Contents

BJPS is an IMS supported journal ISSN 0103-0752 Brazilian Statistical Association

Brazilian Journal of Probability and Statistics

Volume 36 • Number 2 • June 2022

Brazilian Journal of Probability and Statistics

Volume 36 • Number 2 • June 2022

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 36, Number 2, June 2022. Published quarterly by the Brazilian Statistical Association.

POSTMASTER:

Send address changes to Brazilian Journal of Probability and Statistics, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

Brazilian Statistical Association members should send address changes to Rua do Matão, 1010 sala 250A, 05508-090 São Paulo/SP Brazil (address of the BSA office).

Printed in the United States of America

Partial financial support:
CNPq and CAPES (Brazil).

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 199–219
https://doi.org/10.1214/21-BJPS523

A robust partial least squares approach for function-on-function regression

Ufuk Beyaztas[1,a] **and Han Lin Shang**[2,b]

[1]*Department of Statistics, Marmara University, 34722, Kadikoy-Istanbul, Turkey,* [a]*ufuk.beyaztas@marmara.edu.tr*

[2]*Department of Actuarial Studies and Business Analytics, Level 7, 4 Eastern Road, Macquarie University, Sydney, New South Wales 2109, Australia,* [b]*hanlin.shang@mq.edu.au*

Abstract. The function-on-function linear regression model in which the response and predictors consist of random curves has become a general framework to investigate the relationship between the functional response and functional predictors. Existing methods to estimate the model parameters may be sensitive to outlying observations, common in empirical applications. In addition, these methods may be severely affected by such observations, leading to undesirable estimation and prediction results. A robust estimation method, based on iteratively reweighted simple partial least squares, is introduced to improve the prediction accuracy of the function-on-function linear regression model in the presence of outliers. The performance of the proposed method is based on the number of partial least squares components used to estimate the function-on-function linear regression model. Thus, the optimum number of components is determined via a data-driven error criterion. The finite-sample performance of the proposed method is investigated via several Monte Carlo experiments and an empirical data analysis. In addition, a nonparametric bootstrap method is applied to construct pointwise prediction intervals for the response function. The results are compared with some of the existing methods to illustrate the improvement potentially gained by the proposed method.

References

Agostinelli, C. and Library, S. C. M. (2015). wle: Weighted likelihood estimation. R package version 0.9-91.

Aguilera, A. M., Escabias, M., Preda, C. and Saporta, G. (2010). Using basis expansions for estimating functional PLS regression: Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems* **104**, 289–305.

Aguilera, A. M., Escabias, M., Preda, C. and Saporta, G. (2016). Penalized versions of functional PLS regression. *Chemometrics and Intelligent Laboratory Systems* **154**, 80–92.

Aguilera, A. M., Ocana, F. A. and Valderrama, M. J. (1999). Forecasting with unequally spaced data by a functional principal component approach. *Test* **8**, 233–254. MR1707639 https://doi.org/10.1007/BF02595871

Alin, A. and Agostinelli, C. (2017). Robust iteratively reweighted SIMPLS. *Journal of Chemometrics* **31**, e2881.

Beyaztas, U. and Shang, H. L. (2019). Forecasting functional time series using weighted likelihood methodology. *Journal of Statistical Computation and Simulation* **89**, 3046–3060. MR4000283 https://doi.org/10.1080/00949655.2019.1650935

Beyaztas, U. and Shang, H. L. (2020a). On function-on-function regression: Partial least squares approach. *Environmental and Ecological Statistics* **27**, 95–114.

Beyaztas, U. and Shang, H. L. (2020b). Robust bootstrap prediction intervals for univariate and multivariate autoregressive time series models. *Journal of Applied Statistics*. To appear.

Beyaztas, U. and Shang, H. L. (2021). A partial least squares approach for function-on-function interaction regression. *Computational Statistics* **36**, 911–939. MR4255794 https://doi.org/10.1007/s00180-020-01058-z

Chiou, J.-M., Yang, Y.-F. and Chen, Y.-T. (2016). Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis* **146**, 301–312. MR3477667 https://doi.org/10.1016/j.jmva.2015.10.003

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik* **31**, 377–703. MR0405795 https://doi.org/10.1007/BF01437407

Key words and phrases. Basis function, dimension reduction, SIMPLS, weighted likelihood.

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference* **147**, 1–23. MR3151843 https://doi.org/10.1016/j.jspi.2013.04.002

Cummins, D. J. and Andrews, C. W. (1995). Iteratively reweighted partial least squares: A performance analysis by Monte Carlo simulation. *Journal of Chemometrics* **9**, 489–507.

Dayal, B. S. and MacGregor, J. F. (1997). Improved PLS algorithms. *Journal of Chemometrics* **11**, 73–85.

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **18**, 251–263.

Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics* **40**, 322–352. MR3014309 https://doi.org/10.1214/11-AOS958

Escoufier, Y. (1970). Echantillonnage dans une population de variables aléatories réelles. *Publications de I'Institut de Statistique de I'Université de Paris* **19**, 1–47. MR0451482

Febrero-Bande, M., Galeano, P. and Gonzalez-Manteiga, W. (2017). Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. *International Statistical Review* **85**, 61–83. MR3637736 https://doi.org/10.1111/insr.12116

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. New York: Springer. MR2229687

Gervini, D. (2012). Functional robust regression for longitudinal data. Working paper.

Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M. and Stadelmann, M. (2021). expm: Matrix Exponential, Log, 'etc'. R package version 0.999-6.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B* **68**, 109–126. MR2212577 https://doi.org/10.1111/j.1467-9868.2005.00535.x

He, G., Müller, H. G., Wang, J. L. and Yang, W. J. (2010). Functional linear regression via canonical analysis. *Bernoulli* **16**, 705–729. MR2730645 https://doi.org/10.3150/09-BEJ228

Horvath, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. New York: Springer. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Chennai, India: John Wiley & Sons. MR3379106 https://doi.org/10.1002/9781118762547

Hubert, M. and Branden, K. V. (2003). Robust methods for partial least squares regression. *Journal of Chemometrics* **17**, 537–549.

Hullait, H., Leslie, D. S., Pavlidis, N. G. and King, S. (2021). Robust function-on-function regression. *Technometrics* **63**, 396–409. MR4296905 https://doi.org/10.1080/00401706.2020.1802350

Hyndman, R. J. and Shang, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society* **38**, 199–211. MR2750314 https://doi.org/10.1016/j.jkss.2009.06.002

Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics* **19**, 29–45. MR2752026 https://doi.org/10.1198/jcgs.2009.08158

Ivanescu, A. E., Staicu, A.-M., Scheipl, F. and Greven, S. (2015). Penalized function-on-function regression. *Computational Statistics* **30**, 539–568. MR3357075 https://doi.org/10.1007/s00180-014-0548-4

Kalogridis, I. and Aelst, S. V. (2019). Robust functional regression based on principal components. *Journal of Multivariate Analysis* **173**, 393–415. MR3945631 https://doi.org/10.1016/j.jmva.2019.04.003

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Boca Raton: CRC Press. MR3793167

Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *The Annals of Statistics* **22**, 1018–1114. MR1292557 https://doi.org/10.1214/aos/1176325512

Luo, R. and Qi, X. (2019). Interaction model and model selection for function-on-function regression. *Journal of Computational and Graphical Statistics* **28**, 309–322. MR3974882 https://doi.org/10.1080/10618600.2018.1514310

Markatou, M. (1996). Robust statistical inference: Weighted likelihoods or usual M-estimation? *Communications in Statistics Theory and Methods* **25**, 2597–2613. MR1424732 https://doi.org/10.1080/03610929608831858

Maronna, R. A. and Yohai, V. J. (2013). Robust functional linear regression based on splines. *Computational Statistics & Data Analysis* **65**, 46–55. MR3064942 https://doi.org/10.1016/j.csda.2011.11.014

Martines-Hernandez, I. and Genton, M. G. (2020). Recent developments in complex and spatially correlated functional data. *Brazilian Journal of Probability and Statistics* **34**, 204–229. MR4093256 https://doi.org/10.1214/20-BJPS466

Matsui, H. (2020). Quadratic regression for functional response models. *Econometrics and Statistics* **13**, 125–136. MR4058330 https://doi.org/10.1016/j.ecosta.2018.12.003

Matsui, H., Kawano, S. and Konishi, S. (2009). Regularized functional regression modeling for functional response and predictors. *Journal of Math-for-Industry* **1**, 17–25. MR2654326

Müller, H.-G. and Yao, F. (2008). Functional additive models. *Journal of the American Statistical Association* **103**, 1534–1544. MR2504202 https://doi.org/10.1198/016214508000000751

Olsen, N. L., Markussen, B. and Raket, L. L. (2018). Simultaneous inference for misaligned multivariate functional data. *Journal of the Royal Statistical Society Series C* **67**, 1147–1176. MR3873704 https://doi.org/10.1111/rssc.12276

Preda, C. and Saporta, G. (2005). PLS regression on a stochastic process. *Computational Statistics & Data Analysis* **48**, 149–158. MR2134488 https://doi.org/10.1016/j.csda.2003.10.003

Preda, C. and Schiltz, J. (2011). Functional PLS regression with functional response: The basis expansion approach. In *Proceedings of the 14th Applied Stochastic Models and Data Analysis Conference* 1126–1133. Universita di Roma La Spienza.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B* **53**, 539–572. MR1125714

Ramsay, J. O., Graves, S. and Hooker, G. (2020). fda: Functional data analysis. R package version 5.1.9. MR3645102 https://doi.org/10.1007/978-1-4939-7190-9

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. New York: Springer. MR1910407 https://doi.org/10.1007/b98886

Ramsay, J. O. and Silverman, B. W. (2006). *Functional Data Analysis*. New York: Springer. MR2168993

Raña, P., Aneiros, G. and Vilar, J. M. (2015). Detection of outliers in functional time series. *EnvironMetrics* **26**, 178–191. MR3335660 https://doi.org/10.1002/env.2327

Rao, B. L. S. P. (2010). Nonparametric density estimation for functional data by delta sequences. *Brazilian Journal of Probability and Statistics* **24**, 468–478. MR2719697 https://doi.org/10.1214/09-BJPS104

Reiss, P. T. and Odgen, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association* **102**, 984–996. MR2411660 https://doi.org/10.1198/016214507000000527

Serrneels, S., Croux, C., Filzmoser, P. and Espen, P. J. V. (2005). Partial robust M-regression. *Chemometrics and Intelligent Laboratory Systems* **79**, 55–64.

Shin, H. and Lee, S. (2016). An RKHS approach to robust functional linear regression. *Statistica Sinica* **26**, 255–272. MR3468352

Tucker, R. S. (1938). The reasons for price rigidity. *The American Economic Review* **28**, 41–54.

Wakelinc, I. N. and Macfie, H. J. H. (1992). A robust PLS procedure. *Journal of Chemometrics* **6**, 189–198.

Wang, W. (2014). Linear mixed function-on-function regression models. *Biometrics* **70**, 794–801. MR3295740 https://doi.org/10.1111/biom.12207

Wilcox, R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*. Waltham, MA: Elsevier. MR3286430 https://doi.org/10.1016/B978-0-12-386983-8.00001-9

Wold, H. (1974). Causal flows with latent variables: Partings of the ways in the light of NIPALS modelling. *European Economic Review* **5**, 67–86.

Yamanishi, Y. and Tanaka, Y. (2003). Geographically weighted functional multiple regression analysis: A numerical investigation. *Journal of the Japanese Society of Computational Statistics* **15**, 307–317. MR2027947 https://doi.org/10.5183/jjscs1988.15.2\protect\T1\textunderscore307

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33**, 2873–2903. MR2253106 https://doi.org/10.1214/009053605000000660

Younse, A. (2020). Nonparametric discrimination of areal functional data. *Brazilian Journal of Probability and Statistics* **34**, 112–126. MR4058973 https://doi.org/10.1214/18-BJPS418

Zhang, J., Zhou, Y., Cui, X. and Xu, W. (2018). Semiparametric quantile estimation for varying coefficient partially linear measurement errors models. *Brazilian Journal of Probability and Statistics* **32**, 616–656. MR3812385 https://doi.org/10.1214/17-BJPS357

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 220–242
https://doi.org/10.1214/21-BJPS524

Exponential squared loss based robust variable selection of AR models

Yaxin Wu[1], **Yunquan Song**[1,a], **Xijun Liang**[1] **and Yujie Gai**[2]

[1]*College of Science, China University of Petroleum, Qingdao 266580, P.R. China,* [a]*syqfly1980@163.com*
[2]*School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, P.R. China.*

Abstract. Time series analysis is widely used in the fields of economics, ecology and medicine. Robust variable selection procedures through penalized regression have been gaining increased attention. In our work, a robust penalized regression estimator based on exponential squared loss for autoregressive (AR) models is proposed and discussed. The objective model with adaptive Lasso penalty realizes variable selection and parameter estimation simultaneously. Under some regular conditions, we establish the asymptotic and "Oracle" properties of the proposed estimator. In particular, the induced non-convex and non-differentiable mathematical programming problem offers challenges for solving algorithms. To solve this problem efficiently, we specially design a block coordinate descent (BCD) algorithm equipped with concave-convex process (CCCP) and provide a convergence guarantee. Numerical simulation studies are carried out to show that the proposed method is particularly robust and applicable compared with some recent methods when there are different types of noise or different intensity of noise. Furthermore, an application on a dataset of daily minimum temperature in Melbourne over 1981–1990 is performed.

References

Akaike, H. (1977). On entropy maximisation principle. In *Applications of Statistics* (P. R. Krishnaiah, ed.) 27–41. Amsterdam: North Holland. MR0455163

Beck, A. and Teboulle, M. (2009). A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. MR2486527 https://doi.org/10.1137/080716542

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. New York: Springer. MR2839251

Chan, N. H., Ling, S. and Yau, C. Y. (2020). Lasso-based Variable Selection of ARMA Models. *Statistica Sinica.* MR4260750 https://doi.org/10.5705/ss.20

Chen, K. and Chan, K. S. (2011). Subset ARMA selection via the adaptive lasso. *Statistics and its Interface* **4**, 197–205. MR2812815 https://doi.org/10.4310/SII.2011.v4.n2.a14

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

Frank, I. and Friedman, J. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35**, 109–148.

Fu, W. J. (1998). Penalized regression: The bridge versus the lasso. *Journal of Computational and Graphical Statistics* **7**, 397–416. MR1646710 https://doi.org/10.2307/1390712

Hannan, E. J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* 81071-1081. MR0585705

Huber (1981). Interspecific variation in activity and regulation of leaf sucrose phosphate synthetase. *Zeitschrift für Pflanzenphysiologie* **102**, 443–450.

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics* **28**, 13561378. MR1805787 https://doi.org/10.1214/aos/1015957397

Kock, A. B. (2016). Consistent and conservative model selection with the adaptive lasso in stationary and nonstationary autoregressions. *Econometric Theory* **32**, 243–259. MR3442507 https://doi.org/10.1017/S0266466615000304

Key words and phrases. AR model, penalized robust regression, tuning parameter, variable selection.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 1. MR0474644 https://doi.org/10.2307/1913643

Liao, Z. P. and Phillips, P. C. B. (2015). Automated estimation of vector error correction models. *Econometric Theory* **31**, 581–646. MR3348460 https://doi.org/10.1017/S026646661500002X

Ling, S. and McAleer, M. (2010). A general asymptotic theory for time-series models. *Statistica Neerlandica* **64**, 97–111. MR2830968 https://doi.org/10.1111/j.1467-9574.2009.00447.x

Nardi, Y. and Rinaldo, A. (2011). Autoregressive process modeling via the lasso procedure. *Journal of Multivariate Analysis* **102**, 528–549. MR2755014 https://doi.org/10.1016/j.jmva.2010.10.012

Pötscher, B. M. (1983). Order estimation in ARMA-models by Lagrangian multiplier tests. *The Annals of Statistics* **11**, 872–885. MR0707937

Pötscher, B. M. and Srinivasan, S. (1994). A comparison of order estimation procedures for ARMA models. *Statistica Sinica* **4**, 50. MR1282864

Rissanen, J. (1978). Modeling by shortest data description. *Automatica* **14**, 465–471.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464. MR0468014

Song, S. and Bickel, P. J. (2011). Large vector autoregressions. Working paper, Univ. California, Berkeley.

Song, Y., Liang, X., Zhu, Y., et al (2020). Robust variable selection with exponential squared loss for the spatial autoregressive model. *Computational Statistics & Data Analysis* **155**, 107094. MR4161786 https://doi.org/10.1016/j.csda.2020.107094

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **58**, 267–288. MR1379242

Wang, H., Li, G. and Tsai, C. L. (2007). Regression coefficients and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **69**, 63–78. MR2301500 https://doi.org/10.1111/j.1467-9868.2007.00577.x

Wang, X., Jiang, Y. L., Huang, M. and Zhang, H. (2013). Robust variable selection with exponential squared loss. *Journal of the American Statistical Association* **108**, 632–643. MR3174647 https://doi.org/10.1080/01621459.2013.766613

Yuille, A. L. and Rangarajan, A. (2001). The Concave-Convex Procedure. National Institute of Health, R01-EY 12691-01.

Yuille, A. L. and Rangarajan, A. (2003). The concave-convex procedure. *Neural Computation* **15**, 915–936.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* **36**, 1108–1126. MR2418651 https://doi.org/10.1214/07-AOS507

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 243–262
https://doi.org/10.1214/21-BJPS525

Model selection for functional linear regression with hierarchical structure

Sanying Feng, Xinyu Zhang, Hui Liang and Lifang Pei[a]

School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China, [a]*plf5801@zzu.edu.cn*

Abstract. Scalar-on-function regression allows for a scalar response to be dependent on functional predictors; however, not much work has been done when interaction effects between the functional predictors are included. In this paper, we introduce a multiple functional linear regression model with interaction terms. Meanwhile, we enforce the hierarchical structure constraint on the model, that is, interaction terms can be selected into the model only if the associated main effects are in the model. Based on the functional principal component analysis and group smoothly clipped absolute deviation (SCAD) penalty, we propose a new penalized estimation procedure to select the important functional predictors and interactions while automatically obeying the hierarchical structure. With appropriate selection of the tuning parameters, the rates of convergence of the proposed estimators and the consistency of the model selection procedure are established under some regularity conditions. At last, we illustrate the finite sample performance of our proposed methods with some simulation studies and a real data application.

References

Bien, J., Taylor, J. and Tibshirani, R. (2013). A lasso for hierarchical interactions. *The Annals of Statistics* **41**, 1111–1141. MR3113805 https://doi.org/10.1214/13-AOS1096

Cai, T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics* **34**, 2159–2179. MR2291496 https://doi.org/10.1214/009053606000000830

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45**, 11–22. MR1718346 https://doi.org/10.1016/S0167-7152(99)00036-X

Choi, N. H., Li, W. and Zhu, J. (2010). Variable selection with the strong heredity constraint and its oracle property. *Journal of the American Statistical Association* **105**, 354–364. MR2656056 https://doi.org/10.1198/jasa.2010.tm08281

Collazos, J. A., Dias, R. and Zambom, A. Z. (2016). Consistent variable selection for functional regression models. *Journal of Multivariate Analysis* **146**, 63–71. MR3477649 https://doi.org/10.1016/j.jmva.2015.06.007

Cox, D. R. (1984). Interaction. *International Statistical Review* **52**, 1–31. MR0967201 https://doi.org/10.2307/1403235

Fan, J. Q. and Li, R. Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. New York: Springer. MR2229687

Fuchs, K., Scheipl, F. and Greven, S. (2015). Penalized scalar-on-functions regression with interaction term. *Computational Statistics & Data Analysis* **81**, 38–51. MR3257399 https://doi.org/10.1016/j.csda.2014.07.001

Gertheiss, J., Maity, A. and Staicu, A. M. (2013). Variable selection in generalized functional linear models. *Stat* **2**, 86–101. MR4027303 https://doi.org/10.1002/sta4.20

Hall, P. and Hooker, G. (2016). Truncated linear models for functional data. *Journal of the Royal Statistical Society, Series B* **78**, 637–653. MR3506796 https://doi.org/10.1111/rssb.12125

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35**, 70–91. MR2332269 https://doi.org/10.1214/009053606000000957

Hao, N., Feng, Y. and Zhang, H. H. (2018). Model selection for high-dimensional quadratic regression via regularization. *Journal of the American Statistical Association* **113**, 615–625. MR3832213 https://doi.org/10.1080/01621459.2016.1264956

Key words and phrases. Functional linear regression model, hierarchical structure, interaction effect, main effect, model selection, SCAD.

Haris, A., Witten, D. and Simon, N. (2016). Convex modeling of interactions with strong heredity. *Journal of Computational and Graphical Statistics* **25**, 981–1004. MR3572025 https://doi.org/10.1080/10618600.2015.1067217

Horvath, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. New York: Springer. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

Huang, L., Zhao, J., Wang, H. and Wang, S. (2016). Robust shrinkage estimation and selection for functional multiple linear model through LAD loss. *Computational Statistics & Data Analysis* **103**, 384–400. MR3522639 https://doi.org/10.1016/j.csda.2016.05.017

Kong, E., Tong, H. and Xia, Y. (2010). Statistical modelling of nonlinear long-term cumulative effects. *Statistica Sinica* **20**, 1097–1123. MR2730175

Kong, Y., Li, D., Fan, Y. and Lv, J. (2017). Interaction pursuit in high-dimensional multi-response regression via distance correlation. *The Annals of Statistics* **45**, 897–922. MR3650404 https://doi.org/10.1214/16-AOS1474

Lian, H. (2013). Shrinkage estimation and selection for multiple functional regression. *Statistica Sinica* **23**, 51–74. MR3076158

Lim, M. and Hastie, T. (2015). Learning interactions via hierarchical group lasso regularization. *Journal of Computational and Graphical Statistics* **24**, 627–654. MR3397226 https://doi.org/10.1080/10618600.2014.938812

Luo, R. and Qi, X. (2019). Interaction model and model selection for function-on-function regression. *Journal of Computational and Graphical Statistics* **28**, 309–322. MR3974882 https://doi.org/10.1080/10618600.2018.1514310

Matsui, H. and Konishi, K. (2011). Variable selection for functional regression models via the L_1 regularization. *Computational Statistics & Data Analysis* **55**, 3304–3310. MR2825412 https://doi.org/10.1016/j.csda.2011.06.016

McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research* **16**, 285–292. MR0748553 https://doi.org/10.1016/0377-2217(84)90282-0

Pannu, J. and Billor, N. (2017). Robust group-Lasso for functional regression model. *Communications in Statistics Simulation and Computation* **46**, 3356–3374. MR3656106 https://doi.org/10.1080/03610918.2015.1096375

Radchenko, P. and James, G. M. (2010). Variable selection using adaptive nonlinear interaction structures in high dimensions. *Journal of the American Statistical Association* **105**, 1541–1553. MR2796570 https://doi.org/10.1198/jasa.2010.tm10130

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer. MR2168993

She, Y., Wang, Z. and Jiang, H. (2018). Group regularized estimation under structural hierarchy. *Journal of the American Statistical Association* **113**, 445–454. MR3803477 https://doi.org/10.1080/01621459.2016.1260470

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **58**, 267–288. MR1379242

Usset, J., Staicu, A. M. and Maity, A. (2016). Interaction models for functional regression. *Computational Statistics & Data Analysis* **94**, 317–329. MR3412828 https://doi.org/10.1016/j.csda.2015.08.020

Wang, H., Li, R. and Tsai, C. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* **94**, 553–568. MR2410008 https://doi.org/10.1093/biomet/asm053

Wong, C. M., Atkinson, R. W., Anderson, H. R., Hedley, A. J., Ma, S., Chau, P. Y. K. and Lam, T. H. (2002). A tale of two cities: Effects of air pollution on hospital admissions in Hong Kong and London compared. *Environmental Health Perspectives* **110**, 67–77.

Wong, H., Shao, Q. and Ip, W. (2013). Modeling respiratory illnesses with change point: A lesson from the SARS epidemic in Hong Kong. *Computational Statistics & Data Analysis* **57**, 589–599. MR2981111 https://doi.org/10.1016/j.csda.2012.07.029

Wong, R. K. W., Li, Y. and Zhu, Z. (2019). Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association* **114**, 406–418. MR3941264 https://doi.org/10.1080/01621459.2017.1411268

Xia, Y. and Tong, H. (2006). Cumulative effects of air pollution on public health. *Statistics in Medicine* **25**, 3548–3559. MR2252410 https://doi.org/10.1002/sim.2446

Xia, Y., Zhang, W. and Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika* **91**, 661–681. MR2090629 https://doi.org/10.1093/biomet/91.3.661

Xue, K. and Yao, F. (2021). Hypothesis testing in large-scale functional linear regression. *Statistica Sinica* **31**, 1101–1123. MR4286208 https://doi.org/10.5705/ss.20

Yao, F. and Muller, H. G. (2010). Functional quadratic regression. *Biometrika* **97**, 49–64. MR2594416 https://doi.org/10.1093/biomet/asp069

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B* **68**, 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38**, 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

Zhao, P., Rocha, G. and Yu, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics* **37**, 3468–3497. MR2549566 https://doi.org/10.1214/07-AOS584

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 263–286
https://doi.org/10.1214/21-BJPS526

An alternative class of models to position social network groups in latent spaces

Izabel Nolau[1,a] **and Gustavo S. Ferreira**[2,b]

[1]*Department of Statistical Methods, Federal University of Rio de Janeiro, 149 Athos da Silveira Ramos Ave, ZIP 21941-909, Rio de Janeiro, RJ, Brazil,* [a]*nolau@dme.ufrj.br*

[2]*National School of Statistical Sciences—ENCE, Brazilian Institute of Geography and Statistics—IBGE, 106 Andre Cavalcanti St, ZIP 20231-050, Rio de Janeiro, RJ, Brazil,* [b]*gustavo.ferreira@ibge.gov.br*

Abstract. Identifying key nodes, estimating the probability of connection between them, and distinguishing latent groups are some of the main objectives of social network analysis. In this paper, we propose a class of blockmodels to model stochastic equivalence and visualize groups in an unobservable space. In this setting, the proposed method is based on two approaches: latent distances and latent dissimilarities at the group level. The projection proposed in the paper is performed without needing to project individuals, unlike the main approaches in the literature. Our approach can be used in undirected or directed graphs and is flexible enough to cluster and quantify between and within-group tie probabilities in social networks. The effectiveness of the methodology in representing groups in latent spaces was analyzed under artificial datasets and in two case studies.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* **9**, 1981–2014.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, 3–41. MR1349818

Breiger, R. L., Boorman, S. A. and Arabie, P. (1975). An algorithm for blocking relational data, with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology* **12**, 328–383.

Cavallari, S., Cambria, E., Cai, H., Chang, K. C.-C. and Zheng, V. W. (2019). Embedding both finite and infinite communities on graphs [application notes]. *IEEE Computational Intelligence Magazine* **14**, 39–50.

Cavallari, S., Zheng, V. W., Cai, H., Chang, K. C.-C. and Cambria, E. (2017). Learning community embedding with community detection and node embedding on graphs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 377–386.

Celeux, G. (1998). Bayesian inference for mixture: The label switching problem. In *Compstat*, 227–232.

Celeux, G., Hurn, M. and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association* **95**, 957–970. MR1804450 https://doi.org/10.2307/2669477

Choi, D. S., Wolfe, P. J. and Airoldi, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99**, 273–284. MR2931253 https://doi.org/10.1093/biomet/asr053

Doreian, P., Batagelj, V. and Ferligoj, A. (2005). *Generalized Blockmodeling, Vol. 25*. Cambridge University Press.

Faust, K. and Romney, A. K. (1985). Does structure find structure?: A critique of Burt's use of distance as a measure of structural equivalence. *Social Networks* **7**, 77–103.

Fienberg, S. E. and Wasserman, S. S. (1981). Categorical data analysis of single sociometric relations. *Sociological Methodology* **12**, 156–192.

Fosdick, B. K., McCormick, T. H., Murphy, T. B., Ng, T. L. J. and Westling, T. (2019). Multiresolution network models. *Journal of Computational and Graphical Statistics* **28**, 185–196. MR3939381 https://doi.org/10.1080/10618600.2018.1505633

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press. MR2260716

Key words and phrases. Blockmodel, social networks, multidimensional scaling, latent space, visualization.

Handcock, M. S., Raftery, A. E. and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society Series A Statistics in Society* **170**, 301–354. MR2364300 https://doi.org/10.1111/j.1467-985X.2007.00471.x

Hedenfalk, I. A., Ringnér, M., Trent, J. M. and Borg, A. (2002). Gene expression in inherited breast cancer. *Cancer Research* **84**.

Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97**, 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5**, 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* **76**, 33–50.

Krivitsky, P. N. and Handcock, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software* **24**.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E. and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks* **31**, 204–213.

Lorrain, F. and White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology* **1**, 49–80.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 281–297. Berkeley, Calif.: University of California Press. https://projecteuclid.org/euclid.bsmsp/1200512992. MR0214227

Mardia, K. V. (1978). Some properties of clasical multi-dimesional scaling. *Communications in Statistics Theory and Methods* **7**, 1233–1241. MR0514645 https://doi.org/10.1080/03610927808827707

Ng, T. L. J., Murphy, T. B., Westling, T., McCormick, T. H. and Fosdick, B. K. (2018). Modeling the social media relationships of Irish politicians using a generalized latent space stochastic blockmodel. arXiv preprint arXiv:1807.06063. MR4355082 https://doi.org/10.1214/21-aoas1483

Nikkilä, J., Törönen, P., Kaski, S., Venna, J., Castrén, E. and Wong, G. (2002). Analysis and visualization of gene expression data using self-organizing maps. *Neural Networks* **15**, 953–966.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* **96**, 1077–1087. MR1947255 https://doi.org/10.1198/016214501753208735

Oh, M.-S. and Raftery, A. E. (2007). Model-based clustering with dissimilarities: A Bayesian approach. *Journal of Computational and Graphical Statistics* **16**, 559–585. MR2351080 https://doi.org/10.1198/106186007X236127

Papastamoulis, P. and Iliopoulos, G. (2010). An artificial allocations based solution to the label switching problem in Bayesian analysis of mixtures of distributions. *Journal of Computational and Graphical Statistics* **19**, 313–331. MR2758306 https://doi.org/10.1198/jcgs.2010.09008

Papastamoulis, P. and Iliopoulos, G. (2013). On the convergence rate of random permutation sampler and ECR algorithm in missing data models. *Methodology and Computing in Applied Probability* **15**, 293–304. MR3053958 https://doi.org/10.1007/s11009-011-9238-7

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Rodriguez, C. E. and Walker, S. G. (2014). Label switching in Bayesian mixture models: Deterministic relabeling strategies. *Journal of Computational and Graphical Statistics* **23**, 25–45. MR3173759 https://doi.org/10.1080/10618600.2012.735624

Rohe, K., Chatterjee, S., Yu, B., et al (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* **39**, 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

Sampson, S. F. (1968). A novitiate in a period of change: An experimental and case study of social relationships. PhD Thesis, Cornell University.

Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **77**, 647–676. MR3351449 https://doi.org/10.1111/rssb.12081

Snijders, T. A. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification* **14**, 75–100. MR1449742 https://doi.org/10.1007/s003579900004

Tallberg, C. (2004). A Bayesian approach to modeling stochastic blockstructures with covariates. *The Journal of Mathematical Sociology* **29**, 1–23.

Wang, X., Cui, P., Wang, J., Pei, J., Zhu, W. and Yang, S. (2017). Community preserving network embedding. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* **82**, 8–19. MR0883333 https://doi.org/10.1214/21-aoas1483

Wasserman, S. and Anderson, C. (1987). Stochastic a posteriori blockmodels: Construction and assessment. *Social Networks* **9**, 1–36. MR0885874 https://doi.org/10.1016/0378-8733(87)90015-3

White, H. C., Boorman, S. A. and Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology* **81**, 730–780.

Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* **33**, 452–473.

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 287–314
https://doi.org/10.1214/21-BJPS527

A general restricted estimator in binary logistic regression in the presence of multicollinearity

Gargi Tyagi[a] **and Shalini Chandra**[b]

Department of Mathematics & Statistics, Banasthali Vidyapith, Banasthali-304022 Rajasthan, India,
[a]*tyagi.gargi@gmail.com,* [b]*chandrshalini@gmail.com*

Abstract. The presence of multicollinearity adversely affects the inferential properties of the maximum likelihood (ML) estimator in logistic regression model. It is a well-established fact that the use of restrictions lowers the effect of multicollinearity. In this article, an alternative to the ML estimator has been introduced by combining the exact prior information into the logistic $r - k$ class (Lrk) estimator. The estimator is named a logistic restricted $r - k$ class estimator. Another estimator, logistic restricted PCR estimator, is also developed as a special case of the LRrk estimator. The asymptotic mean squared error (MSE) matrix properties of the estimators are studied and necessary and sufficient conditions are derived. Further, a Monte Carlo simulation study is performed to compare the performance of the estimators in terms of the scalar MSE and the prediction MSE. It is found that the proposed estimators perform better than the existing estimators in most of the cases considered. Moreover, a numerical example has also been presented for comparing the performance of the estimators.

References

Abdel-Fattah, M. A. (2020). On a new class of binomial ridge-type regression estimators. *Communications in Statistics—Simulation and Computation.*

Aguilera, A. M., Escabias, M. and Valderrama, M. J. (2006). Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis* **50**, 1905–1924. MR2225551 https://doi.org/10.1016/j.csda.2005.03.011

Asar, Y., Arashi, M. and Wu, J. (2017). Restricted ridge estimator in the logistic regression model. *Communications in Statistics Simulation and Computation* **46**, 6538–6544. MR3740796 https://doi.org/10.1080/03610918.2016.1206932

Baksalary, J. K. and Trenkler, G. (1991). Nonnegative and positive definiteness of matrices modified by two matrices of rank one. *Linear Algebra and Its Applications* **151**, 169–184. MR1102148 https://doi.org/10.1016/0024-3795(91)90362-Z

Baye, M. R. and Parker, D. F. (1984). Combining ridge and principal component regression: A money demand illustration. *Communications in Statistics Theory and Methods* **13**, 197–205. MR0746211 https://doi.org/10.1080/03610928408828675

Duffy, D. E. and Santner, T. J. (1989). On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models. *Communications in Statistics Theory and Methods* **18**, 959–980. MR1001630 https://doi.org/10.1080/03610928908829944

Farebrother, R. W. (1972). Principal component estimators and minimum mean square error criteria in regression analysis. *Review of Economics and Statistics* **54**, 332–336.

Farebrother, R. W. (1976). Further results on the mean square error of ridge regression. *Journal of the Royal Statistical Society, Series B* **38**, 248–250. MR0653156

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics* **12**, 55–67. MR4165988 https://doi.org/10.1080/00401706.2020.1742207

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013). *Applied Logistic Regression 398*. New York: Wiley. MR3287463 https://doi.org/10.1002/9781118596333.ch21

Jadhav, N. H. (2020). On linearized ridge logistic estimator in the presence of multicollinearity. *Computational Statistics* **35**, 667–687. MR4102948 https://doi.org/10.1007/s00180-019-00935-6

Key words and phrases. Logistic regression, multicollinearity, restrictions, mean squared error, predicted mean squared error.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Pearson-Prentice Hall. MR2372475

Le Cessie, S. and Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Applied Statistics* **41**, 191–201.

Lee, A. H. and Silvapulle, M. J. (1988). Ridge estimation in logistic regression. *Communications in Statistics Simulation and Computation* **17**, 1231–1257. MR0971585 https://doi.org/10.1080/03610918808812723

Lukman, A. F., Emmanuel, A., Clement, O. A. and Ayinde, K. (2020). A modified ridge-type logistic estimator. *Iranian Journal of Science and Technology* **44**, 437–443. MR4085334 https://doi.org/10.1007/s40995-020-00845-z

Mackinon, M. J. and Puterman, M. L. (1989). Collinearity in generalized linear models. *Communications in Statistics Theory and Methods* **18**, 3463–3472. MR1031129 https://doi.org/10.1080/03610928908830102

Månsson, K., Kibria, B. M. G. and Shukur, G. (2016). A restricted Liu estimator for binary regression models and its application to an applied demand system. *Journal of Applied Statistics* **43**, 1119–1127. MR3460556 https://doi.org/10.1080/02664763.2015.1092110

Mansson, K. and Shukur, G. (2011). On ridge parameters in logistic regression. *Communications in Statistics Theory and Methods* **40**, 3366–3381. MR2860826 https://doi.org/10.1080/03610926.2010.500111

Marx, B. D. (1992). A continuum of principal component generalized linear regressions. *Computational Statistics & Data Analysis* **13**, 385–393. MR1173329 https://doi.org/10.1016/0167-9473(92)90113-T

Marx, B. D. and Smith, E. P. (1990). Principal component estimation for generalized linear regression. *Biometrika* **77**, 23–31. MR1049405 https://doi.org/10.1093/biomet/77.1.23

Massy, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association* **60**, 234–256.

Myers, R. H., Montogomery, D. C., Vining, G. G. and Robinson, T. J. (2010). *Gerneralized Linear Models with Applications in Engineering and the Sciences*. New Jersey: Wiley. MR2683129 https://doi.org/10.1002/9780470556986

Özkale, M. R. (2009). Principal components regression estimator and a test for the restrictions. *Statistics* **43**, 541–551. MR2588267 https://doi.org/10.1080/02331880802605460

Özkale, M. R. (2019). The red indicator and corrected VIFs in generalized linear models. *Communications in Statistics—Simulation and Computation.* MR4343317 https://doi.org/10.1080/03610918.2019.1639740

Özkale, M. R. and Arıcan, E. (2016). A new biased estimator in logistic regression model. *Statistics* **50**, 233–253. MR3452984 https://doi.org/10.1080/02331888.2015.1123711

Özkale, M. R. and Kaçıranlar, S. (2007). The restricted and unrestricted two parameter estimators. *Communications in Statistics Theory and Methods* **36**, 2707–2725. MR2413601 https://doi.org/10.1080/03610920701386877

Rao, C. R. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives*. New York: Springer. MR1354840 https://doi.org/10.1007/978-1-4899-0024-1

Schaefer, R. L. (1986). Alternative estimators in logistic regression when the data are collinear. *Journal of Statistical Computation and Simulation* **25**, 75–91.

Schaefer, R. L., Roi, L. D. and Wolfe, R. A. (1984). A ridge logistic estimator. *Communications in Statistics Theory and Methods* **13**, 99–113.

Şiray, G. U., Toker, S. and Kaçiranlar, S. (2015). On the restricted Liu estimator in the logistic regression model. *Communications in Statistics Simulation and Computation* **44**, 217–232. MR3238559 https://doi.org/10.1080/03610918.2013.771742

Smith, E. P. and Marx, B. D. (1990). Ill-conditioned information matrices, generalized linear models and estimation of the effects of acid rain. *EnvironMetrics* **1**, 57–71.

Wang, S. (1994). *The Inequalities of Matrices*. Hefei: The Education of Anhui Press.

Webster, J. T., Gunst, R. F. and Mason, R. L. (1974). Latent root regression analysis. *Technometrics* **16**, 513–522. MR0362741 https://doi.org/10.2307/1267602

Weissfeld, L. A. and Sereika, S. M. (1991). A multicollinearity diagnostic for generalized linear models. *Communications in Statistics Theory and Methods* **20**, 1183–1198. MR1117321 https://doi.org/10.1080/03610929108830558

Wu, J., Asar, Y. and Arashi, M. (2018). On the restricted almost unbiased Liu estimator in the logistic regression model. *Communications in Statistics Theory and Methods* **47**, 4389–4401. MR3819789 https://doi.org/10.1080/03610926.2017.1376082

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 315–348
https://doi.org/10.1214/21-BJPS528

Convergence of partial sum processes to stable processes with application for aggregation of branching processes

Mátyás Barczy[1,a], **Fanni K. Nedényi**[1,b] **and Gyula Pap**[2]

[1]*MTA-SZTE Analysis and Stochastics Research Group, Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H–6720 Szeged, Hungary,* [a]*barczy@math.u-szeged.hu,* [b]*nfanni@math.u-szeged.hu*
[2]*Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H–6720 Szeged, Hungary*

Abstract. We provide a generalization of Theorem 1 in Bartkiewicz et al. (2011) in the sense that we give sufficient conditions for weak convergence of finite dimensional distributions of the partial sum processes of a strongly stationary sequence to the corresponding finite dimensional distributions of a non-Gaussian stable process instead of weak convergence of the partial sums themselves to a non-Gaussian stable distribution. As an application, we describe the asymptotic behaviour of finite dimensional distributions of aggregation of independent copies of a strongly stationary subcritical Galton–Watson branching process with regularly varying immigration having index in $(0, 1) \cup (1, 4/3)$ in a so-called iterated case, namely when first taking the limit as the time scale and then the number of copies tend to infinity.

References

Araujo, A. and Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0576407

Barczy, M., Basrak, B., Kevei, P., Pap, G. and Planinić, H. (2021). Statistical inference of subcritical strongly stationary Galton–Watson processes with regularly varying immigration. *Stochastic Processes and Their Applications* **132**, 33–75. MR4168330 https://doi.org/10.1016/j.spa.2020.10.004

Barczy, M., Bősze, Zs. and Pap, G. (2020). On tail behaviour of stationary second-order Galton–Watson processes with immigration. *Modern Stochastics: Theory and Applications* **7**, 315–338. MR4159152 https://doi.org/10.15559/20-vmsta161

Barczy, M., K. Nedényi, F. and Pap, G. (2018). On aggregation of multitype Galton–Watson branching processes with immigration. *Modern Stochastics: Theory and Applications* **5**, 53–79. MR3784038 https://doi.org/10.15559/18-vmsta95

Barczy, M., K. Nedényi, F. and Pap, G. (2019a). Convergence of partial sum processes to stable processes with application for aggregation of branching processes. Available at 1906.04999.

Barczy, M., K. Nedényi, F. and Pap, G. (2019b). On aggregation of subcritical Galton–Watson branching processes with regularly varying immigration. Available at 1906.00373. MR4174658 https://doi.org/10.1007/s10986-020-09492-8

Barczy, M., K. Nedényi, F. and Pap, G. (2020). On aggregation of subcritical Galton–Watson branching processes with regularly varying immigration. *Lithuanian Mathematical Journal* **60**, 425–451. MR4174658 https://doi.org/10.1007/s10986-020-09492-8

Bartkiewicz, K., Jakubowski, A., Mikosch, T. and Wintenberger, O. (2011). Stable limits for sums of dependent infinite variance random variables. *Probability Theory and Related Fields* **150**, 337–372. MR2824860 https://doi.org/10.1007/s00440-010-0276-9

Basrak, B. and Kevei, P. (2020). Limit theorems for branching processes with immigration in a random environment. Available at 2002.00634.

Basrak, B., Krizmanić, D. and Segers, J. (2012). A functional limit theorem for dependent sequences with infinite variance stable limits. *Annals of Probability* **40**, 2008–2033. MR3025708 https://doi.org/10.1214/11-AOP669

Basrak, B., Kulik, R. and Palmowski, Z. (2013). Heavy-tailed branching process with immigration. *Stochastic Models* **29**, 413–434. MR3175851 https://doi.org/10.1080/15326349.2013.838508

Basrak, B. and Segers, J. (2009). Regularly varying multivariate time series. *Stochastic Processes and Their Applications* **119**, 1055–1080. MR2508565 https://doi.org/10.1016/j.spa.2008.05.004

Key words and phrases. Multivariate regular variation, strong stationarity, stable process, Galton–Watson branching processes with immigration, iterated aggregation.

Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-Memory Processes. Probabilistic Properties and Statistical Methods*. Heidelberg: Springer. MR3075595 https://doi.org/10.1007/978-3-642-35512-7

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications.* **27**. Cambridge: Cambridge University Press. MR1015093

Buraczewski, D., Damek, E. and Mikosch, T. (2016). *Stochastic Models with Power-Law Tails. The Equation $X = AX + B$*. *Springer Series in Operations Research and Financial Engineering.* Switzerland: Springer. MR3497380 https://doi.org/10.1007/978-3-319-29679-1

Cattiaux, P. and Manou-Abi, M. (2014). Limit theorems for some functionals with heavy tails of a discrete time Markov chain. *ESAIM Probabilités Et Statistique* **18**, 468–482. MR3333999 https://doi.org/10.1051/ps/2013043

Davis, R. and Hsing, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability* **23**, 879–917. MR1334176

Kallenberg, O. (2017). *Random Measures, Theory and Applications*. *Probability Theory and Stochastic Modelling* **77**. Cham: Springer. MR3642325 https://doi.org/10.1007/978-3-319-41598-7

Kesten, H., Kozlov, M. V. and Spitzer, F. L. (1975). A limit law for random walk in a random environment. *Compositio Mathematica* **30**, 145–168. MR0380998

Klenke, A. (2014). *Probability Theory: A Comprehensive Course*, 2nd ed. *Universitext.* London: Springer. MR3112259 https://doi.org/10.1007/978-1-4471-5361-0

Kulik, R. and Soulier, P. (2020). *Heavy-Tailed Time Series*. *Springer Series in Operations Research and Financial Engineering.* New York: Springer. MR4174389 https://doi.org/10.1007/978-1-0716-0737-4

Lin, Z. and Lu, C. (1996). *Limit Theory for Mixing Dependent Random Variables*. *Mathematics and Its Applications.* **378**. Dordrecht: Kluwer Academic Publishers; Science Press Beijing, New York. MR1486580

Lindskog, F. (2004). Multivariate extremes and regular variation for stochastic processes. PhD thesis, Swiss Federal Institute of Technology Zürich. Diss. ETH No. 15319. MR2715640

Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge University Press. With a prologue by Peter W. Glynn. MR2509253 https://doi.org/10.1017/CBO9780511626630

Mikosch, T. and Wintenberger, O. (2014). The cluster index of regularly varying sequences with applications to limit theory for functions of multivariate Markov chains. *Probability Theory and Related Fields* **159**, 157–196. MR3201920 https://doi.org/10.1007/s00440-013-0504-1

Mikosch, T. and Wintenberger, O. (2016). A large deviations approach to limit theory for heavy-tailed time series. *Probability Theory and Related Fields* **166**, 233–269. MR3547739 https://doi.org/10.1007/s00440-015-0654-4

Mikosch, T. and Wintenberger, O. (2022+). *Extremes for Time Series*. Berlin: Springer. Forthcoming.

Pilipauskaitė, V. and Surgailis, D. (2014). Joint temporal and contemporaneous aggregation of random-coefficient AR(1) processes. *Stochastic Processes and Their Applications* **124**, 1011–1035. MR3138604 https://doi.org/10.1016/j.spa.2013.10.004

Quine, M. P. (1970). The multi-type Galton–Watson process with immigration. *Journal of Applied Probability* **7**, 411–422. MR0263168 https://doi.org/10.1017/s0021900200034975

Resnick, S. I. (1986). Point processes, regular variation and weak convergence. *Advances in Applied Probability* **18**, 66–138. MR0827332 https://doi.org/10.2307/1427239

Resnick, S. I. (2007). *Heavy-Tail Phenomena*. *Springer Series in Operations Research and Financial Engineering.* New York: Springer. MR2271424

Roitershtein, A. and Zhong, Z. (2013). On random coefficient INAR(1) processes. *Science China. Mathematics* **56**, 177–200. MR3016591 https://doi.org/10.1007/s11425-012-4547-z

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge University Press. Translated from the 1990 Japanese original, Revised by the author. MR1739520

Tyran-Kamińska, M. (2010). Convergence to Lévy stable processes under some weak dependence conditions. *Stochastic Processes and Their Applications* **120**, 1629–1650. MR2673968 https://doi.org/10.1016/j.spa.2010.05.010

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 349–368
https://doi.org/10.1214/21-BJPS529

Approximations related to the sums of *m*-dependent random variables

Amit N. Kumar[1,a], **Neelesh S. Upadhye**[2,b] **and P. Vellaisamy**[3,c]

[1]*Department of Mathematical Sciences, Indian Institute of Technology (BHU), Varanasi-221005, India,* [a]*amit.kumar2703@gmail.com*

[2]*Department of Mathematics, Indian Institute of Technology Madras, Chennai-600036, India,* [b]*neelesh@iitm.ac.in*

[3]*Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai-400076, India,* [c]*pv@math.iitb.ac.in*

Abstract. In this paper, we mainly focus on the sums of non-negative integer-valued 1-dependent random variables and its approximation to the power series distribution. We first discuss some relevant results for power series distribution such as the Stein operator, uniform and non-uniform bounds on the solution of the Stein equation. Using Stein's method, we obtain error bounds for the approximation problem considered. The obtained results can also be applied to the sums of m-dependent random variables via appropriate rearrangements of random variables. As special cases, we discuss two applications, namely, 2-runs and (k_1, k_2)-runs, and compare our bounds with existing bounds.

References

Balakrishnan, N. and Koutras, M. V. (2002). *Runs and Scans with Applications*. New York: Wiley. MR1882476

Barbour, A. D. and Chen, L. H. Y. (2014). Stein's (magic) method. Preprint. Available at arXiv:1411.1179.

Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*. Oxford: Oxford Univ. Press. MR1163825

Barbour, A. D. and Xia, A. (1999). Poisson perturbation. *ESAIM Probabilités Et Statistique* **3**, 131–150. MR1716120 https://doi.org/10.1051/ps:1999106

Brown, T. C. and Xia, A. (2001). Stein's method and birth-death processes. *Annals of Probability* **29**, 1373–1403. ISSN 0091-1798. https://doi.org/10.1214/aop/1015345606

Čekanavičius, V. and Vellaisamy, P. (2015). Discrete approximations for sums of m-dependent random variables. *ALEA – Latin American Journal of Probability and Mathematical Statistics* **12**, 765–792. ISSN 1980-0436. MR3446037

Daly, F., Lefèvre, C. and Utev, S. (2012). Stein's method and stochastic orderings. *Advances in Applied Probability* **44**, 343–372. ISSN 0001-8678. https://doi.org/10.1239/aap/1339878715

Edwin, T. K. (2014). Power series distributions and zero-inflated models. PhD thesis, University of Nairobi.

Eichelsbacher, P. and Reinert, G. (2008). Stein's method for discrete Gibbs measures. *The Annals of Applied Probability* **18**, 1588–1618. MR2434182 https://doi.org/10.1214/07-AAP0498

Fu, J. C. and Johnson, B. C. (2009). Approximate probabilities for runs and patterns in I.I.D. and Markov-dependent multistate trials. *Advances in Applied Probability* **41**, 292–308. ISSN 0001-8678. MR2514955 https://doi.org/10.1239/aap/1240319586

Godbole, A. P. (1993). Approximate reliabilities of m-consecutive-k-out-of-n: Failure systems. *Statistica Sinica* **3**, 321–327. ISSN 1017-0405. MR1243390

Godbole, A. P. and Schaffner, A. A. (1993). Improved Poisson approximations for word patterns. *Advances in Applied Probability* **25**, 334–347. MR1212615 https://doi.org/10.2307/1427656

Hess, K. T., Liewald, A. and Schmidt, K. D. (2002). An extension of Panjer's recursion. *ASTIN Bulletin* **32**, 283–297. MR1942940 https://doi.org/10.2143/AST.32.2.1030

Huang, W. T. and Tsai, C. S. (1991). On a modified binomial distribution of order k. *Statistics & Probability Letters* **11**, 125–131. MR1092971 https://doi.org/10.1016/0167-7152(91)90129-F

Kumar, A. N. and Upadhye, N. S. (2020). On discrete Gibbs measure approximation to runs, *Comm. Statist. Theory Methods*. To appear.

Key words and phrases. Power series distribution, m-dependent random variables, Stein's method, runs.

Kumar, A. N., Vellaisamy, P. and Viens, F. (2021). Poisson approximation to the convolution of power series distributions, *Probab. Math. Statist.* To appear.

Ley, C., Reinert, G. and Swan, Y. (2017). Stein's method for comparison of univariate distributions. *Probability Surveys* **14**, 1–52. MR3595350 https://doi.org/10.1214/16-PS278

Lin, Z.-Y. and Liu, W. (2012). m-dependence approximation for dependent random variables. In *Probability Approximations and Beyond*. *Lect. Notes Stat.* **205**, 117–133. Berlin: Springer. MR3289380 https://doi.org/10.1007/978-1-4614-1966-2\protect\T1\textunderscore9

Mattner, L. and Roos, B. (2007). A shorter proof of Kanter's Bessel function concentration bound. *Probability Theory and Related Fields* **139**, 191–205. MR2322695 https://doi.org/10.1007/s00440-006-0043-0

Noack, A. (1950). A class of random variables with discrete distributions. *The Annals of Mathematical Statistics* **21**, 127–132. ISSN 0003-4851. https://doi.org/10.1214/aoms/1177729894

Panjer, H. H. and Wang, S. (1995). Computational aspects of Sundt's generalized class. *ASTIN Bulletin* **25**, 5–17.

Patil, G. P. (1962). Certain properties of the generalized power series distribution. *Annals of the Institute of Statistical Mathematics* **14**, 179–182. ISSN 0020-3157. https://doi.org/10.1007/BF02868639

Reinert, G. (2005). Three general approaches to Stein's method. An introduction to Stein's method. In *A Program in Honour of Charles Stein: Tutorial Lecture Notes* (A. D. Barbour and L. H. Y. Chen, eds.) 183–221. Singapore: World Scientific. MR2235451 https://doi.org/10.1142/9789812567680\protect\T1\textunderscore0004

Röllin, A. (2008). Symmetric and centered binomial approximation of sums of locally dependent random variables. *Electronic Journal of Probability* **13**, 756–776. ISSN 1083-6489. MR2399295 https://doi.org/10.1214/EJP.v13-503

Soon, S. Y. T. (1996). Binomial approximation for dependent indicators. *Statistica Sinica* **6**, 703–714. ISSN 1017-0405. MR1410742

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory*, 583–602. MR0402873

Sundt, B. and Jewell, W. S. (1981). Further results on recursive evaluation of compound distributions. *ASTIN Bulletin* **12**, 27–39. MR0632573 https://doi.org/10.1017/S0515036100006802

Upadhye, N. S., Čekanavičius, V. and Vellaisamy, P. (2017). On Stein operators for discrete approximations. *Bernoulli* **23**, 2828–2859. MR3648047 https://doi.org/10.3150/16-BEJ829

Upadhye, N. S. and Kumar, A. N. (2018). Pseudo-binomial approximation to (k_1, k_2)-runs. *Statistics & Probability Letters* **141**, 19–30. ISSN 0167-7152. https://doi.org/10.1016/j.spl.2018.05.016

Vellaisamy, P. (2004). Poisson approximation for (k_1, k_2)-events via the Stein–Chen method. *Journal of Applied Probability* **41**, 1081–1092. MR2122802 https://doi.org/10.1017/s0021900200020842

Wang, X. and Xia, A. (2008). On negative binomial approximation to k-runs. *Journal of Applied Probability* **45**, 456–471. MR2426844 https://doi.org/10.1239/jap/1214950360

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 369–384
https://doi.org/10.1214/22-BJPS530

Consistency of nearest neighbor estimator of density function for m-END samples

Wei Wang and Yi Wu[a]

Applied Mathematics Center, School of Big Data and Artificial Intelligence, Chizhou University, Chizhou, P.R. China, [a]*wuyi8702@163.com*

Abstract. In this paper, we mainly study the consistency of the nearest neighbor estimator of the density function based on m-extended negatively dependent samples. The weak consistency, strong consistency, uniformly strong consistency and the convergence rate are established under some mild conditions. The results obtained in this paper extend and improve some existing ones in the literature.

References

Boente, G. and Fraiman, R. (1988). Consistency of a nonparametric estimate of a density function for dependent variables. *Journal of Multivariate Analysis* **25**, 90–99. MR0935296 https://doi.org/10.1016/0047-259X(88)90154-6

Chai, G. X. (1989). Consistency of nearest neighbor density estimator of stationary processes. *Acta Mathematica Sinica* **32**, 423–432. MR1044399

Hu, T. C., Chiang, C. Y. and Taylor, R. L. (2009). On complete convergence for arrays of rowwise m-negatively associated random variables. *Nonlinear Analysis* **71**, 1075–1081. MR2671899 https://doi.org/10.1016/j.na.2009.01.104

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *The Annals of Statistics* **11**, 286–295. MR0684886 https://doi.org/10.1214/aos/1176346079

Lan, C. F. and Wu, Q. Y. (2014a). Rate of strong consistency of nearest neighbor estimator of density function for END samples. *Journal of Mathematics (PRC)* **3**, 665–671. MR3236698

Lan, C. F. and Wu, Q. Y. (2014b). Uniform strong consistency rate of nearest neighbor estimator of density function for END samples. *Journal of Jilin University (Science Edition)* **32**, 631–634. MR3236698

Lehmann, E. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics* **37**, 1137–1153. MR0202228 https://doi.org/10.1214/aoms/1177699260

Liu, L. (2009). Precise large deviations for dependent random variables with heavy tails. *Statistics & Probability Letters* **79**, 1290–1298. MR2519013 https://doi.org/10.1016/j.spl.2009.02.001

Liu, Y. and Zhang, Y. (2010). The consistency and asymptotic normality of nearest neighbor density estimator under φ-mixing condition. *Acta Mathematica Sinica* **30**, 733–738. MR2675782 https://doi.org/10.1016/S0252-9602(10)60074-4

Loftsgarden, D. O. and Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics* **36**, 1049–1051. MR0176567 https://doi.org/10.1214/aoms/1177700079

Shen, A. T. (2016). Complete convergence for weighted sums of END random variables and its application to nonparametric regression models. *Journal of Nonparametric Statistics* **28**, 702–715. MR3555453 https://doi.org/10.1080/10485252.2016.1225050

Wang, X. J. and Hu, H. S. (2015). The consistency of the nearest neighbor estimator of the density function based on WOD samples. *Journal of Mathematical Analysis and Applications* **429**, 497–512. MR3339087 https://doi.org/10.1016/j.jmaa.2015.04.016

Wang, X. J., Wu, Y. and Hu, S. H. (2016). Exponential probability inequality for m-END random variables and its applications. *Metrika* **79**, 127–147. MR3451373 https://doi.org/10.1007/s00184-015-0547-7

Wang, X. J., Zheng, L. L., Xu, C. and Hu, S. H. (2015). Complete consistency for the estimator of nonparametric regression models based on extended negatively dependent errors. *Statistics* **49**, 396–407. MR3325366 https://doi.org/10.1080/02331888.2014.888431

WU, Y. and Wang, X. J. (2019). On consistency of the nearest neighbor estimator of the density function and its applications. *Acta Mathematica Sinica* **35**, 703–720. MR3943510 https://doi.org/10.1007/s10114-019-8099-9

Key words and phrases. Nearest neighbor estimator, m-extended negatively dependent samples, weak consistency, strong consistency.

Wu, Y. and Wang, X. J. (2021). Strong laws for weighted sums of m-extended negatively dependent random variables and its applications. *Journal of Mathematical Analysis and Applications* **494**, 124566. MR4153252 https://doi.org/10.1016/j.jmaa.2020.124566

Wu, Y. F., Ordonez, M. C. and Volodin, A. (2014). Complete convergence and complete moment convergence for arrays of rowwise end random variables. *Glasnik Matematički* **49**, 449–468. MR3287069 https://doi.org/10.3336/gm.49.2.16

Xu, W. F., Wu, Y., Zhang, R., Jiang, H. L. and Wang, X. J. (2018). The mean consistency of the weighted estimator in the fixed design regression models based on m-END errors. *Journal of Mathematical Inequalities* **12**, 765–775. MR3857361 https://doi.org/10.7153/jmi-2018-12-58

Yang, S. C. (2003). Consistency of nearest neighbor estimator of density function for negative associated samples. *Acta Mathematicae Applicatae Sinica* **26**, 385–395. MR2022206

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 385–407
https://doi.org/10.1214/22-BJPS531

Limit theorems for quasi-arithmetic means of random variables with applications to point estimations for the Cauchy distribution

Yuichi Akaoka[1,a], **Kazuki Okamura**[2,c] **and Yoshiki Otobe**[1,b]

[1]*Shinshu University, 3-1-1, Asahi, Matsumoto, Nagano, Japan,* [a]*18ss101b@gmail.com,* [b]*otobe@math.shinshu-u.ac.jp*

[2]*Shizuoka University, 836, Ohya, Suruga-ku, Shizuoka, Japan,* [c]*okamura.kazuki@shizuoka.ac.jp*

Abstract. We establish some limit theorems for quasi-arithmetic means of random variables. This class of means contains the arithmetic, geometric and harmonic means. Our feature is that the generators of quasi-arithmetic means are allowed to be complex-valued, which makes considerations for quasi-arithmetic means of random variables which could take negative values possible. Our motivation for the limit theorems is finding simple estimators of the parameters of the Cauchy distribution. By applying the limit theorems, we obtain some closed-form unbiased strongly-consistent estimators for the joint of the location and scale parameters of the Cauchy distribution, which are easy to compute and analyze.

References

Aczél, J. (1948). On mean values. *Bulletin of the American Mathematical Society* **54**, 392–400. MR0024482 https://doi.org/10.1090/S0002-9904-1948-09016-4

Akaoka, Y. (2020). Parameter estimation using complex valued moments for Cauchy distributions. Master's thesis, Department of mathematics, Shinshu University.

Akaoka, Y., Okamura, K. and Otobe, Y. (2021). Confidence disc for Cauchy distributions. Preprint.

Akaoka, Y., Okamura, K. and Otobe, Y. (2022). Bahadur efficiency of the maximum likelihood estimator and one-step estimator for quasi-arithmetic means of the Cauchy distribution. *Annals of the Institute of Statistical Mathematics*. https://doi.org/10.1007/s10463-021-00818-y

Arslan, O. and Kent, J. T. (1998). A note on the maximum likelihood estimators for the location and scatter parameters of a multivariate Cauchy distribution. *Communications in Statistics Theory and Methods* **27**, 3007–3014. MR1659367 https://doi.org/10.1080/03610929808832269

Auderset, C., Mazza, C. and Ruh, E. A. (2005). Angular Gaussian and Cauchy estimation. *Journal of Multivariate Analysis* **93**, 180–197. MR2119770 https://doi.org/10.1016/j.jmva.2004.01.007

Bai, Z. D. and Fu, J. C. (1987). On the maximum-likelihood estimator for the location parameter of a Cauchy distribution. *Canadian Journal of Statistics* **15**, 137–146. MR0905141 https://doi.org/10.2307/3315202

Balmer, D. W., Boulton, M. and Sack, R. A. (1974). Optimal solutions in parameter estimation problems for the Cauchy distribution. *Journal of the American Statistical Association* **69**, 238–242. MR0375583 https://doi.org/10.1080/01621459.1974.10480162

Barczy, M. and Burai, P. (2022). Limit theorems for Bajraktarević and Cauchy quotient means of independent identically distributed random variables. *Aequationes mathematicae* **96** 279–305. MR4405550 https://doi.org/10.1007/s00010-021-00813-x

Barnett, V. D. (1966). Order statistics estimators of the location of the Cauchy distribution. *Journal of the American Statistical Association* **61**, 1205–1218. MR0205363 https://doi.org/10.1080/01621459.1966.10482205

Besbeas, P. and Morgan, B. J. T. (2001). Integrated squared error estimation of Cauchy parameters. *Statistics & Probability Letters* **55**, 397–401. MR1877644 https://doi.org/10.1016/s0167-7152(01)00153-5

Bhattacharya, R. and Patrangenaru, V. (2002). Nonparametic estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference* **108**, 23–35. MR1947389 https://doi.org/10.1016/S0378-3758(02)00268-9

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics* **31**, 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

Key words and phrases. Point estimation, Cauchy distribution, quasi-arithmetic mean.

Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *The Annals of Statistics* **33**, 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

Bloch, D. (1966). A note on the estimation of the location parameter of the Cauchy distribution. *Journal of the American Statistical Association* **61**, 852–855. MR0205366 https://doi.org/10.1080/01621459.1966.10480912

Boos, D. D. (1981). Minimum distance estimators for location and goodness of fit. 76 663–670. MR0629752

Cane, G. J. (1974). Linear estimation of parameters of the Cauchy distribution based on sample quantiles. *Journal of the American Statistical Association* **69**, 243–245. MR0400508 https://doi.org/10.1080/01621459.1974.10480163

Chan, L. K. (1970). Linear estimation of the location and scale parameters of the Cauchy distribution based on sample quantiles. *Journal of the American Statistical Association* **65**, 851–859. MR0400508

Cohen Freue, G. V. (2007). The Pitman estimator of the Cauchy location parameter. *Journal of Statistical Planning and Inference* **137**, 1900–1913. MR2323872 https://doi.org/10.1016/j.jspi.2006.05.002

Copas, J. B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62**, 701–704. MR0388627 https://doi.org/10.1093/biomet/62.3.701

de Carvalho, M. (2016). Mean, what do you mean? *American Statistician* **70**, 270–274. MR3535513 https://doi.org/10.1080/00031305.2016.1148632

de Finetti, B. (1931). Sul concetto di media. *Gionale dell'Instituto Italiano degli Attuarii* **2**, 369–396.

Evard, J. C. and Jafari, F. (1992). A complex Rolle's theorem. *The American Mathematical Monthly* **99**, 858–861. MR1191706 https://doi.org/10.2307/2324123

Ferguson, T. S. (1962). A representation of the symmetric bivariate Cauchy distribution. *The Annals of Mathematical Statistics* **33**, 1256–1266. MR0143281 https://doi.org/10.1214/aoms/1177704357

Ferguson, T. S. (1978). Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4. *Journal of the American Statistical Association* **73**, 211–213. MR0686406 https://doi.org/10.1080/01621459.1978.10480031

Gabrielsen, G. (1982). On the unimodality of the likelihood for the Cauchy distribution: Some comments. *Biometrika* **69**, 677–678. MR0695217 https://doi.org/10.1093/biomet/69.3.677

Gürtler, N. and Henze, N. (2000). Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Annals of the Institute of Statistical Mathematics* **52**, 267–286. MR1763563 https://doi.org/10.1023/a:1004113805623

Haas, G., Bain, L. and Antle, C. (1970). Inferences for the Cauchy distribution based on maximum likelihood estimators. *Biometrika* **57**, 403–408. MR2618313 https://doi.org/10.1093/biomet/57.2.403

Hardy, G. H., Littlewood, J. E. and Polya, G. (1952). *Inequalities*, 2nd ed. Cambridge: Cambridge University Press. MR0046395

Higgins, J. J. and Tichenor, D. M. (1977). Window estimates of location and scale with application to the Cauchy distribution. *Applied Mathematics and Computation* **3**, 113–126. MR0431495 https://doi.org/10.1016/0096-3003(77)90024-8

Higgins, J. J. and Tichenor, D. M. (1978). Efficiencies of window estimates of parameters of the Cauchy distribution. *Applied Mathematics and Computation* **4**, 157–165.

Hinkley, D. V. (1978). Likelihood inference about location and scale parameters. *Biometrika* **65**, 253–261. MR0625372 https://doi.org/10.2307/3315228

Howlader, H. A. and Weiss, G. (1988). On Bayesian estimation of the Cauchy parameters. Sankhyā: The Indian (1960–2002). *Journal of Statistics, Series B* **50**, 350–361. MR1065314 https://doi.org/10.2307/25052554

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol. 1*, 2nd ed. New York: Wiley. MR1326603

Kolmogorov, A. N. (1930). Sur la notion de la moyenne. *Atti Della Accademia Nazionale dei Lincei* **12**, 388–391.

Kravchuk, O. Y. (2005). Rank test of location optimal for hyperbolic secant distribution. *Communications in Statistics Theory and Methods* **34**, 1617–1630. MR2168514 https://doi.org/10.1081/sta-200063236

Kravchuk, O. Y. and Pollett, P. K. (2012). Hodges–Lehmann scale estimator for Cauchy distribution. *Communications in Statistics Theory and Methods* **41**, 3621–3632. MR2967869 https://doi.org/10.1080/03610926.2011.563016

Mardia, K. V., Southworth, H. R. and Taylor, C. C. (1999). On bias in maximum likelihood estimators. 76 31–39. MR1673338 https://doi.org/10.1016/S0378-3758(98)00176-1

Matsui, M. (2020). Asymptotics of maximum likelihood estimation for stable law with continuous parameterization. *Communications in Statistics Theory and Methods* **50**, 3695–3712. MR4282606 https://doi.org/10.1080/03610926.2019.1710199

Matsui, M. and Takemura, A. (2005). Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Annals of the Institute of Statistical Mathematics* **57**, 183–199. MR2165616 https://doi.org/10.1007/bf02506887

McCullagh, P. (1992). Conditional inference and Cauchy models. *Biometrika* **79**, 247–259. MR1185127 https://doi.org/10.1093/biomet/79.2.247

McCullagh, P. (1993). On the distribution of the Cauchy maximum-likelihood estimator. *Proceedings of the Royal Society of London Series A* **440**, 475–479. MR1232841 https://doi.org/10.1098/rspa.1993.0028

McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation. *The Annals of Statistics* **24**, 787–808. MR1394988 https://doi.org/10.1214/aos/1032894465

Nagumo, M. (1930). Über eine klasse der mittelwerte. *Japanese Journal of Mathematics* **7**, 71–79.

Ogawa, J. (1962a). *Estimation of the Location and Scale Parameters by Sample Quantiles (for Large Samples)*, 47–55. New York: John Wiley.

Ogawa, J. (1962b). Distribution and moments of order statistics, 2, 11–19. Wiley publications in statistics.

Okamura, K. and Otobe, Y. (2021). Characterizations of the maximum likelihood estimator of the Cauchy distribution. *Lobachevskii Journal of Mathematics.* to appear.

Onen, B. H., Dietz, D. C., Yen, V. C. and Moore, A. H. (2001). Goodness-of-fit tests for the Cauchy distribution. *Computational Statistics* **16**, 97–107. MR1854194 https://doi.org/10.1007/s001800100053

Pakes, A. G. (1999). On the convergence of moments of geometric and harmonic means. *Statistica Neerlandica* **53**, 96–110. MR1705350 https://doi.org/10.1111/1467-9574.00100

Press, S. J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association* **67**, 842–846. MR0362666 https://doi.org/10.1080/01621459.1972.10481302

Reeds, J. A. (1985). Asymptotic number of roots of Cauchy location likelihood equations. *The Annals of Statistics* **13**, 775–784. MR0790572 https://doi.org/10.1214/aos/1176349554

Rothenberg, T. J., Fisher, F. M. and Tilanus, C. B. (1964). A note on estimation from a Cauchy sample. *Journal of the American Statistical Association* **59**, 460–463. MR0166872 https://doi.org/10.1080/01621459.1964.10482170

Rublik, F. (2001). A quantile goodness-of-fit test for Cauchy distribution, based on extreme order statistics. *Applications of Mathematics* **46**, 339–351. MR1925192 https://doi.org/10.1023/a:1013704326683

Saleh, A. K. M. E., Hassanein, K. M. and Brown, E. F. (1985). Optimum spacings for the joint estimation and tests of hypothesis of location and scale parameters of the Cauchy distribution. *Communications in Statistics Theory and Methods* **14**, 247–254. MR0788797 https://doi.org/10.1080/03610928508828908

Vaughan, D. C. (1992). On the Tiku-Suresh method of estimation. *Communications in Statistics Theory and Methods* **21**, 451–469. MR1158570 https://doi.org/10.1080/03610929208830788

Zhang, J. (2009). A highly efficient L-estimator for the location parameter of the Cauchy distribution. *Computational Statistics* **25**, 97–105. MR2586726 https://doi.org/10.1007/s00180-009-0163-y

Zolotarev, V. M. (1986). *One-Dimensional Stable Distributions*. Providence: American Mathematical Society. MR0854867 https://doi.org/10.1090/mmono/065

Brazilian Journal of Probability and Statistics
2022, Vol. 36, No. 2, 408–419
https://doi.org/10.1214/22-BJPS532
This research was funded, in whole or in part, by [Slovak Research and Development Agency; Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, APVV-18-0052 and VEGA 1/0006/19].

A heteroscedasticity diagnostic of a regression analysis with copula dependent random variables

Ayyub Sheikhi[1,a], Fereshteh Arad[1,b] and Radko Mesiar[2,c]

[1]*Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran,* [a]*sheikhy.a@uk.ac.ir,* [b]*fereshteh_arad@math.uk.ac.ir*
[2]*Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, STU Bratislava, Slovakia,* [c]*radko.mesiar@stuba.sk*

Abstract. One of the most important assumptions in multiple regression analysis is the independence of the explanatory variables, however, this assumption is violated in several situations. In this work, we investigate regression equations when this independence does not hold and the explanatory variables are connected by many of elliptical copulas. We apply the proposed regression equation to study its heteroscedasticity diagnostic and using simulated data we also assess our regression model. A cross-validation procedure is carried out to ensure the unbiasedness of the results. Also, a real data analysis is presented as an application.

References

Acar, E. F., Azimaee, P. and Hoque, Md. E. (2019). Predictive assessment of copula models. *Canadian Journal of Statistics* **47**, 8–26. MR3919892 https://doi.org/10.1002/cjs.11468

Alqawba, M., Diawara, N. and Kim, J. (2019). Copula directional dependence of discrete time series marginals. *Communications in Statistics-Simulation and Computation*, 1–18. MR4336363 https://doi.org/10.1080/03610918.2019.1630434

Ando, T., Konishi, S. and Imoto, S. (2008). Nonlinear regression modeling via regularized radial basis function networks. *Journal of Statistical Planning and Inference* **138**, 3616–3633. MR2450101 https://doi.org/10.1016/j.jspi.2005.07.014

Awad, M. and Khanna, R. (2015). Support vector regression. *Efficient learning machines*, 67–80.

Bates, D. M. and Watts, D. G. (2007). *Nonlinear Regression Analysis and Its Applications, Vol. 2*. New York: Wiley. MR1060528 https://doi.org/10.1002/9780470316757

Bennafla, D., Bouchentouf, A., Rabhi, A. and Sabri, Kh. (2016). On the recursive estimation using copula function in the regression model. *New Trends in Mathematical Sciences* **4**, 25. MR3455640 https://doi.org/10.20852/ntmsci.2016115601

Bentler, P. (1985). A new look at the statistical identification model. *IEEE Transactions on Automatic Control* **19**, 716–723.

Chang, B. and Joe, H. (2019). Prediction based on conditional distributions of vine copulas symmetrical linear models. *Computational Statistics & Data Analysis* **139**, 45–63. MR3952615 https://doi.org/10.1016/j.csda.2019.04.015

Crane, G. J. and van der Hoek, J. (2008). Conditional expectation formulae for copulas. *Australian & New Zealand Journal of Statistics* **50**, 53–67. MR2414655 https://doi.org/10.1111/j.1467-842X.2007.00499.x

Cysneiros, F., Cordeiro, G. and Cysneiros, A. (2010). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation* **80**, 451–461. MR2604169 https://doi.org/10.1080/00949650802706420

Cysneiros, F. J. A., Paula, G. A. and Galea, M. (2007). Heteroscedastic symmetrical linear models. *Statistics & Probability Letters* **70**, 1084–1090. MR2395064 https://doi.org/10.1016/j.spl.2007.01.012

Frahm, G., Junker, M. and Szimayer, A. (2003). Elliptical copulas: Applicability and limitations. *Statistics & Probability Letters* **63**, 275–286. MR1986327 https://doi.org/10.1016/S0167-7152(03)00092-0

Key words and phrases. Gaussian copula, t-copula, regression, heteroscedasticity.

Hoang, Q., Khandelwal, P. and Ghosh, S. (2019). Robust predictive model using copulas. *Data-Enabled Discovery and Applications* **3**, 8.

Kersting, K., Plagemann, C., Pfaff, P., Burgard, W. L., Krzyżak, A. and Yuille, A. (2007) *Proceedings of the 24th International Conference on Machine Learning*, 393–400.

Kim, Li, S. Y. and Spiegelman, D. (2016). A semiparametric copula method for Cox models with covariate measurement error. *Lifetime Data Analysis* **22**, 16. MR3447180 https://doi.org/10.1007/s10985-014-9315-7

Kole, E., Koedijk, K. and Verbeek, M. (2007). Selecting copulas for risk management. *Journal of Banking & Finance* **31**, 2405–2423.

Konishi, S. (2014). *Introduction to Multivariate Analysis: Linear and Nonlinear Modeling*. Boca Raton: CRC Press. MR3222571

Kumar, P. and Shoukri, M. M. (2007). Copula based prediction models: An application to an aortic regurgitation study. *BMC Medical Research Methodology* **7**, 21.

Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A. E. and Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical Engineering Online* **6**, 23.

Marsh, L. C. and Cormier, D. R. (2001). Spline regression models. *Sage.*

Masarotto, G., Varin, C., et al (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics* **6**, 1517–1549. MR2988457 https://doi.org/10.1214/12-EJS721

Mesiar, R., Sheikhi, A. and Komorníková, M. (2019). Random noise and perturbation of copulas. *Kybernetika* **55**, 422–434. MR4014595 https://doi.org/10.14736/kyb-2019-2-0422

Noh, H., Ghouch, El, A. and Bouezmarni, T. P. (2013). Copula-based regression estimation and inference. *IEEE Transactions on Automatic Control* **108**, 676–688. MR3174651 https://doi.org/10.1080/01621459.2013.783842

Pitt, M., Chan, D. and Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika* **93**, 537–554. MR2261441 https://doi.org/10.1093/biomet/93.3.537

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, Th., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M., et al. (2015). Package 'VineCopula'. package version 2.

Seber, G. A. F. and Wild, Ch. J. (2003). *Nonlinear Regression, Vol. 62*, 63. Hoboken. New Jersey: John Wiley & Sons. MR0986070 https://doi.org/10.1002/0471725315

Sheikhi, A. and Mesiar, R. (2020). Copula-based measurement error models. *Iranian Journal of Fuzzy Systems* **17**, 29–38. MR4155893

Wang, C. and Neal, R. M. (2012). Gaussian process regression with heteroscedastic or non-Gaussian residuals. arXiv preprint. Available at 1212.6246. MR3295214

Xu, L., Krzyżak, A. and Yuille, A. (1994). On radial basis function nets and kernel regression: Statistical consistency, convergence rates, and receptive field size. *Neural Networks* **7**, 609–628.

AIMS AND SCOPE

The Brazilian Journal of Probability and Statistics aims to publish high quality research papers in applied probability, applied statistics, computational statistics, mathematical statistics, probability theory and stochastic processes.

More specifically, the following types of contributions will be considered:

(i) Original articles dealing with methodological developments, comparison of competing techniques or their computational aspects;
(ii) Original articles developing theoretical results;
(iii) Articles that contain novel applications of existing methodologies to practical problems. For these papers the focus is in the importance and originality of the applied problem, as well as, applications of the best available methodologies to solve it.
(iv) Survey articles containing a thorough coverage of topics of broad interest to probability and statistics. The journal will occasionally publish book reviews, invited papers and essays on the teaching of statistics.

GENERAL INFORMATION

Journal Homepage: http://www.imstat.org/journals-and-publications/brazilian-journal-of-probability-and-statistics/

Submissions: Manuscripts for *Brazilian Journal of Probability and Statistics* should be submitted online. Authors may access the Electronic Journals Management System (EJMS) at http://www.e-publications.org/ims/submission.

Preparation of Manuscripts: http://www.imstat.org/journals-and-publications/brazilian-journal-of-probability-and-statistics/brazilian-journal-of-probability-and-statistics-manuscript-submission/

Correspondence. Mail concerning membership, subscriptions, nonreceipt claims, copyright permissions, advertising or back issues should be sent to the IMS Dues and Subscription Office, PO Box 729, Middletown, Maryland 21769, USA. Mail concerning submissions or editorial content should be sent to the Editor at nancy@ime.unicamp.br. Mail concerning the production of this journal should be sent to: Geri Mattson at bjps@mattsonpublishing.com.

Individual and Organizational Memberships:
http://www.imstat.org/individual-membership/

Individual and General Subscriptions:
http://www.imstat.org/orders/

Electronic Access to IMS Journals:
http://www.imstat.org/journals-and-publications/electronic-access/

The *Brazilian Journal of Probability and Statistics* is an IMS supported journal:
http://www.imstat.org/journals-and-publications/ims-supported-journals/