Contents

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

Brazilian Journal of Probability and Statistics

Volume 37 • Number 3 • September 2023

Brazilian Journal of Probability and Statistics

Volume 37 • Number 3 • September 2023

www.imstat.org/bjps

ISSN 0103-0752 (Print) ISSN 2317-6199 (Online), Volume 37, Number 3, September 2023. 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
2023, Vol. 37, No. 3, 463–492
https://doi.org/10.1214/23-BJPS577

Tuning parameter selection in fused lasso signal approximator with false discovery rate control

Won Son[1,a], **Johan Lim**[2,b] **and Donghyeon Yu**[3,c]

[1]*Department of Statistics, Dankook University, Gyeonggi-do 16890, Korea,* [a]*son.won@dankook.ac.kr*
[2]*Department of Statistics, Seoul National University, Seoul 08826, Korea,* [b]*johanlim@stats.snu.ac.kr*
[3]*Department of Statistics, Inha University, Incheon 22212, Korea,* [c]*dyu@inha.ac.kr*

Abstract. The fused lasso signal approximator (FLSA) obtains sparse and blocky estimates of the piecewise constant mean model with two tuning parameters for the total variation (TV)-norm and ℓ_1-norm penalties. The FLSA can be divided into the fusion procedure for finding block structures and the soft-thresholding procedure for identifying non-zero block signals. In this paper, we first prove that Bayesian information criterion-type criteria guarantee that the FLSA obtains the minimally over-fitted block estimates. Second, we propose a new procedure to select the soft-thresholding level that controls the false discovery rate of the estimated signals for identifying non-zero signals under the aimed level based on the preliminary test statistics. We show that the soft-thresholded fusion estimators improve the preliminary test statistics regarding false discovery rates. We apply the FLSA with the proposed selection procedure to the COVID-19 pandemic dataset in Korea to identify the change points.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**, 289–300. MR1325392

Braun, J. V., Braun, R. and Müller, H. G. (2000). Multiple changepoint fitting via quasi-likelihood with application to DNA sequence segmentation. *Biometrika* **87**, 301–314. MR1782480 https://doi.org/10.1093/biomet/87.2.301

Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**, 759–771. MR2443189 https://doi.org/10.1093/biomet/asn034

Efron, B. (2004). Large-scale simultaneous hypothesis testing. *Journal of the American Statistical Association* **99**, 96–104. MR2054289 https://doi.org/10.1198/016214504000000089

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

Friedman, J., Hastie, T., Höfling, H. and Tibshirani, R. (2007). Pathwise coordinate optimization. *Annals of Applied Statistics* **1**, 302–332. MR2415737 https://doi.org/10.1214/07-AOAS131

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics* **42**, 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

Guo, J., James, G., Levina, E., Michailidis, G. and Zhu, J. (2010). Principal component analysis with sparse fused loadings. *Journal of Computational and Graphical Statistics* **19**, 930–946. MR2791262 https://doi.org/10.1198/jcgs.2010.08127

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point esetimation with a total variation penalty. *Journal of Amerian Statistical Association* **105**, 1480–1493. MR2796565 https://doi.org/10.1198/jasa.2010.tm09181

Hoefling, H. (2010). A path algorithm for the fused lasso signal approximator. *Journal of Computational and Graphical Statistics* **19**, 984–1006. MR2791265 https://doi.org/10.1198/jcgs.2010.09208

Hütter, J. C. and Rigollet, P. (2016). Optimal rates for total variation denoising. *Proceedings of Machine Learning Research* **49**, 1115–1146.

Key words and phrases. False discovery rate, fused lasso signal approximator, generalized information criteria, piecewise constant mean model, tuning parameter selection.

Lee, J. and Chen, J. (2019). A penalized regression approach for DNA copy number study using the sequencing data. *Statistical Applications in Genetics and Molecular Biology* **18**, 20180001. MR3987905 https://doi.org/10.1515/sagmb-2018-0001

Lee, J. and Chen, J. (2020). A modified information criterion for tuning parameter selection in 1d fused lasso for inference on multiple change points. *Journal of Statistical Computation and Simulation* **90**, 1496–1519. MR4090365 https://doi.org/10.1080/00949655.2020.1732379

Lin, K., Sharpnack, J., Rinaldo, A. and Tibshirani, R. J. (2016). Approximate recovery in changepoint problems, from ℓ_2 estimation error rates. arXiv preprint. Available at arXiv:1606.06746.

Lin, K., Sharpnack, J., Rinaldo, A. and Tibshirani, R. J. (2017). A sharp error analysis for the fused lasso, with application to approximate changepoint screening. In *Advances in Neural Information Processing Systems*, 6884–6893.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *The Annals of Statistics* **25**, 387–413. MR1429931 https://doi.org/10.1214/aos/1034276635

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics* **12**, 758–765. MR0740928 https://doi.org/10.1214/aos/1176346522

Niu, Y. S., Hao, N. and Zhang, H. (2016). Multiple change-point detection: A selective overview. *Statistical Science* **31**, 611–623. MR3598742 https://doi.org/10.1214/16-STS587

Olshen, A. B., Venkatraman, E., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics* **5**, 557–572.

Pan, J. and Chen, J. (2006). Application of modified information criterion to multiple change point problems. *Journal of Multivariate Analysis* **97**, 2221–2241. MR2301636 https://doi.org/10.1016/j.jmva.2006.05.009

Qian, J. and Jia, J. (2016). On stepwise pattern recovery of the fused lasso. *Computational Statistics & Data Analysis* **94**, 221–237. MR3412821 https://doi.org/10.1016/j.csda.2015.08.013

Rinaldo, A. (2009). Properties and refinements of the fused lasso. *The Annals of Statistics* **37**, 2922–2952. MR2541451 https://doi.org/10.1214/08-AOS665

Rinaldo, A. (2014). Corrections to properties and refinements of the fused lasso. available at. https://www.stat.cmu.edu/~arinaldo/Fused_Correction.pdf. MR2541451 https://doi.org/10.1214/08-AOS665

Rojas, C. R. and Wahlberg, B. (2014). On change point detection using the fused lasso method. arXiv preprint. Available at arXiv:1401.5408.

Rojas, C. R. and Wahlberg, B. (2015). How to monitor and mitigate stair-casing in ℓ_1 trend filtering. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3946–3950.

Seong, H., Hyun, H. J., Yun, J. G., Noh, J. Y., Cheong, H. J., Kim, W. J. and Song, J. Y. (2021). Comparison of the second and third waves of the COVID-19 pandemic in South Korea: Importance of early public health intervention. *International Journal of Infectious Diseases* **104**, 742–745.

Son, W. and Lim, J. (2019). Modified path algorithm of fused lasso signal approximator for consistent recovery of change points. *Jounal of Statistical Planning and Inference* **200**, 223–238. MR3907280 https://doi.org/10.1016/j.jspi.2018.10.003

Son, W., Lim, J. and Yu, D. (2022). Path algorithms for fused lasso signal approximator with application to COVID-19 spread in Korea. *International Statistical Review*, Advance online publication. https://doi.org/10.1111/insr.12521

Son, W.,, Lim, J., and Yu, D. (2023). Supplement to "Tuning parameter selection in fused lasso signal approximator with false discovery rate control." https://doi.org/10.1214/23-BJPS577SUPP

Sun, W., Wang, H. J. and Fuentes, M. (2016). Fused adaptive lasso for spatial and temporal quantile function estimation. *Technometrics* **58**, 127–137. MR3463163 https://doi.org/10.1080/00401706.2015.1017115

Tang, L. and Song, P. X. (2016). Fused lasso approach in regression coefficients clustering—learning parameter heterogeneity in data integration. *Journal of Machine Learning Research* **17**, 1–23. MR3543519

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society, Series B* **67**, 91–108. MR2136641 https://doi.org/10.1111/j.1467-9868.2005.00490.x

Tibshirani, R. and Wang, P. (2008). Spatial smoothing and hot sopt detection for CGH data using the fused lasso. *Biostatistics* **9**, 18–29.

Tibshirani, R. J. and Taylor, J. (2011). The solution path of the generalized lasso. *The Annals of Statistics* **39**, 1335–1371. MR2850205 https://doi.org/10.1214/11-AOS878

Wang, F., Madrid, O., Yu, Y. and Rinaldo, A. (2022). Denoising and change point localisation in piecewise-constant high-dimensional regression coefficients. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, Vol. 151*, 4309–4338.

Yao, Y. C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statistics & Probability Letters* **6**, 181–189. MR0919373 https://doi.org/10.1016/0167-7152(88)90118-6

Yao, Y. C. and Au, S. T. (1989). Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics, Series A* **51**, 370–381. MR1175613

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

Zhang, N. R. and Siegmund, D. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63**, 22–32. MR2345571 https://doi.org/10.1111/j.1541-0420.2006.00662.x

Zhang, Y., Li, R. and Tsai, C. L. (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association* **105**, 312–323. MR2656055 https://doi.org/10.1198/jasa.2009.tm08013

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 3, 493–512
https://doi.org/10.1214/23-BJPS578

Variable selection for an improved INAR(1) model with explanatory variables using 2SPCLS

Ye Liu[1,a] **and Dehui Wang**[2,b]

[1]*School of Mathematics, Jilin University, Changchun, China,* [a]*lye20@mails.jlu.edu.cn*
[2]*School of Mathematics and Statistics, Liaoning University, Shenyang, China,* [b]*wangdehui@lnu.edu.cn*

Abstract. In the era of Big Data, a tremendous amount of usable data is produced every day. Count data is an essential component. The first-order integer-valued autoregressive (INAR(1)) model is one of the most effective tools for evaluating count data. In this study, we provide an improved INAR(1) model with explanatory variables. This model can well characterize the type of data where the variance of innovation is influenced by other time-varying factors. We introduce a two-step penalized conditional least squares (2SPCLS) method for unknown parameter estimation and variable selection. This method facilitates the selection of explanatory variables in the model, allowing us to more effectively address a modeling challenge. The asymptotical properties have been thoroughly investigated. This paper demonstrates, via a simulation study, that the 2SPCLS approach can accurately and effectively select the zero parameters. Finally, we perform a real-time series analysis, suggesting that this method can be used to solve problems in real life.

References

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* **8**, 261–275. MR0903755 https://doi.org/10.1111/j.1467-9892.1987.tb00438.x

Alzahrani, N., Neal, P., Spencer, S., Mckinley, T. J. and Touloupou, P. (2018). Model selection for time series of count data. *Computational Statistics & Data Analysis* **122**, 33–44. MR3765813 https://doi.org/10.1016/j.csda.2018.01.002

Barreto-Souza, W., Ndreca, S., Silva, R. B. and Silva, R. W. (2022). Modified Galton–Watson processes with immigration under an alternative offspring mechanism.

Brännäs, K. (1995). Explanatory variables in the AR(1) count data model. *Umea Economic Studies* **381**.

Chen, C. S., Chen, C. and Liu, T. C. (2022). Dust storms increase the risk of age-related aacular degeneration. *International Journal of Environmental Research and Public Health* **19**, 7403.

Ding, X. and Wang, D. (2016). Empirical likelihood inference for INAR(1) model with explanatory variables. *Journal of the Korean Statistical Society* **45**, 623–632. MR3566167 https://doi.org/10.1016/j.jkss.2016.05.004

Efron, B., Hastie, T., Johnstone, I. and Tibshiran, R. (2004). Least angle regression. *The Annals of Statistics* **32**, 407–451. MR2060166 https://doi.org/10.1214/009053604000000067

Enciso-Mora, V., Neal, P. and Rao, T. S. (2009). Integer valued AR processes with explanatory variables. *Sankhya Series B* **71**, 248–263. MR2639306

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

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability* **20**, 822–835. MR0968000 https://doi.org/10.2307/1427362

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction*. New York: Springer. MR0671255

Steutel, F. W. and Harn, K. V. (1978). Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**, 893–899. MR0542141

Tan, S. C. and San Lau, J. P. (2014). Time series clustering: A superior alternative for market basket analysis. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*, 241–248.

Key words and phrases. Variable selection, two-step conditional least squares, time-series analysis, integer-valued autoregressive.

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

Tibshirani, R. (1997). The lasso method for aariable selection in the Cox model. *Statistics in Medicine* **16**, 385–395.

Tweedie, R. L. (1975). Sufficient conditions for regularity, recurrence and ergocidicity of Markov processes. *Stochastic Processes and Their Applications* **3**, 385–403. MR0391269 https://doi.org/10.1017/S0305004100051562

Wang, X. (2020). Variable selection for first-order Poisson integer-valued autoregressive model with covariables. *Australian & New Zealand Journal of Statistics* **62**, 278–295. MR4156543 https://doi.org/10.1111/anzs.12295

Wang, X., Wang, D. and Yang, K. (2020). Integer-valued time series model order shrinkage and selection via penalized quasi-likelihood approach. *Metrika* **84**, 713–750. MR4274636 https://doi.org/10.1007/s00184-020-00799-7

Winahju, W. S. and Mukarromah, A. (2012). Modeling Dengue cases using Poisson INAR. *Procedia Engineering* **50**, 837–847.

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
2023, Vol. 37, No. 3, 513–533
https://doi.org/10.1214/23-BJPS579

Interest rate modeling with generalized Langevin equations

Markus Hess[a]

Department of Mechanical and Process Engineering, University of Kaiserslautern-Landau, Gottlieb-Daimler-Straße, 67663 Kaiserslautern, Germany, [a]*Markus-Hess@gmx.net*

Abstract. In this paper, we present an arithmetic short rate model based on generalized Langevin equations. The innovative feature of the model is that it accounts for memory effects in interest rate markets via the involved Langevin processes. In this setup, we provide a representation for the related zero-coupon bond price and infer its risk-neutral time dynamics. We also deduce the associated forward rate dynamics, the latter being of Heath–Jarrow–Morton type. We further establish a measure change to the risk-adjusted forward measure and propose a market-consistent calibration procedure. We finally derive a pricing formula for a European call option written on the zero-coupon bond by Fourier transform methods.

References

Agram, N. and Øksendal, B. (2015). Malliavin calculus and optimal control of stochastic Volterra equations. *Journal of Optimization Theory and Applications* **167**, 1070–1094. MR3424704 https://doi.org/10.1007/s10957-015-0753-5

Barndorff-Nielsen, O., Benth, F. and Veraart, A. (2013). Modelling energy spot prices by volatility modulated Lévy driven Volterra processes. *Bernoulli* **19**, 803–845. MR3079297 https://doi.org/10.3150/12-BEJ476

Barrat, J. and Hansen, J. (2003). Brownian motion, diffusion and the Langevin equation. In *Basic Concepts for Simple and Complex Liquids*, 228–256. Cambridge: Cambridge University Press.

Benth, F., Khedher, A. and Vanmaele, M. (2020). Pricing of commodity derivatives on processes with memory. *Risks* **8**, 8010008.

Björk, T., Kabanov, Y. and Runggaldier, W. (1997). Bond market structure in the presence of marked point processes. *Mathematical Finance* **7**, 211–239. MR1446647 https://doi.org/10.1111/1467-9965.00031

Black, F. and Karasinski, P. (1991). Bond and option pricing when short rates are lognormal. *Financial Analysts Journal* **47**, 52–59.

Brace, A., Gatarek, D. and Musiela, M. (1997). The market model of interest rate dynamics. *Mathematical Finance* **7**, 127–155. MR1446645 https://doi.org/10.1111/1467-9965.00028

Brigo, D. and Mercurio, F. (2001). *Interest Rate Models—Theory and Practice*. Berlin, Heidelberg: Springer. MR1846525 https://doi.org/10.1007/978-3-662-04553-4

Carr, P. and Madan, D. (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance* **2**, 61–73.

Coffey, W. and Kalmykov, Y. (2012). *The Langevin Equation—with Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering, Vol. 27*, 3rd ed. Singapore: World Scientific. MR3236656 https://doi.org/10.1142/8195

Comte, F. and Renault, E. (1996). Long memory continuous time models. *Journal of Econometrics* **73**, 101–149. MR1410003 https://doi.org/10.1016/0304-4076(95)01735-6

Cox, J., Ingersoll, J. and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–407. MR0785475 https://doi.org/10.2307/1911242

Dothan, L. (1978). On the term structure of interest rates. *Journal of Financial Economics* **6**, 59–69.

Farias, R., Ramos, R. and Da Silva, L. (2009). Stochastic Langevin equations: Markovian and non-Markovian dynamics. *Physical Review E* **80**, 031143.

Filipovic, D. (2009). *Term-Structure Models—a Graduate Course*. Berlin, Heidelberg: Springer. MR2553163 https://doi.org/10.1007/978-3-540-68015-4

Fox, R. (1977). The generalized Langevin equation with Gaussian fluctuations. *Journal of Mathematical Physics* **18**, 2331–2335. MR0456151 https://doi.org/10.1063/1.523242

Key words and phrases. Short rate, forward rate, zero-coupon bond price, option pricing, forward measure, market-consistent calibration, generalized/retarded Langevin equation.

Heath, D., Jarrow, R. and Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica* **60**, 77–105.

Hess, M. (2020). A pure-jump mean-reverting short rate model. *Modern Stochastics: Theory and Applications* **7**, 113–134. MR4120611 https://doi.org/10.15559/20-vmsta152

Hess, M. (2023). The stochastic Leibniz formula for Volterra integrals under enlarged filtrations. *Stochastic Models*, 1–28. https://doi.org/10.1080/15326349.2023.2173233

Hull, J. and White, A. (1990). Pricing interest-rate-derivative securities. *The Review of Financial Studies* **3**, 573–592.

Jeanblanc, M., Yor, M. and Chesney, M. (2009). *Mathematical Methods for Financial Markets*. London: Springer. MR2568861 https://doi.org/10.1007/978-1-84628-737-4

Küchler, U. and Mensch, B. (1992). Langevin's stochastic differential equation extended by a time-delayed term. *Stochastics An International Journal of Probability and Stochastic Processes* **40**, 23–42. MR1275126 https://doi.org/10.1080/17442509208833780

Lipster, R. and Shiryaev, A. (1984). *Statistics of Random Processes I, General Theory*, 2nd ed. Berlin: Springer. MR1800857

Morino, L. and Runggaldier, W. (2014). On multicurve models for the term structure. In *Nonlinear Economic Dynamics and Financial Modeling, Vol. 275*. Cham: Springer. MR3329961

Pavliotis, G. (2014). The Langevin equation. In *Stochastic Processes and Applications*. *Texts in Applied Mathematics* **60**. New York: Springer. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7

Pelsser, A. (2005). *Efficient Methods for Valuing Interest Rate Derivatives*, 3rd ed. London: Springer. MR1779538 https://doi.org/10.1007/978-1-4471-3888-4

Takahashi, M. (1996). Non-ideal Brownian motion, generalized Langevin equation and its application to the security market. *Financial Engineering and the Japanese Markets* **3**, 87–119.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5**, 177–188.

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 3, 534–551
https://doi.org/10.1214/23-BJPS580

L-moments of asymmetric generalized distributions obtained through quantile splicing

Brenda Mac'Oduol[1,a], Narayanaswamy Balakrishnan[1,2,d], Paul van Staden[1,b] and Robert King[1,c]

[1]*Department of Statistics, University of Pretoria, Pretoria, South Africa,* [a]*brenda.omachar@up.ac.za,* [b]*paul.vanstaden@up.ac.za,* [c]*robert.king.newcastle@gmail.com*
[2]*Department of Mathematics and Statistics, McMaster University, Ontario, Canada,* [d]*bala@mcmaster.ca*

Abstract. Balakrishnan et al. (*Communications in Statistics Simulation and Computation* **46** (2017) 4082–4097) proposed a skew logistic distribution by making use of the cumulative distribution function (CDF) of the folded logistic distribution. They made use of moments of order statistics from the standard folded logistic distribution to obtain the single and product moments of order statistics from the skew logistic distribution. Subsequently, Mac'Oduol et al. (*Communications in Statistics—Theory and Methods* **49** (2020) 4413–4429) proposed quantile splicing for the construction of two-piece distributions using quantile functions of symmetric distributions as building blocks. This paper presents the derivation of a general formula for the *L*-moments of such two-piece distributions. In addition, quantile splicing and its results are then specialized to the Tukey lambda distribution, and an example is used to illustrate the results developed.

References

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). *A First Course in Order Statistics*. New York: Wiley. MR1178934

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica* **46**, 199–208. MR0877720

Balakrishnan, N., Dai, Q. and Liu, K. (2017). A skew logistic distribution as an alternative to the model of van Staden and King. *Communications in Statistics Simulation and Computation* **46**, 4082–4097. MR3656158 https://doi.org/10.1080/03610918.2015.1096383

Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables*. New York: Wiley. MR0095553

Castillo, E. and Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association* **440**, 1609–1620. MR1615270 https://doi.org/10.2307/2965432

Castillo, E., Hadi, A. S. Balakrishnan, N. and Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Hoaglin, D. C. (1983). Letter values: A set of selected order statistics. In *Understanding Robust and Explanatory Data Analysis* (D. C. Hoaglin, F. Mosteller and J. W. Tukey, eds.) 33–57. New York: Wiley. MR2056572 https://doi.org/10.1214/ss/1076102418

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis, Vol. 1*. New York: Wiley. MR1100924

Gilchrist, W. (2000). *Statistical Modelling with Quantile Functions*. Florida: CRC Press.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B* **52**, 105–124. MR1049304

Kang, L., Damien, P. and Walker, S. (2021). On a transform for modeling skewness. *Brazilian Journal of Probability and Statistics* **35**, 335–350. MR4255159 https://doi.org/10.1214/20-bjps477

Karvanen, J. and Nuutinen, A. (2008). Characterizing the generalized lambda distribution by *L*-moments. *Computational Statistics & Data Analysis* **52**, 1971–1983. MR2418484 https://doi.org/10.1016/j.csda.2007.06.021

Mac'Oduol, B. V., van Staden, P. J. and King, R. A. R. (2020). Asymmetric generalizations of symmetric univariate probability distributions obtained through quantile splicing. *Communications in Statistics—Theory and Methods* **49**, 4413–4429. MR4130872 https://doi.org/10.1080/03610926.2019.1601219

Key words and phrases. Two-piece distribution, folded distribution, *l*-moments, order statistics, tukey lambda distribution.

Tukey, J. W. (1960). The practical relationship between the common transformations of percentages of counts and of amounts. Technical Report 36. New Jersey: Princeton.

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 3, 552–565
https://doi.org/10.1214/23-BJPS581

An enhanced design of nonparametric modified EWMA sign control chart using repetitive sampling

Ambreen Shafqat[1,4,a], Faisal Shahzad[1,b], Muhammad Aslam[2,c] and Rafael Perez Abreu[3,d]

[1]*Department of Statistics and Financial Mathematics, Nanjing University of Science and Technology, Nanjing, China,* [a]*amber_shafqat@yahoo.com,* [b]*faisalshahzad@qq.com*
[2]*Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia,* [c]*aslam_ravian@hotmail.com*
[3]*Department of Probability and Statistics, Centro de Investigación en Matemáticas (CIMAT), Guanajuato, Mexico,* [d]*rabreu@cimat.mx*
[4]*Department of Urology, Roswell Park Cancer Research Institute, Buffalo, USA*

Abstract. The exponential weighted moving average (EWMA) control chart is commonly used in industry to monitor process performance for small changes in objective values. In everyday life, a large amount of data is generated by a technique where a monitoring statistic displays an undefined probability distribution; in this case, nonparametric control charts are used to examine nonconformities from the procedure aim. We aim for an existing nonparametric modified arcsine EWMA (NPMASE) sign control chart in this study. This control chart is accessible based on a single sample; however, the repetitive sampling scheme is less well-known and has received less attention, but it outperforms other sampling schemes. An NPMASE control chart based on repetitive sampling (namely RSMASE) is used here to improve the detectability of small process shifts. The performance of the resulting chart is examined in terms of popular run-length properties such as mean run-length (ARL), median run-length (MDRL), and standard deviation run-length (SDRL). The early RSMASE chart demonstrates the efficiency of shift detection abilities, followed by the NPMASE and nonparametric arcsine EWMA (NPASE) control charts. An actual-life application based on a soft-drink beverage data set for the industrial implementation of the newly designed chart is also explained.

References

Abbasi, S. A. (2012). A new nonparametric EWMA sign control chart. *Expert Systems with Applications* **39**, 8503.

Abbasi, S. A., Miller, A. and Riaz, M. (2013). Nonparametric progressive mean control chart for monitoring process target. *Quality and Reliability Engineering International* **29**, 1069–1080.

Al-Omari, A. I. and Bouza, C. N. (2014). Review of ranked set sampling: Modifications and applications. *Revista Investigacion Operacional* **3**, 215–240. MR3266914

Altukife, F. S. (2003a). A new nonparametric control charts based on the observations exceeding the grand median. *Pakistan Journal of Statistics* **19**, 343–353. MR2009729

Altukife, F. S. (2003b). Nonparametric control charts based on sum of ranks. *Pakistan Journal of Statistics* **19**, 291–300. MR2009724

Amin, R. W., Reynolds, M. R. Jr. and Baker, S. T. (1995). Nonparametric quality control charts based on the sign statistic. *Communications in Statistics—Theory and Methods* **24**, 1597–1624. MR1349928 https://doi.org/10.1080/03610929508831574

Aslam, M., Ali, R. M., Azam, M., Ahmad, L. and Jun, C. H. (2020). Design of a sign chart using a new EWMA statistic. *Communications in Statistics—Theory and Methods* **49**, 1299–1310. MR4055519 https://doi.org/10.1080/03610926.2018.1563163

Aslam, M., Azam, M. and Jun, C.-H. (2001). A mixed repetitive sampling plan based on process capability index. *Applied Mathematical Modelling* **37**, 10027–10035. MR3125518 https://doi.org/10.1016/j.apm.2013.05.058

Key words and phrases. EWMA, MEWMA, sign statistic, binomial distribution, repetitive sampling.

Aslam, M., Azam, M. and Jun, C.-H. (2014). A new exponentially weighted moving average sign chart using repetitive sampling. *Journal of Process Control* **24**, 1149–1153.

Bakir, S. T. (2004). A distribution-free Shewhart quality control chart based on signed-ranks. *Quality Engineering* **16**, 613–623. MR2256249 https://doi.org/10.1080/03610920500498907

Bakir, S. T. (2006). Distribution free quality control charts based in sign rank like statistics. *Communications in Statistics—Theory and Methods* **35**, 743–757. MR2256249 https://doi.org/10.1080/03610920500498907

Bakir, S. T. and Reynolds, M. R. Jr. (1979). A nonparametric procedure for process control based on within-group ranking. *Technometrics* **21**, 175–183. MR2627346

Balamurali, S. and Jun, C-H. (2006). Repetitive group sampling procedure for variables inspection. *Journal of Applied Statistics* **33**, 327–338. MR2226541 https://doi.org/10.1080/02664760500446010

Cavalcanti, A. B. and Cordeiro, G. M. (2006). An improved u chart for attributes. *Brazilian Journal of Probability and Statistics* **20**. MR2345489

Chakraborti, S., Van der Lann, P. and Bakir, S. T. (2001). Nonparametric control charts: An overview and some results. *Journal of Quality Technology* **33**, 304–315.

Croasdale, R. (1974). Control charts for a double-sampling scheme based on average production run lengths. *International Journal of Production Research* **12**, 585–592.

Haq, A. (2017). A new nonparametric EWMA control chart for monitoring process variability. *Quality and Reliability Engineering International* **33**, 1499–1512. MR3945946 https://doi.org/10.1080/03610918.2017.1422750

Haq, A., Brown, J. and Moltchanova, E. (2016). New synthetic EWMA and synthetic CUSUM control charts for monitoring the process mean. *Quality and Reliability Engineering International* **32**, 269–290.

Khan, N., Aslam, M. and Jun, C. H. (2017). Design of a control chart using a modified EWMA statistic. *Quality and Reliability Engineering International* **33**, 1095–1104. MR2130897 https://doi.org/10.1002/0471714682

Lee, L. Y., Khoo, M. B. C., Teh, S. Y. and Lee, M. H. (2015). A variable sampling interval synthetic Xbar chart for the process mean. *PLoS ONE* **10**, 0126331.

Liu, L., Tsung, F. and Zhang, J. (2014). Adaptive nonparametric CUSUM scheme for detecting unknown shifts in location. *International Journal of Production Economics* **52**, 1592–1606.

Malela-Majika, J.-C., Adeoti, O. A. and Rapoo, E. (2018). An EWMA control chart based on the Wilcoxon rank-sum statistic using repetitive sampling. *International Journal of Quality and Reliability Management* **35**, 711–728.

Montgomery, D. C. (2004). The DMAIC process. In: *Introduction to Statistical Quality Control*, 6th ed, 45–60. Hoboken, NJ: Wiley.

Page, E. E. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–115. MR0088850 https://doi.org/10.1093/biomet/41.1-2.100

Riaz, M. (2015). A sensitive non-parametric EWMA control chart. *Journal of the Chinese Institute of Engineers* **38**, 208–219.

Riaz, M. and Abbasi, S. A. (2016). Nonparametric double EWMA control chart for process monitoring. *Revista Colombiana de Estadāstica* **39**, 167–184. MR3530174 https://doi.org/10.15446/rce.v39n2.58914

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics* **1**, 239–250.

Shafqat, A., Huang, Z. and Aslam, M. (2021). Design of X-bar control chart based on inverse Rayleigh distribution under repetitive group sampling. *Ain Shams Engineering Journal* **12**, 943–953.

Shafqat, A., Huang, Z., Aslam, M. and Nawaz, M. S. (2020). A nonparametric repetitive sampling DEWMA control chart based on linear prediction. *IEEE Access* **8**, 74977–74990.

Shafqat, A., Hussain, J., Al-Nasser, A. D. and Aslam, M. (2018). Attribute control chart for some popular distributions. *Communications in Statistics—Theory and Methods* **47**, 1978–1988. MR3757724 https://doi.org/10.1080/03610926.2017.1335414

Shafqat, A., Sabir, Q. U., Yang, S. F., Aslam, M., Albassam, M. and Abbas, K. (2023). Monitoring and comparing air and green house gases emissions of various countries. *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 1–24.

Shewhart, W. A. (1931). *Economic Control of Quality Manufactured Product*. New York: Van Nostrand. MR1522840 https://doi.org/10.2307/2301530

Tsai, T.-R., Chiang, Y.-C. and Wu, S.-J. (2004). Conditional maximum likelihood estimation for control charts in the presence of correlation. *Brazilian Journal of Probability and Statistics* **18**, 151–162. MR2347940

Tsai, T.-R. and Wu, S.-J. (2008). Adjusted weighted standard deviation r chart for skewed distributions. *Brazilian Journal of Probability and Statistics* **22**, 9–22. MR2439900

Wald, A. (1943). *Sequential Analysis of Statistical Data: Theory*. Statistical Research Group, Columbia University.

Xie, X. and Qiu, P. (2023). Control charts for dynamic process monitoring with an application to air pollution surveillance. *Annals of Applied Statistics* **17**, 47–66. MR4539021 https://doi.org/10.1214/22-aoas1615

Yang, S., Lin, J. and Cheng, S. W. (2011). A new nonparametric EWMA sign control chart. *Expert Systems with Applications* **38**, 239–6243.

Yang, S.-F. and Arnold, B. C. (2016). A new approach for monitoring process variance. *Journal of Statistical Computation and Simulation* **86**, 2749–2765. MR3515120 https://doi.org/10.1080/00949655.2015.1125901

Yang, S.-F. and Cheng, S. W. (2011). A new non-parametric CUSUM mean chart. *Quality and Reliability Engineering International* **27**, 867–875.

Yue, J. and Liu, L. (2017). Multivariate nonparametric control chart with variable sampling interval. *Applied Mathematical Modelling* **52**, 603–612. MR3716312 https://doi.org/10.1016/j.apm.2017.08.005

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 3, 566–595
https://doi.org/10.1214/23-BJPS582

Inference for a competing risks model with Burr XII distributions under generalized progressive hybrid censoring

Prakash Chandra[1,2,a], **Amulya Kumar Mahto**[3,c] **and Yogesh Mani Tripathi**[2,b]

[1]*Bihar Mausam Sewa Kendra, Sardar Patel Bhawan, Patna-800022 (Bihar), India*
[2]*Department of Mathematics, Indian Institute of Technology Patna, Bihta-801106, India,*
[a]*prakash_2021ma16@iitp.ac.in,* [b]*yogesh@iitp.ac.in*
[3]*Mehta Family School of Data Sciences and Artificial Intelligence, Indian Institute of Technology Guwahati, Assam-781039, India,* [c]*akmahto@iitg.ac.in*

Abstract. This article considers the inference for a competing risks model with a partially observed failure cause when latent failure times follow Burr XII distributions. Inference is obtained under a generalized progressive hybrid censoring. Estimations of unknown parameters under different restrictions are provided using frequentist and Bayesian approaches. Subsequently, interval estimators are also derived. Bayesian estimators are developed for order-restricted parameters and are compared with corresponding likelihood estimators. The case of unrestricted parameters is considered as well. The performance of all estimators is evaluated based on a simulation study, and a real data set is also presented for illustrative purposes.

References

Ahmed, E., Ali Alhussain, Z., Salah, M., Haj Ahmed, H. and Eliwa, M. (2020). Inference of progressively type-II censored competing risks data from Chen distribution with an application. *Journal of Applied Statistics* **47**, 2492–2524. MR4149567 https://doi.org/10.1080/02664763.2020.1815670

Asgharzadeh, A. and Abdi, M. (2012). Confidence intervals for the parameters of the Burr type XII distribution based on records. *International Journal of Statistics and Economics* **8**, 96–104. MR2815499

Balakrishnan, N. (2007). Progressive censoring methodology: An appraisal. *Test* **16**, 211–259. MR2393645 https://doi.org/10.1007/s11749-007-0061-y

Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods, and Applications*. Boston, MA: Birkhäuser. MR1763020 https://doi.org/10.1007/978-1-4612-1334-5

Balakrishnan, N. and Cramer, E. (2014). *The Art of Progressive Censoring*. *Statistics for Industry and Technology*. Boston, MA: Birkhäuser. MR3309531 https://doi.org/10.1007/978-0-8176-4807-7

Balakrishnan, N. and Sandhu, R. (1995). A simple simulational algorithm for generating progressive type-II censored samples. *American Statistician* **49**, 229–230.

Burr, I. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics* **13**, 215–232. MR0006644 https://doi.org/10.1214/aoms/1177731607

Childs, A., Chandrasekar, B., Balakrishnan, N. and Kundu, D. (2003). Exact likelihood inference based on type-I and type-II hybrid censored samples from the exponential distribution. *Annals of the Institute of Statistical Mathematics* **55**, 319–330. MR2001867 https://doi.org/10.1023/A:1026326020003

Cho, Y., Sun, H. and Lee, K. (2015). Exact likelihood inference for an exponential parameter under generalized progressive hybrid censoring scheme. *Statistical Methodology* **23**, 18–34. MR3278799 https://doi.org/10.1016/j.stamet.2014.09.002

Cox, D. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **21**, 411–421. MR0114280

Crowder, M. (2001). *Classical Competing Risks*. Boca Raton: Chapman and Hall.

David, H. and Moeschberger, M. (1978). *The Theory of Competing Risks*. Oxford: Charles Griffin. MR0592960

Dey, S. and Dey, T. (2014). On progressively censored generalized inverted exponential distribution. *Journal of Applied Statistics* **41**, 2557–2576. MR3262679 https://doi.org/10.1080/02664763.2014.922165

Epstein, B. (1954). Truncated life tests in the exponential case. *The Annals of Mathematical Statistics* **25** 555–564. MR0064372 https://doi.org/10.1214/aoms/1177728723

Key words and phrases. Competing risks model, generalized progressive hybrid censoring, maximum likelihood estimate, Bayes estimate, credible interval, order restriction.

Gupta, R. and Kundu, D. (1998). Hybrid censoring schemes with exponential failure distribution. *Communications in Statistics - Theory and Methods* **27**, 3065–3083.

Han, D. and Balakrishnan, N. (2010). Inference for a simple step-stress model with competing risks for failure from the exponential distribution under time constraint. *Computational Statistics & Data Analysis* **54**, 2066–2081. MR2719742 https://doi.org/10.1016/j.csda.2010.03.015

Kayal, T., Tripathi, Y., Rastogi, M. and Asgharzadeh, A. (2017). Inference for Burr XII distribution under type I progressive hybrid censoring. *Communications in Statistics - Simulation and Computation* **46**, 7447–7465. MR3764980 https://doi.org/10.1080/03610918.2016.1241405

Klein, J. and Basu, A. (1981). Weibull accelerated life tests when there are competing causes of failure. *Communications in Statistics - Theory and Methods* **A10**, 2073–2100. MR0629567 https://doi.org/10.1080/03610928108828174

Koley, A. and Kundu, D. (2017). On generalized progressive hybrid censoring in presence of competing risks. *Metrika* **80**, 401–426. MR3635469 https://doi.org/10.1007/s00184-017-0611-6

Kundu, D. (2007). On hybrid censored Weibull distribution. *Journal of Statistical Planning and Inference* **137**, 2127–2142. MR2325420 https://doi.org/10.1016/j.jspi.2006.06.043

Kundu, D. and Joarder, A. (2006). Analysis of type-II progressively hybrid censored data. *Computational Statistics & Data Analysis* **50**, 2509–2528. MR2227317 https://doi.org/10.1016/j.csda.2005.05.002

Kundu, D. and Pradhan, B. (2011). Bayesian analysis of progressively censored competing risks data. *Sankhya Series B* **73**, 276–296. MR2924196 https://doi.org/10.1007/s13571-011-0024-x

Lin, C., Chou, C. and Huang, Y. (2012). Inference for the Weibull distribution with progressive hybrid censoring. *Computational Statistics & Data Analysis* **56**, 451–467. MR2853746 https://doi.org/10.1016/j.csda.2011.09.002

Lin, C. and Huang, Y. (2012). On progressive hybrid censored exponential distribution. *Journal of Statistical Computation and Simulation* **82**, 689–709. MR2916085 https://doi.org/10.1080/00949655.2010.550581

Lodhi, C., Tripathi, Y. M. and Rastogi, M. K. (2021). Estimating the parameters of a truncated normal distribution under progressive type II censoring. *Communications in Statistics - Simulation and Computation* **50**, 2757–2781. MR4315977 https://doi.org/10.1080/03610918.2019.1614619

Lodhi, C., Tripathi, Y. M. and Wang, L. (2021). Inference for a general family of inverted exponentiated distributions with partially observed competing risks under generalized progressive hybrid censoring. *Journal of Statistical Computation and Simulation* **91**, 2503–2526. MR4294198 https://doi.org/10.1080/00949655.2021.1901290

Mahto, A., Tripathi, Y., Sultana, F. and Rastogi, M. K. (2022). Statistical inference for a Gumbel type-II distribution under hybrid censoring. *Journal of Statistical Computation and Simulation* **92**, 2290–2316. MR4445041 https://doi.org/10.1080/00949655.2022.2028282

Mahto, A. K., Dey, S. and Tripathi, Y. M. (2020). Statistical inference on progressive-stress accelerated life testing for the logistic exponential distribution under progressive type-II censoring. *Quality and Reliability Engineering International* **36**, 112–124.

Mondal, S. and Kundu, D. (2019). Point and interval estimation of Weibull parameters based on joint progressively censored data. *Sankhya Series B* **81**, 1–25. MR4029335 https://doi.org/10.1007/s13571-017-0134-1

Mousa, M. and Jaheen, Z. (2002). Statistical inference for the Burr model based on progressively censored data. *Computers & Mathematics with Applications* **43**, 1441–1449. MR1906365 https://doi.org/10.1016/S0898-1221(02)00110-4

Pena, E. and Gupta, A. (1990). Bayes estimation for the Marshall–Olkin exponential distribution. *Journal of the Royal Statistical Society, Series B* **52**, 379–389. MR1064424

Pradhan, B. and Kundu, D. (2013). Inference and optimal censoring schemes for progressively censored Birnbaum–Saunders distribution. *Journal of Statistical Planning and Inference* **143**, 1098–1108. MR3029236 https://doi.org/10.1016/j.jspi.2012.11.007

Rastogi, M. K. and Tripathi, Y. M. (2012). Estimating the parameters of a Burr distribution under progressive type II censoring. *Statistical Methodology* **9**, 381–391. MR2871439 https://doi.org/10.1016/j.stamet.2011.10.002

Rastogi, M. K. and Tripathi, Y. M. (2013). Inference on unknown parameters of a Burr distribution under hybrid censoring. *Statistical Papers* **54**, 619–643. MR3072892 https://doi.org/10.1007/s00362-012-0452-3

Samanta, D. and Kundu, D. (2021). Bayesian inference of a dependent competing risk data. *Journal of Statistical Computation and Simulation* **91**, 3069–3086. MR4319636 https://doi.org/10.1080/00949655.2021.1917575

Sen, T., Bhattacharya, R., Tripathi, Y. M. and Pradhan, B. (2020). Inference and optimum life testing plans based on type-II progressive hybrid censored generalized exponential data. *Communications in Statistics - Simulation and Computation* **49**, 3254–3282. MR4186561 https://doi.org/10.1080/03610918.2018.1538456

Singh, K., Mahto, A. K. and Tripathi, Y. M. (2023). On partially observed competing risks model for Chen distribution under generalized progressive hybrid censoring. *Statistica Neerlandica*.

Sultana, F., Tripathi, Y. M., Rastogi, M. K. and Wu, S. (2018). Parameter estimation for the Kumaraswamy distribution based on hybrid censoring. *American Journal of Mathematical and Management Sciences* **37**, 243–261.

Tadikamalla, P. (1980). A look at the Burr and related distributions. *International Statistical Review/Revue Internationale De Statistique* **48**, 337–344. MR0601301 https://doi.org/10.2307/1402945

Wang, L. (2018). Inference of progressively censored competing risks data from Kumaraswamy distributions. *Journal of Computational and Applied Mathematics* **343**, 719–736. MR3813577 https://doi.org/10.1016/j.cam.2018.05.013

Wang, L. and Li, H. (2020). Inference for exponential competing risks data under generalized progressive hybrid censoring. *Communications in Statistics - Simulation and Computation* **51**, 1255–1271. MR4393670 https://doi.org/10.1080/03610918.2019.1667388

Wang, L., Tripathi, Y. M. and Lodhi, C. (2020). Inference for Weibull competing risks model with partially observed failure causes under generalized progressive hybrid censoring. *Journal of Computational and Applied Mathematics* **368**, 112537. MR4038685 https://doi.org/10.1016/j.cam.2019.112537

Wu, Q. and Vos, P. (2018). Chapter 6—Inference and prediction. In *Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications, Vol. 38* 111–172.

Brazilian Journal of Probability and Statistics
2023, Vol. 37, No. 3, 596–618
https://doi.org/10.1214/23-BJPS583

Simultaneous outlier detection and variable selection for spatial Durbin model

Yi Cheng and Yunquan Song[a]

College of Science, China University of Petroleum, Qingdao 266580, P.R. China, [a]*statistics99@163.com*

Abstract. With the continuous development of economy and technology, the application of spatial data has become increasingly widespread. Handling complex spatial data, outlier detection has become an important problem in the study of spatial models. This article proposes a method for simultaneously performing outlier detection and variable selection in the spatial Durbin model. This method combines relevant theories of spatial statistics and enables accurate identification and location of outliers, as well as variable selection of estimation coefficients, by modeling and analyzing spatial data. The experimental results indicate that the proposed method effectively detects outliers in spatial data while maintaining accuracy, and has high interpretability and generalization value. Furthermore, a practical case is presented to demonstrate the method's effectiveness in real-world scenarios.

References

AECE and BVPA (2021). Outlier detection and quasi-periodicity optimization algorithm: Frequency domain based outlier detection (FOD)—ScienceDirect. *European Journal of Operational Research* **291**, 560–574. MR4204447 https://doi.org/10.1016/j.ejor.2020.01.014

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* **60**, 255–265. MR0326953 https://doi.org/10.1093/biomet/60.2.255

Angiulli, F., Basta, S. and Pizzuti, C. (2005). Distance-based detection and prediction of outliers. *IEEE Transactions on Knowledge and Data Engineering* **18**, 145–160.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Berlin: Springer.

Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations. *Journal of the American Statistical Association* **96**, 939–967. MR1946364 https://doi.org/10.1198/016214501753208942

Bekti (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* **70**, 120–126. MR0413393

Bekti (2013). Maximum likelihood estimation for spatial Durbin model. *Journal of Mathematics and Statistics* **9**, 169–174.

Bekti (2020). Energy investment, economic growth and carbon emissions in China—Empirical analysis based on spatial Durbin model. *Energy Policy* **140**, 111425.

Breunig, M. M., Kriegel, H. P., Ng, R. T., et al (1999). Optics-of: Identifying local outliers. In *Principles of Data Mining and Knowledge Discovery: Third European Conference*, 262–270.

Breunig, M. M., Kriegel, H. P., Ng, R. T., et al (2000). LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, 93–104.

Du, X., Zuo, E., Chu, Z., et al (2023). Fluctuation-based outlier detection. *Scientific Reports* **13**, 2408.

Foster, D. P. and George, E. I. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics* **22**, 1947–1975. MR1329177 https://doi.org/10.1214/aos/1176325766

Ghiringhelli, C., Piras, G., Arbia, G. and Mira, A. (2022). Recursive estimation of the spatial error model. *Geographical Analysis* **55**, 90–106. https://doi.org/10.1111/gean.12317

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B, Methodological* **41**, 190–195. MR0547244

Hodge, V. and Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review* **22**, 85–126.

Huang, Y., Chen, C., Lei, L., et al (2022). Impacts of green finance on green innovation: A spatial and nonlinear perspective. *Journal of Cleaner Production* **365**, 132548.

Key words and phrases. Spatial Durbin model, outlier detection, variable selection.

Jin, L., Dai, X., Shi, A. and Shi, L. (2013). Detection of outliers in mixed regressive-spatial autoregressive models. *Communications in Statistics - Theory and Methods* **45**, 5179–5192. MR3520303 https://doi.org/10.1080/03610926.2014.941493

Kelejian, H. H. and Prucha, I. R. (2013). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* **17**, 99–121.

Kong, D., Bondell, H. D. and Wu, Y. (2018). Fully efficient robust estimation, outlier detection and variable selection via penalized regression. *Statistica Sinica*, 1031–1052.

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **72**, 1899–1925. MR3489437

Lee, L. F. (2014). Outlier detection in general spatial model. *Journal of Systems Science and Mathematical Sciences*.

Liu, X., Chen, J. and Cheng, S. (2018). A penalized quasi-maximum likelihood method for variable selection in the spatial autoregressive model. *Spatial Statistics* **25**, 86–104. MR3809257 https://doi.org/10.1016/j.spasta.2018.05.001

Logan, B. F., Mallows, C. L., Rice, S. O., et al (1973). Limit distributions of self-normalized sums. *Annals of Probability* **1**, 788–809.

Mur, J. and Angulo, A. (2006). The spatial Durbin model and the common factor tests. *Spatial Economic Analysis* **1**, 207–226.

Navarro, J., Isaac, M. D. D., Fernandez, R. R., et al (2022). Triangle-based outlier detection. *Pattern Recognition Letters* **2022**, 156.

On detecting spatial outliers (2008). LOF: Identifying density-based local outliers. *Geoinformatica* **12**, 455–475.

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

She, Y. and Owen, A. B. (2011). Outlier detection using nonconvex penalized regression. *Journal of the American Statistical Association* **106**, 626–639. MR2847975 https://doi.org/10.1198/jasa.2011.tm10390

Smirnov, O. and Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. *Computational Statistics & Data Analysis* **35**, 301–319. MR1819042 https://doi.org/10.1016/S0167-9473(00)00018-9

Stone, M. (1974). Cross-validation and multinomial prediction. *Biometrika* **61**, 509–515. MR0415896 https://doi.org/10.1093/biomet/61.3.509

Tibshirani, J. and Manning, C. D. (2013). Robust logistic regression using shift parameters. (long version). arXiv preprint. Available at arXiv:1305.4987. MR2634010

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

Ying, D. (2018). Estimation for the spatial autoregressive threshold model. *Economics Letters* **171**, 172–175.

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

Zhang, M., Li, Y., Lu, J. and Shi, L. (2021). Outlier detection and accommodation in meta-regression models. *Communications in Statistics—Theory and Methods* **50**, 1728–1744. MR4241917 https://doi.org/10.1080/03610926.2019.1652321

Zhu, D., Liu, Y., Yao, X., et al (2021). Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. *Geoinformatica* **2021**, 1–32.

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
2023, Vol. 37, No. 3, 619–642
https://doi.org/10.1214/23-BJPS584

Bivariate log-symmetric models: Distributional properties, parameter estimation and an application to public spending data

Roberto Vila[1,2,a], **Narayanaswamy Balakrishnan**[2,d], **Helton Saulo**[1,3,b] and **Ana Protazio**[1,c]

[1]*Department of Statistics, University of Brasilia, Brasilia, Brazil,* [a]*rovig161@gmail.com,* [b]*heltonsaulo@gmail.com,* [c]*ana.protazio@gmail.com*
[2]*Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada,* [d]*bala@mcmaster.ca*
[3]*Department of Mathematics, University of Texas at Arlington, Arlington, Texas, USA*

Abstract. The bivariate Gaussian distribution has been a key model for many developments in statistics. However, many real-world phenomena produce data that follow asymmetric distributions, and consequently bivariate normal model becomes inappropriate in such situations. Bidimensional log-symmetric models have attractive properties and can be considered as good alternatives in such situations. In this paper, we discuss bivariate log-symmetric distributions and their characterizations. We establish several distributional properties and also discuss the maximum likelihood estimation of model parameters. A Monte Carlo simulation study is performed for examining the performance of the developed parameter estimation method. A real data set is finally analyzed to illustrate the proposed model and the associated inferential method.

References

Abdous, B., Fougères, A.-L. and Ghoudi, K. (2005). Extreme behaviour for bivariate elliptical distributions. *Canadian Journal of Statistics*, 317–334. MR2193978 https://doi.org/10.1002/cjs.5540330302

Balakrishnan, N. and Lai, C.-D. (2009). *Continuous Bivariate Distributions*, 2nd ed. New York: Springer. MR2840643 https://doi.org/10.1007/b101765

Berk, R. H. (1972). Consistency and asymptotic mormality of MLE's for xxponential models. *The Annals of Mathematical Statistics* **43**, 193–204. MR0298810

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton, New Jersey: Princeton University Press. MR0016588

Crow, E. L. and Shimizu, K., eds. (1988). *Lognormal Distributions: Theory and Applications*. New York: Marcel Dekker. MR0939191

Cunha, D. R., Divino, J. A. and Saulo, H. (2022). On a log-symmetric quantile Tobit model applied to female labor supply data. *Journal of Applied Statistics* **49**, 4225–4253. MR4504461 https://doi.org/10.1080/02664763.2021.1976120

Cunha, D. R., Divino, J. A. and Saulo, H. (2023). Zero-adjusted log-symmetric quantile regression models. *Computational Economics*. https://doi.org/10.1007/s10614-023-10420-4

Dasilva, A., Saulo, H., Vila, R., Fiorucci, J. A. and Pal, S. (2023). Parametric quantile autoregressive moving average models with exogenous terms. *Statistical Papers*. https://doi.org/10.1007/s00362-023-01459-4

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65**, 457–487. MR0521817 https://doi.org/10.1093/biomet/65.3.457

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. London, England: Chapman and Hall. MR1071174 https://doi.org/10.1007/978-1-4899-2937-2

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153–161. MR0518832 https://doi.org/10.2307/1912352

James, B. R. (2004). *Probabilidade: Um Curso em Nível Intermediário*, Projeto Euclides. Brazil. MR0651713

Key words and phrases. Bivariate Log-symmetric Models, Monte Carlo simulation, Maximum likelihood method, R software.

Jones, M. C. (2008). On reciprocal symmetry. *Journal of Statistical Planning and Inference* **138**, 3039–3043. MR2526220 https://doi.org/10.1016/j.jspi.2007.11.006

Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions—Vol. 1*, 2nd ed. New York: Wiley. MR1326603

Kotz, S., Kozubowski, T. J. and Podgórski, K. (2001). *The Laplace Distribution and Generalizations*. New York: Wiley. MR1935481 https://doi.org/10.1007/978-1-4612-0173-1

Mäkeläinen, T., Schmidt, K. and Styan, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *The Annals of Statistics* **9**, 758–767. MR0619279

Nadarajah, S. and Lyu, J. (2022). New bivariate and multivariate log-normal distributions as models for insurance data. *Results in Applied Mathematics* **14**, 100246. MR4384605 https://doi.org/10.1016/j.rinam.2022.100246

Puig, P. (2008). A note on the harmonic law: A two-parameter family of distributions for ratios. *Statistics & Probability Letters* **78**, 320–326. MR2387553 https://doi.org/10.1016/j.spl.2007.07.024

Rohatgi, V. K. and Saleh, A. K. Md. E. (2015). *An Introduction to Probability Theory and Mathematical Statistics*, 3rd ed. Hoboken, New Jersey: Wiley. MR3364976 https://doi.org/10.1002/9781118799635

Saulo, H., Dasilva, A., Leiva, V., Sánchez, L. and de la Fuente-Mella, H. (2022). Log-symmetric quantile regression models. *Statistica Neerlandica* **76**, 124–163. MR4423263 https://doi.org/10.1111/stan.12243

Saulo, H., Vila, R., Cordeiro, S. S. and Leiva, V. (2023). Bivariate symmetric Heckman models and their characterization. *Journal of Multivariate Analysis* **193**, 105097. MR4517028 https://doi.org/10.1016/j.jmva.2022.105097

Vanegas, L. H. and Paula, G. A. (2016). Log-symmetric distributions: Statistical properties and parameter estimation. *Brazilian Journal of Probability and Statistics* **30**, 196–220. MR3481101 https://doi.org/10.1214/14-BJPS272

Vila, R., Balakrishnan, N., Saulo, H. and Zörnig, P. (2023). Family of bivariate distributions on the unit square: Theoretical properties and applications. *Journal of Applied Statistics*. https://doi.org/10.1080/02664763.2023.2232127

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: https://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 https://www.e-publications.org/ims/submission/.

Preparation of Manuscripts: https://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:
https://www.imstat.org/individual-membership/

Individual and General Subscriptions:
https://imstat.org/journals-and-publications/subscriptions-and-orders/

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

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