronbach's alpha reliability coefficient. y shows the value after the Bonett transformation has been applied. The mean effect size can be estimated using either fixed-effect or random-effects models (Şen and Yıldırım, 2020; Borenstein et al., 2021). Moreover, when the aim is to generalize findings to a broader population, the random-effects model should be preferred (Schmid et al., 2021). Therefore, in this study, the mean effect size was calculated using the random-effects model. The meta-analysis was conducted using raw alpha coefficients within the Hunter-Schmidt framework implemented in Jamovi 2.6.19 program. The Hunter-Schmidt method corrects for measurement error and sampling variance, which partially addresses reliability-related differences across studies. Although transformed alpha coefficients (Bonett, 2002) can be used to improve distributional properties, the Hunter-Schmidt estimator is widely recommended for reliability generalization meta-analyses. Model fit was evaluated using maximum

likelihood based AIC and BIC statistics.

Interpretation of the magnitude of effect sizes was based on the classification proposed by Cohen et al. (2007). According to this classification, effect sizes are considered:

- Weak: between 0.00 and ± 0.10
- Small: between ± 0.10 and 0.30
- Medium: between ± 0.30 and 0.50
- Large: between ± 0.50 and 0.80
- Very large: above ± 0.80

For estimating effect size, Cohen's d statistic calculated based on the difference between group means is used. Cohen's d is computed as given in equation 5 (Cohen, 1988):

$$Cohen's d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2}}} \quad (5)$$

2.2.3. Publication bias assessment

Possible publication bias in the included studies (Rothstein et al., 2005) may arise either from the inability to comprehensively review the existing literature (Dinçer, 2021) or from researchers' tendency to prefer studies reporting large effect sizes (Göçen and Şen, 2021). Such bias can influence the results of a meta-analysis (Yumuşak and Korkmaz, 2021) and alter the estimated effect size (Thornton and Lee, 2000). In this context, publication bias in the present study was examined using a funnel plot, Rosenthal's (1979) fail-safe N , and Egger's regression test (Egger et al., 1997).

3. Results and Discussion

This reliability generalization meta-analysis includes 23 studies conducted between 2018 and 2024 that reported reliability coefficients based on their own samples. The total number of items used across these studies is 25574. The reported Cronbach's alpha coefficients ranged from 0.270 to 0.987. Summary information regarding the included studies is presented in Table 1.

3.1. Publication Bias Results

An asymmetric pattern observed in the funnel plot may indicate potential publication bias (Borenstein et al., 2021). As shown in Figure 1, the Cronbach's alpha values of the studies are not symmetrically distributed around the mean transformed alpha value. This suggests that publication bias may be present in the current study. However, it should be noted that an asymmetric funnel plot does not necessarily confirm publication bias, and caution is required when interpreting the graph (Üstün and Eryılmaz, 2014).

In this study, the possibility of publication bias was assessed using the funnel plot, Rosenthal's (1979) fail-safe N , and Egger's linear regression test (Egger et al., 1997). The empty lower-right corner of the triangle in the funnel plot shown in Figure 1 indicates the need for further evaluation with additional tests to determine the presence of publication bias.

Table 1. Information about the Studies

Author (Year)	Cronbach alpha	n	Item
Mikkola et al. (2021)	0.77	4316	84
Salonen et al. (2021)	0.78	15371	17
Jolivald et al. (2022)	0.39	25	26
Lavan et al. (2023)	0.96	327	35
de Lima et al. (2024)	0.77	52	30
Doane and Sarenbo (2019)	0.783	100	87
Alves et al. (2022)	0.943	50	12
Barrios et al. (2022)	0.9468	156	35
Fonseca et al. (2023)	0.6	30	12
Nose and Kakinuma (2019)	0.73	1000	19
Della Rocca et al. (2023)	0.9575	127	12
Farooq et al. (2023)	0.682	264	3
Klinck et al. (2018)	0.598	54	9
Lianou et al. (2020)	0.987	273	22
van Houtert et al. (2019)	0.27	501	28
Riggio et al. (2021)	0.767	31	6
Lanci et al. (2022)	0.842	420	15
van Loon et al. (2020)	0.98	59	11
Howard et al. (2024)	0.957	23	13
Borrelli et al. (2023)	0.8	1175	23
van Dierendonck et al. (2020)	0.955	264	34
Broseghini et al. (2023)	0.748	803	62
Bell and Rogers (2021)	0.89	153	30

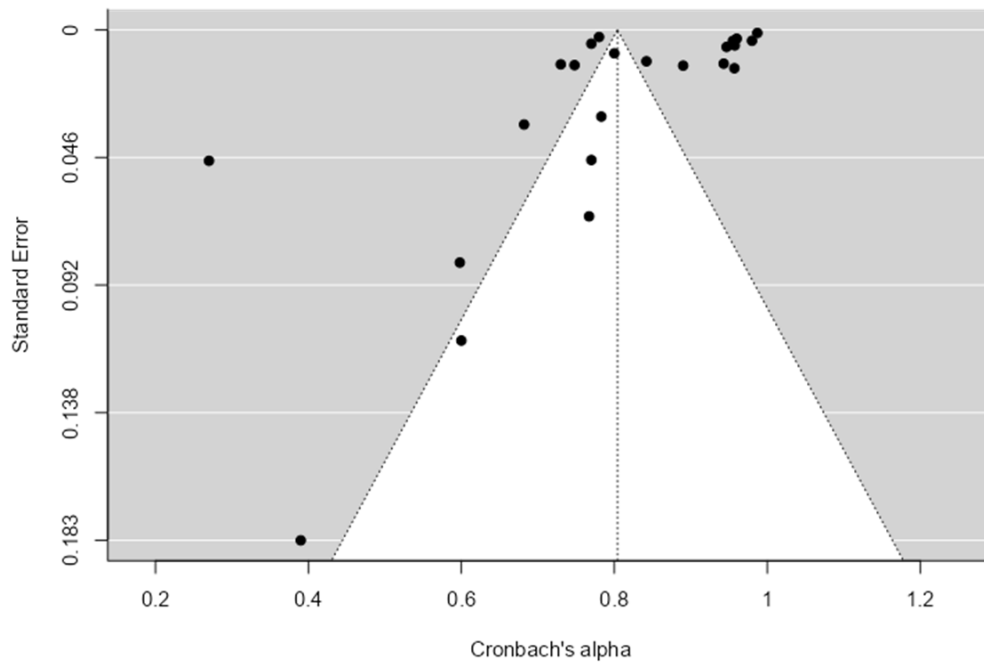


Figure 1. Funnel plot.

3.2. Fail-Safe N and Egger's Test Results

The fail-safe N provides an alternative method to assess the significance of the p-value obtained from a meta-analysis. A fail-safe N indicating $P < 0.05$ suggests that the study is strong and possesses a reliable effect size (Eser et al., 2021). As presented in Table 2, Egger's regression test yielded $P < 0.001$, indicating that there may be statistically significant asymmetry in the data. However, Kendall's Tau ($P > 0.05$) suggests that publication bias may be somewhat controlled.

Additionally, Rosenberg's fail-safe N was calculated as 3 394 611, indicating a very large number of studies would be needed to nullify the observed effect size, further supporting the robustness of the results.

Table 2. Publication bias assessment

Test Name	Value	P
Fail-Safe N	3394611	< 0.001
Kendalls Tau	-0.123	0.432
Egger's Regression	-4.082	< 0.001

Note: Fail-safe N calculation using the rosenthal approach.

3.3. Heterogeneity Results

Information regarding the heterogeneity statistics is presented in Table 3.

Examination of Table 3 shows that Cochran's Q-test was statistically significant ($Q(df = 22) = 8050.308, P < 0.001$), indicating a high level of heterogeneity among the studies. The I^2 statistic represents the proportion of observed variation in effect sizes that is attributable to sampling error. In the present study, $I^2 = 99.56\%$. It should be noted that I^2 is a ratio and provides an absolute measure of heterogeneity. The obtained I^2 value is very high, confirming substantial heterogeneity among the included studies.

Given the observed heterogeneity, a random-effects model was employed for the analysis. Although the restricted maximum likelihood (REML) method was preferred due to its advantages as an estimation procedure, model statistics comparing REML with the conventional maximum likelihood (ML) method are presented in Table 4.

Table 3. Heterogeneity statistics

Statistics	Value
Tau	0.079
Tau ²	0.0062 (SE= 0.0028)
I^2	99.56%
H^2	228.926
R^2	,
df	22
Q	8050.308
P	<0.001

Table 4. Model fit statistics and information criteria

	Maximum-Likelihood	Restricted Maximum-Likelihood
log-likelihood	-5.127	-6.668
Deviance	164.413	13.337
AIC	14.254	17.337
BIC	16.525	19.519
AICc	14.854	17.968

Table 4 presents the model fit statistics and information criteria obtained from the random-effects meta-analytic model estimated using the Hunter-Schmidt estimator in Jamovi. Given the presence of between-study heterogeneity, model fit was evaluated based on

maximum likelihood estimation. The ML-based results yielded a log-likelihood value of -5.127 and relatively low information criteria values (AIC=14.254, BIC=16.525, AICc=14.854), indicating an adequate model fit. Restricted maximum likelihood estimates resulted in lower log-likelihood values, as expected, and were therefore not used for information-criterion-based model comparisons. Overall, the ML-based information criteria support the appropriateness of the specified random-effects model.

3.4. Model Fit and Random-Effects Analysis

As shown in Table 4, the model fit statistics for the restricted maximum likelihood (REML) method were smaller than those obtained using the conventional maximum likelihood (ML) method. Considering that better model fit is indicated by lower values of log-likelihood, AIC, and BIC (Fabozzi et al., 2014), it can be concluded that the REML method provides a superior model fit.

Information regarding the random-effects model is presented in Table 5.

Table 5. Random-effects model ($k = 23$) in reliability generalization

	Intercept
Estimate	0.824
SE	0.0178
Z	46.3
P	<0.001
CI Lower Bound	0.789
CI Upper Bound	0.859

Note: Tau² estimator: Restricted Maximum-Likelihood.

3.5. Random-Effects Model Results for Reliability Generalization

Based on the random-effects model analysis conducted for the reliability generalization, the mean reliability coefficient (effect size, EF) was found to be 0.824, with a 95% confidence interval ranging from 0.789 to 0.859 (Table 5).

According to these results, the effect size representing the relationship between animals' personality traits and welfare is considered strong (EF=0.824, $P<0.001$). The obtained effect size (EF=0.824) corresponds to a strong effect according to the classification proposed by Cohen et al. (2007).

A forest plot illustrating the distribution of reliability coefficients across the included studies is presented in Figure 3.

In Figure 3, the black squares represent the effect sizes (reliability coefficients) of the individual studies, while the horizontal lines adjacent to the squares indicate the upper and lower bounds of the 95% confidence intervals for the effect sizes. Additionally, the percentage weight shown on the left side of the forest plot numerically represents each study's contribution to the meta-analysis results.

The studies with the largest black squares, and thus the greatest weights in the random-effects model, are those by Salonen et al. (2021), Lavan et al. (2023), Lianou et al. (2020), van Loon et al. (2020), and van Dierendonck (2020). Therefore, these studies can be considered to have the largest effect sizes. Conversely, the study by Jolivald et al. (2022), represented by the smallest black square, has the smallest effect size.

Regarding the numerical weights, Salonen et al. (2021), Lavan et al. (2023), Lianou et al. (2020), van Loon et al. (2020), and van Dierendonck (2020) each have the highest weight at 4.65%, while Jolivald et al. (2022) has the smallest weight at 2.07%.

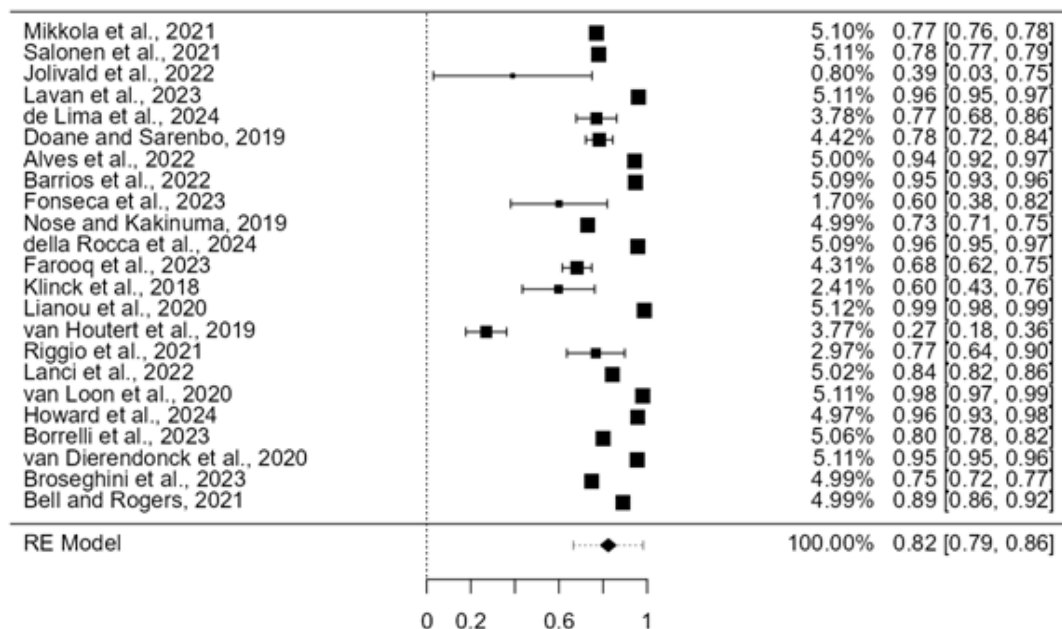


Figure 3. Forest plot.

4. Conclusion

This reliability generalization study included research reporting Cronbach's alpha reliability coefficients in the context of animal welfare and animals' behavioral traits. A limitation of the present study is that it was conducted solely based on Cronbach's alpha values. Another limitation is the scope of the literature review conducted. The alpha reliability estimates exhibited very high heterogeneity among the included studies ($I^2 = 99.56\%$). This indicates that the reliability of scores related to animal welfare and personality traits varies depending on the context, sample, or characteristics of the study, and therefore cannot be generalized to different populations.

The overall effect size of Cronbach's alpha coefficients across all studies was 0.824, with a 95% confidence interval ranging from 0.789 to 0.859. Based on these results, it can be concluded that the reliability of measurements obtained from the surveys in these studies is generally sufficient for the respective samples. In other words, the observed effect size indicates a good level of reliability.

Although most studies reported high reliability coefficients, a few studies with lower reliability estimates likely involved small sample sizes, which may explain their lower representation in the meta-analysis. Considering the overall effect on measurements related to animal welfare and behavioral traits, the findings of this study are expected to be valuable for both researchers and practitioners. A carefully conducted reliability generalization provides a clearer understanding of the reliability of scores obtained from scales and therefore holds practical importance for researchers.

In conclusion, this reliability generalization meta-analysis demonstrates that, on average, measurements of animal welfare and personality traits exhibit good reliability. Researchers intending to use this scale are advised to conduct reliability estimates based on their own collected data.

Author Contributions

The percentages of the author' contributions are presented below. The author reviewed and approved the final version of the manuscript.

	Ş.Ç.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declares that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

- Alvarez, L., Nava, R. A., Ramírez, A., Ramírez, E., & Gutiérrez, J. (2009). Physiological and behavioural alterations in disbudded goat kids with and without local anaesthesia. *Applied Animal Behaviour Science*, 117(3-4), 190-196. <https://doi.org/10.1016/j.applanim.2009.01.001>
- Alves, J. C., Santos, A., Jorge, P., Lavrador, C., & Carreira, L. M. (2022). Evaluation of four clinical metrology instruments for the assessment of osteoarthritis in dogs. *Animals*, 12(20), 2808. <https://doi.org/10.3390/ani12202808>
- Barnes, L. L., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger state-trait anxiety inventory. *Educational and Psychological Measurement*, 62(4), 603-618. <https://doi.org/10.1177/0013164402062004005>
- Barrios, C. L., Gornall, V., Bustos-López, C., Cirac, R., & Calvo, P. (2022). Creation and validation of a tool for evaluating caregiver burnout syndrome in owners of dogs (Canis lupus familiaris) diagnosed with behavior disorders. *Animals*, 12(9), 1185. <https://doi.org/10.3390/ani12091185>
- Bell, C., & Rogers, S. (2021). Attitudes of the equestrian public towards equine end-of-life decisions. *Animals*, 11(6), 1776. <https://doi.org/10.3390/ani11061776>
- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 62(4), 570-589. <https://doi.org/10.1177/0013164402062004003>
- Bergner, R. M. (2020). What is personality? Two myths and a definition. *New Ideas in Psychology*, 57, 100759. <https://doi.org/10.1016/j.newideapsych.2019.100759>
- Bonett, D. G. (2002). Sample size requirements for estimating

- intraclass correlations with desired precision. *Statistics in Medicine*, 21(9), 1331-1335. <https://doi.org/10.1002/sim.1108>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.
- Bornmann, L., Mutz, R., & Daniel, H. D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PloS one*, 5(12), e14331. <https://doi.org/10.1371/journal.pone.0014331>
- Borrelli, C., Riggio, G., Howell, T. J., Piotti, P., Diverio, S., Albertini, M., Mongillo, P., Marinelli, L., Baragli, P., & Mariti, C. (2022). The cat-owner relationship: validation of the Italian C/DORS for cat owners and correlation with the LAPS. *Animals*, 13(1), 69. <https://doi.org/10.3390/ani13010069>
- Broseghini, A., Guérineau, C., Lööke, M., Mariti, C., Serpell, J., Marinelli, L., & Mongillo, P. (2023). Canine Behavioral Assessment and Research Questionnaire (C-BARQ): Validation of the Italian Translation. *Animals*, 13(7), 1254. <https://doi.org/10.3390/ani13071254>
- Carere, C., & Maestripieri, D. (2013). *Animal personalities: behavior, physiology, and evolution*. University of Chicago Press.
- Caspi, A., Harrington, H., Milne, B., Amell, J. W., Theodore, R. F., & Moffitt, T. E. (2003). Children's behavioral styles at age 3 are linked to their adult personality traits at age 26. *Journal of Personality*, 71(4), 495-514. <https://doi.org/10.1111/1467-6494.7104001>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th edition). Taylor & Francis.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3): 297-334. <https://doi.org/10.1007/BF02310555>
- de Lima, M. T., Trindade, P. H., Pinho, R. H., Oliveira, A. R., Gil, J. C., Almeida, T. R., & Luna, S. P. (2024). Validation of the Portuguese version of the short-form Glasgow composite measure pain scale (CMPs-SF) according to COSMIN and GRADE guidelines. *Animals*, 14(6), 831. <https://doi.org/10.3390/ani14060831>
- Della Rocca, G., Schievano, C., Di Salvo, A., Hielm-Björkman, A. K., & Della Valle, M. F. (2023). Psychometric testing and validation of the Italian version of the Helsinki Chronic Pain Index (I-HCPI) in dogs with pain related to osteoarthritis. *Animals*, 14(1), 83. <https://doi.org/10.3390/ani14010083>
- Diñer, S. (2021). *Eğitim bilimlerinde uygulamalı meta analiz* (3. Baskı). Pegem Akademi.
- Doane, M., & Sarenbo, S. (2019). A modified combined C-BARQ and QoL for both the companion dog and its owner. An embryo to a companion dog welfare assessment?. *Applied Animal Behaviour Science*, 213, 91-106. <https://doi.org/10.1016/j.applanim.2019.02.012>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629-634.
- Eser, M. T., Yurtçu, M., & Aksu, G. (2021). *R programlama dili ve Jamovi ile meta analiz uygulamaları*. Pegem Akademi.
- Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The basics of financial econometrics: Tools, concepts, and asset management applications*. John Wiley & Sons.
- Farooq, U., Idris, M., Sajjad, N., & Afzal, M. A. (2023). Assessing total erythrocyte count as a potential attribute for estimating hemoglobin in Cholistani cattle. *Tropical animal health and production*, 55(5), 306. <https://doi.org/10.1007/s11250-023-03744-8>
- Fonseca, M. W., Trindade, P. H. E., Pinho, R. H., Justo, A. A., Tomacheuski, R. M., Silva, N. E. D. O. F. D., Gonçalves, H. C., & Luna, S. P. L. (2023). Development and validation of the unesp-botucatu goat acute pain scale. *Animals*, 13(13), 2136. <https://doi.org/10.3390/ani13132136>
- Göçen, A., & Şen, S. (2021). Spiritual leadership and organizational citizenship behavior: A meta-analysis. *Sage Open*, 11(3), 21582440211040777. <https://doi.org/10.1177/21582440211040777>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539-1558. <https://doi.org/10.1002/sim.1186>
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557-560.
- Howard, D. L., Lancaster, B., & de Grauw, J. (2024). Development and preliminary validation of an equine brief pain inventory for owner assessment of chronic pain due to osteoarthritis in horses. *Animals*, 14(2), 181. <https://doi.org/10.3390/ani14020181>
- Huedo-Medina, T. B., Sánchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index?. *Psychological Methods*, 11(2), 193. <https://psycnet.apa.org/doi/10.1037/1082-989X.11.2.193>
- Huxley, J. N., & Whay, H. R. (2006). Current attitudes of cattle practitioners to pain and the use of analgesics in cattle. *Veterinary Record*, 159(20), 662-668. <https://doi.org/10.1136/vr.159.20.662>
- İnal, H., Yılmaz Kogar, E., Demirdüzen, E., & Gelbal, S. (2017). Cronbach's coefficient alpha: A meta-analysis study. *Hacettepe University Journal of Education*, 32(1), 18-32.
- Jolivald, A., Yarnell, K., Hall, C., & Ijichi, C. (2022). Do you see what I see? Investigating the validity of an equine personality questionnaire. *Applied Animal Behaviour Science*, 248, 105567. <https://doi.org/10.1016/j.applanim.2022.105567>
- Klinck, M. P., Gruen, M. E., del Castillo, J. R., Guillot, M., Thomson, A. E., Heit, M., Lascelles, B. D. X., & Troncy, E. (2018). Development and preliminary validity and reliability of the montreal instrument for cat arthritis testing, for use by caretaker/owner, MI-CAT (C), via a randomised clinical trial. *Applied Animal Behaviour Science*, 200, 96-105. <https://doi.org/10.1016/j.applanim.2017.11.013>
- Lanci, A., Benedetti, B., Freccero, F., Castagnetti, C., Mariella, J., van Loon, J. P., & Padalino, B. (2022). Development of a composite pain scale in foals: a pilot study. *Animals*, 12(4), 439. <https://doi.org/10.3390/ani12040439>
- Lavan, R. P., Tahir, M., O'Donnell, C., Bellenger, A., de Bock, E., & Koochaki, P. (2023). Development and Validation of a Canine Health-Related Quality of Life Questionnaire and a Human-Canine Bond Questionnaire for Use in Veterinary Practice. *Animals*, 13(20), 3255. <https://doi.org/10.3390/ani13203255>
- Li, A., & Bagger, J. (2007). The Balanced inventory of desirable responding (BIDR) a reliability generalization study. *Educational and Psychological Measurement*, 67(3), 525-544. <https://doi.org/10.1177/0013164406292087>
- Lianou, D. T., Chatziprodromidou, I. P., Vasileiou, N. G., Michael, C. K., Mavrogianni, V. S., Politis, A. P., Kordalis, N. G., Billinis, C., Giannakopoulos, A., Papadopoulos, E., Giannenas, I., Ioannidi, K. S., Katsafadou, A. I., Gougoulis, D. A., Lacasta, D.,

- Caroprese, M., & Fthenakis, G. C. (2020). A detailed questionnaire for the evaluation of health management in dairy sheep and goats. *Animals*, 10(9), 1489. <https://doi.org/10.3390/ani10091489>
- Mikkola, S., Salonen, M., Hakanen, E., Sulkama, S., & Lohi, H. (2021). Reliability and validity of seven feline behavior and personality traits. *Animals*, 11(7), 1991. <https://doi.org/10.3390/ani11071991>
- Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural equation Modeling*, 2(3): 255-273. <https://doi.org/10.1080/10705519509540013>
- Nose, I., & Kakinuma, M. (2019). Japanese version of the Dog Personality Scale: Reliability and validity based on owner and dog instructor evaluation and behavioral observation. *Applied Animal Behaviour Science*, 214, 72-78. <https://doi.org/10.1016/j.applanim.2019.03.008>
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5(3): 343-335. <https://psycnet.apa.org/doi/10.1037/1082-989X.5.3.343>
- Özdamar, K. (2013). *Paket programlar ile istatistiksel veri analizi*. Nisan Kitabevi.
- Özdemir, V., Yıldırım, Y., & Tan, Ş. (2020). A meta-analytic reliability generalization study of the Oxford Happiness Scale in Turkish sample. *Journal of Measurement and Evaluation in Education and Psychology*, 11(4), 374-404. <https://doi.org/10.21031/epod.766266>
- Riggio, G., Piotti, P., Diverio, S., Borrelli, C., Di Iacovo, F., Gazzano, A., Howell, T. J., Pirrone, F., & Mariti, C. (2021). The dog-owner relationship: Refinement and validation of the Italian C/DORS for dog owners and correlation with the LAPS. *Animals*, 11(8), 2166. <https://doi.org/10.3390/ani11082166>
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86: 638-641. <https://psycnet.apa.org/doi/10.1037/0033-2909.86.3.638>
- Rothstein HR, Sutton AJ, Borenstein M. 2005. *Publication bias in meta-analysis*. In Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Ed.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (1-7). John Wiley & Sons.
- Salonen, M., Mikkola, S., Hakanen, E., Sulkama, S., Puurunen, J., & Lohi, H. (2021). Reliability and validity of a dog personality and unwanted behavior survey. *Animals*, 11(5), 1234. <https://doi.org/10.3390/ani11051234>
- Schmid CH, Stijnen T, White IR. 2021. *General themes in meta-analysis*. In Schmid, C. H., Stijnen, T., & White I. R. (Ed.), *Handbook of meta-analysis* (19-26). CRC Press.
- Sen, S. (2022). A reliability generalization meta-analysis of runco ideational behavior scale. *Creativity Research Journal*, 34(2), 178-194. <https://doi.org/10.1080/10400419.2021.1960719>
- Şen, S., & Yıldırım, İ. (2020). *CMA ile meta-analiz uygulamaları*. Anı Yayıncılık.
- Stamps, J., & Groothuis, T. G. (2010). The development of animal personality: relevance, concepts and perspectives. *Biological Reviews*, 85(2), 301-325. <https://doi.org/10.1111/j.1469-185X.2009.00103.x>
- Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of Clinical Epidemiology*, 53(2), 207-216. [https://doi.org/10.1016/S0895-4356\(99\)00161-4](https://doi.org/10.1016/S0895-4356(99)00161-4)
- Underwood, W. J. (2002). Pain and distress in agricultural animals. *Journal of the American Veterinary Medical Association*, 221(2), 208-211.
- Üstün, U., & Eryılmaz, A. (2014). Etkili araştırma sentezleri yapabilmek için bir araştırma yöntemi: Meta-analiz. *Education & Science*, 39(174), 1-32.
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1), 6-20. <https://doi.org/10.1177/0013164498058001002>
- Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, 44(3), 159-168. <https://doi.org/10.1177/0748175611409845>
- van Dierendonck, M. C., Burden, F. A., Rickards, K., & van Loon, J. P. (2020). Monitoring acute pain in donkeys with the equine utrecht university scale for donkeys composite pain assessment (equus-donkey-compass) and the equine Utrecht University scale for donkey facial assessment of pain (Equus-Donkey-Fap). *Animals*, 10(2), 354. <https://doi.org/10.3390/ani10020354>
- van Houtert, E. A., Endenburg, N., Wijnker, J. J., Rodenburg, T. B., van Lith, H. A., & Vermetten, E. (2019). The translation and validation of the Dutch Monash dog-owner relationship scale (MDORS). *Animals*, 9(5), 249. <https://doi.org/10.3390/ani9050249>
- van Loon, J., Verhaar, N., van Den Berg, E., Ross, S., & de Grauw, J. (2020). Objective assessment of acute pain in foals using a facial expression-based pain scale. *Animals*, 10(9), 1610. <https://doi.org/10.3390/ani10091610>
- Vicent, M., Rubio-Aparicio, M., Sánchez-Meca, J., & González, C. (2019). A reliability generalization meta-analysis of the child and adolescent perfectionism scale. *Journal of Affective Disorders*, 245, 533-544. <https://doi.org/10.1016/j.jad.2018.11.049>
- Wolf, M., & Weissing, F. J. (2012). Animal personalities: consequences for ecology and evolution. *Trends in Ecology & Evolution*, 27(8), 452-461.
- Yumuşak, H., & Korkmaz, M. (2021). Liderlik ve okul etkililiği düzeyleri arasındaki ilişki: Bir meta analiz çalışması. *Erciyes Journal of Education*, 5(2), 121-148. <https://doi.org/10.32433/eje.980997>
- Zumbo, B. D. (1999). A glance at coefficient alpha with an eye towards robustness studies: Some mathematical notes and a simulation model. *Edgeworth Series in Quantitative Behavioral Science*, 99(1), 1-13.



SENSORY CHARACTERISTICS ANALYSIS OF HONEY USING GENERALIZED PROCRUSTES ANALYSIS

Hasan ÖNDER^{1*}, Uğur ŞEN², Cem TIRINK³

¹Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye

²Ondokuz Mayıs University, Agricultural Faculty, Department of Agricultural Biotechnology, 55139, Samsun, Türkiye

³Iğdır University, Agricultural Faculty, Department of Animal Science, 76000, Iğdır, Türkiye

Abstract: Sensory evaluation of honey constitutes a complex analytical chain encompassing visual, olfactory, gustatory, and tactile assessments. In food science, the primary objective of sensory profiling is to align product characteristics with consumer demands to optimize industrial production. This study utilizes Generalized Procrustes Analysis (GPA) to evaluate the assessments of 20 trained panelists regarding the organoleptic and physicochemical attributes of both comb and extracted honey. The variables analyzed included flavor profiles, chemical constituents, aroma, color, viscosity, and overall aesthetic presentation. Results showed that color, view and viscosity were found effective variables for comb honey and sucrose, taste, color and viscosity were found significant variables for extracted honey. Consensus of panelists was found 92% for comb honey and 89% for extracted honey. The results demonstrate that Generalized Procrustes Analysis is a robust and successful methodology for the sensory quantification of honey, providing a high-fidelity mapping of product characteristics and evaluator consistency.

Keywords: Generalized Procrustes analysis, Honey, Honey quantification, Sensory evaluation

*Corresponding author: Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye

E mail: hasanonder@gmail.com (H. ÖNDER)

Hasan ÖNDER <https://orcid.org/0000-0002-8404-8700>

Uğur ŞEN <https://orcid.org/0000-0001-6058-1140>

Cem TIRINK <https://orcid.org/0000-0001-6902-5837>

Received: March 16, 2026

Accepted: May 23, 2026

Published: June 15, 2026

Cite as: Önder, H., Şen, U., & Tırink, C. (2026). Sensory characteristics analysis of honey using generalized Procrustes analysis. *Black Sea Journal of Statistics*, 2(1), 9-11.

1. Introduction

Honey is a natural, high-energy food produced by bees from plant nectar or secretions of living organisms, which they combine with their unique substances to create a rich paste. In addition to simple sugars like glucose and fructose, it contains enzymes, amino acids, vitamins, and phenolic compounds, giving it antioxidant, antibacterial, and anti-inflammatory properties. With a wide range of medicinal uses, from wound healing to gastrointestinal health, honey also plays a critical role in the food industry as a natural sweetener and preservative. From a biodiversity perspective, the pollination activity of bees in honey production is an indispensable element for agricultural sustainability and global food security (Bogdanov et al., 2008; Da Silva et al., 2016).

Honey's sensory profile consists of parameters such as color, aroma, taste, and texture, which vary widely depending on the nectar source, production region, and processing techniques. From a sensory perspective, honey's color can range from water white to dark amber, while its viscosity and crystallization directly affect its texture. These characteristics are the most important factors determining a consumer's perception of quality. Biochemically, honey is rich in phenolic compounds, flavonoids, and enzymes. These components give honey

antioxidant, antimicrobial, and anti-inflammatory properties, making it indispensable in apitherapy (treatment with bee products). Furthermore, honey's low water activity and high osmotic pressure allow it to act as a natural preservative, increasing its value in the food industry (Bogdanov et al., 2008; Escuredo and Seijo, 2024).

In sensory analyses, panelists often have different scaling habits or differences in terminological perception when evaluating products. Generalized Procrustes Analysis (GPA) minimizes "panelist variations" resulting from these individual differences, creating a common level of consensus. GPA combines data from different panelists into a common configuration using mathematical transformations such as rotation, reflection, and scaling. This method allows for the objective mapping of fundamental similarities and differences between products, even when panelists use their own descriptive terminology, especially in free choice profiling techniques. In complex matrix foods like honey, it is a critical tool for identifying the interaction of sensory dimensions such as color, viscosity, and aroma, and for determining panelist agreement (Gower, 1975; Lawless and Heymann, 2010).

In this study, evaluation of 20 panelists on the characteristics of honey (comb honey and extracted



honey) such as taste and chemical variables, smelling, color, viscosity, view and fancy were analyzed by Generalized Procrustes.

2. Materials and Methods

Pure comb and extracted flower honeys and sucrose additive honey produced in Ondokuz Mayıs University in 2014 was used as data. 20 panelist were evaluated the sensorial characteristics such as smelling, color, viscosity, view and fancy. Idiogrid software was used to analyze the data.

The matching procedure of Generalized Procrustes Analysis is based on minimizing the distances between different evaluators and the same test object. The differences in distances for individual matrices may be stated as given in Equation 1.

$$\sum_{k=1}^K \|\tau(X_k) - \tau(X_1)\| \quad (1)$$

$\tau(X_k)$ and $\tau(X_1)$ represent a certain transformation of the matrices(τ) and M is the total of the squares of the elements. The points produced by transformation (τ) do not keep relative distances. At the first step the minimization procedure is performed with using Equations 2 and 3.

$$\sum_{k=1}^K \|\tau(X_k) - Y\| \quad (2)$$

$$Y = K^{-1} \sum_{k=1}^K \tau(X_k) \quad (3)$$

Y is the mean of the transformed individual data matrices ($\tau(X_k)$) and is stated as the consensus matrix. The transformations used in the Procrustes analysis are translation, rotation and scaling. These transformations are given in Equation 4.

$$\tau(X_k) = p_k X_k H_k + T_k \quad (4)$$

p_k , is the isotropic scaling factor, H_k is the rotation matrix and T_k is translation. The minimizing criteria for Generalized Procrustes Analysis are the totals of all quadratic distances between the transformed individual matrices and is stated as given in Equation 5 (Ser, 2019);

$$\begin{aligned} \sum_{k=1}^K \|p_k X_k H_k - p_1 X_1 H_1\| \\ = K \sum_{k=1}^K \|Y - p_k X_k H_k\| \end{aligned} \quad (5)$$

3. Results and Discussion

Principal Component Analysis (PCA) results were given in Table 1.

Both for comb and extracted honey two component exactly explain the total variance.

The results showed that 1) Color, view and viscosity were found effective variables for comb honey and sucrose. 2) Taste, color and viscosity were found significant variables for extracted honey. 3) Consensus of panelists was found 92% for comb honey and 89% for

extracted honey. Results indicated that Generalized Procrustes analysis can be successfully applied to honey sensory evaluation.

Table 1. PCA results

Comb Honey	Eigenvalue	Variance Explained (%)
PC-1	6.25	62.48
PC-2	3.75	37.52
Extracted Honey		
PC-1	7.18	71.77
PC-2	2.83	28.23

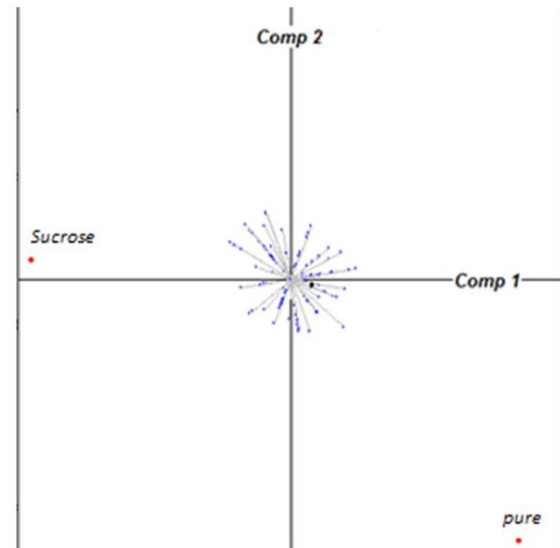


Figure 1. PCA results for comb honey.

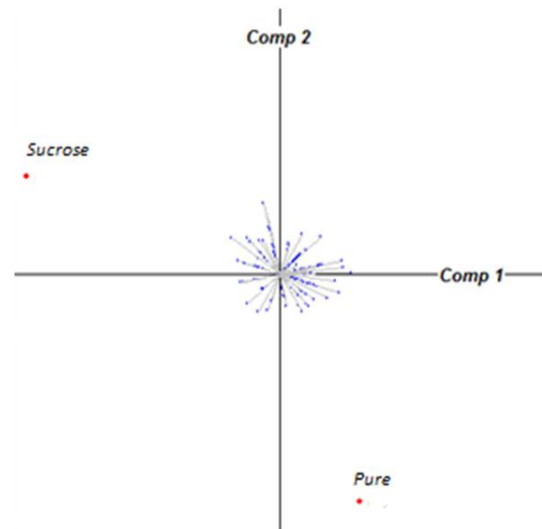


Figure 2. PCA results for extracted honey.

4. Conclusion

The results of this study showed that the Generalized Procrustes Analysis can be used successfully for the sensory analysis of both comb and extracted honey. For the comb honey the most important variables were color, view and viscosity while the most important variables were taste, color and viscosity for the extracted honey.

This study successfully demonstrates that Generalized Procrustes Analysis (GPA) is a highly robust and sophisticated methodology for the sensory and organoleptic quantification of complex food matrices like honey. The mathematical alignment process of GPA efficiently minimized individual panelist variations, yielding remarkably high consensus rates of 92% for comb honey and 89% for extracted honey. These high consensus percentages confirm the reliability of the trained panel and the fidelity of the sensory mapping. Our findings reveal distinct sensory drivers for different honey types. While visual and structural attributes—namely color, view, and viscosity—emerged as the most effective discriminative variables for comb honey, the quality perception of extracted honey was predominantly driven by a combination of chemical and sensory attributes, specifically sucrose content, taste, color, and viscosity. This differentiation highlights that consumers and evaluators process the quality of comb and extracted honey through fundamentally different sensory tracks. From an industrial and economic perspective, these insights offer valuable guidelines for honey producers and quality control laboratories. By understanding which specific parameters dominate product profiles, producers can optimize processing, packaging, and sorting techniques to better match consumer expectations and safeguard authenticity against adulteration. Future research should extend the application of GPA to honeys from different botanical origins and geographical regions to establish a broader baseline for sensory standardization in apicultural products.

Author Contributions

The percentage of the authors' contributions is presented below. All authors reviewed and approved the final version of the manuscript.

	H.Ö.	U.Ş	C.T.
C	50	25	25
D	50	25	25
S	50	25	25
DCP	50	25	25
DAI	50	25	25
L	50	25	25
W	50	25	25
CR	50	25	25
SR	50	25	25
PM	50	25	25
FA	50	25	25

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declare that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

Acknowledgements

This study has been presented in International Forum on Food, Agriculture & Biological Sciences organized in Helsinki, Finland in April 27-30, 2026.

References

- Bogdanov, S., Jurendic, T., Sieber, R., & Gallmann, P. (2008). Honey for nutrition and health: a review. *Journal of the American College of Nutrition*, 27, 677-689. <https://doi.org/10.1080/07315724.2008.10719745>
- Da Silva, P. M., Gauche, C., Gonzaga, L. V., Costa, A. C. O., & Fett, R. (2016). Honey: Chemical composition, stability and authenticity. *Food Chemistry*, 196, 309-323. <https://doi.org/10.1016/j.foodchem.2015.09.051>
- Escuredo, O., & Seijo, M. C. (2024). Authenticity of honey: Characterization, bioactivities and sensorial properties series II. *Foods*, 13, 2079. <https://doi.org/10.3390/foods13132079>
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40, 33-51. <https://doi.org/10.1007/BF02291478>
- Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food: principles and practices*. Springer Science & Business Media.
- Ser, G. (2019). Using Generalized Procrustes Analysis for evaluation of sensory characteristic data of lamb meat. *Turkish Journal of Agriculture -Food Science and Technology*, 7(6), 840-844. <https://doi.org/10.24925/turjaf.v7i6.840-844.2214>



UZAKLIK TEMELLİ REGRESYON MODELLERİNİN PEKİN ÖRDEKLERİNE AİT VÜCUT AĞIRLIKLARI ÜZERİNE İNCELENMESİ

Burcu KURNAZ^{1*}

¹Ondokuz Mayıs University, Faculty of Agriculture, Department of Animal Science, 55139, Samsun, Türkiye

Özet: Bu çalışmada klasik doğrusal regresyona alternatif bir yöntem olan uzaklık temelli regresyon modellerinin incelenmesi amaçlanmıştır. Analizde 100 adet erkek Pekin ördeklerine ait farklı günlerde ölçülen vücut ağırlıkları kullanılmış, Öklid, Gower ve Manhattan uzaklıklarına dayalı regresyon modelleri incelenmiştir. Analizler R yazılımında 'dbstats' paketi kullanılarak gerçekleştirilmiştir. Elde edilen modeller belirleme kriterleri olan AIC, BIC ve GCV değerleri bakımından karşılaştırılmıştır. En düşük AIC (1499,281), BIC (1504,492) ve GCV (182596,7) değerleri Gower uzaklığına ait modelden elde edilmiştir. Ayrıca uzaklık temelli regresyon modellerinin AIC değerleri, klasik doğrusal regresyon modelinden daha düşük olduğu belirlenmiştir. Sonuç olarak Gower uzaklığının incelenen veri yapısı için daha uygun olduğu ve uzaklık temelli regresyonun klasik regresyon varsayımlarının sağlanamadığı durumlarda alternatif bir yöntem olarak kullanılabilmesi sonucuna varılmıştır.

Anahtar kelimeler: Uzaklık temelli regresyon, dbstats paketi, R yazılımı, Model belirleme kriterleri

Investigation of Distance-Based Regression Models on Body Weights of Pekin Ducks

Abstract: This study aims to examine distance-based regression models, an alternative method to classical linear regression. The analysis used body weight measurements of 100 male Peking ducks taken on different days, and regression models based on Euclidean, Gower, and Manhattan distances were investigated. The analyses were performed using the 'dbstats' package in R software. The obtained models were compared in terms of their identification criteria: AIC, BIC, and GCV values. The lowest AIC (1499.281), BIC (1504.492), and GCV (182596.7) values were obtained from the Gower distance model. Furthermore, it was determined that the AIC values of the distance-based regression models were lower than those of the classical linear regression model. In conclusion, it was found that the Gower distance is more suitable for the data structure under investigation, and that distance-based regression can be used as an alternative method when the assumptions of classical regression cannot be met.

Keywords: Distance based regression, Package dbstats, R software, Model selection criteria

*Sorumlu yazar (Corresponding author): Ondokuz Mayıs University, Faculty of Agriculture, Department of Animal Science, 55139, Samsun, Türkiye

E mail: burcu2039@hotmail.com (B. KURNAZ)

Burcu KURNAZ  <https://orcid.org/0000-0001-5613-6992>

Gönderi: 19 Nisan 2026

Kabul: 22 Mayıs 2026

Yayınlanma: 15 Haziran 2026

Received: April 19, 2026

Accepted: May 22, 2026

Published: June 15, 2026

Cite as: Kurnaz, B. (2026). Investigation of distance-based regression models on body weights of Pekin ducks. *Black Sea Journal of Statistics*, 2(1): 12-14.

1. Giriş

Regresyon analiz yöntemleri istatistiğin ilişkiler ile ilgili kısmını oluşturmakta ve bir sebep-sonuç ilişkisini incelemektedir. Bir veya birden fazla sebebin sonucu hangi yönde ve ne ölçüde etkilediğini inceleyen bir ilişki yöntemidir ve bu ilişkiler sebepten sonuca tek yönlüdür (Kurnaz ve Önder, 2025). Basit ve çoklu doğrusal regresyon analizi sonucunda elde edilecek olan regresyon modeline ait parametre kestirimlerinin güvenilir olabilmesi için modelle ilgili bazı varsayımların sağlanabilmesi gerekmektedir (Alpar, 2010).

Parametre tahmin yöntemlerinde varsayımların sağlanamadığı durumlar için çözüm olarak geliştirilen az sayıdaki modellerden biri Uzaklık Temelli Regresyon yöntemleridir. Bu yöntemlerin amacı kategorik veya gerçek değerli ve kategorik açıklayıcı değişkenlerin bir karışımı dahil olmak üzere, ölçüm değer tahmin edicileri ile problemleri doğru bir şekilde ele almaktır. Uzaklık

temelli regresyon, karışık tipte açıklayıcı değişkenler kullanıldığında doğrusal regresyon modellerinde parametre tahmini için alternatif bir yöntemdir. Uzaklık temelli regresyon klasik doğrusal regresyona benzer, ancak açıklayıcı değişkenler ham değerler yerine uzaklık ölçülerine göre ölçülmektedir (Solà ve Avellana, 2002).

Bu çalışmada 100 adet Pekin ördeğine ait veriler üzerinde Euclidean, Gower ve Manhattan uzaklık ölçülerinin incelenmesi amaçlanmıştır.

2. Materyal ve Yöntem

Bu çalışmada materyal olarak, Ondokuz Mayıs Üniversitesi Ziraat Fakültesi Araştırma ve Uygulama Çiftliği'nde 2023 yılında yapılan bir çalışmada bildirilen erkek Pekin ördeğine ait 1., 21., 42. Ve 72. gün vücut ağırlıkları değerleri kullanılarak simülasyon ile üretilen veri kullanılmıştır (Önder vd., 2025).

Uzaklık temelli regresyon, klasik regresyon yöntemlerine



alternatif olarak geliştirilen ve gözlemlerin ham değerleri yerine aralarındaki uzaklık veya benzerliklerden yararlanılan bir yöntemdir. Bu yöntemde açıklayıcı ve yanıt değişkenlerine ait uzaklık matrisleri oluşturularak bu matrisler arasındaki ilişki modellenmektedir. Sürekli ve kategorik değişkenlerin birlikte bulunduğu karma veri yapılarına uygulanabilmesi yöntemin önemli avantajlarından olmaktadır. Analizde Öklid, Manhattan ve Gower gibi farklı uzaklık ölçüleri kullanılabilmekte, modelin istatistiksel anlamlılığı ise genellikle pseudo-F istatistiği ile değerlendirilmektedir. Özellikle klasik doğrusal regresyon varsayımlarının sağlanamadığı

durumlarda esnek bir alternatif olarak kullanılabilir. Bu çalışmada normal dağılımlı Pekin ördkelerine ait veriler için Öklid, Manhattan ve Gower uzaklık ölçüleri kullanılmış ve elde edilen modeller AIC, BIC ve GCV kriterlerine göre karşılaştırılmıştır. Bu kriterlerde daha düşük olan değerler modelin daha uygun olduğunun göstergesi olarak belirlenmiştir.

Verilerin analizinde R yazılımı kullanılmış, uzaklık temelli regresyon modelleri *dbstats* paketinde yer alan *dbglm* fonksiyonu ile oluşturulmuştur.

Tablo1. Veriye ait tanımlayıcı istatistikler (Ort. $\pm$ S. sapma)

Ölçüm zamanı	Canlı ağırlık
1. gün	51,52 $\pm$ 3.78
21. gün	948,86 $\pm$ 148,2
42. gün	3104,34 $\pm$ 319,78
70. gün	4350,97 $\pm$ 425,19
n	100

Tablo 2. Uzaklık ölçüleri için model belirleme kriter değerleri

Uzaklık	AIC	BIC	GCV
Öklid	1499,397	1504,607	182807,4
Gower	1499,281	1504,492	182596,7
Manhattan	1499,427	1504,637	182863,2

AIC= Akaike bilgi kriteri, BIC= Bayesian bilgi kriteri, GCV= genelleştirilmiş çapraz geçerlilik.

3. Bulgular

Çalışmada uygulanan analizler sonucu elde edilen değerler aşağıdaki Tablo 1. ve Tablo 2.'de verilmiştir. Elde edilen bulgular sonucunda uzaklık ölçüleri arasında en düşük AIC (1499,281), BIC (1504,492) ve GCV (182596,7) değerleri Gower uzaklık ölçüsüne ait olduğu belirlenmiştir (Tablo 2.). Klasik doğrusal regresyon modelinin AIC değeri 1503 olarak bulunmuş olup en yüksek AIC değerine sahip olduğu gözlemlenmiştir. Buna göre incelenen veri yapısı için en uygun modelin Gower uzaklığına ait regresyon modelinin olduğu belirlenmiştir.

4. Tartışma ve Sonuç

Klasik doğrusal regresyon modelinin AIC değeri 1503 olarak bulunurken uzaklık temelli modellerde bu değerlerin yaklaşık 1499 düzeyinde olduğu belirlenmiştir. AIC değerinin daha düşük olması model uyumunun daha iyi olduğunu gösterdiğinden, uzaklık temelli modellerin klasik doğrusal regresyon modeline göre daha uygun sonuç verdiği söylenebilir. Bu sonuç, klasik regresyon varsayımlarının sağlanmasında güçlük yaşanan veya farklı yapıdaki değişkenlerin birlikte incelendiği durumlarda uzaklık temelli regresyonun alternatif bir yöntem olarak kullanılabileceğini göstermektedir. Lichstein (2007) de uzaklık temelli yöntemlerin farklı veri türlerine uygulanabilen esnek bir yapıya sahip olduğunu ifade etmiştir.

Bu çalışmada, erkek pekin öredkelerine ait veriler

kullanılarak Öklid, Gower ve Manhattan uzaklık ölçülerine dayalı regresyon modelleri oluşturulmuş ve elde edilen modeller AIC, BIC ve GCV kriterleri bakımından karşılaştırılmıştır. En düşük AIC, BIC ve GCV değerleri Gower uzaklığına dayalı modelden elde edilmiş, bu nedenle incelenen veri yapısı için en uygun uzaklık ölçüsünün Gower olduğu belirlenmiştir. Öklid ve Manhattan uzaklıklarına ait sonuçların Gower uzaklığına yakın olduğu gözlemlenmiştir. Ayrıca uzaklık temelli regresyon modellerinin AIC değerlerinin klasik doğrusal regresyon modelinden daha düşük olduğu görülmüştür. Sonuç olarak uzaklık temelli regresyonun, özellikle klasik regresyon varsayımlarının sağlanamadığı veya farklı yapıdaki verilerin birlikte incelendiği durumlarda alternatif bir yöntem olarak kullanılabileceği sonucuna varılmıştır. Gelecekte yapılacak çalışmalarda farklı uzaklık ölçülerinin, çeşitli veri yapılarında ve örnek büyüklüklerinde uygulanıp değerlendirilmesi önerilmektedir.

Katkı Oranı Beyanı

Yazarın katkı yüzdeleri aşağıda verilmiştir. Yazar makaleyi incelemiş ve onaylamıştır.

	B.K.
K	100
T	100
Y	100
VTI	100
VAY	100
KT	100
YZ	100
KI	100
GR	100
PY	100

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= yazım, KI= kritik inceleme, GR= gönderim ve revizyon, PY= proje yönetimi.

Çatışma Beyanı

Yazar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Etik Onay Beyanı

Bu araştırmada hayvanlar ve insanlar üzerinde herhangi bir çalışma yapılmadığı için etik kurul onayı alınmamıştır.

Kaynaklar

- Alpar, R. (2010). *Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlik*. Detay yayıncılık.
- Kurnaz, B., & Önder, H. (2025). Examination of distance based regression methods for different data structures in animal science. *Black Sea Journal of Engineering and Science*, 8(2), 354362. <https://doi.org/10.34248/bsengineering.1599606>
- Lichstein, J. W. (2007). Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology*, 188(2), 117-131.
- Önder, H., Yılmaz, M., Şahin, M., Şen, U., Bozkurt, S., Erensoy, K., Gök, İ., Tolun, T., & Uçar, A. (2025). Classification and comparison of eight different growth Curve methods for Pekin duck. *Veterinary Medicine and Science*, 11, e70476. <https://doi.org/10.1002/vms3.70476>
- Solà, C. A., & Avellana, C. M. C. (2002). Recent statistical methods based on distances. *Contributions to Science*, 183-192.



COMPARATIVE ANALYSIS OF REGRESSION METHODS IN SMALL-SAMPLE AGRICULTURAL TIME SERIES: THE CASE OF TURKISH BEEKEEPING DATA

Mehmet BEYAZ^{1*}, Lütfi BAYYURT¹

¹Tokat Gaziosmanpaşa University, Faculty of Agriculture, Department of Animal Science, 60250, Tokat, Türkiye

Abstract: Accurate modeling of agricultural time series with a limited number of observations is a significant challenge, as multicollinearity and small sample sizes may reduce the reliability of conventional regression methods. In this study, the predictive performance of five different regression techniques was compared for modeling real honey prices using Turkish beekeeping data covering the period 2015–2025. Ordinary Least Squares (OLS) regression, Ridge Regression, Elastic Net Regression, Principal Component Regression (PCR), and Partial Least Squares Regression (PLSR) were evaluated using a nested leave-one-out cross-validation approach. Model performance was compared based on the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and predicted residual sum of squares (PRESS). Among the evaluated models, PLSR demonstrated the highest predictive performance under cross-validation, with an R^2 of 0.972, RMSE of 1.937, MAE of 1.563, and PRESS of 41.290. Correlation analysis revealed substantial multicollinearity among the explanatory variables, supporting the use of latent-variable-based regression methods. The findings indicate that PLSR provides a robust and reliable approach for modeling small-sample agricultural time series and represents a potentially applicable method for agricultural price forecasting and decision-making processes in the beekeeping sector.

Keywords: Agricultural time series, Partial Least Squares Regression, Nested cross-validation, Multicollinearity, Small-sample regression

*Corresponding author: Tokat Gaziosmanpaşa University, Faculty of Agriculture, Department of Animal Science, 60250, Tokat, Türkiye

E mail: mehmet.beyaz6622@gop.edu.tr (M. BEYAZ)

Mehmet BEYAZ  <https://orcid.org/0009-0005-3352-4454>

Lütfi BAYYURT  <https://orcid.org/0000-0003-2613-9302>

Received: February 07, 2026

Accepted: April 19, 2026

Published: June 15, 2026

Cite as: Beyaz, M., & Bayyurt, L. (2026). Comparative analysis of regression methods in small-sample agricultural time series: The case of Turkish beekeeping data. *Black Sea Journal of Statistics*, 2(1), 15-27.

1. Introduction

With the increasing adoption of digital technologies in agricultural production systems, the volume of data obtained from sensor networks, remote sensing applications, farm management record systems, and official statistics has increased substantially. These developments have accelerated the use of data-driven modeling approaches in various fields, including production process monitoring, yield forecasting, disease management, and economic analysis (Guindani et al., 2024). In particular, reliable forecasting of agricultural commodity prices is of great importance for enabling producers to plan production, develop marketing strategies, mitigate market risks, and formulate policies aimed at ensuring food supply security (Jin et al., 2024). However, agricultural prices are shaped by the combined effects of numerous interrelated factors, such as production volume, climatic conditions, input costs, supply-demand dynamics, and macroeconomic indicators, making price forecasting a statistically complex problem (Avinash et al., 2024).

In recent years, a substantial proportion of studies on agricultural price forecasting have shifted toward

artificial intelligence, deep learning, and hybrid modeling approaches (Guindani et al., 2024). However, the vast majority of these studies rely on high-frequency time series, such as daily, weekly, or monthly data, and datasets containing thousands of observations. In contrast, as in many countries, key indicators of agricultural production in Türkiye are predominantly published as annual official statistics. Consequently, researchers, particularly in long-term economic analyses, are often required to work with datasets consisting of a limited number of observations. The overfitting of forecasting models to the training data, overestimation of model performance, and reduced generalizability to new observations under small sample sizes are among the major methodological challenges associated with such studies (Vabalas et al., 2019; Pavlou et al., 2024; Ghasemzadeh et al., 2024).

Another important problem encountered in small-sample agricultural datasets is the high degree of linear dependence among explanatory variables. Since variables such as the number of beekeeping operations, production volume, hive population, yield, and climatic indicators represent different components of the same



production system, strong correlations among these variables are expected. Such dependence among explanatory variables can lead to unstable coefficient estimates, inflated standard errors, and excessive sensitivity of the model to small changes in the data in ordinary least squares regression (Dormann et al., 2013; Schreiber-Gregory, 2018). The effects of multicollinearity become particularly pronounced in datasets where the number of observations approaches the number of explanatory variables, potentially leading to poorer out-of-sample predictive performance of conventional regression models (Pavlou et al., 2024).

As small sample sizes and multicollinearity limit the performance of conventional linear regression models, alternative regression approaches have been developed in recent years to mitigate these problems. Regularization-based methods, such as Ridge Regression and Elastic Net, aim to control regression coefficients, reduce prediction variance, and improve model generalizability (Hoerl & Kennard, 1970; Zou and Hastie, 2005). Similarly, dimensionality-reduction methods, including Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR), can mitigate the adverse effects of multicollinearity by representing highly correlated explanatory variables through a smaller number of components (Jolliffe, 1982; Wold et al., 1984). Although the performance of these methods across different data structures has been investigated in numerous studies, the relative effectiveness of each approach has been reported to vary depending on the data structure, sample size, and degree of association among the explanatory variables (James et al., 2021; Hastie et al., 2021). Therefore, comparing different regression methods under the same conditions is particularly important for small-sample agricultural datasets.

Reliable evaluation of model performance depends not only on the selection of an appropriate regression method but also on the use of an appropriate validation strategy. In small-sample datasets, conducting model development and model evaluation using the same data can result in an overestimation of predictive performance (Wainer and Cawley, 2021). Similarly, applying preprocessing procedures such as scaling, variable selection, or dimensionality reduction to the entire dataset before the validation process may lead to information leakage and consequently to a biased assessment of model performance (Parvande et al., 2020; Moscovich and Rosset, 2022). Therefore, particularly in the fields of biometry, bioinformatics, and applied statistics, the use of validation approaches that allow model development and performance evaluation to be conducted independently has been increasingly recommended in recent years (Yates et al., 2023; Pavlou et al., 2024).

Recent research on agricultural commodity price forecasting indicates that studies have largely focused on high-frequency time series or large-scale datasets. These

studies have primarily concentrated on the predictive performance of machine learning, deep learning, and hybrid modeling approaches (Guindani et al., 2024). In contrast, relatively few studies have systematically compared conventional linear regression with regularization- and dimensionality-reduction-based regression methods under the same validation strategy using small-sample agricultural datasets derived from annual official statistics. Furthermore, a substantial proportion of existing studies have evaluated model performance based on training data or single-level cross-validation results, without adequately accounting for the potentially optimistic estimates of predictive performance that may arise under small-sample conditions (Wainer and Cawley, 2021; Pavlou et al., 2024). This indicates that methodological uncertainty remains regarding which regression approach can provide more reliable and generalizable results for small-sample agricultural time series.

In this study, the performance of Ordinary Least Squares (OLS) Regression, Ridge Regression, Elastic Net, Principal Component Regression (PCR), and Partial Least Squares Regression (PLSR) was compared for forecasting real honey prices using beekeeping indicators from Türkiye for the period 2015–2025. All models were evaluated under the same data preprocessing procedure and a nested Leave-One-Out Cross-Validation (Nested LOOCV) framework designed to prevent information leakage. Thus, the objective of the study was not only to compare the predictive accuracy of different regression methods but also to provide a methodological framework for reliable model evaluation under conditions of small sample size and strong dependence among explanatory variables. In this respect, the study is expected to contribute to the literature on the selection of appropriate regression methods for small-sample agricultural time series.

2. Materials and Methods

2.1. Materials

The data used in this study were obtained from official beekeeping and precipitation statistics published by the Turkish Statistical Institute (TÜİK) and the Turkish State Meteorological Service (MGM) for the period 2015–2025. The dataset includes annual observations on the number of beekeeping operations, the number of modern and traditional beehives, honey production, beeswax production, annual precipitation, honey yield, and honey prices. The nominal honey prices used as the dependent variable in the analyses were converted into real honey prices using the Consumer Price Index (CPI) to eliminate the effects of inflation across years. All data used in the study are publicly available and were obtained from the official TÜİK database. The variables used in the analysis and their units of measurement are presented in Table 1.

Table 1. Descriptive statistics of the variables included in the regression analyses

Variable	Unit	Mean	SD	CV (%)	Minimum	Maximum
Real honey price	TRY kg ⁻¹	49.17	12.21	24.84	37.32	73.14
Enterprises	Count	88,715.91	7,457.25	8.41	80,675	100,399
Modern hives	Count	8,206,061.91	497,584.87	6.06	7,525,652	8,969,387
Traditional hives	Count	228,272.00	25,757.18	11.28	194,406	277,089
Honey production	t	106,538.82	7,758.87	7.28	95,492.31	118,297.46
Beeswax production	t	3,989.36	464.66	11.65	3,258.09	4,755.53
Annual rainfall	mm	555.66	55.74	10.03	502.30	659.40
Honey yield	kg hive ⁻¹	12.68	1.25	9.86	10.66	14.32
Beeswax ratio	%	3.74	0.33	8.82	3.35	4.40

Descriptive statistics were calculated from annual observations collected between 2015 and 2025 (n=11). SD= standard deviation, CV= coefficient of variation, TRY= Turkish Lira.

2.2. Methods

2.2.1. Data preprocessing and descriptive statistics

Prior to the regression analyses, the dataset was examined for missing observations, duplicate records, and inconsistent data entries, and only complete and consistent observations were included in the analyses. The nominal honey prices used as the dependent variable were converted into real prices using the Consumer Price Index (CPI) to eliminate the effects of inflation across years. To assess the general characteristics of the dataset, the mean, standard deviation, minimum value, maximum value, and coefficient of variation (CV, %) were calculated for all variables. These descriptive statistics were used not only to characterize the central tendency and dispersion of the variables but also to facilitate comparisons of relative variability among variables measured on different scales.

Because the independent variables were measured on different scales, z-score standardization was applied prior to the analyses to eliminate the effects of scale differences in the regression models, as shown in equation 1.

$$z_i = \frac{x_i - \bar{x}}{s} \quad (1)$$

Here, x_i denotes the corresponding observation value, μ represents the variable mean, and σ denotes the standard deviation. Following standardization, all variables were rescaled to have a mean of zero and a standard deviation of one. This procedure contributes to obtaining more stable and comparable model estimates, particularly for scale-sensitive methods such as Ridge Regression, Elastic Net, PCR, and PLSR (Sharma et al., 2024; Tilevik, 2026). Finally, standardized boxplots were generated and visually examined to assess the overall structure of the data distributions and identify potential outliers.

2.2.2. Trend analysis

The non-parametric Mann-Kendall trend test was applied to determine the temporal trends of the variables included in the study. This test is widely used in environmental, agricultural, and hydrological studies because it does not require any assumption regarding the normality of the data distribution and provides reliable

results for small-sample time series (Meshram et al., 2020; Gadedjisso-Tossou et al., 2021; Yue and Pilon, 2004). The Mann-Kendall test statistic was calculated as shown in equation 2:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (2)$$

Here, n denotes the number of observations, while x_i and x_j represent the i th and j th observation values, respectively, and sgn denotes the sign function. A positive S value indicates an increasing trend, whereas a negative S value indicates a decreasing trend. In addition to determining the direction of the trend, the Theil-Sen slope estimator was used to quantify its magnitude. Since this method is based on the median of the slope estimates calculated from all pairs of observations, it is less sensitive to outliers and provides reliable slope estimates, particularly for short time series (Sen, 1968; Bal, 2024). The slope coefficient was calculated using equation 3:

$$\beta = \text{Median}\left(\frac{x_j - x_i}{j - i}\right), i < j \quad (3)$$

Here, β represents the median rate of change of the variable per unit of time. Positive values indicate an increasing trend, whereas negative values indicate a decreasing trend. In this study, the statistical significance of the trend was assessed using the Mann-Kendall test, while the direction and magnitude of the trend were jointly interpreted using the Theil-Sen slope estimator. Thus, the temporal variation characteristics of all variables included in the study were evaluated independently of parametric assumptions.

2.2.3. Correlation analysis and assessment of multicollinearity

Prior to applying the regression models, Pearson correlation analysis was performed to determine the linear relationships among the variables. The Pearson correlation coefficient is one of the most widely used parametric methods for assessing the direction and strength of the linear relationship between two

continuous variables (Šverko et al., 2022). The correlation coefficient was calculated using equation 4:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

Here, r denotes the Pearson correlation coefficient, x_i and y_i represent the observation values, and $\bar{x}$ and $\bar{y}$ denote the means of the corresponding variables. The correlation coefficient ranges from -1 to $+1$, with increasing absolute values indicating a stronger linear relationship between the variables. A high degree of linear association among the independent variables may reduce the stability of regression coefficient estimates and complicate model interpretation. Therefore, the problem of multicollinearity was assessed using the Variance Inflation Factor (VIF) (Kyriazos and Poga, 2023; Faraz et al., 2023; Şahin 2026). The VIF was calculated using equation 5:

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

Here, R_i^2 denotes the coefficient of determination of the auxiliary regression model constructed for the corresponding independent variable using the other independent variables. An increase in the VIF value indicates a higher degree of linear association between the corresponding variable and the other explanatory variables. In this study, VIF values were used to assess multicollinearity among the explanatory variables prior to the application of the regression models. Correlation analysis and VIF assessment were conducted solely to examine the structure of the data, and no variables were excluded from the analyses. The primary reason for this approach is that the methods compared in the study, including Ridge Regression, Elastic Net, Principal Component Regression (PCR), and Partial Least Squares Regression (PLSR), are regression techniques specifically developed to address multicollinearity. Thus, all models were compared using the same set of explanatory variables, ensuring a fair evaluation across the methods.

2.2.4. Regression models

To compare the performance of different regression approaches in forecasting real honey prices, a total of five regression methods were evaluated, comprising one conventional and four modern approaches. These included Ordinary Least Squares (OLS) Regression, Ridge Regression, Elastic Net Regression, Principal Component Regression (PCR), and Partial Least Squares Regression (PLSR). These methods differ in their approaches to addressing multicollinearity among explanatory variables. While the OLS model represents the conventional linear regression approach, Ridge and Elastic Net are regularization-based methods. In contrast, PCR and PLSR are dimensionality-reduction-based regression methods that perform modeling by transforming explanatory variables into lower-dimensional components (Önder et al., 2025; Tirink and

Önder., 2022). The same dependent and independent variables were used in all regression models to enable direct comparison among the methods. Model hyperparameters were optimized using only the training data, and all performance evaluations were conducted under the same validation strategy. Thus, differences among the methods were attributable solely to their respective modeling approaches. The following subsections provide detailed descriptions of each regression method in terms of its theoretical foundations, mathematical formulation, and implementation in the present study.

Ordinary Least Squares Regression

Ordinary Least Squares (OLS) Regression is the most fundamental regression method used to model the linear relationship between a dependent variable and one or more independent variables. The OLS approach estimates the regression coefficients by minimizing the sum of squared errors between the observed values and the values predicted by the model (Burton., 2021; Mahaboob et al., 2018). The general linear regression model is expressed by equation 6.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (6)$$

Here, Y denotes the dependent variable, X represents the independent variables, β_0 denotes the intercept, β represents the regression coefficients, and ε denotes the error term. In the OLS method, the regression coefficients are estimated as the values that minimize the Residual Sum of Squares (RSS), as calculated in equation 7.

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

Minimizing this objective function yields the regression coefficients in matrix form, as shown in equation 8.

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (8)$$

Here, X denotes the matrix of explanatory variables, Y represents the dependent variable vector, and $\hat{\beta}$ denotes the estimated regression coefficients. For OLS regression to provide reliable results, several fundamental assumptions are expected to be satisfied, including linearity, independence of error terms, constant variance (homoscedasticity), approximately normally distributed error terms, and the absence of severe multicollinearity among the explanatory variables (Aslam., 2014). Particularly in datasets with high multicollinearity, the variance of the regression coefficients may increase, thereby reducing the predictive stability of the model. In this study, the OLS model was used as the baseline reference model without regularization or dimensionality reduction. The results were evaluated by comparison with the Ridge Regression, Elastic Net, Principal Component Regression (PCR), and Partial Least Squares Regression (PLSR) models under the same validation strategy.

Ridge Regression

Ridge Regression is a regularization-based regression method developed to improve the stability of OLS regression in situations where a high degree of multicollinearity exists among explanatory variables. The Ridge approach shrinks coefficient estimates by adding a penalty term based on the squared regression coefficients (L_2 penalty term) to the residual sum of squares, thereby reducing coefficient variance and improving model generalizability (Owen 2007; Hastie et al., 2022). The objective function for Ridge Regression is presented in equation 9.

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (9)$$

Here, λ denotes the regularization parameter. When $\lambda=0$, the model is equivalent to the conventional OLS regression, whereas as the value of λ increases, the regression coefficients are increasingly shrunk toward zero. However, because Ridge Regression does not set coefficients exactly to zero, all explanatory variables included in the model are retained. In Ridge Regression, the coefficient estimates are obtained in matrix form as shown in equation 10.

$$\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T y \quad (10)$$

Here, I denotes the identity matrix. The addition of the penalty term reduces numerical instability that may arise when taking the inverse of the $X^T X$ matrix and substantially mitigates the adverse effects of multicollinearity on regression coefficients (Wang et al., 2021). The performance of Ridge Regression depends largely on the selection of the regularization parameter (λ). While very small λ values produce results similar to those of the OLS model, excessively large values may cause the coefficients to be shrunk excessively, resulting in underfitting. Therefore, in this study, the optimal λ value was determined exclusively using the training data within the inner loop of the Nested Leave-One-Out Cross-Validation (Nested LOOCV) procedure. Ridge Regression was applied in this study particularly to mitigate the effects of the multicollinearity identified among the explanatory variables on model performance and to evaluate the predictive performance of regularization-based approaches. Model performance was compared with that of the other regression methods under the same validation procedure.

Elastic Net Regression

Elastic Net Regression is a hybrid regularization method that combines the advantages of Ridge and LASSO regression. By combining the stability of Ridge Regression against multicollinearity with the variable selection capability of the LASSO method, Elastic Net can provide effective results, particularly in datasets containing highly correlated explanatory variables (Zou and Hastie, 2005). Therefore, Elastic Net is widely preferred in regression problems where strong

relationships exist among explanatory variables. The objective function of Elastic Net is expressed in equation 11.

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \left[(1 - \alpha) \sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j| \right] \right\} \quad (11)$$

Here, λ denotes the penalty parameter that determines the level of regularization, while α represents the mixing parameter that controls the contribution of the L_1 and L_2 penalty terms to the model. When $\alpha=0$, the model is equivalent to Ridge Regression, whereas when $\alpha=1$, it becomes equivalent to LASSO Regression. One of the key advantages of Elastic Net is its ability to jointly account for groups of highly correlated variables. While the LASSO method often selects only one variable from a group of highly correlated variables, Elastic Net can retain these variables simultaneously in the model and produce more stable coefficient estimates (Zou and Hastie, 2005; Hastie et al., 2022). This characteristic provides an important advantage in studies involving explanatory variables that may exhibit strong interrelationships, such as agricultural production data. In this study, the Elastic Net hyperparameters, λ and α , were optimized exclusively using the training data within the inner loop of the Nested LOOCV approach. The model fitted using the optimal parameters was evaluated in the outer validation loop, and its performance was compared with that of the other regression methods using the same evaluation metrics.

Principal Component Regression

Principal Component Regression (PCR) is a regression method that combines Principal Component Analysis (PCA) with linear regression to reduce the effects of multicollinearity among explanatory variables. In PCR, highly correlated explanatory variables are first transformed into new components that are linearly uncorrelated with one another, and the regression model is subsequently fitted using these components (Lee et al., 2015; Xu and Hsu., 2019).

The principal components are expressed as linear combinations of the explanatory variables, as shown in equation 12.

$$PC_k = a_{1k}X_1 + a_{2k}X_2 + \dots + a_{pk}X_p \quad (12)$$

Here, PC_j denotes the j th principal component, p_{jk} represents the loading coefficients associated with the corresponding component, and X_k denotes the explanatory variables. The principal components are orthogonal to one another and are ordered according to the proportion of the total variance in the explanatory variables that they explain. In the PCR method, the regression model is fitted using the selected principal components rather than the original explanatory variables, as expressed in equation 13.

$$y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + \dots + \beta_m PC_m + \varepsilon \quad (13)$$

Here, k denotes the number of principal components included in the regression model. Thus, the problem of multicollinearity arising from highly correlated variables is substantially reduced. The performance of PCR depends largely on the number of principal components included in the model. Using too few components may result in the loss of important information, whereas including too many components may increase model complexity (Boggy and Woolf, 2010). Therefore, in this study, the optimal number of components was determined exclusively using the training data within the inner loop of the Nested Leave-One-Out Cross-Validation approach. A major limitation of PCR is that the principal components are constructed solely based on the variance in the explanatory variables. Consequently, components with high variance but weak associations with the dependent variable may be included in the model, whereas some components containing important information about the dependent variable may be overlooked (Agarwal et al., 2019). Therefore, the performance of PCR was compared with that of the PLSR method, which also incorporates information from the dependent variable.

Partial Least Squares Regression

Partial Least Squares Regression (PLSR) is a dimensionality-reduction-based regression method widely used for data structures characterized by multicollinearity and limited sample sizes. In addition to summarizing the variance in the explanatory variables, PLSR aims to maximize the covariance between the latent components and the dependent variable (Mevik and Wehrens, 2007; Mehmood et al., 2012). Owing to this characteristic, PLSR can provide more effective results than PCR for predicting the dependent variable. In the PLSR method, the explanatory variable matrix X and the dependent variable vector Y are simultaneously decomposed into lower-dimensional latent variables, as shown in equation 14.

$$\begin{aligned} X &= TP^T + E \\ y &= Uq^T + F \end{aligned} \quad (14)$$

Here, T and U denote the score matrices of the explanatory and dependent variables, respectively; P and Q represent the loading matrices; and E and F denote the residual matrices. The primary objective of the PLSR algorithm is to identify linear components that maximize the covariance between the latent components derived from the explanatory and dependent variables. Thus, each component incorporates not only the variance in the explanatory variables but also information relevant to explaining the dependent variable (Helland, 1990). This characteristic can improve model performance, particularly in datasets with highly correlated explanatory variables. In the PLSR model, the dependent variable is predicted using the linear structure given in equation 15.

$$\hat{y} = XB \quad (15)$$

Here, B denotes the matrix of regression coefficients estimated by the PLSR algorithm. In PCR, the principal components are constructed solely according to the variance in the explanatory variables, whereas in PLSR, information from the dependent variable is also incorporated into the component construction process. Therefore, PLSR may provide superior predictive performance, particularly in studies characterized by highly correlated explanatory variables and a limited number of observations (Mehmood et al., 2012; Cheng and Sun, 2017). In this study, the optimal number of components to be used in the PLSR model was determined exclusively using the training data within the inner loop of the Nested Leave-One-Out Cross-Validation approach. The resulting optimal model was then evaluated in the outer validation loop, and its performance was compared with that of the other regression methods using the same evaluation metrics.

2.2.5. Nested Leave-One-Out Cross-Validation

To obtain an unbiased assessment of the generalizable predictive performance of the regression models, a Nested Leave-One-Out Cross-Validation (Nested LOOCV) approach was used for all analyses. This method is considered one of the most reliable resampling approaches because it separates model development from model evaluation and reduces optimistic estimates of performance arising from hyperparameter optimization (Varma and Simon, 2006; Cawley and Talbot, 2006). Nested LOOCV consists of a two-stage validation structure. The outer loop is used to evaluate the model's independent test performance, whereas the inner loop is used exclusively for hyperparameter optimization. Thus, the test data were not involved in the model development process at any stage, thereby preventing data leakage. In this study, because the dataset consisted of 11 annual observations, one observation was held out as the test set in each iteration of the outer loop, while the remaining observations constituted the training set. A second Leave-One-Out Cross-Validation procedure was then applied to the training set to determine the optimal hyperparameters for each model. The regularization parameter (λ) in Ridge Regression, λ and the mixing parameter (α) in the Elastic Net model, and the optimal number of components in the PCR and PLSR models were optimized exclusively using the training data. The OLS model, which does not involve hyperparameters, was fitted directly to the training data. At the end of the inner loop, the optimal model was refitted using the entire training set from the outer loop, and predictions were generated for the previously unseen test observation. This process was repeated for each observation so that all predictions were obtained from independent test observations. Thus, model selection and performance evaluation were completely separated, allowing the actual generalization performance of the models to be assessed (Cawley and

Talbot, 2006). The Nested LOOCV approach was preferred because the dataset contained a limited number of observations. While the Leave-One-Out method uses the largest possible training set in each iteration, the nested validation structure reduces bias arising from model selection and provides more reliable estimates of predictive performance, particularly in small-sample studies (Vabalas ve ark., 2019).

2.2.6. Model performance metrics

The predictive performance of the regression models was compared using the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), prediction error sum of squares (PRESS), and predicted coefficient of determination (Predicted R^2) to jointly assess explanatory power and prediction error. The coefficient of determination (R^2) was calculated to evaluate the explanatory power of the models. To assess the magnitude of prediction error, the root mean squared error (RMSE) was calculated using the following equation. The mean absolute error (MAE) was calculated to determine the average absolute deviation of the model predictions. The prediction error sum of squares (PRESS) was used to assess the total prediction error obtained during the cross-validation process. Finally, the predicted coefficient of determination (Predicted R^2) was calculated to evaluate the explanatory power of the models for observations outside the training sample. In this study, all regression models were evaluated under the same Nested LOOCV procedure, and performance comparisons were based exclusively on the R^2 , RMSE, MAE, PRESS, and Predicted R^2 values obtained from out-of-sample predictions. The best-performing model was identified as the method that achieved the highest R^2 and Predicted R^2 values while providing the lowest RMSE, MAE, and PRESS values. All data preprocessing, statistical analyses, and regression modeling were performed using the Python 3.12 programming language. Pandas and NumPy were used for data manipulation, SciPy for statistical analyses, Scikit-learn for regression modeling and cross-validation procedures, and Matplotlib for graphical visualization. All analyses were conducted within the same computational environment, and all regression models were evaluated using the same data preprocessing steps and validation strategy.

3. Results

3.1. Descriptive Statistics and Data Distribution

The descriptive statistics for the variables used in the study are presented in Table 1. During the study period, the mean real honey price was 49.17 TL kg⁻¹, with a standard deviation of 12.21 and a coefficient of variation (CV) of 24.84%. Among the variables examined, the highest variability was observed in real honey prices. This finding suggests that honey prices may be influenced not only by production-related indicators but also by market conditions and economic factors. Evaluating real prices after adjusting for inflation allows

actual changes in prices across years to be compared more reliably.

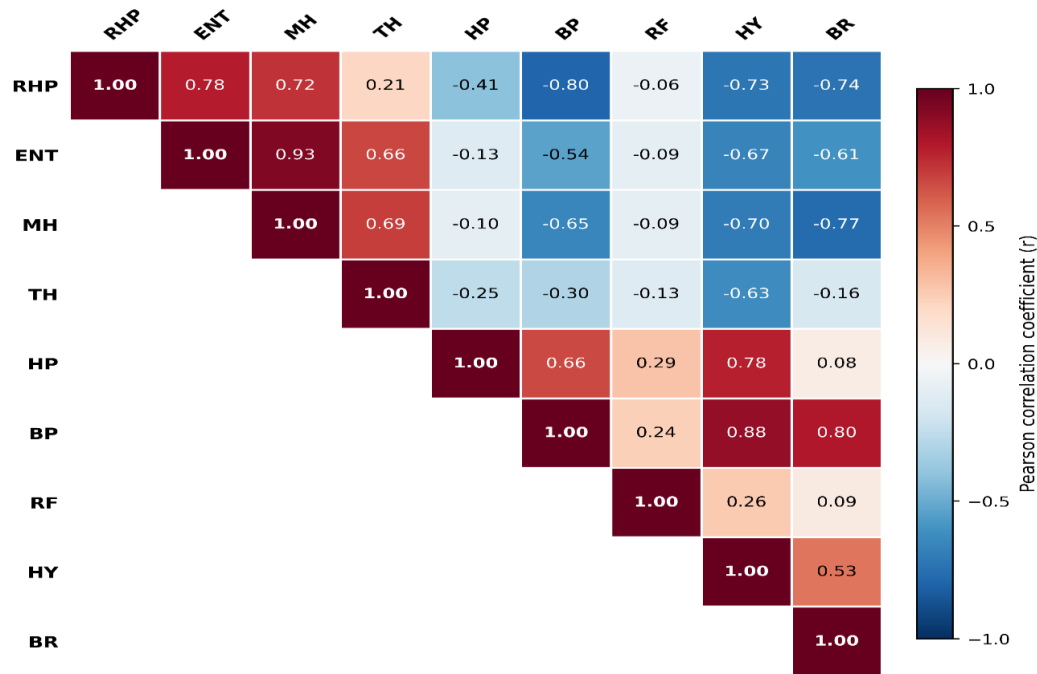
The mean number of beekeeping operations was 88.715, with a coefficient of variation of 8.41%, indicating relatively stable variation in the number of operations throughout the study period. The mean number of modern beehives was 8.206.061, with a coefficient of variation of 6.06%, representing one of the lowest levels of variability among all variables. In contrast, the mean number of traditional beehives was 228.272, with a coefficient of variation of 11.28%, indicating a higher degree of variability compared with modern beehives.

The mean honey production was 106.538.82 tons, with a coefficient of variation of 7.28%. Similarly, the mean honey yield per hive was 12.68 kg hive⁻¹, with a coefficient of variation of 9.86%, indicating that production efficiency remained relatively stable throughout the study period. The mean beeswax production was 3.989.36 tons, with a coefficient of variation of 11.65%, while the mean beeswax ratio was 3.74%, with a coefficient of variation of 8.82%. The mean annual precipitation was 555.66 mm, with a coefficient of variation of 10.03%. These findings indicate that production and environmental indicators generally exhibited low to moderate levels of variability.

Overall, the dataset consisted of explanatory variables exhibiting varying levels of variability. In particular, the relatively low variability observed in variables such as the number of modern and traditional beehives, production volume, and honey yield indicates a certain degree of stability across annual observations. In contrast, the higher variability of real honey prices compared with the other variables suggests that, in addition to production-related indicators, economic and sectoral factors may also influence price formation. The standardized distributions of the variables are presented in Figure 1. Examination of the boxplots based on z-scores indicates that the overall distributions of the variables were balanced and that there were no apparent outlying observations likely to substantially affect the results of the analyses. Furthermore, standardization enabled variables measured in different units to be evaluated on a common scale, thereby eliminating potential effects arising from differences in scale in the regression models.

3.2. Assessment of Temporal Trends

The annual changes in all variables examined in the study are presented in Figure 2. Overall, the graphs indicate that the number of beekeeping operations in the sector increased over the study period. The number of operations generally exhibited an upward trend throughout the years. This increase was relatively continuous and consistent, suggesting a stable expansion of the production capacity of the sector. Similarly, a clear increase in the number of modern beehives was observed throughout the study period.



RHP, real honey price; ENT, enterprises; MH, modern hives; TH, traditional hives; HP, honey production; BP, beeswax production; RF, annual rainfall; HY, honey yield; BR, beeswax ratio.

Figure 1. Upper-triangular Pearson correlation matrix among the variables included in the regression analyses.

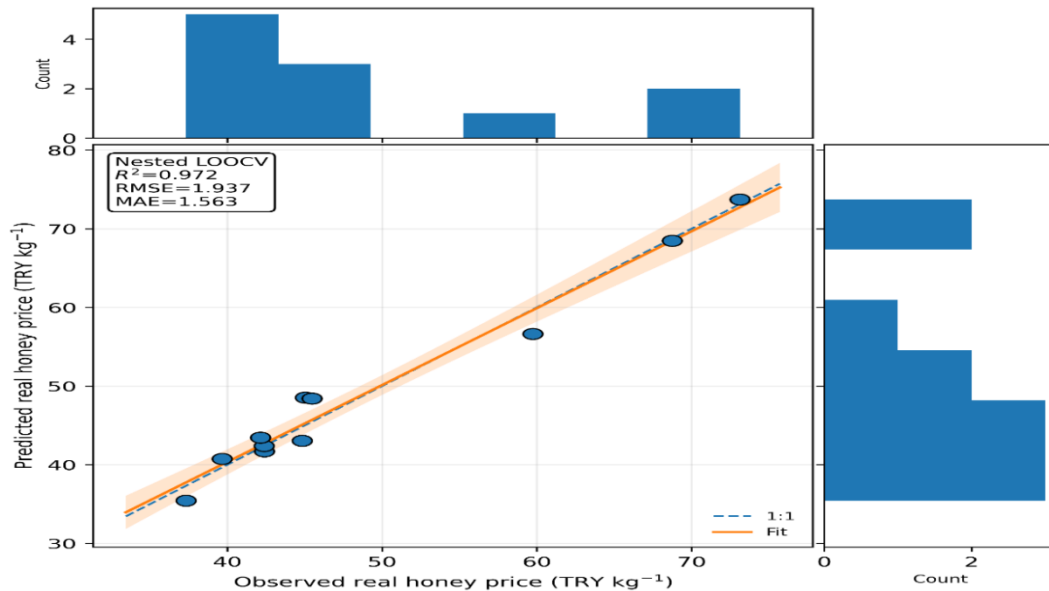


Figure 2. Observed versus nested leave-one-out cross-validated predictions of real honey price obtained from the partial least squares regression model.

The more consistent increase in the number of modern beehives compared with the growth in the number of beekeeping operations suggests that the use of modern beehives became increasingly widespread during the study period. In contrast, a pronounced decreasing trend was observed in the number of traditional beehives. This finding suggests a shift from traditional production practices toward modern beehive systems in Turkish beekeeping. When the trends in the numbers of modern and traditional beehives are considered together, they indicate a structural transformation in the production system.

Honey production remained relatively stable throughout the study period, despite limited year-to-year fluctuations. Honey yield per hive similarly did not exhibit substantial fluctuations. This indicates that variations in production volume are influenced not only by the number of beehives but also by environmental conditions and management practices. In contrast, real honey prices exhibited a more volatile pattern than the other variables. In particular, the increases observed in some years suggest that, in addition to production-related indicators, market conditions and economic factors may also play an important role in price

formation. The real price series obtained after adjusting for inflation provides a more comparable basis for evaluating price changes alongside production-related indicators.

Overall, the trend analyses revealed an increasing trend in modern production indicators and a decreasing trend in traditional production indicators. In contrast, production and yield indicators exhibited more limited changes. Although the findings indicate trends in different directions across production-related indicators during the study period, they suggest that real honey prices cannot be explained by a single production indicator alone. Rather, production capacity, environmental conditions, and structural changes within the sector should be considered jointly. Therefore, in the subsequent stage of the study, the predictive performance of different regression methods was compared.

3.3. Comparison of Regression Models

The Nested Leave-One-Out Cross-Validation results for the five regression methods are presented in Table 2. Overall, the models exhibited substantial differences in predictive performance. Although the conventional Ordinary Least Squares (OLS) model produced a relatively high R^2 value in terms of apparent fit, it showed lower performance than the other methods in out-of-sample validation. This finding suggests that the small sample size and high correlation among the explanatory variables may have limited the out-of-sample predictive performance of the conventional linear regression model. Ridge Regression produced more balanced predictions than the OLS model by shrinking the regression coefficients through the L_2 penalty term. This regularization approach particularly reduced the effects of multicollinearity and resulted in a marked improvement in cross-validation performance. However, because the Ridge model retains all variables in the model, it may not completely eliminate the shared information structure among highly correlated variables in some cases.

Although the Elastic Net model combines the Ridge and LASSO approaches, it did not provide a clear advantage over the Ridge model for this dataset. This may be because most of the explanatory variables in the dataset carry shared information, and variable selection may not have substantially improved predictive performance. The

PCR method largely eliminated the problem of multicollinearity by transforming the explanatory variables into orthogonal principal components. Consequently, it achieved better predictive performance than the OLS model. However, because the principal components in PCR are constructed solely based on the variance in the explanatory variables, some information that is important for explaining the dependent variable may not have been fully captured by the model.

The highest predictive performance was achieved by the PLSR model. The PLSR model yielded the highest Predicted R^2 (0.972) along with the lowest RMSE (1.937), MAE (1.563), and PRESS (41.290) values. These results indicate that PLSR achieved superior out-of-sample predictive performance for this dataset by simultaneously accounting for the shared variance among the explanatory variables and their covariance with the dependent variable. The superior performance of the PLSR model may be particularly attributable to the highly correlated structure of the dataset. Variables such as the number of modern beehives, number of beekeeping operations, honey production, and honey yield exhibited close relationships with one another. By effectively summarizing this shared information structure through latent components, PLSR provided higher out-of-sample predictive accuracy than the other methods. The observed and predicted real honey prices and the distribution of residuals for the best-performing PLSR model are presented in Figure 3. The close agreement between the observed and predicted values indicates that the model successfully represented the underlying structure of the data. The random distribution of residuals around zero further suggests that there was no pronounced systematic error in the model and that the prediction errors did not exhibit any clear systematic pattern.

Overall, the regularization- and dimensionality-reduction-based methods were found to achieve better out-of-sample predictive performance than conventional linear regression for this dataset, which was characterized by a small sample size and substantial multicollinearity. In particular, the PLSR method emerged as the best-performing approach among the methods evaluated, both in terms of predictive accuracy and model stability.

Table 2. Predictive performance of the evaluated regression models under nested leave-one-out cross-validation (LOOCV)

Model	Apparent R^2	LOOCV R^2 (95% CI)	RMSE (95% CI)	MAE (95% CI)	PRESS
PLSR	0.996	0.972 (0.606–0.992)	1.937 (1.204–2.513)	1.563 (0.898–2.255)	41.290
Ridge	0.996	0.962 (0.315–0.982)	2.259 (1.706–2.722)	2.040 (1.449–2.605)	56.121
PCR	0.996	0.961 (0.299–0.987)	2.304 (1.566–2.910)	1.970 (1.285–2.670)	58.403
Elastic Net	0.996	0.960 (0.367–0.981)	2.315 (1.769–2.753)	2.061 (1.403–2.652)	58.940
OLS	0.998	0.950 (0.140–0.973)	2.604 (1.844–3.201)	2.277 (1.526–3.009)	74.568

Apparent R^2 was calculated using the complete dataset. LOOCV R^2 , RMSE, and MAE are reported together with bootstrap 95% confidence intervals. PRESS = Prediction Residual Error Sum of Squares. Higher LOOCV R^2 and lower RMSE, MAE, and PRESS values indicate better predictive performance.

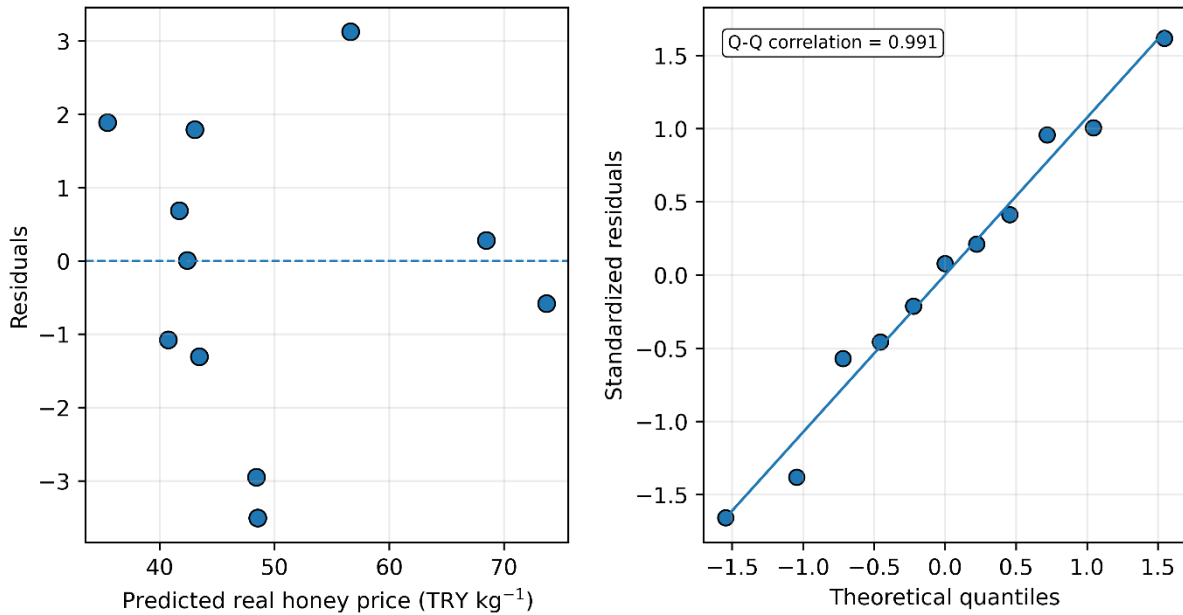


Figure 3. Residual diagnostics of the optimal partial least squares regression model based on nested leave-one-out cross-validation: (A) Residuals versus predicted values illustrating the absence of systematic prediction errors. (B) Normal Q-Q plot of standardized residuals for assessing residual normality.

4. Discussion

In this study, the predictive performance of OLS, Ridge Regression, Elastic Net, PCR, and PLSR models was compared using small-sample annual data from the Turkish beekeeping sector covering the 2015–2025 period. The Nested Leave-One-Out Cross-Validation results showed that the PLSR model provided the highest out-of-sample explanatory power while also achieving the lowest prediction errors. This finding indicates that conventional linear regression may not always be the most appropriate predictive approach for data structures characterized by substantial multicollinearity among explanatory variables and a limited number of observations. Similarly, methodological studies have reported that regularization- and dimensionality-reduction-based regression methods can provide more stable and generalizable predictions under such data conditions (Dormann et al., 2013; Hastie et al., 2021; James et al., 2021; Mehmood et al., 2012).

The superior performance of the PLSR model compared with the other methods in this study is consistent with its theoretical properties. Whereas PCR constructs latent components solely based on the variance in the explanatory variables, PLSR also accounts for the covariance between the explanatory and dependent variables during the component construction process (Wold et al., 1984; Helland, 1990). Consequently, the resulting components are not intended solely for dimensionality reduction but also represent linear combinations that contain the greatest amount of information relevant to prediction. This property can enhance the predictive performance of PLSR, particularly in data structures characterized by high correlations among explanatory variables, and helps explain its

widespread use in multivariate regression problems (Mehmood et al., 2012; Hastie et al., 2021). The highest out-of-sample explanatory power achieved by PLSR in the present study is also consistent with this theoretical expectation.

Although the dataset used in this study consisted of only 11 annual observations, the high out-of-sample performance of the PLSR model is noteworthy. In datasets with small sample sizes, increasing model complexity raises the risk of overfitting and may cause performance metrics based solely on the training data to overestimate model performance (Varma and Simon, 2006; Cawley and Talbot, 2006). In the present study, the use of Nested Leave-One-Out Cross-Validation separated the model selection and model performance evaluation processes, thereby reducing selection bias and enabling a more reliable assessment of out-of-sample predictive performance (Cawley and Talbot, 2006; Yates et al., 2023). Similarly, Pavlou et al. (2024) emphasized that the use of appropriate validation strategies in small-sample datasets is critical for preventing overly optimistic estimates of model performance. Therefore, the validation approach employed in this study is considered to have enhanced the reliability of the reported performance results.

Although the OLS model showed a high degree of fit on the training data, it performed worse than the PLSR and Ridge models in out-of-sample validation. This finding can be explained by the fact that multicollinearity increases the variance of regression coefficient estimates and reduces model stability in data structures with a limited number of observations (Dormann et al., 2013). Ridge Regression, in contrast, produced more stable results than the OLS model by reducing prediction

variance through the application of L_2 regularization to the coefficients (Hoerl and Kennard, 1970). However, because the Ridge model retains all explanatory variables in the model, it does not eliminate the shared information structure among highly correlated variables but rather stabilizes the coefficient estimates (Hastie et al., 2021). Indeed, both methodological studies and comparative applications have shown that Ridge Regression can provide more reliable and lower-variance predictions than conventional linear regression in data structures characterized by multicollinearity (James et al., 2021; Hastie et al., 2021).

Although the Elastic Net model combines the characteristics of Ridge and LASSO methods, it underperformed relative to PLSR in this study. This outcome suggests that since a large majority of the explanatory variables in the dataset are highly correlated, preserving the shared information structure for predictive performance may be more crucial than variable selection. Elastic Net yields particularly effective results in high-dimensional data structures and problems containing numerous irrelevant or weak explanatory variables (Zou and Hastie, 2005; Hastie et al., 2021). Conversely, in data structures where the number of explanatory variables is limited and all variables are biologically or economically meaningful, PLSR's ability to construct components that account for the shared information structure between the explanatory variables and the dependent variable can enhance predictive performance (Helland, 1990; Mehmood et al., 2012). The findings obtained in our study are aligned with this methodological expectation.

5. Conclusion

In this study, the performances of OLS, Ridge Regression, Elastic Net, PCR, and PLSR models in predicting real honey prices were compared using small-sample annual time-series data covering the 2015–2025 period for the Turkish beekeeping sector. All models were evaluated under the same data preprocessing workflow and a Nested Leave-One-Out Cross-Validation approach. The results demonstrated that the PLSR model was the most successful approach among the evaluated methods, providing the highest out-of-sample explanatory power along with the lowest prediction errors. In contrast, the OLS model exhibited lower generalizability, while Ridge Regression, Elastic Net, and PCR methods yielded better results than classical linear regression but fell short of PLSR's performance. The findings reveal that the PLSR method is a reliable and effective alternative for agricultural time series characterized by small sample sizes and multicollinearity. Furthermore, the Nested Leave-One-Out Cross-Validation approach utilized in this study allowed for an unbiased evaluation of the models' out-of-sample performance. In future studies, it is considered that predictive performance could be further enhanced by incorporating longer time series, regional data, and climatic variables into the models.

Author Contributions

The percentage of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	M.B.	L.B.
C	50	50
D	50	50
S	50	50
DCP	50	50
DAI	50	50
L	50	50
W	50	50
CR	50	50
SR	50	50
PM	50	50
FA	50	50

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declare that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

- Agarwal, A., Shah, D., Shen, D., & Song, D. (2019). On robustness of principal component regression. *Advances in Neural Information Processing Systems*, 32.
- Aslam, M. (2014). Using heteroscedasticity-consistent standard errors for the linear regression model with correlated regressors. *Communications in Statistics—Simulation and Computation*, 43(10), 2353–2373.
- Avinash, G., Ramasubramanian, V., Ray, M., Paul, R. K., Godara, S., Nayak, G. H. H., Kumar, R. R., Manjunatha, B., Dahiya, S., & Iquebal, M. A. (2024). Hidden Markov guided deep learning models for forecasting highly volatile agricultural commodity prices. *Applied Soft Computing*, 158, 111557. <https://doi.org/10.1016/j.asoc.2024.111557>
- Bal, A. (2024). Improving the robustness of the Theil-Sen estimator using a simple heuristic-based modification. *Symmetry*, 16(6), 698. <https://doi.org/10.3390/sym16060698>
- Boggy, G. J., & Woolf, P. J. (2010). A mechanistic model of PCR for accurate quantification of quantitative PCR data. *PLOS One*, 5(8), e12355.
- Burton, A. L. (2021). *OLS (linear) regression*. In *The Encyclopedia of Research Methods in Criminology and Criminal Justice* (Vol. 2, pp. 509–514).
- Cawley, G. C., & Talbot, N. L. C. (2006). Gene selection in cancer classification using sparse logistic regression with Bayesian regularization. *Bioinformatics*, 22(19), 2348–2355. <https://doi.org/10.1093/bioinformatics/btl386>
- Cheng, J. H., & Sun, D. W. (2017). Partial least squares regression (PLSR) applied to NIR and HSI spectral data modeling to predict chemical properties of fish muscle. *Food*

- Engineering Reviews*, 9(1), 36–49. <https://doi.org/10.1007/s12393-016-9147-1>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Faraz, A., Tirink, C., Önder, H., Şen, U., Ishaq, H. M., Tauqir, N. A., Waheed, A., & Nabeel, M. S. (2023). Usage of the XGBoost and MARS algorithms for predicting body weight in Kajli sheep breed. *Tropical Animal Health and Production*, 55(4), 276. <https://doi.org/10.1007/s11250-023-03700-6>
- Gadedjisso-Tossou, A., Adjegan, K. I., & Kablan, A. K. M. (2021). Rainfall and temperature trend analysis by Mann–Kendall test and significance for rainfed cereal yields in Northern Togo. *Sci*, 3(1), 17. <https://doi.org/10.3390/sci3010017>
- Ghasemzadeh, H., Hillman, R. E., & Mehta, D. D. (2024). Toward generalizable machine learning models in speech, language, and hearing sciences: Estimating sample size and reducing overfitting. *Journal of Speech, Language, and Hearing Research*, 67(3), 753–781. <https://doi.org/10.1044/2023.JSLHR-23-00273>
- Guindani, L. G., Oliveira, G. A., & Ribeiro, M. H. D. M. (2024). Exploring current trends in agricultural commodities forecasting methods through text mining: Developments in statistical and artificial intelligence methods. *Heliyon*, 10(23), e40568. <https://doi.org/10.1016/j.heliyon.2024.e40568>
- Hastie, T., Montanari, A., Rosset, S., & Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, 50(2), 949–986.
- Hastie, T., Tibshirani, R., & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Helland, I. S. (1990). Partial least squares regression and statistical models. *Scandinavian Journal of Statistics*, 17(2), 97–114.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer.
- Jin, Y., Li, W., & Gil, J. M. (2024). Forecasting fish prices with an artificial neural network model during the tuna fraud. *Journal of Agriculture and Food Research*, 18, 101340. <https://doi.org/10.1016/j.jafr.2024.101340>
- Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Applied Statistics*, 31(3), 300–303. <https://doi.org/10.2307/2348005>
- Klein, N. (2024). Distributional regression for data analysis. *Annual Review of Statistics and Its Application*, 11, 321–346. <https://doi.org/10.1146/annurev-statistics-040722-053607>
- Kyriazos, T., & Poga, M. (2023). Dealing with multicollinearity in factor analysis: The problem, detections, and solutions. *Open Journal of Statistics*, 13(3), 404–424.
- Lee, H., Park, Y. M., & Lee, S. (2015). Principal component regression by principal component selection. *Communications for Statistical Applications and Methods*, 22(2), 173–180.
- Mahaboob, B., Venkateswarlu, B., Narayana, C., & Balasiddamuni, P. (2018). A treatise on ordinary least squares estimation of parameters of linear model. *International Journal of Engineering & Technology*, 7(4.10), 518.
- Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 118, 62–69. <https://doi.org/10.1016/j.chemolab.2012.07.010>
- Meshram, S. G., Kahya, E., Meshram, C., Ghorbani, M. A., Ambade, B., & Mirabbasi, R. (2020). Long-term temperature trend analysis associated with agriculture crops. *Theoretical and Applied Climatology*, 140(3–4), 1139–1159. <https://doi.org/10.1007/s00704-020-03137-z>
- Mevik, B. H., & Wehrens, R. (2007). The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, 18(2), 1–23. <https://doi.org/10.18637/jss.v018.i02>
- Moscovich, A., & Rosset, S. (2022). On the cross-validation bias due to unsupervised preprocessing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(4), 1474–1502. <https://doi.org/10.1111/rssb.12537>
- Önder, H., Tirink, C., Yakubets, T., Getya, A., Matvieiev, M., Kononenko, R., Şen, U., Özkan, Ç. Ö., Tolun, T., & Kaya, F. (2025). Predicting live weight for female rabbits of meat crosses from body measurements using LightGBM, XGBoost and support vector machine algorithms. *Veterinary Medicine and Science*, 11(1), e70149. <https://doi.org/10.1002/vms3.70149>
- Owen, A. B. (2007). A robust hybrid of lasso and ridge regression. *Contemporary Mathematics*, 443, 59–72.
- Parvande, S., Yeh, H. W., Paulus, M. P., & McKinney, B. A. (2020). Consensus features nested cross-validation. *Bioinformatics*, 36(10), 3093–3098. <https://doi.org/10.1093/bioinformatics/btaa046>
- Pavlou, M., Omar, R. Z., & Ambler, G. (2024). Penalized regression methods with modified cross-validation and bootstrap tuning produce better prediction models. *Biometrical Journal*, 66(5), e202300245. <https://doi.org/10.1002/bimj.202300245>
- Şahin, M. (2026). Estimate of albumen height in Atak-S hen eggs using ridge regression. *Black Sea Journal of Agriculture*, 9(4), 520–525.
- Schreiber-Gregory, D. N. (2018). Ridge regression and multicollinearity: An in-depth review. *Model Assisted Statistics and Applications*, 13(4), 359–365. <https://doi.org/10.3233/MAS-180446>
- Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Sharma, P., Villegas-Diaz, R., & Fennell, A. (2024). Predicting grapevine physiological parameters using hyperspectral remote sensing integrated with hybrid convolutional neural network and ensemble stacked regression. *Remote Sensing*, 16(14), 2626. <https://doi.org/10.3390/rs16142626>
- Šverko, Z., Vrankić, M., Vlahinić, S., & Rogelj, P. (2022). Complex Pearson correlation coefficient for EEG connectivity analysis. *Sensors*, 22(4), 1477. <https://doi.org/10.3390/s22041477>
- Tilevik, A. (2026). *PCR, PLS, and Lasso Regression*. In *Multivariate Statistics and Machine Learning in R for Beginners: With Applications in Biology and Medicine* (pp. 279–297). Springer Nature Switzerland.
- Tirink, C., & Önder, H. (2022). Comparison of M, MM and LTS estimators in linear regression in the presence of outlier. *Turkish Journal of Veterinary & Animal Sciences*, 46(3), 420–

- 428.
- Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS One*, 14(11), e0224365. <https://doi.org/10.1371/journal.pone.0224365>
- Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7, 91. <https://doi.org/10.1186/1471-2105-7-91>
- Wainer, J., & Cawley, G. C. (2021). Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Systems with Applications*, 182, 115222. <https://doi.org/10.1016/j.eswa.2021.115222>
- Wang, X., Wang, X., Ma, B., Li, Q., & Shi, Y. Q. (2021). High precision error prediction algorithm based on ridge regression predictor for reversible data hiding. *IEEE Signal Processing Letters*, 28, 1125–1129.
- Wold, S., Ruhe, A., Wold, H., & Dunn, W. J., III. (1984). The collinearity problem in linear regression: The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3), 735–743. <https://doi.org/10.1137/0905052>
- Xu, J., & Hsu, D. J. (2019). On the number of variables to use in principal component regression. *Advances in Neural Information Processing Systems*, 32.
- Yates, L. A., Aandahl, Z., Richards, S. A., & Brook, B. W. (2023). Cross validation for model selection: A review with examples from ecology. *Ecological Monographs*, 93(1), e1557. <https://doi.org/10.1002/ecm.1557>
- Yue, S., & Pilon, P. (2004). A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. *Hydrological Sciences Journal*, 49(1), 21–37. <https://doi.org/10.1623/hysj.49.1.21.53996>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>



BIBLIOMETRIC MAPPING OF ZERO-INFLATED STATISTICAL MODELS

Memduh AK^{1*}, Cem TIRINK¹

¹Iğdır University, Faculty of Agriculture, Department of Animal Science, 76000, Iğdır, Türkiye

Abstract: In this study, the evolution of scientific literature published on zero-inflated models was examined using bibliometric methods. Records obtained from the Web of Science Core Collection database were analyzed in the R environment using bibliometrix and related packages. The main analysis period was set as 1997–2025, and a total of 1,067 publications were evaluated. Annual scientific output, citation trends, leading sources, and authors were examined, along with keyword co-occurrence networks, temporal trends, emerging research topics, thematic structure, and reference co-citation networks. The findings indicate that the literature on zero-inflated models has grown rapidly, particularly after 2010, reaching its highest annual publication count in 2024. While zero inflation, overdispersion, zero-inflated Poisson, hurdle models, and the negative binomial model form the core conceptual framework of the field, Bayesian inference, marginalized models, INLA, and deep learning have emerged as topics gaining increasing prominence in recent years. Citation network analysis highlighted the central role of seminal works by Lambert, Mullahy, Hall, and Ridout in the field's intellectual structure. The results indicate that zero-inflated models constitute a dynamic, interdisciplinary field of statistical research that continues to diversify methodologically.

Keywords: Bibliometric analysis, Zero-inflated models, Zero-inflated Poisson, Overdispersion, Count data, Science mapping, Thematic analysis

*Corresponding author: Iğdır University, Faculty of Agriculture, Department of Animal Science, 76000, Iğdır, Türkiye

E mail: memduh.ak.33@gmail.com (M. AK)

Memduh AK  <https://orcid.org/0000-0001-5894-8986>

CemTIRINK  <https://orcid.org/0000-0001-6902-5837>

Received: April 03, 2026

Accepted: May 21, 2026

Published: June 15, 2026

Cite as: Ak, M., & Tirink, C. (2026). Bibliometric mapping of zero-inflated statistical models. *Black Sea Journal of Statistics*, 2(1), 28-43.

1. Introduction

Count data is a type of data commonly encountered in many scientific disciplines that indicates how many times an observed event occurs within a specific time frame, area, or experimental unit. Examples of this type of data include the number of diseases or symptoms, the frequency of traffic accidents, insurance claims, species abundance, the number of harmful organisms, production defects, the frequency of healthcare utilization, and the number of technological applications. One of the most fundamental models in the analysis of count data is Poisson regression. However, the fundamental assumption of the Poisson distribution—that the mean equals the variance—is often not met in real-world data structures. In particular, overdispersion—defined as the observed variance being significantly larger than the mean—poses a significant modeling problem in count data analysis. While negative binomial regression is one of the primary alternatives developed to address this issue, in some datasets, a significant portion of the overdispersion stems from the fact that the number of observed zero values is much higher than predicted by the standard Poisson or negative binomial distributions (Cameron & Trivedi, 2013; Hilbe, 2011). This phenomenon is referred to in the literature as “excess zeros” or “zero inflation”.

When zeros occur at a higher rate than expected in the normal counting process, this can affect not only the goodness-of-fit of classical counting models but also parameter estimates and the scientific conclusions drawn from those estimates. The fundamental idea behind zero-inflated models is that not all observed zeros may stem from the same data-generation mechanism. In this approach, a portion of the zero observations is treated as “structural zeros”, arising from a separate condition where the occurrence of the event in question is impossible or extremely unlikely in practice. In contrast, the remaining zeros are considered “sampling zeros”, which occur randomly from the normal counting distribution. Consequently, a zero-inflated model is generally expressed as a mixture of a component that produces zeros and a counting component that can produce both positive counts and some zeros. This approach allows zeros to be modeled not merely as a distributional issue but as an indicator of different data-generation mechanisms that may underlie the observed process (Lambert, 1992; Ridout et al., 1998; Young et al., 2022a). Lambert's (1992) Zero-Inflated Poisson (ZIP) regression model, developed for production errors, has become one of the most fundamental studies of this approach in modern statistical literature and has served as an important starting point for subsequent



methodological developments.

However, the conceptual foundation for modeling census data with excessive zeros had begun to take shape even before Lambert's (1992) work. Mullahy (1986) emphasized that standard census models might be insufficient to explain heterogeneity in data-generation processes and made significant contributions to modified census models, particularly through hurdle models. Although hurdle and zero-inflated models can be applied to similar data structures, they are not conceptually the same. In hurdle models, zero and positive outcomes are generated by two separate processes, and the positive count distribution is truncated at zero; therefore, zeros cannot be generated from the count component. In zero-inflated models, however, zeros can arise both from a specific zero-generation mechanism and from the normal process of a Poisson, negative binomial, or other count distribution. This distinction is not merely a difference in mathematical parameterization but is important for the interpretation of model parameters and the scientific justification of model selection (Mullahy, 1986; Feng, 2021).

Although the ZIP model is the most common zero-inflated approach, the fact that the mean-variance equality assumption of the Poisson component is not satisfied for nonzero observations has necessitated the use of more flexible distributions. For this reason, model families based on the zero-inflated negative binomial (ZINB), the zero-inflated generalized Poisson, and other discrete distributions have been developed. In particular, using the ZIP model when there is additional excess dispersion among nonzero observations can lead to significant biases in parameter estimates. Ridout, Hinde, and Demétrio (2001) demonstrated the significance of this problem through a score test they developed to compare the ZIP model with the ZINB alternative. Therefore, a high proportion of zeros in a dataset does not, by itself, indicate that the ZIP model is sufficient; it is necessary to evaluate both the structure of the zeros and the distributional characteristics of the positive counts (Ridout et al., 2001; Hilbe, 2011; Cameron & Trivedi, 2013).

Another noteworthy aspect of the development of models inflated with zeros is the growing importance of "the relationship between model selection and the scientific explanation of the data-generation mechanism". The fact that a dataset contains a large number of zeros does not always imply the existence of two distinct biological, economic, or physical states. Unobserved heterogeneity, sampling design, low event rates, missing explanatory variables, or differences in the tails of the distribution can also lead to high zero rates. Therefore, selecting a zero-inflated model based solely on information criteria or goodness-of-fit results may lead to the assumption of an incorrect data generation process. For example, in the traffic safety literature, the theoretical validity of zero-inflated models implying a binary structure, such as "completely safe" and "unsafe"

road segments, has long been debated (Lord et al., 2005). Similarly, it has been shown that it is inappropriate to automatically use the Vuong test as a zero-inflated test in model comparisons (Wilson, 2015). These discussions highlight that the family of zero-inflated models should be evaluated not merely as statistical tools that provide a better fit, but as models that carry explicit assumptions about the data-generating mechanism.

In parallel with methodological advances, zero-inflated models have spread to a wide variety of application areas. Following Lambert's (1992) classic application to production errors, these models have been extensively used, particularly in biometrics, epidemiology, dentistry, ecology, agricultural sciences, transportation, insurance, and the social sciences. Böhning et al. (1999) addressed the high zero rates in the epidemiology of dental caries using the ZIP model; subsequent studies have shown that zero-inflated regressions have become increasingly common in dental epidemiological research (Preisser et al., 2012). However, Preisser et al. (2012) pointed out that interpreting parameters related to "sensitive" and "non-sensitive" groups, which are defined as latent in traditional zero-inflated models, as general population effects could be misleading. This issue became one of the key justifications for the subsequent development of marginalized zero-inflated models, leading to approaches that enable direct estimation of population-level effects via the marginal mean (Preisser et al., 2016). Thus, the literature has progressed from merely modeling zero inflation to more accurately linking model parameters to the scientific questions in applied research.

Ecology and environmental sciences are another important area of application for zero-inflated models. Since it is quite common for a species not to be observed at certain sampling points in studies of species presence and abundance, the proportion of zero values can be much higher than expected under the classical Poisson distribution. Chiogna and Gaetan (2007) extended the zero-inflated Poisson structure in animal abundance data using semiparametric methods, thereby enabling the modeling of nonlinear relationships with environmental variables. Arab et al. (2008), on the other hand, analyzed data with high zero rates obtained with different fishing gears in fish abundance studies on the Missouri River, using a Bayesian ZIP model that incorporates random effects. These examples demonstrate that zero inflation is not merely a theoretical distribution problem but a modeling issue directly related to the sampling process and ecological mechanisms. Today, methodological and computational research continues addressing zero inflation and over-dispersion together in ecological survey data (Sidumo et al., 2024).

Similarly, in agricultural sciences, there may be production, disease, pest, adaptation, or technology-use data that contain many zeros. Hall (2000) modeled within-subject correlation and between-subject heterogeneity in repeated measurements by adding random effects to Poisson and binomial models inflated

for zeros and applied this approach to experimental data on garden crops. Isgin et al. (2008), on the other hand, used ZIP and ZINB models to model the intensity of farmers' adoption of precision agriculture technologies. These studies demonstrate that the zero-inflated approach is important not only for correcting for excess zeros in event counts but also for examining, within the same model, the distinct mechanisms underlying "participation/non-participation in a process" and "the magnitude of the event count when participation occurs." The use of zero-inflated models in insurance and transportation sciences also clearly demonstrates the field's interdisciplinary nature. The fact that many policyholders report no claims during the observation period has created a natural setting for applying zero inflation. Boucher et al. (2007) examined the use of these models in risk classification by comparing zero-inflated, hurdle, and various hybrid Poisson models in auto insurance portfolios. In the transportation and traffic safety literature, the fact that many intersections or road segments record no accidents during certain periods has led to the widespread use of ZIP and ZINB models (Lord et al., 2005). Subsequent studies have addressed unobserved heterogeneity and complex dependency structures in greater detail through stochastic-parameter, multivariate, and Bayesian zero-inflated models. This broad scope of application demonstrates that zero-inflated models have evolved from being methods specific to a single discipline into a general-purpose family of statistical models.

The expansion of the literature on zero-inflated models is not limited to their application in different fields. It has also led to the incorporation of features such as increasingly complex data structures, repeated measurements, clustered observations, spatial and temporal dependencies, multivariate counts, and random effects into these models. While studies such as those by Hall (2000) and Min and Agresti (2005) focused on random effects and repeated measurements, subsequent research has expanded into spatial models, count time series, multivariate models, and Bayesian hierarchical structures. Young, Roemmele, and Shi (2022b), by reviewing the literature beyond classical univariate zero-inflated regressions, demonstrated that significant progress has been made, particularly regarding correlated counts, discrete time series, spatial models, and multivariate zero-inflated structures. This progress demonstrates that, over time, the literature on zero-inflated models has evolved from a simple Poisson mixture model into a much broader family of models.

However, the growing popularity of zero-inflated methods has also sparked new debates about which model to prefer in which situations. Although ZIP, ZINB, and hurdle models are often compared as alternatives in practice, it should not be overlooked that they do not represent the same scientific mechanism. Feng (2021) emphasized that although zero-inflated and hurdle models can be used on similar datasets with a high

proportion of zeros, their mechanisms for generating zeros differ. Similarly, findings in the dental literature regarding the misinterpretation of model parameters have shown that appropriate model selection cannot be reduced solely to fit criteria such as AIC, BIC, or log-likelihood (Preisser et al., 2012). For this reason, in the contemporary literature, model selection is evaluated in conjunction with the source of zeros, outliers, the structure of dependence, the population for which the effects of explanatory variables are to be interpreted, and the model's compatibility with the scientific question. Such a development suggests that the field is becoming increasingly diverse not only in the number of publications but also in its conceptual and methodological dimensions.

As the literature has grown more extensive and interdisciplinary, it has become difficult for traditional narrative-style compilations alone to reveal the field's full structure. While comprehensive reviews such as those by Young et al. (2022a, 2022b) explain in detail the theoretical and computational development of zero-inflated model families, bibliometric methods are needed to simultaneously and quantitatively reveal a research field's patterns of scientific output over time, its pioneering sources, influential authors, collaborations between countries and institutions, conceptual clusters, intellectual foundations, and current research frontiers. Bibliometric analysis enables the evaluation of a research field's development through quantitative indicators and relationship networks by systematically examining large volumes of scientific publication data (Donthu et al., 2021). The science mapping approach, on the other hand, focuses not only on the volume of output in a field but also on its "conceptual, intellectual, and social structure" (Cobo et al., 2011).

In bibliometric science mapping, keyword co-occurrence networks help identify the conceptual structure and thematic clusters of a research field; co-citation analyses reveal the intellectual knowledge base underpinning the field; bibliographic coupling analyses identify closely related current research fronts; while country and author collaboration networks enable an examination of the field's social organization. In addition, thematic maps, thematic evolution analyses, and temporal keyword studies can be used to assess how research topics have become more central, diversified, or lost significance over time. While tools such as VOSviewer have popularized the visual mapping of bibliometric networks (van Eck & Waltman, 2010), the 'bibliometrix' R package has brought together bibliometric performance analysis and science mapping methods within a programmable and integrated working environment (Aria & Cuccurullo, 2017). The capabilities of these tools provide a suitable analytical framework for examining the scientific evolution of statistical methods, particularly those with long historical development and broad interdisciplinary reach.

It would not be accurate to claim that the bibliometric perspective on zero-inflated models has not been fully explored. Indeed, Tajuddin and Ismail (2025) conducted a three-decade bibliometric study examining “inflated” count distributions. While this study provides an important macro-level overview of the scientific development of inflated-count models, it also leaves room for a more focused bibliometric framework, specifically for zero-inflated models. In particular, a more detailed science mapping approach could examine how zero-inflated Poisson, zero-inflated negative binomial, and their numerous extensions have spread across different disciplines; how the field’s conceptual clusters, key references, and current research frontiers are structured; which themes have emerged over time; and how international collaboration has taken shape.

At this point, the scope of the present study differs from the previous bibliometric research in several respects. First, the study focuses directly on “zero-inflated models” rather than the broader category of “inflated count distributions,” thereby examining the literature on statistical methods related to zero inflation as a more homogeneous field of research. Second, the “Web of Science Core Collection” is used as the data source, and the analysis period is updated through 2025 for publication years. Third, rather than relying solely on performance indicators such as publication, source, author, and country frequencies, the study aims to evaluate complementary science mapping techniques collectively, including keyword co-occurrence networks, temporal keyword structures, thematic maps and thematic evolution, co-citation networks, bibliographic matching, and international collaboration networks. In this way, it will be possible to reveal not only “how much” the zero-inflated model literature has grown, but also “the statistical ideas around which it has developed, the studies upon which it is built, the research frontiers toward which it is directed, and the disciplines and research communities through which this development has taken place”.

In this context, the primary objective of this study is to systematically examine the evolution of the scientific literature published on zero-inflated models using bibliometric and science mapping methods. To this end, the study aims to: (i) identify trends in scientific output and citations by year; (ii) identify the most productive and influential journals, authors, and countries; (iii) examine international research collaborations; (iv) determine the conceptual structure and thematic clusters of the field through author keywords; (v) assess changes in themes over time; (vi) to uncover the field’s intellectual foundation through co-citation analysis, (vii) to identify current research frontiers through bibliographic matching, and (viii) to identify emerging research topics. In this regard, the study aims to evaluate the development of zero-inflated models in the statistical literature, spanning more than thirty years, within a comprehensive framework and to provide a systematic

reference on the field’s current structure and future potential research directions for both researchers developing methods and scientists applying these models in practical fields.

2. Material and Methods

2.1. Research Design

This study was conducted using bibliometric analysis and the science mapping approach to examine the evolution over time of scientific studies published on zero-inflated models, as well as their structure of scientific output, conceptual orientations, intellectual foundations, and the social structure among researchers. While bibliometric analysis enables the quantitative examination of large volumes of scientific publication data, the science mapping approach reveals the conceptual, intellectual, and social structures of research fields through networks (Cobo et al., 2011a; Donthu et al., 2021).

The study did not limit itself to examining only basic performance indicators; instead, it combined author-keyword co-occurrence analysis, thematic mapping, thematic evolution, co-citation analysis, bibliographic matching, and cross-national scientific collaboration analyses. Thus, the “zero-inflated” model literature was evaluated under different, complementary bibliometric dimensions: performance structure, conceptual structure, intellectual structure, and social structure. In designing the bibliometric analysis process, we drew upon the general science mapping approaches proposed by Aria and Cuccurullo (2017), Cobo et al. (2011a, 2011b), and Donthu et al. (2021).

2.2. Data Source and Literature Review Strategy

Bibliographic data were obtained from the Web of Science Core Collection (WoSCC) database. A key factor in selecting Web of Science was its provision of detailed bibliographic information on publications, as well as essential fields required for scientific mapping analyses—such as citation records, authors, institutions, countries, keywords, source journals, and cited references.

The literature review was conducted in August 2026. The search was performed in the Topic (TS) field of the Web of Science database. Since the topic field encompasses terms found in publication titles, abstracts, and keywords, the aim was to capture as broad a range as possible of both methodological and applied studies on zero-inflated models.

The final Web of Science query used in the study was constructed as follows:

TS=(“zero-inflated model” OR “zero inflated model*” OR “zero-inflated regression” OR “zero inflated regression*” OR “zero-inflated distribution” OR “zero inflated distribution*” OR “zero-inflated count” OR “zero inflated count*”)

When determining the search terms, both hyphenated and non-hyphenated forms were included separately in the query. By adding an asterisk (*) at the end of the terms, the study aimed to capture singular and plural

forms as well as different word forms derived from the same root. For example, the phrase “zero-inflated model” covers both “zero-inflated model” and “zero-inflated models.”

Since one of the primary objectives of the study was to reveal the interdisciplinary spread of zero-inflated models, no field restrictions were applied based on Web of Science Categories. Similarly, no restrictions were imposed regarding document type or publication language during the initial phase. Thus, the dataset was designed to include studies that use zero-inflated models across statistics, biometrics, health sciences, ecology, agriculture, economics, insurance, transportation, and other applied fields.

The records obtained from WoSCC were exported as a Plain Text File (.txt) using the “Full Record and Cited References” option. Due to export limitations, the records were downloaded as multiple files and merged into a single bibliographic dataset during the analysis phase.

2.3. Determining the Study Period

No restriction on the starting year was applied in the literature review, and all studies were retrieved starting from the earliest record on the subject found in Web of Science. It was determined that the earliest relevant publications in the dataset date back to 1997.

Since the year 2026 had not yet ended when the search was conducted, the number of publications for 2026 does not represent a full calendar year. To prevent the change in the number of publications for an incomplete year from causing an artificial decrease in keyword frequencies and other temporal indicators, the period 1997–2025 was designated as the primary bibliometric analysis period.

Records for the year 2026 were not deleted during the data cleaning phase but were preserved as a separate dataset. While these records can be used as indicators to show current publication activity, when necessary, they have not been included in annual scientific production trends, thematic evolution, emerging-topic analyses, or other temporal comparisons involving complete years. This was done to ensure comparability over the years.

2.4. Data Preprocessing and Data Cleaning

Text files obtained from Web of Science were converted into a bibliographic data frame in the R environment using the `convert2df()` function of the `‘bibliometrix’` package (Aria & Cuccurullo, 2017). To prevent the analysis results of identical publications, which might appear in different export files, from being artificially inflated, automatic record deletion was not applied during the initial data import; instead, duplicate records were identified through a controlled cleaning process.

In the data cleaning process, the UT (Unique Article Identifier) field, assigned uniquely to each record by Web of Science, was used first. If multiple records had the same UT value, only a single record was retained. To prevent publications lacking UT information from being mistakenly identified as duplicates, these records were tracked using temporary unique sequence numbers.

In the second stage, Digital Object Identifier (DOI) values were checked. DOI values were converted to lowercase, unnecessary leading and trailing spaces were removed, and records with the same normalized DOI value were considered duplicates. Records lacking DOI information were not automatically deleted.

In addition, the publication title was converted to lowercase, punctuation was removed, and the title and publication year were evaluated together to report potential title-year duplicates. However, because identical or highly similar titles may represent different scientific records, automatic record deletion was not performed solely on the basis of title similarity. This was done to reduce the risk of false-positive duplicate deletions during data cleaning.

The publication year (PY) and total citation count (TC) fields were converted to numerical data types; the dataset was also checked to verify the presence of basic bibliographic fields such as author, title, source, document type, author keywords, Keywords Plus, abstract, address, cited references, WoS categories, DOI, and UT. Missing metadata rates were calculated separately for each bibliographic field.

2.5. Standardization of Author Keywords

In the analyses examining the conceptual structure of the research field, the author’s keywords (Author Keywords; DE) were used as the basis. Since the same concept may appear under different spellings, resulting in multiple separate nodes in co-occurrence networks, the keywords were standardized prior to analysis.

During this process, keywords were converted to lowercase, hyphen characters were standardized, unnecessary punctuation was removed, and excess spaces were eliminated. Additionally, some terms that clearly represent the same scientific concept were combined under a common form. For example:

• “ZIP,” “ZIP model,” “zero-inflated Poisson model,” and similar usages were grouped under “zero-inflated Poisson.”

• “ZINB,” “zero-inflated negative binomial model,” and similar expressions were grouped under “zero-inflated negative binomial.”

• “ZIB” and similar terms have been grouped under “zero-inflated binomial.”

• terms such as “overdispersion” and “over-dispersed” have been grouped under “overdispersion,”

• Moreover, terms such as “excess zero” and “excess zeros” have been grouped under “excess zeros.”

All original and standardized keywords have also been stored in a conversion table, thereby ensuring the traceability of the standardization processes.

Since there is a high likelihood that general terms such as “zero-inflated models,” “zero-inflated regression,” “zero-inflated count data,” and “zero-inflated distribution,” as a direct result of the search query, would excessively dominate the co-occurrence network, these terms were retained in the descriptive keyword frequency analyses. However, they were excluded from the detailed scientific

mapping networks created to identify the subthemes of the research field better. The aim was to reduce the artificial centralization effect that could arise from the structure of the search query itself.

2.6. Bibliometric Performance Analysis

A bibliometric performance analysis was conducted to quantitatively assess the scientific development of the literature on zero-inflated models. The performance analysis is based on an evaluation of the productivity and impact levels of key bibliometric units, such as publications, authors, sources, and countries, that contribute to a specific research field. While science mapping techniques reveal the conceptual, intellectual, and social relationships within a research field, performance analysis enables determination of the field's overall size, rate of development, and key scientific actors. For this reason, performance analysis and science mapping are considered two complementary core approaches in bibliometric research (Aria & Cuccurullo, 2017; Donthu et al., 2021). Furthermore, the systematic evaluation of bibliometric indicators helps to more objectively identify prominent researchers, publication sources, and scientific contributions within a broad and fragmented literature (Zupic & Čater, 2015).

In this context, the study first examined changes in scientific output over time using models adjusted for zero-inflated data. In addition, annual scientific output was assessed using yearly publication counts, and temporal trends were visualized using LOESS smoothing. The total number of documents was calculated for each publication year to determine the field's historical development trend, and annual scientific output was visualized. In bibliometric studies, changes in the annual number of publications are used as a key indicator to assess the development of a research field within the scientific community, researchers' changing interests over time, and the field's maturation process (Donthu et al., 2021). However, since the year 2026 had not yet concluded at the time of data collection for this study, a direct comparison of the number of publications for that year with previous years could be misleading. For this reason, records for 2026 have been retained in the dataset and shown separately in the annual output graph; however, only the period from 1997 to 2025, comprising completed publication years, was considered in assessing long-term trends. To visually illustrate the general publication trend in the field, a LOESS curve was fitted to the annual publication counts for 1997–2025. The LOESS curve was used solely for descriptive visualization of the publication trend and was not evaluated as a predictive model for future outcomes.

In addition to the volume of scientific output, the citation performance of the literature was also examined. For this purpose, the total number of citations, the average number of citations per publication, and the median number of citations were calculated for each publication year. Citation analysis is widely used in bibliometric research to assess the extent to which scientific studies

are utilized by subsequent research and the degree to which they have become visible in the bibliographic landscape (Zupic & Čater, 2015; Donthu et al., 2021). However, it was noted that the total number of citations can be significantly influenced by the publication's age. Since older studies have had more time to accumulate citations, a direct comparison of absolute citation counts across different publication years is not sufficient on its own. Therefore, in addition to total citation counts, annual average and median citation values were also reported to enable a more balanced assessment of citation distribution. Furthermore, citation metrics were interpreted not as direct measures of scientific quality, but as indicators of the usage and bibliometric visibility of publications within the relevant literature.

To identify the primary publication channels in the research field, the performance of source journals was also evaluated. For each source, the number of published documents, the total number of citations for these documents, and the average citations per publication were calculated; the top 20 sources with the highest output were examined in detail. While the number of publications indicates a specific journal's quantitative contribution to a model literature inflated by zero-citation papers, the total and average citation values provide complementary information regarding the bibliographic visibility of studies published in these sources within the literature. Since evaluating sources based solely on the number of publications could overlook journals that publish a small number of high-impact studies, productivity and citation indicators were considered together. The combined evaluation of multiple indicators in bibliometric performance analysis helps mitigate the limitations that may arise from classifying key scientific actors in a research field based on a single metric (Donthu et al., 2021).

Similarly, productivity and citation metrics were calculated at the author level to identify the field's leading researchers. For each author, the total number of publications, the total number of citations for those publications, the year of the first publication, and the year of the most recent publication were determined. The scientific activities of the top 20 most productive authors were additionally visualized, taking the time dimension into account; thus, not only the total number of publications but also the periods during which the researchers were active in the model literature, which is inflated by zero-citation papers, were evaluated. While bibliometric analyses enable the identification of key researchers influencing the development of a research field, the inclusion of citation metrics provides complementary information regarding the visibility of this productivity within the literature (Aria & Cuccurullo, 2017; Donthu et al., 2021). However, because citations for a study in multi-authored publications are attributed to all authors, the author-level citations were interpreted as the total bibliographic visibility of the publications in

which the relevant authors appeared, rather than as independent or fractional citation shares.

Country-based performance indicators were also calculated to assess the study's geographic distribution and level of internationalization. Country information was derived from author addresses found in Web of Science records and standardized using Bibliometrix (Aria & Cuccurullo, 2017). In addition to the total number of publications and the total citation counts of relevant publications for each country, the Multiple Country Publication (MCP) indicator was calculated to assess the level of international research collaboration. Publications involving researchers from more than one country were classified as international co-publications, and the MCP ratio for each country was obtained by dividing the number of international co-publications by that country's total number of publications. This made it possible to distinguish between countries with high publication output and those with high international collaboration intensity. Co-authorship and country-based collaboration indicators are key components of bibliometric studies in assessing which researcher communities and geographic centers produce scientific knowledge (Zupic & Čater, 2015; Donthu et al., 2021).

Consequently, within the scope of bibliometric performance analysis, annual scientific output, annual citation dynamics, leading publication sources, the most productive authors, and countries' scientific output and international collaboration indicators were evaluated collectively. These indicators were used to determine the quantitative development of the "zero-inflated models" literature and its key scientific actors; inferences regarding the conceptual and intellectual structure of the research field were based on keyword co-occurrence, thematic mapping, co-citation, and bibliographic matching analyses conducted in subsequent stages. Thus, by evaluating performance indicators in conjunction with science mapping techniques, the study aimed to comprehensively examine not only the production and impact levels of the literature but also its scientific structure and transformations over time (Aria & Cuccurullo, 2017; Donthu et al., 2021).

All analyses were conducted as part of a bibliometric analysis using the "bibliometrix" package (v 5.4.1) in the R software (v. 2026.09.0+174) (Aria and Cuccurullo, 2017).

3. Results and Discussion

A literature search conducted in the Web of Science Core Collection database yielded 1,110 bibliographic records related to models inflated with zeros. After checking for duplicate records using the Web of Science unique record number (UT), DOI, and title-year combinations, no records were removed from the dataset, and all 1,110 unique records were retained. Since the year 2026 had not yet concluded at the time of the search, 43 publications from that year were excluded from the main temporal analyses; thus, the 1,067 publications spanning

the 1997–2025 period, which covers completed publication years, formed the core bibliometric analysis dataset.

The dataset covering the 1997–2025 period comprises 1,067 documents published in 577 scientific sources. A total of 3,306 different authors were listed in these publications, and 31,943 cited references were identified. The annual compound growth rate of scientific output was calculated to be 17.1%, while the average age of the documents was 8.39 years. The average number of citations per document across the publications examined was 33.37, and the average annual number of citations per document was 3.097. In terms of author collaboration, there were an average of 3.70 co-authors per document, while 66 studies were published by a single author.

An analysis of document types revealed that the literature consists largely of research articles. During the main analysis period, 980 records were classified directly as "article," while 13 studies fell under the "article; proceedings paper" category, six under "article; book chapter," and two under "article; early access." Additionally, the dataset includes 40 proceedings papers, 16 reviews, 3 letters, 3 meeting abstracts, 2 corrections, and 2 editorial materials. These results indicate that scientific output in the literature on zero-inflated models has primarily taken the form of original research articles. The most productive sources in terms of the number of publications in the literature on zero-inflated models are presented in Table 2 and Figure 3. With a total of 35 publications, "Statistics in Medicine" ranked first. This journal was followed by "Communications in Statistics–Theory and Methods" with 23 publications, the "Journal of Applied Statistics" with 20, "Statistical Methods in Medical Research" with 19, and the "Korean Journal of Applied Statistics" with 16. "Biometrics" and the "Journal of Statistical Computation and Simulation" were also among the leading sources, each with 15 publications.

When the number of publications and citation metrics were evaluated together, significant differences were identified between publication productivity and citation visibility. The 35 publications in "Statistics in Medicine", the most productive journal, received a total of 703 citations, averaging 20.09 citations per publication. In contrast, "Accident Analysis and Prevention", ranked eighth with only 13 publications, demonstrated the most striking citation performance, with a total of 2,526 citations and an average of 194.31 citations per publication.

Other sources with high average citation counts per publication included "Statistical Modeling" (49.91), "Ecological Modeling" (48.13), "Environmetrics" (39.78), "Computational Statistics & Data Analysis" (35.42), and "Caries Research" (34.50). It was observed that studies on zero-inflated models were published both in pure statistics and biometrics journals and in sources focused on applied fields such as health, traffic safety, ecology, and forestry.

The distribution of scientific publications on zero-inflated models over the years shows that the field produced relatively little in its early years but has grown significantly, especially since the late 2000s (Figure 1). While there was only one publication in 1997, no records

were found in 1998, and one publication was published in each of 1999 and 2000. The annual number of publications increased to 13 in 2005, 18 in 2008, and 28 in 2010.

Table 1. Basic bibliometric characteristics of literature on zero-inflated models.

Bibliometric indicator	Values
Analysis period	1997–2025
Number of publications in the main analysis	1067
Number of excluded publications from 2026	43
Total retrieved records	1.110
Number of sources (journals, books, etc.)	577
Total number of authors	3.306
Author appearances	3.948
Number of author keywords	3.208
Number of author keywords plus	2.435
Number of cited references	31.943
Annual scientific production growth rate (%)	17.10
Average document age (years)	8.39
Average citations per document	33.37
Average annual citations per document	3.097
Average co-authors per document	3,70
Number of single-authored documents	66
Number of unique authors of single-authored documents	64

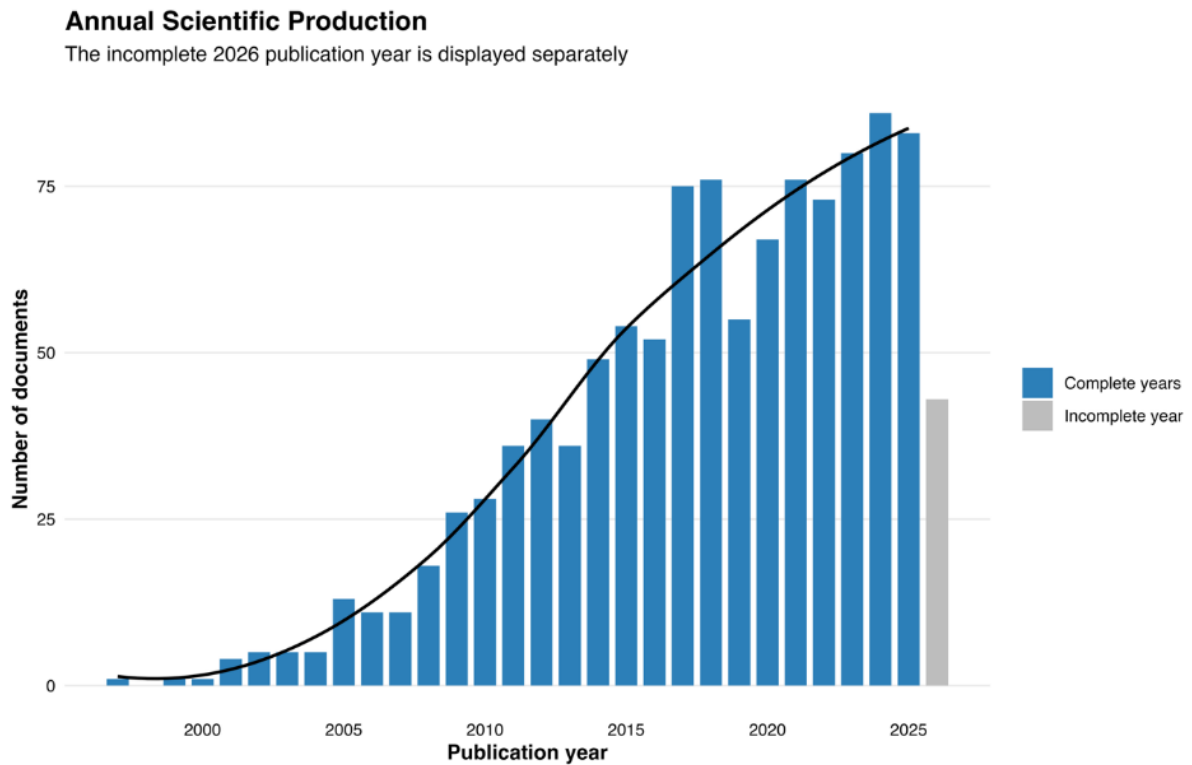


Figure 1. Annual scientific output based on zero-inflated models.

The increase became more pronounced after 2010. The annual number of publications was 36 in 2011, 40 in 2012, 54 in 2015, 75 in 2017, and 76 in 2018. Although a temporary decrease in the number of publications was observed in 2019, production increased again, reaching its highest level in the analysis period with 76 publications in 2021, 80 in 2023, and 86 in 2024. The 83 studies published in 2025 showed that this high publication volume continued in the last completed year as well.

Citation dynamics have shown a more heterogeneous structure compared to the growth in publication output (Figure 2). Publications in the main analysis period received a total of 35,606 citations based on their current citation status. While some older publication years have

higher average citation levels, as might be expected, the average values are lower due to the shorter citation accumulation period for those publications.

The year 2017, in particular, has a remarkable citation profile. 75 studies published this year received a total of 13,217 citations, with an average citation count of 176.23. In contrast, the median citation counts of only 11 indicates a strongly right-skewed distribution and that the annual average is significantly influenced by a small number of publications with very high citation counts. Similarly, in 2008, the average citation count was 159.56 while the median was 28.5; and in 2005, the average was 141.15 while the median was 71. Therefore, it seems necessary to consider both the mean and median values when evaluating the citation effect.

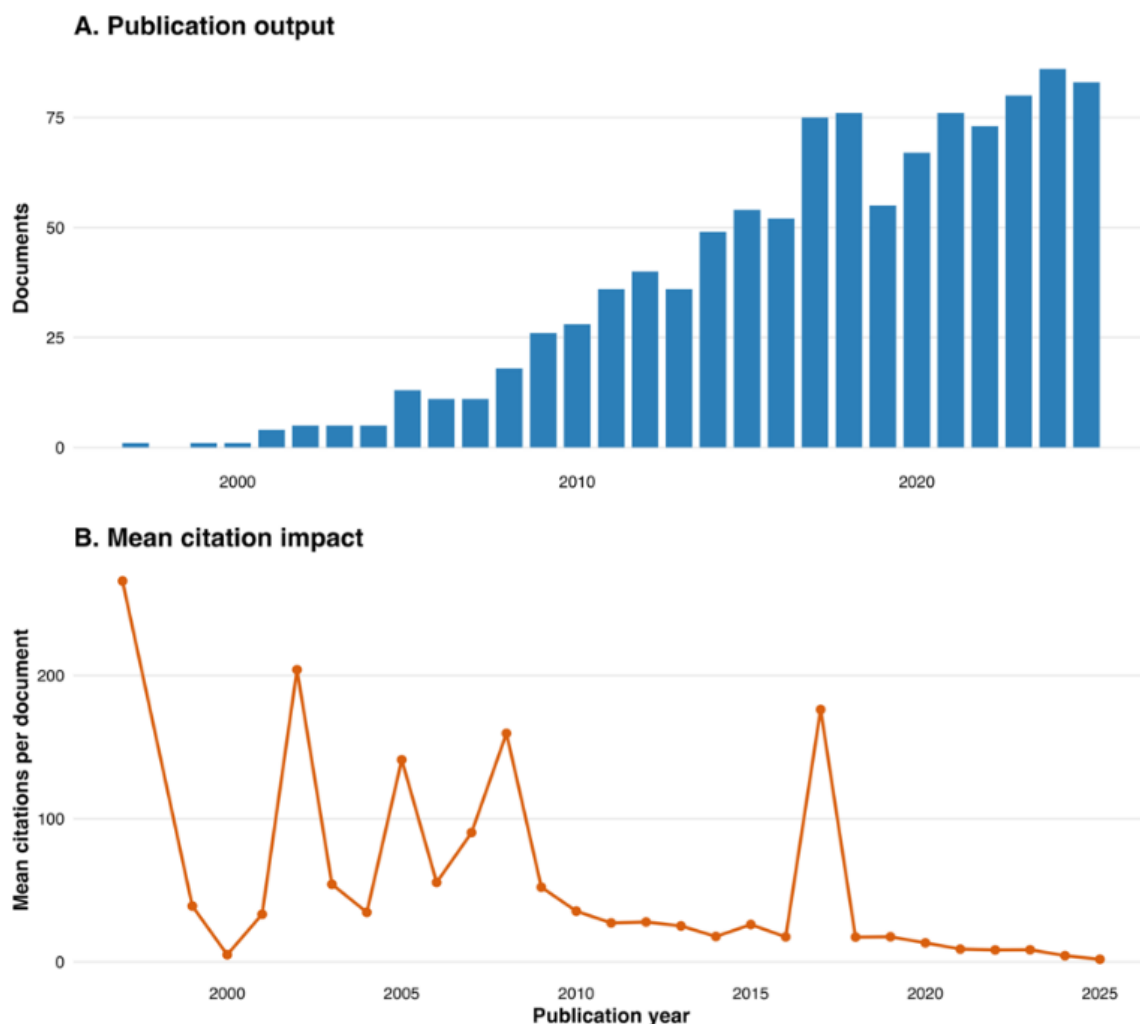


Figure 2. The sources that have produced the most publications in the literature on zero-inflated models

In terms of publication sources, it was determined that the literature on zero-inflated models is not concentrated in a single journal but is distributed across statistics, biometrics, and applied sciences. The highest number of publications was in the journal *Statistics in Medicine*, with 35 studies. This was followed by *Communications in Statistics-Theory and Methods* (n=23), *Journal of Applied*

Statistics (n=20), *Statistical Methods in Medical Research* (n=19), and *Korean Journal of Applied Statistics* (n=16) (Table 2).

Significant differences were found between publication productivity and citation visibility. For example, *Accident Analysis and Prevention*, despite ranking eighth with only 13 publications, received 2,526 citations, averaging

194.31 citations per publication. In contrast, Statistics in Medicine, which ranked first in terms of publication volume, had an average of 20.09 citations per publication. Statistical Modeling (49.91), Ecological Modeling (48.13), and Environmetrics (39.78) were also among the sources with high average citation counts per publication.

In terms of author-based research, Hsu WW was identified as the most prolific researcher with 10

publications. Ganjali M and Todem D followed with nine each, Yang J with eight, and Aldirawi H and Faes C with seven each. However, a strong correlation between publication count and citation visibility was not found at the author level. For example, Mannering F, despite being represented by five publications, received a total of 1,434 citations for these publications. Similarly, Lee J's five publications received 365 citations, and Gelfand AE's six publications received 258 citations (Table 3).

Table 2. Top 10 most productive scientific sources

Journals	Documents	Total Citations	Mean Citations
Statistics in Medicine	35	703	20.09
Communications in Statistics-Theory and Methods	23	176	7.65
Journal of Applied Statistics	20	229	11.45
Statistical Methods in Medical Research	19	361	19.00
Korean Journal of Applied Statistics	16	21	1.31
Biometrics	15	190	12.67
Journal of Statistical Computation and Simulation	15	417	27.80
Accident Analysis and Prevention	13	2526	194.31
Biometrical Journal	13	366	28.15
Plos One	13	415	31.92

Table 3. The top 10 authors and their bibliometric indicators

Author	Documents	Total Citations	First Year	Last Year
Hsu WW	10	53	2012	2024
Ganjali M	9	46	2012	2025
Todem D	9	50	2012	2023
Yang J	8	57	2018	2023
Aldirawi H	7	55	2018	2023
Faes C	7	83	2014	2018
Chen T	6	41	2015	2022
Gelfand AE	6	258	2002	2025
Hardin JW	6	135	2009	2019
Kim K	6	39	2012	2023

An examination of standardized author keywords revealed that the core concepts in the research area centered on zero inflation, overdispersion, and alternative counting models (Figure 3). The most common keyword was "zero-inflated models," used in 266 publications. This was followed by "zero inflation" (n=96), "overdispersion" (n=85), "count data" (n=64), "zero-inflated Poisson" (n=58), "hurdle model" (n=52), "negative binomial" (n=45), and "zero-inflated distribution" (n=45). Additionally, the terms zero-inflated negative binomial (n=28), Bayesian inference (n=27), excess zeros (n=25), Poisson regression (n=23), and EM algorithm (n=22) also stood out.

This distribution shows that the field is not solely limited to the general concept of "zero inflation" and also encompasses methodological axes such as overdispersion accompanying zero inflation, alternative distributions, hurdle structures, estimation algorithms, and Bayesian inference.

The co-occurrence network revealed the connections between concepts in greater detail and identified six communities (Figure 4). At the center of the network were, in particular, zero inflation, overdispersion, count data, zero-inflated Poisson, hurdle model, and negative binomial. The fact that these terms have large node sizes and numerous connections indicates that they form the fundamental conceptual backbone of the research field.

While one cluster includes concepts such as zero inflation, overdispersion, count data, excess zeros, and distribution-related topics, a separate cluster encompasses zero-inflated Poisson, zero-inflated negative binomial, variable selection, and generalized estimating equations. Another distinct cluster has formed around Bayesian inference, Gibbs sampling, Markov chain Monte Carlo, hierarchical models, and data augmentation. Thus, it has been observed that Bayesian computational approaches form a relatively independent methodological framework within the literature on zero-inflated models.

A temporal overlay network, based on the average publication years of keywords, revealed that methodological interest in the zero-inflated model literature has changed over time (Figure 5). While the average publication years of some classical terms such as score test, Poisson regression, negative binomial distribution, and random effects are concentrated in earlier periods, terms such as simulation, under

dispersion, variable selection, model selection, and logistic regression are associated with more recent publications. The fact that high-frequency core concepts such as “zero inflation” and “count data” have remained central even in recent years indicates that these topics are not merely historical but are still actively used in the current literature.

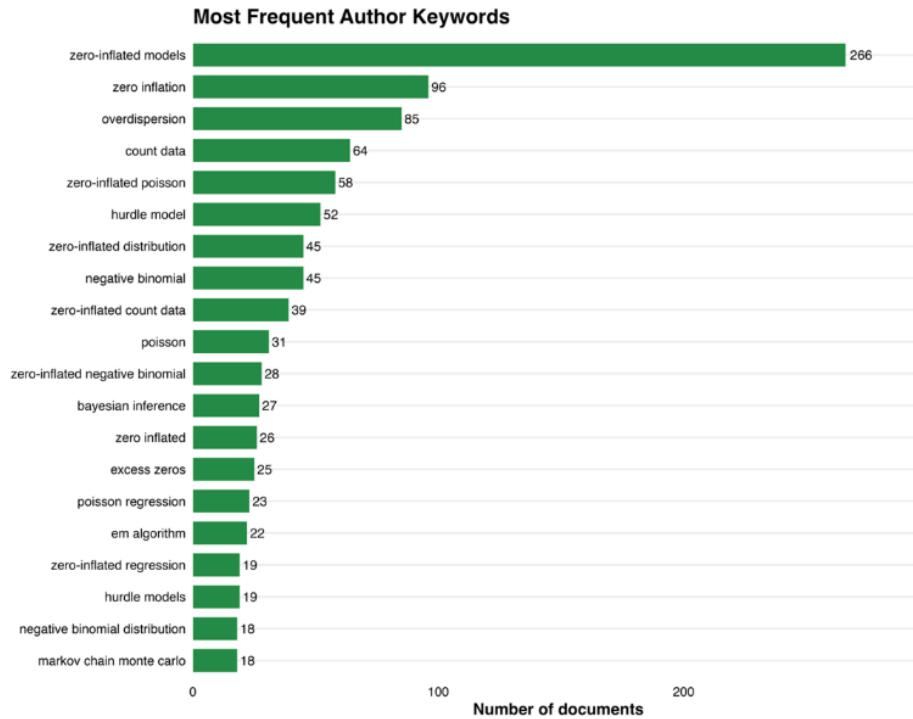


Figure 3. Most frequently used author keywords.

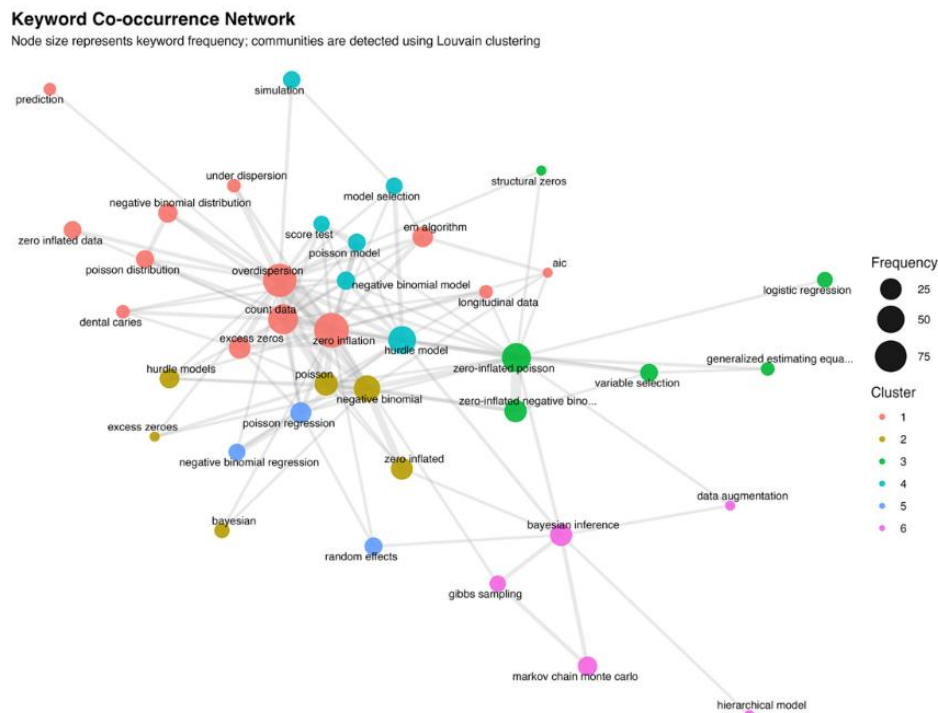


Figure 4. Keyword co-occurrence network.

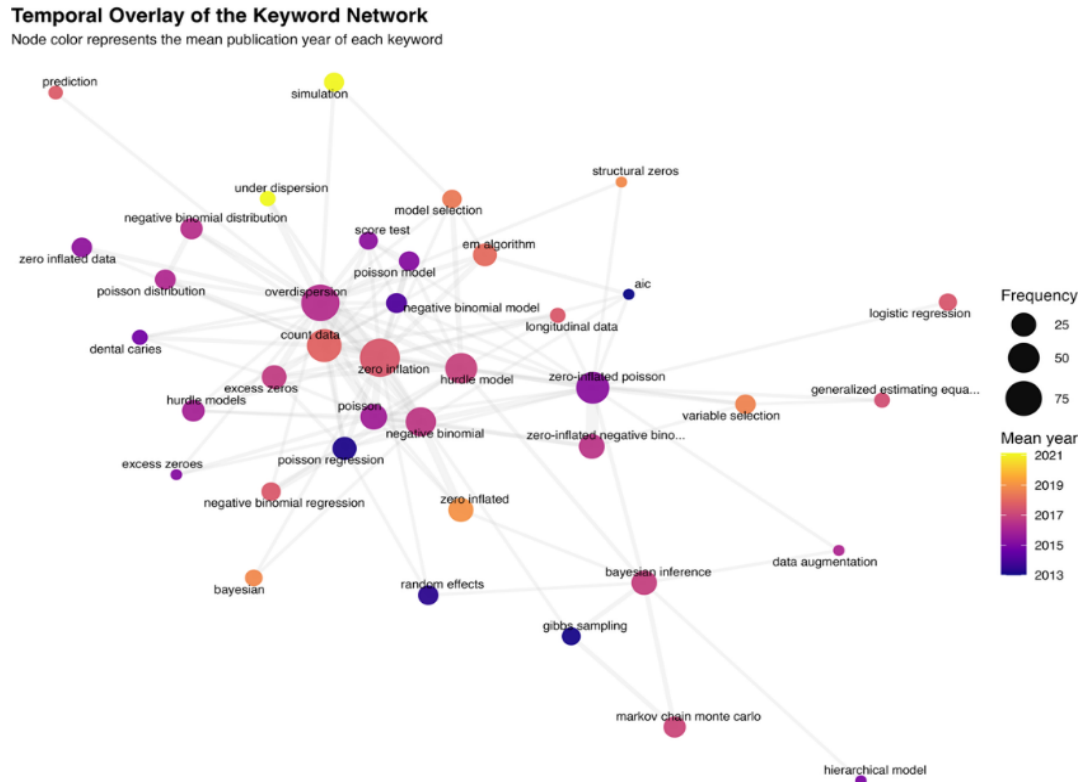


Figure 5. Temporal overlay of the keyword network.

Thematic map analysis revealed that the literature is organized around nine distinct thematic clusters (Table 4; Figure 6). The cluster with the highest total frequency was the zero-inflated/hurdle models cluster, which also exhibited the highest centrality value (0.430) with a frequency of 454. This cluster included concepts such as the EM algorithm, simulation, model selection, auto insurance, the microbiome, longitudinal data, and mixed models. The theme's high centrality indicates that it is strongly connected to other research themes.

The second-largest theme was zero inflation/overdispersion, with a frequency of 343 and a centrality of 0.289. Count data, excess zeros, the negative binomial model, the Poisson model, and the score test constituted the prominent sub-concepts of this cluster. On the thematic map, this cluster was located in the Basic Themes region, demonstrating that zero inflation and

overdispersion are fundamental concepts linking the broader field. The "zero-inflated Poisson/zero-inflated count data" cluster, with 202 frequencies, stood out for its high density (16,932) and was located in the "core themes" region of the thematic map. This cluster was observed alongside more specific methodological concepts such as ZINB, variable selection, generalized estimating equations, maximum likelihood, and estimation. The negative binomial/Poisson cluster was another advanced theme characterized by a high combination of centrality and density. The Bayesian inference/Markov chain Monte Carlo cluster formed a distinct thematic structure with 99 occurrences; it was positioned alongside random effects, Gibbs sampling, hierarchical models, Bayesian analysis, and data augmentation.

Table 4. Clusters identified as a result of the thematic analysis

Cluster	Main theme	Frequency	Centrality	Density
1	Zero-inflated model / Hurdle model	454	0.430	13.629
2	Zero inflation / Overdispersion	343	0.289	11.175
3	Zero-inflated Poisson / Zero-inflated count data	202	0.188	16.932
4	Negative binomial / Poisson	147	0.196	14.546
5	Zero-inflated distribution / Negative binomial distribution	104	0.168	13.003
6	Bayesian inference / Markov chain Monte Carlo	99	0.186	13.123
7	Poisson regression / Zero-inflated regression	72	0.159	9.477
8	Zero inflated data / Count regression	27	0.037	13.776
9	Logistic regression / Maximum likelihood estimation	37	0.090	16.051

The network also shows strong connections to Vuong (1989), Akaike (1974), Zeileis et al. (2008), Cameron (1998), McCullagh (1998), Nelder (1972), and other key sources on count models, model selection, and generalized linear models. This structure demonstrates that the literature on zero-inflated models has developed on an intellectual foundation grounded not only in ZIP/ZINB studies but also in the broader literature on count regression, model comparison, maximum likelihood, and generalized linear models.

4. Conclusion

This study comprehensively evaluated the development of scientific literature on zero-inflated models using data from the Web of Science Core Collection, through bibliometric performance analysis and science mapping techniques. An analysis of 1,067 publications from the main analysis period (1997–2025) revealed that the literature on zero-inflated models, which was initially a limited research area, has grown significantly over time. In particular, the rapid increase in publication output after 2010 and the fact that a significant portion of total publications was produced during the 2021–2025 period demonstrate that the zero-inflated model family continues to be important in current statistical research and applied sciences. The fact that the highest publication output occurred in 2024 and continued at a similar level in 2025 indicates that the field has not yet stabilized and that scientific interest persists.

Citation results show that the increase in publication output does not progress in a perfectly parallel manner with scientific visibility. It was determined that in some publication years, a small number of studies with very high citation counts significantly influenced the annual average citation values. This situation became particularly evident in 2017, and the high average citation value that year stemmed largely from a few highly influential methodological studies. This finding indicates that, when evaluating bibliometric performance, not only average citations count but also median values and the distribution of citations across publications must be taken into account.

The distribution of scientific sources has revealed that “zero-inflated” models exhibit a strong interdisciplinary structure that extends beyond statistical journals. While statistical and biometric journals such as “Statistics in Medicine”, “Communications in Statistics–Theory and Methods”, “Journal of Applied Statistics”, and “Statistical Methods in Medical Research” demonstrate high publication productivity, it was determined that some studies published in applied fields such as traffic safety, ecology, and environmental sciences possess a very high citation impact. Similarly, author analyses have shown that there is no direct equivalence between the number of publications and bibliometric impact. While some researchers achieve high citation visibility with a limited number of publications, others have contributed to the

development of literature through sustained publication output spread over longer periods.

Keyword analyses have shown that the conceptual core of the zero-inflated model literature centers on fundamental concepts such as zero inflation, overdispersion, count data, zero-inflated Poisson, hurdle model, negative binomial, and zero-inflated negative binomial. The co-occurrence network has revealed that these fundamental concepts are strongly interconnected and that different methodological frameworks have emerged within the field.

In particular, the clustering of terms such as Bayesian inference, Markov chain Monte Carlo, Gibbs sampling, hierarchical models, and data augmentation within the same conceptual framework indicates that Bayesian computation and hierarchical modeling approaches have evolved into a distinct research strand within the zero-inflated model literature.

Temporal analyses have shown that the research field has not only grown but has also undergone a methodological transformation. Alongside earlier concepts such as the score test, classical Poisson regression, and the negative binomial distribution, it has been observed that topics such as deep learning, marginalized models, count regression, INLA, Bayesian inference, variable selection, and under dispersion have gained prominence in recent years. In particular, the recent emergence of terms like deep learning and marginalized models indicates that zero-inflated data structures are now being addressed using more flexible and computationally intensive methods. However, the relative decline observed in certain keywords should not be interpreted as these topics having completely disappeared from the literature; rather, it should be viewed as an indication that the research agenda is diversifying toward new methodological problems and application areas.

The results of the thematic map demonstrate that the field possesses both a strong core of fundamental concepts and advanced methodological subthemes. The inclusion of the concepts of zero inflation and overdispersion among the core themes highlights that these two problems form the conceptual backbone of the research field; meanwhile, the presence of clusters related to zero-inflated Poisson, zero-inflated count data, and negative binomial models among the driving themes indicates that classical and extended count models continue to play central roles in the field’s current development. The fact that Bayesian methods form a distinct thematic structure is another indication of the field’s methodological diversity.

The co-citation analysis, meanwhile, has revealed that the literature on zero-inflated models possesses a strong and identifiable intellectual foundation. Pioneering works by researchers such as Lambert, Mullahy, Hall, Böhning, and Ridout have been at the center of the co-citation network; furthermore, it has been observed that classical methodological sources on model selection,

generalized linear models, over-dispersion, and count regression also constitute important components of the literature's knowledge base. Thus, it has been determined that research on zero-inflated models has not developed solely around a specific regression model; rather, it has progressed in an integrated manner with the broader literature on count data modeling, distribution theory, model comparison, and computational statistics.

When the study's overall findings are evaluated collectively, it is concluded that, over the course of approximately thirty years, the literature on zero-inflated models has evolved into a broader, more flexible, and interdisciplinary field of research than the classical ZIP and hurdle approaches. While the field's fundamental problems—excess zeros and excess dispersion—remain significant, current research is shifting toward more advanced methods such as Bayesian modeling, marginalized approaches, variable selection, complex dependency structures, and machine learning. This transformation suggests that future research will focus not only on achieving better model fit but also on more accurately defining the data-generation mechanisms underlying zeros, enhancing model interpretability, and developing new methods suitable for high-dimensional or complex data structures.

Unlike previous bibliometric evaluations in the literature, which were conducted within the broader framework of “inflated count distributions,” this study has focused directly on zero-inflated models and assessed the field's scientific output, conceptual structure, thematic development, and intellectual knowledge base using an integrated science-mapping approach. Therefore, it is believed that the results obtained could serve as a useful reference for both statisticians developing new methods for zero-inflated models and researchers applying these methods in fields such as bibliometrics, health sciences, ecology, agriculture, traffic safety, and other applied disciplines, in assessing the current state of the field and its potential future directions.

However, the findings should be evaluated within the framework of certain limitations. The fact that the study is based solely on Web of Science Core Collection records may have resulted in studies indexed in other bibliographic databases but not included in the WoS being excluded from the analysis. Furthermore, bibliometric indicators reflect publication and citation behavior but do not directly measure the methodological quality of the studies. Since keyword-based analyses are sensitive to the terminology used by authors, the standardization of terms that are conceptually similar but expressed in different ways may influence the results to a certain extent. In future research, comparing different bibliographic databases, using more detailed topic modeling techniques, and combining bibliometric findings with systematic content analysis could contribute to a more detailed understanding of the

methodological evolution of zero-inflated model literature.

In conclusion, zero-inflated models can be regarded as a mature yet still evolving dynamic field of statistical research. While classical ZIP, ZINB, and hurdle models continue to form the foundation of the field, emerging new methodological trends indicate that the problem of zero-inflated data will continue to be reexamined in the context of different data structures and scientific disciplines.

Author Contributions

The percentage of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	M.A	C.T
C	50	50
D	50	50
S	50	50
DCP	50	50
DAI	50	50
L	50	50
W	50	50
CR	50	50
SR	50	50
PM	50	50
FA	50	50

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declare that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

- Arab, A., Wildhaber, M. L., Wikle, C. K., & Gentry, C. N. (2008). Zero-inflated modeling of fish catch per unit area resulting from multiple gears: Application to channel catfish and shovelnose sturgeon in the Missouri River. *North American Journal of Fisheries Management*, 28(4), 1044–1058. <https://doi.org/10.1577/M06-250.1>
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Bekalo, D. B., & Kebede, D. T. (2021). Zero-inflated models for count data: An application to number of antenatal care service visits. *Annals of Data Science*, 8, 683–708. <https://doi.org/10.1007/s40745-021-00328-x>
- Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L., & Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*,

- 162(2), 195–209. <https://doi.org/10.1111/1467-985X.00130>
- Boucher, J.-P., Denuit, M., & Guillén, M. (2007). Risk classification for claim counts: A comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal*, 11(4), 110–131. <https://doi.org/10.1080/10920277.2007.10597487>
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 378–400. <https://doi.org/10.32614/RJ-2017-066>
- Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22, 155–205. <https://doi.org/10.1007/BF02019280>
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (2nd ed.). Cambridge University Press.
- Chiogna, M., & Gaetan, C. (2007). Semiparametric zero-inflated Poisson models with application to animal abundance studies. *Environmetrics*, 18(3), 303–314. <https://doi.org/10.1002/env.830>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011a). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402. <https://doi.org/10.1002/asi.21525>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011b). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Feng, C. X. (2021). A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of Statistical Distributions and Applications*, 8, Article 8. <https://doi.org/10.1186/s40488-021-00121-4>
- Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56(4), 1030–1039. <https://doi.org/10.1111/j.0006-341X.2000.01030.x>
- Hilbe, J. M. (2012). *Negative binomial regression* (2nd ed.). Cambridge University Press.
- Isgin, T., Bilgic, A., Forster, D. L., & Batte, M. T. (2008). Using count data models to determine the factors affecting farmers' quantity decisions of precision farming technology adoption. *Computers and Electronics in Agriculture*, 62(2), 231–242. <https://doi.org/10.1016/j.compag.2008.01.004>
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25. <https://doi.org/10.1002/asi.5090140103>
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1–14. <https://doi.org/10.1080/00401706.1992.10485228>
- Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accident Analysis & Prevention*, 37(1), 35–46. <https://doi.org/10.1016/j.aap.2004.02.004>
- Min, Y., & Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, 5(1), 1–19. <https://doi.org/10.1191/1471082X05st084oa>
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3), 341–365. [https://doi.org/10.1016/0304-4076\(86\)90002-3](https://doi.org/10.1016/0304-4076(86)90002-3)
- Preisser, J. S., Das, K., Long, D. L., & Divaris, K. (2016). Marginalized zero-inflated negative binomial regression with application to dental caries. *Statistics in Medicine*, 35(10), 1722–1735. <https://doi.org/10.1002/sim.6804>
- Preisser, J. S., Stamm, J. W., Long, D. L., & Kincade, M. E. (2012). Review and recommendations for zero-inflated count regression modeling of dental caries indices in epidemiological studies. *Caries Research*, 46(4), 413–423. <https://doi.org/10.1159/000338992>
- R Core Team. (2026). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ridout, M., Demétrio, C. G. B., & Hinde, J. (1998). Models for count data with many zeros. In *Proceedings of the 19th International Biometric Conference*. Cape Town, South Africa.
- Ridout, M., Hinde, J., & Demétrio, C. G. B. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, 57(1), 219–223. <https://doi.org/10.1111/j.0006-341X.2001.00219.x>
- Sidumo, B., Sonono, E., & Takaidza, I. (2024). Count regression and machine learning techniques for zero-inflated overdispersed count data: Application to ecological data. *Annals of Data Science*, 11, 803–817. <https://doi.org/10.1007/s40745-023-00464-6>
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269. <https://doi.org/10.1002/asi.4630240406>
- Tajuddin, R. R. M., & Ismail, N. (2025). Inflation properties in count data distributions: A three-decade bibliometric analysis. *Sains Malaysiana*, 54(1), 303–312. <https://doi.org/10.17576/jsm-2025-5401-24>
- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84, 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Wilson, P. (2015). The misuse of the Vuong test for non-nested models to test for zero-inflation. *Economics Letters*, 127, 51–53. <https://doi.org/10.1016/j.econlet.2015.01.002>
- Young, D. S., Roemmele, E. S., & Shi, X. (2022b). Zero-inflated modeling part II: Zero-inflated models for complex data structures. *WIREs Computational Statistics*, 14(2), Article e1540. <https://doi.org/10.1002/wics.1540>
- Young, D. S., Roemmele, E. S., & Yeh, P. (2022a). Zero-inflated modeling part I: Traditional zero-inflated count regression models, their applications, and computational tools. *WIREs Computational Statistics*, 14(1), Article e1541. <https://doi.org/10.1002/wics.1541>
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, 27(8), 1–25. <https://doi.org/10.18637/jss.v027.i08>
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429–472. <https://doi.org/10.1177/1094428114562629>