ISSN 1932-6157 (print)
ISSN 1941-7330 (online)

THE ANNALS *of* APPLIED STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

continued

Vol. 17, No. 2—June 2023

ISSN 1932-6157 (print)
ISSN 1941-7330 (online)

THE ANNALS *of* APPLIED STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

Vol. 17, No. 2—June 2023

THE ANNALS OF APPLIED STATISTICS Vol. 17, No. 2, pp. 901–1800 June 2023

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

The Annals of Applied Statistics **[ISSN 1932-6157 (print); ISSN 1941-7330 (online)],** Volume 17, Number 2, June 2023. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage pending at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Applied Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 901–928
https://doi.org/10.1214/22-AOAS1629

HOSPITAL QUALITY RISK STANDARDIZATION VIA APPROXIMATE BALANCING WEIGHTS

BY LUKE J. KEELE[1,a], ELI BEN-MICHAEL[2,c], AVI FELLER[3,d], RACHEL KELZ[1,b] AND LUKE MIRATRIX[4,e]

[1]*Hospital of the University of Pennsylvania, University of Pennsylvania,* [a]*luke.keele@gmail.com,* [b]*Rachel.Kelz@pennmedicine.upenn.edu*

[2]*Department of Statistics and Institute for Quantitative Social Science, Harvard University,* [c]*ebenmichael@fas.harvard.edu*

[3]*Goldman School of Public Policy, University of California, Berkeley,* [d]*afeller@berkeley.edu*

[4]*Graduate School of Education, Harvard University,* [e]*lmiratrix@g.harvard.edu*

Comparing outcomes across hospitals, often to identify underperforming hospitals, is a critical task in health services research. However, naive comparisons of average outcomes, such as surgery complication rates, can be misleading because hospital case mixes differ—a hospital's overall complication rate may be lower simply because the hospital serves a healthier population overall. In this paper we develop a method of "direct standardization" where we reweight each hospital patient population to be representative of the overall population and then compare the weighted averages across hospitals. Adapting methods from survey sampling and causal inference, we find weights that directly control for imbalance between the hospital patient mix and the target population, even across many patient attributes. Critically, these balancing weights can also be tuned to preserve sample size for more precise estimates. We also derive principled measures of statistical uncertainty and use outcome modeling and Bayesian shrinkage to increase precision and account for variation in hospital size. We demonstrate these methods using claims data from Pennsylvania, Florida, and New York, estimating standardized hospital complication rates for general surgery patients. We conclude with a discussion of how to detect low performing hospitals.

REFERENCES

ABADIE, A. and IMBENS, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *J. Bus. Econom. Statist.* **29** 1–11. MR2789386 https://doi.org/10.1198/jbes.2009.07333

ATHEY, S., IMBENS, G. W. and WAGER, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 597–623. MR3849336 https://doi.org/10.1111/rssb.12268

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2021). The augmented synthetic control method. *J. Amer. Statist. Assoc.* **116** 1789–1803. MR4353714 https://doi.org/10.1080/01621459.2021.1929245

BEN-MICHAEL, E., HIRSCHBERG, D., FELLER, A. and ZUBIZARRETA, J. (2021). The balancing act for causal inference.

BLOOM, H. S., RAUDENBUSH, S. W., WEISS, M. and PORTER, K. E. (2016). Using multi-site experiments to study cross-site variation in treatment effects: A hybrid approach with fixed intercepts and a random treatment coefficient. *J. Res. Educ. Eff.* 1–66.

CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76**.

CENTERS FOR MEDICARE AND MEDICAID SERVICES (2021). Quality payment program website. Available at https://qpp.cms.gov/about/qpp-overview.

Key words and phrases. Risk adjustment, weighting, direct standardization.

CENTERS FOR MEDICARE & MEDICAID SERVICES (2021a). Facility-level 7-day hospital visits after general surgery procedures performed at ambulatory surgical centers. Technical report, Washington, D.C.

CENTERS FOR MEDICARE & MEDICAID SERVICES (2021b). Patient-mix coefficients for January 2022 (3Q20 through 1Q21 Discharges) publicly reported HCAHPS results. Available at http://https://www.hcahpsonline.org/ (accessed February 11, 2022).

D'AMOUR, A. and FRANKS, A. (2019). Covariate reduction for weighted causal effect estimation with deconfounding scores.

DECKER, M. R., DODGION, C. M., KWOK, A. C., HU, Y.-Y., HAVLENA, J. A., JIANG, W., LIPSITZ, S. R., KENT, K. C. and GREENBERG, C. C. (2014). Specialization and the current practices of general surgeons. *J. Am. Coll. Surg.* **218** 8–15.

DEMING, W. E. and STEPHAN, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Stat.* **11** 427–444. MR0003527 https://doi.org/10.1214/aoms/1177731829

DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804

DEVILLE, J. C., SÄRNDAL, C. E. and SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *J. Amer. Statist. Assoc.* **88** 1013–1020. https://doi.org/10.1080/01621459.1993.10476369

ELIXHAUSER, A., STEINER, C., HARRIS, D. R. and COFFEY, R. M. (1998). Comorbidity measures for use with administrative data. *Med. Care* **36** 8–27. https://doi.org/10.1097/00005650-199801000-00004

FLEISS, J. L., LEVIN, B. and PAIK, M. C. (2003). *Statistical Methods for Rates and Proportions*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2001202 https://doi.org/10.1002/0471445428

GEORGE, E. I., ROČKOVÁ, V., ROSENBAUM, P. R., SATOPÄÄ, V. A. and SILBER, J. H. (2017). Mortality rate estimation and standardization for public reporting: Medicare's Hospital Compare. *J. Amer. Statist. Assoc.* **112** 933–947. MR3735351 https://doi.org/10.1080/01621459.2016.1276021

GOLDSTEIN, H. and SPIEGELHALTER, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *J. Roy. Statist. Soc. Ser. A* **159** 385–409.

HAINMUELLER, J. (2011). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20** 25–46. https://doi.org/10.1093/pan/mpr025

HAZLETT, C. (2020). Kernel balancing: A flexible non-parametric weighting procedure for estimating causal effects. *Statist. Sinica* **30** 1155–1189. MR4257528 https://doi.org/10.5705/ss.20

HEDGES, L. V. and PIGOTT, T. D. (2001). The power of statistical tests in meta-analysis. *Psychol. Methods* **6** 203–217.

HIRSHBERG, D. A., MALEKI, A. and ZUBIZARRETA, J. (2019). Minimax linear estimation of the retargeted mean. arXiv preprint arXiv:1901.10296.

HIRSHBERG, D. A. and WAGER, S. (2021). Augmented minimax linear estimation. *Ann. Statist.* **49** 3206–3227. MR4352528 https://doi.org/10.1214/21-aos2080

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

HULL, P. (2018). Estimating hospital quality with quasi-experimental data. SSRN 3118358.

IEZZONI, L. I. (2012). *Risk Adjustment for Measuring Health Care Outcomes*, 4th ed. Health Administration Press, Chicago, IL.

IMBENS, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.

JACOB, P. E., MURRAY, L. M., HOLMES, C. C. and ROBERT, C. P. (2017). Better together? Statistical learning in models made of modules. arXiv:1708.08719.

KALLUS, N. (2020). Generalized optimal matching methods for causal inference. *J. Mach. Learn. Res.* **21** Paper No. 62, 54. MR4095341

KEELE, L. J., BEN-MICHAEL, E., FELLER, A., KELZ, R. and MIRATRIX, L. (2023a). Supplement to "Hospital quality risk standardization via approximate balancing weights." https://doi.org/10.1214/22-AOAS1629SUPPA

KEELE, L. J., BEN-MICHAEL, E., FELLER, A., KELZ, R. and MIRATRIX, L. (2023b). Supplement to "Hospital quality risk standardization via approximate balancing weights." https://doi.org/10.1214/22-AOAS1629SUPPB

KITAGAWA, E. M. (1955). Components of a difference between two rates. *J. Amer. Statist. Assoc.* **50** 1168–1194.

KRUMHOLZ, H. M., WANG, Y., MATTERA, J. A., WANG, Y., HAN, L. F., INGBER, M. J., ROMAN, S. and NORMAND, S.-L. T. (2006). An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation* **113** 1683–1692. https://doi.org/10.1161/CIRCULATIONAHA.105.611186

LI, F., MORGAN, K. L. and ZASLAVSKY, A. M. (2018). Balancing covariates via propensity score weighting. *J. Amer. Statist. Assoc.* **113** 390–400. MR3803473 https://doi.org/10.1080/01621459.2016.1260466

LOHR, S. L. (2010). *Sampling: Design and Analysis*, 2nd ed. Brooks/Cole, Cengage Learning, Boston, MA. MR3057878

LONGFORD, N. T. (2019). Performance assessment as an application of causal inference. *J. Roy. Statist. Soc. Ser. A*.

MCCAFFREY, D. F., LOCKWOOD, J., KORETZ, D., LOUIS, T. A. and HAMILTON, L. (2004). Models for value-added modeling of teacher effects. *J. Educ. Behav. Stat.* **29** 67–101.

MEDICARE.GOV (2013). Hospital compare. Available at https://www.medicare.gov/hospitalcompare/search.html (accessed February 14, 2022).

MIRATRIX, L. W. and FELLER, A. (2020). Reatment effect distributions in multi-site trials.

NATIONAL QUALITY FORUM (2014). Risk adjustment for socioeconomic status or other sociodemographic factors. Technical report, Washington, D.C.

NORMAND, S.-L. T. and SHAHIAN, D. M. (2007). Statistical and clinical aspects of hospital outcomes profiling. *Statist. Sci.* **22** 206–226. MR2408959 https://doi.org/10.1214/088342307000000096

NORMAND, S.-L. T., ASH, A. S., FIENBERG, S. E., STUKEL, T. A., UTTS, J. and LOUIS, T. A. (2016). League tables for hospital comparisons. *Annu. Rev. Stat. Appl.* **3** 21–50.

PADDOCK, S. M., RIDGEWAY, G., LIN, R. and LOUIS, T. A. (2006). Flexible distributions of triple-goal estimates in two-stage hierarchical models. *Comput. Statist. Data Anal.* **50** 3243–3262. MR2239666 https://doi.org/10.1016/j.csda.2005.05.008

PIMENTEL, S. D., KELZ, R. R., SILBER, J. H. and ROSENBAUM, P. R. (2015). Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. *J. Amer. Statist. Assoc.* **110** 515–527. MR3367244 https://doi.org/10.1080/01621459.2014.997879

POTTHOFF, R. F., WOODBURY, M. A. and MANTON, K. G. (1992). "Equivalent sample size" and "equivalent degrees of freedom" refinements for inference using survey weights under superpopulation models. *J. Amer. Statist. Assoc.* **87** 383–396. MR1173805

RIDGEWAY, G. and MACDONALD, J. M. (2009). Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *J. Amer. Statist. Assoc.* **104** 661–668. MR2751446 https://doi.org/10.1198/jasa.2009.0034

ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.

ROBINS, J. M. and ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90** 122–129. MR1325119

RUBIN, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics* 185–203.

RUBIN, D. B. (2008). For objective causal inference, design trumps analysis. *Ann. Appl. Stat.* **2** 808–804. MR2516795 https://doi.org/10.1214/08-AOAS187

SHEN, W. and LOUIS, T. A. (1998). Triple-goal estimates in two-stage hierarchical models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 455–471. MR1616061 https://doi.org/10.1111/1467-9868.00135

SILBER, J. H., ROSENBAUM, P. R. and ROSS, R. N. (1995). Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? *J. Amer. Statist. Assoc.* **90** 7–18.

SILBER, J. H., ROSENBAUM, P. R., ROSS, R. N., LUDWIG, J. M., WANG, W., NIKNAM, B. A., MUKHERJEE, N., SAYNISCH, P. A., EVEN-SHOSHAN, O. et al. (2014a). Template matching for auditing hospital cost and quality. *Health Serv. Res.* **49** 1446–1474.

SILBER, J. H., ROSENBAUM, P. R., ROSS, R. N., LUDWIG, J. M., WANG, W., NIKNAM, B. A., SAYNISCH, P. A., EVEN-SHOSHAN, O., KELZ, R. R. et al. (2014b). A hospital-specific template for benchmarking its cost and quality. *Health Serv. Res.* **49** 1475–1497.

SILBER, J. H., ROSENBAUM, P. R., ROSS, R. N., LUDWIG, J. M., WANG, W., NIKNAM, B. A., HILL, A. S., EVEN-SHOSHAN, O., KELZ, R. R. et al. (2016a). Indirect standardization matching: Assessing specific advantage and risk synergy. *Health Serv. Res.* **51** 2330–2357.

SILBER, J. H., ROSENBAUM, P. R., WANG, W., LUDWIG, J. M., CALHOUN, S., GUEVARA, J. P., ZORC, J. J., ZEIGLER, A. and EVEN-SHOSHAN, O. (2016b). Auditing practice style variation in pediatric inpatient asthma care. *JAMA, J. Am. Med. Assoc. Southeast Asia, Suppl., Pediatr.* **170** 878–886.

WANG, Y. and ZUBIZARRETA, J. R. (2020). Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika* **107** 93–105. MR4064142 https://doi.org/10.1093/biomet/asz050

WEISS, M. J., BLOOM, H. S., VERBITSKY-SAVITZ, N., GUPTA, H., VIGIL, A. E. and CULLINAN, D. N. (2017). How much do the effects of education and training programs vary across sites? Evidence from past multisite randomized trials. *J. Res. Educ. Eff.* **10** 843–876.

YIU, S. and SU, L. (2018). Covariate association eliminating weights: A unified weighting framework for causal effect estimation. *Biometrika* **105** 709–722. MR3842894 https://doi.org/10.1093/biomet/asy015

ZHAO, Q. (2019). Covariate balancing propensity score by tailored loss functions. *Ann. Statist.* **47** 965–993. MR3909957 https://doi.org/10.1214/18-AOS1698

ZHAO, Q. and PERCIVAL, D. (2017). Entropy balancing is doubly robust. *J. Causal Inference* **5** Art. No. 20160010. MR4323812 https://doi.org/10.1515/jci-2016-0010

ZUBIZARRETA, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *J. Amer. Statist. Assoc.* **110** 910–922. MR3420672 https://doi.org/10.1080/01621459.2015.1023805

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 929–960
https://doi.org/10.1214/22-AOAS1636

ESTIMATION OF GAUSSIAN DIRECTED ACYCLIC GRAPHS USING PARTIAL ORDERING INFORMATION WITH APPLICATIONS TO DREAM3 NETWORKS AND DAIRY CATTLE DATA

BY SYED RAHMAN[1,a], KSHITIJ KHARE[2,b], GEORGE MICHAILIDIS[2,c], CARLOS MARTÍNEZ[3,d] AND JUAN CARULLA[4,e]

[1]*Data Scientist, Apple Inc.*, [a]*shr264@ufl.edu*

[2]*Department of Statistics, University of Florida*, [b]*kdkhare@stat.ufl.edu*, [c]*gmichail@ufl.edu*

[3]*Centro de Investigación Tibaitata, Corporación Colombiana de Investigación Agropecuaria*, [d]*cmartinez@agrosavia.co*

[4]*Departamento de Ciencias para la Producción Animal, Universidad Nacional de Colombia*, [e]*jecarullaf@unal.edu.co*

Estimating a directed acyclic graph (DAG) from observational data represents a canonical learning problem and has generated a lot of interest in recent years. Research has focused mostly on the following two cases: when no information regarding the ordering of the nodes in the DAG is available and when a domain-specific complete ordering of the nodes is available. In this paper, motivated by a recent application in dairy science, we develop a method for DAG estimation for the middle scenario, where partition-based partial ordering of the nodes is known based on domain-specific knowledge. We develop an efficient algorithm that solves the posited problem, coined Partition-DAG. Through extensive simulations, using the DREAM3 Yeast networks, we illustrate that Partition-DAG effectively incorporates the partial ordering information to improve both speed and accuracy. We then illustrate the usefulness of Partition-DAG by applying it to recently collected dairy cattle data, and inferring relationships between various variables involved in dairy agroecosystems.

REFERENCES

ALTOMARE, D., CONSONNI, G. and LA ROCCA, L. (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics* **69** 478–487. MR3071066 https://doi.org/10.1111/biom.12018

ARAGAM, B., AMINI, A. A. and ZHOU, Q. (2016). Learning directed acyclic graphs with penalized neighbourhood regression. arXiv.

ARAGAM, B. and ZHOU, Q. (2015). Concave penalized estimation of sparse Gaussian Bayesian networks. *J. Mach. Learn. Res.* **16** 2273–2328. MR3450507

BARGO, F., MULLER, L. D., KOLVER, E. S. and DELAHOY, J. E. (2003). Invited review: Production and digestion of supplemented dairy cows on pasture. *J. Dairy Sci.* **86** 1–42.

CAO, X., KHARE, K. and GHOSH, M. (2019). Posterior graph selection and estimation consistency for high-dimensional Bayesian DAG models. *Ann. Statist.* **47** 319–348. MR3909935 https://doi.org/10.1214/18-AOS1689

CHICKERING, D. M. (2002). Optimal structure identification with greedy search. *J. Mach. Learn. Res.* **3** 507–554.

COLOMBO, D. and MAATHIUS, M. H. (2014). Order-independent constraint-based causal structure learning. *J. Mach. Learn. Res.* **15** 3921–3962.

CONSONNI, G., LA ROCCA, L. and PELUSO, S. (2017). Objective Bayes covariate-adjusted sparse graphical model selection. *Scand. J. Stat.* **44** 741–764. MR3687971 https://doi.org/10.1111/sjos.12273

DILLON, P. (2006). Achieving high dry-matter intake from pastures with grazing dairy cows. In *Fresh Herbage for Dairy Cattle* (A. Elgersma, J. Dijkstra and S. Tamminga, eds.) Springer, Netherlands. https://doi.org/10.1007/978-1-4020-5452-5_1

ELGERSMA, A., DIJKSTRA, J. and TAMMINGA, S. (2006). *Fresh Herbage for Dairy Cattle*. Springer, Netherlands.

ELLIS, B. and WONG, W. H. (2008). Learning causal Bayesian network structures from experimental data. *J. Amer. Statist. Assoc.* **103** 778–789. MR2524009 https://doi.org/10.1198/016214508000000193

Key words and phrases. Sparse covariance estimation, Gaussian DAG, partially ordered variables.

EMMERT-STREIB, F., DEHMER, M. and HAIBE-KAINS, B. (2014). Gene regulatory networks and their applications: Understanding biological and medical problems in terms of networks. *Front*. *Cell Dev*. *Biol*. **2** 38.

GÁMEZ, J. A., MATEO, J. L. and PUERTA, J. M. (2011). Learning Bayesian networks by hill climbing: Efficient methods based on progressive restriction of the neighborhood. *Data Min*. *Knowl*. *Discov*. **22** 106–148. MR2764554 https://doi.org/10.1007/s10618-010-0178-6

GAMEZ, J. A., MATEO, J. L. and PUERTA, J. M. (2012). One iteration chc algorithm for learning Bayesian networks: An effective and efficient algorithm for high dimensional problems. *Prog*. *Artif*. *Intell*. **1** 329–346.

GEIGER, D. and HECKERMAN, D. (2013). Learning Gaussian networks. Available at arXiv:1302.6808.

HAUSER, A. and BÜHLMANN, P. (2012). Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *J*. *Mach*. *Learn*. *Res*. **13** 2409–2464. MR2973606

HECKERMAN, D., GEIGER, D. and CHICKERING, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Mach*. *Learn*. **20** 197–243.

HOYER, P. O., HYVARINEN, A., SCHEINES, R., SPIRTES, P., RAMSEY, J., LACERDA, G. and SHIMIZU, S. (2008). Causal discovery of linear acyclic models with arbitrary distributions. In *Proc*. 24*th Conference on Uncertainty in Artificial Intelligence* (*UAI*2008) 282–289.

HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. MR2277742 https://doi.org/10.1093/biomet/93.1.85

JALVINGH, A. W. (1992). The possible role of existing models in on-farm decision support in dairy cattle and swine production. *Livest*. *Prod*. *Sci*. **31** 355–365.

KALISCH, M. and BUHLMANN, P. (2007). Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *J*. *Mach*. *Learn*. *Res*. **8** 613–636.

KHARE, K., OH, S.-Y. and RAJARATNAM, B. (2015). A convex pseudolikelihood framework for high dimensional partial correlation estimation with convergence guarantees. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **77** 803–825. MR3382598 https://doi.org/10.1111/rssb.12088

KHARE, K., OH, S., RAHMAN, S. and RAJARATNAM, B. (2017). A convex framework for high-dimensional sparse Cholesky based covariance estimation. *Mach*. *Learn*. **108** 2061–2086.

LAM, W. and BACCHUS, F. (1994). Learning Bayesian belief networks: An approach based on the MDL principle. *Comput*. *Intell*. **10** 269–293.

LE, T., HOANG, T., LI, J. and LIU, L. (2015). A fast pc algorithm for high dimensional causal discovery with multi-core pcs. *IEEE/ACM Trans*. *Comput*. *Biol*. *Bioinform*. **16**.

LI, C., SHEN, X. and PAN, W. (2020). Likelihood ratio tests for a large directed acyclic graph. *J*. *Amer*. *Statist*. *Assoc*. **115** 1304–1319. MR4143467 https://doi.org/10.1080/01621459.2019.1623042

LIU, H., ROEDER, K. and WASSERMAN, L. (2010). Stability approach to regularization selection (stars) for high dimensional graphical models. In *Advances in Neural Information Processing Systems* **23**.

MARBACH, D., SCHAFFTER, T., MATTIUSSI, C. and FLOREANO, D. (2009). Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *J*. *Comput*. *Biol*. **16** 229–239. https://doi.org/10.1089/cmb.2008.09TT

MARBACH, D., PRILL, R. J., SCHAFFTER, T., MATTIUSSI, C., FLOREANO, D. and STOLOVITZKY, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. *PNAS* **107** 6286–6291. https://doi.org/10.1073/pnas.0913357107

MARBACH, D., COSTELLO, J. C., KÜFFNER, R., VEGA, N. M., PRILL, R. J., CAMACHO, D. M., ALLISON, K. R., ADERHOLD, A., BONNEAU, R. et al. (2012). Wisdom of crowds for robust gene network inference. *Nat*. *Methods* **9** 796.

MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). *SparseNet*: Coordinate descent with nonconvex penalties. *J*. *Amer*. *Statist*. *Assoc*. **106** 1125–1138. MR2894769 https://doi.org/10.1198/jasa.2011.tm09738

MEEK, C. (1995). Causal inference and causal explanation with background knowledge. In *UAI*'95: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* 403–410.

MEEK, C. (1997). Graphical models: Selecting causal and statistical models Technical report, Ph.D. Thesis, Carnegie Mellon Univ.

MOROTA, G., VENTURA, R., SLIVA, F., KOYAMA, M. and FERNANDO, S. (2018). Machine learning and data mining advance predictive big data analysis in precision animal agriculture. In *Big Data Analytics and Precision Animal Agriculture Symposium* Univ. Nebraska–Lincoln.

NANDY, P., HAUSER, A. and MAATHUIS, M. H. (2018). High-dimensional consistency in score-based and hybrid structure learning. *Ann*. *Statist*. **46** 3151–3183. MR3851768 https://doi.org/10.1214/17-AOS1654

PERKOVIC, E., KALISCH, M. and MAATHUIS, M. H. (2017). Interpreting and using cpdags with background knowledge. In *Proceedings UAI* 2017.

PETERS, J., BÜHLMANN, P. and MEINSHAUSEN, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **78** 947–1012. MR3557186 https://doi.org/10.1111/rssb.12167

PRILL, R. J., MARBACH, D., SAEZ-RODRIGUEZ, J., SORGER, P. K., ALEXOPOULOS, L. G., XUE, X., CLARKE, N. D., ALTAN-BONNET, G. and STOLOVITZKY, G. (2010). Towards a rigorous assessment of systems biology models: The DREAM3 challenges. *PLoS ONE* **5** 1–18. https://doi.org/10.1371/journal.pone.0009202

RAHMAN, S., KHARE, K., MICHAILIDIS, G., MARTINEZ, C. and CARULLA, J. (2023a). Supplement A to "Estimation of Gaussian directed acyclic graphs using partial ordering information with applications to DREAM3 networks and dairy cattle data." https://doi.org/10.1214/22-AOAS1636SUPPA

RAHMAN, S., KHARE, K., MICHAILIDIS, G., MARTINEZ, C. and CARULLA, J. (2023b). Supplement B to "Estimation of Gaussian directed acyclic graphs using partial ordering information with applications to DREAM3 networks and dairy cattle data." https://doi.org/10.1214/22-AOAS1636SUPPB

RAHMAN, S., KHARE, K., MICHAILIDIS, G., MARTINEZ, C. and CARULLA, J. (2023c). Supplement C to "Estimation of Gaussian directed acyclic graphs using partial ordering information with applications to DREAM3 networks and dairy cattle data." https://doi.org/10.1214/22-AOAS1636SUPPC

RAHMAN, S., KHARE, K., MICHAILIDIS, G., MARTÍNEZ, C. and CARULLA, J. (2023d). Supplement to "Estimation of Gaussian directed acyclic graphs using partial ordering information with applications to DREAM3 networks and dairy cattle data." https://doi.org/10.1214/22-AOAS1636SUPPD

RAMSEY, J. D. (2015). Scaling up greedy causal search for continuous variables Technical report, Center for Causal Discovery.

RASKUTTI, G. and UHLER, C. (2018). Learning directed acyclic graph models based on sparsest permutations. *Stat* **7** e183. MR3796726 https://doi.org/10.1002/sta4.183

SCHEINES, R., SPIRTES, P., GLYMOUR, C., MEEK, C. and RICHARDSON, T. (1998). The tetrad project: Constraint based aids to causal model specification. *Multivar. Behav. Res.* **33** 65–117.

SHIMIZU, S., HOYER, P. O., HYVÄRINEN, A. and KERMINEN, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* **7** 2003–2030. MR2274431

SHIMIZU, S., INAZUMI, T., SOGAWA, Y., HYVÄRINEN, A., KAWAHARA, Y., WASHIO, T., HOYER, P. O. and BOLLEN, K. (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *J. Mach. Learn. Res.* **12** 1225–1248. MR2804599

SHOJAIE, A. and MICHAILIDIS, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* **97** 519–538. MR2672481 https://doi.org/10.1093/biomet/asq038

SHOJAIE, A., JAUHIAINEN, A., KALLITSIS, M. and MICHAILIDIS, G. (2014a). Inferring regulatory networks by combining perturbation screens and steady state gene expression profiles. *PLoS ONE* **9** e82393.

SHOJAIE, A., JAUHIAINEN, A., KALLITSIS, M. and MICHAILIDIS, G. (2014b). Inferring regulatory networks by combining perturbation screens and steady state gene expression profiles. *PLoS ONE* **9** e82393.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2001). *Causation, Prediction, and Search*. MIT Press, Cambridge.

THORNLEY, J. H. M. and FRANCE, J. (2007). *Mathematical Models in Agriculture. Quantitative Methods for Plant, Animal and Ecological Sciences*, 2nd ed. Cromwell Press, Trowbridge, UK.

TSAMARDINOS, I., BROWN, L. E. and ALIFERIS, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Mach. Learn.* **65** 31–78.

TSENG, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109** 475–494. MR1835069 https://doi.org/10.1023/A:1017501703105

TSOUMAKAS, G., KATAKIS, I. and VLAHAVAS, I. P. (2010). Mining multi-label data. In *Data Mining and Knowledge Discovery Handbook* (O. Maimon and L. Rokach, eds.) 667–685. Springer, Heidelberg, Germany.

VAN DE GEER, S. and BÜHLMANN, P. (2013). ℓ_0-penalized maximum likelihood for sparse directed acyclic graphs. *Ann. Statist.* **41** 536–567. MR3099113 https://doi.org/10.1214/13-AOS1085

WANG, Y., SOLUS, L., YANG, K. and UHLER, C. (2017). Permutation-based causal inference algorithms with interventions. In *Advances in Neural Information Processing Systems* 5822–5831.

YANG, K., KATCOFF, A. and UHLER, C. (2018). Characterizing and learning equivalence classes of causal dags under interventions. *Proc. Mach. Learn. Res.* **80** 5537–5546.

ZHOU, Q. (2011). Multi-domain sampling with applications to structural inference of Bayesian networks. *J. Amer. Statist. Assoc.* **106** 1317–1330. MR2896838 https://doi.org/10.1198/jasa.2011.ap10346

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 961–984
https://doi.org/10.1214/22-AOAS1637

ROBUST SENSIBLE ADVERSARIAL LEARNING OF DEEP NEURAL NETWORKS FOR IMAGE CLASSIFICATION

BY JUNGEUM KIM[a] AND XIAO WANG[b]

Department of Statistics, Purdue University, [a]*kim2712@purdue.edu,* [b]*wangxiao@purdue.edu*

The idea of robustness is central and critical to modern statistical analysis. However, despite the recent advances of deep neural networks (DNNs), many studies have shown that DNNs are vulnerable to adversarial attacks. Making imperceptible changes to an image can cause DNN models to make the wrong classification with high confidence, such as classifying a benign mole as a malignant tumor and a stop sign as a speed limit sign. The tradeoff between robustness and standard accuracy is common for DNN models. In this paper we introduce sensible adversarial learning and demonstrate the synergistic effect between pursuits of standard natural accuracy and robustness. Specifically, we define a sensible adversary, which is useful for learning a robust model, while keeping high natural accuracy. We theoretically establish that the Bayes classifier is the most robust multiclass classifier with the $0-1$ loss under sensible adversarial learning. We propose a novel and efficient algorithm that trains a robust model using implicit loss truncation. We apply sensible adversarial learning for large-scale image classification to a handwritten digital image dataset, called MNIST, and an object recognition colored image dataset, called CIFAR10. We have performed an extensive comparative study to compare our method with other competitive methods. Our experiments empirically demonstrate that our method is not sensitive to its hyperparameter and does not collapse even with a small model capacity while promoting robustness against various attacks and keeping high natural accuracy. The sensible adversarial learning software is available as a Python package at https://github.com/JungeumKim/SENSE.

REFERENCES

BALAJI, Y., GOLDSTEIN, T. and HOFFMAN, J. (2019). Instance adaptive adversarial training: Improved accuracy tradeoffs in neural nets. arXiv preprint. Available at arXiv:1910.08051.

BIGGIO, B., CORONA, I., MAIORCA, D., NELSON, B., ŠRNDIĆ, N., LASKOV, P., GIACINTO, G. and ROLI, F. (2013). Evasion attacks against machine learning at test time. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 387–402. Springer, Berlin.

CARLINI, N., ATHALYE, A., PAPERNOT, N., BRENDEL, W., RAUBER, J., TSIPRAS, D., GOODFELLOW, I., MADRY, A. and KURAKIN, A. (2019). On evaluating adversarial robustness. arXiv preprint. Available at arXiv:1902.06705.

CARMON, Y., RAGHUNATHAN, A., SCHMIDT, L., DUCHI, J. C. and LIANG, P. S. (2019). Unlabeled data improves adversarial robustness. In *Advances in Neural Information Processing Systems* 11190–11201.

DALVI, N., DOMINGOS, P., SANGHAI, S. and VERMA, D. (2004). Adversarial classification. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 99–108.

DING, G. W., WANG, L. and JIN, X. (2019). AdverTorch v0.1: An adversarial robustness toolbox based on pytorch. arXiv preprint. Available at arXiv:1902.07623.

DING, G. W., SHARMA, Y., LUI, K. Y. C. and HUANG, R. (2020). Mma training: Direct input space margin maximization through adversarial training. In *International Conference on Learning Representations*.

GAO, R., CAI, T., LI, H., HSIEH, C.-J., WANG, L. and LEE, J. D. (2019). Convergence of adversarial training in overparametrized neural networks. In *Advances in Neural Information Processing Systems* 13009–13020.

GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

Key words and phrases. Adversarial learning, Bayes classifier, classification, deep neural networks, implicit loss truncation, robustness.

GOODFELLOW, I. J., SHLENS, J. and SZEGEDY, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint. Available at arXiv:1412.6572.

HE, K., ZHANG, X., REN, S. and SUN, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778.

HUBER, P. J. (1972). The 1972 Wald lecture. Robust statistics: A review. *Ann. Math. Stat.* **43** 1041–1067. MR0314180 https://doi.org/10.1214/aoms/1177692459

HUBER, P. (2002). John W. Tukey's contributions to robust statistics. *Ann. Statist.* **30** 1640–1648.

IDELBAYEV, Y. (2018). Proper resnet implementation for cifar10/cifar100 in pytorch. https://github.com/akamaster/pytorch_resnet_cifar10.

KIM, J. and WANG, X. (2023). Supplement to "Robust Sensible adversarial learning of deep neural networks for image classification." https://doi.org/10.1214/22-AOAS1637SUPPA, https://doi.org/10.1214/22-AOAS1637SUPPB

KRIZHEVSKY, A. and HINTON, G. (2009). Learning multiple layers of features from tiny images Technical report, Citeseer.

KURAKIN, A., GOODFELLOW, I. and BENGIO, S. (2016). Adversarial examples in the physical world. arXiv preprint. Available at arXiv:1607.02533.

LAMB, A., VERMA, V., KANNALA, J. and BENGIO, Y. (2019). Interpolated adversarial training: Achieving robust neural networks without sacrificing too much accuracy. In *Proceedings of the* 12*th ACM Workshop on Artificial Intelligence and Security* 95–103.

LECUN, Y., CORTES, C. and BURGES, C. (2010). Mnist handwritten digit database. *AT&T Labs* **2** 18. [Online]. Available at Http://yann.Lecun.Com/exdb/mnist.

MADRY, A., MAKELOV, A., SCHMIDT, L., TSIPRAS, D. and VLADU, A. (2018). Towards deep learning models resistant to adversarial attacks. *ICLR*.

MOOSAVI-DEZFOOLI, S.-M., FAWZI, A. and FROSSARD, P. (2016). Deepfool: A simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2574–2582.

NICOLAE, M.-I., SINN, M., TRAN, M. N., RAWAT, A., WISTUBA, M., ZANTEDESCHI, V., BARACALDO, N., CHEN, B., LUDWIG, H. et al. (2018). Adversarial robustness toolbox v0.8.0. *CoRR*. Available at arXiv:1807.01069.

RAGHUNATHAN, A., XIE, S. M., YANG, F., DUCHI, J. C. and LIANG, P. (2019). Adversarial training can hurt generalization. arXiv preprint. Available at arXiv:1906.06032.

RAUBER, J., BRENDEL, W. and BETHGE, M. (2017). Foolbox v0.8.0: A python toolbox to benchmark the robustness of machine learning models. CoRR. Available at arXiv:1707.04131.

SCHMIDT, L., SANTURKAR, S., TSIPRAS, D., TALWAR, K. and MADRY, A. (2018). Adversarially robust generalization requires more data. In *Advances in Neural Information Processing Systems* 5014–5026.

SHAFAHI, A., NAJIBI, M., GHIASI, A., XU, Z., DICKERSON, J., STUDER, C., DAVIS, L. S., TAYLOR, G. and GOLDSTEIN, T. (2019). Adversarial training for free! arXiv preprint. Available at arXiv:1904.12843.

SPRINGENBERG, J. T., DOSOVITSKIY, A., BROX, T. and RIEDMILLER, M. (2014). Striving for simplicity: The all convolutional net. arXiv preprint. Available at arXiv:1412.6806.

STANFORTH, R., FAWZI, A., KOHLI, P. et al. (2019). Are labels required for improving adversarial robustness? arXiv preprint. Available at arXiv:1905.13725.

STIGLER, S. M. (2010). The changing history of robustness. *Amer. Statist.* **64** 277–281. MR2758558 https://doi.org/10.1198/tast.2010.10159

STUTZ, D., HEIN, M. and SCHIELE, B. (2019). Disentangling adversarial robustness and generalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 6976–6987.

SUGGALA, A. S., PRASAD, A., NAGARAJAN, V. and RAVIKUMAR, P. (2019). Revisiting adversarial risk. In *Proceedings of Machine Learning Research. Proceedings of Machine Learning Research* **89** 2331–2339. PMLR.

SZEGEDY, C., ZAREMBA, W., SUTSKEVER, I., BRUNA, J., ERHAN, D., GOODFELLOW, I. and FERGUS, R. (2013). Intriguing properties of neural networks. arXiv preprint. Available at arXiv:1312.6199.

TRAMÈR, F., KURAKIN, A., PAPERNOT, N., GOODFELLOW, I., BONEH, D. and MCDANIEL, P. (2017). Ensemble adversarial training: Attacks and defenses. arXiv preprint. Available at arXiv:1705.07204.

TSIPRAS, D., SANTURKAR, S., ENGSTROM, L., TURNER, A. and MADRY, A. (2018). There is no free lunch in adversarial robustness (but there are unexpected benefits).

TSIPRAS, D., SANTURKAR, S., ENGSTROM, L., TURNER, A. and MADRY, A. (2019). Robustness may be at odds with accuracy. In *International Conference on Learning Representations*.

TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* 448–485. Stanford Univ. Press, Stanford, CA. MR0120720

WALD, A. (1945). Statistical decision functions which minimize the maximum risk. *Ann. of Math.* (2) **46** 265–280. MR0012402 https://doi.org/10.2307/1969022

WANG, Y., ZOU, D., YI, J., BAILEY, J., MA, X. and GU, Q. (2020). Improving adversarial robustness requires revisiting misclassified examples. In *International Conference on Learning Representations*.

XIE, C. and YUILLE, A. (2020). Intriguing properties of adversarial training at scale.

YIN, D., KANNAN, R. and BARTLETT, P. (2019). Rademacher complexity for adversarially robust generalization. In *Proceedings of the* 36*th International Conference on Machine Learning*. *Proceedings of Machine Learning Research* **97** 7085–7094. PMLR.

ZHANG, Z. and SABUNCU, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. In *Advances in Neural Information Processing Systems* 8778–8788.

ZHANG, H., YU, Y., JIAO, J., XING, E., GHAOUI, L. E. and JORDAN, M. (2019). Theoretically principled trade-off between robustness and accuracy. In *Proceedings of the* 36*th International Conference on Machine Learning*. *Proceedings of Machine Learning Research* **97** 7472–7482. PMLR.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 985–1016
https://doi.org/10.1214/22-AOAS1654

ESTIMATING THE EFFECTS OF A CALIFORNIA GUN CONTROL PROGRAM WITH MULTITASK GAUSSIAN PROCESSES

BY ELI BEN-MICHAEL[1,a], DAVID ARBOUR[2,b], AVI FELLER[3,5,c], ALEXANDER FRANKS[4,d] AND STEVEN RAPHAEL[5,e]

[1]*Heinz College and Department of Statistics & Data Science, Carnegie Mellon University,* [a]*ebenmichael@cmu.edu*
[2]*Adobe Research,* [b]*arbour@adobe.com*
[3]*Department of Statistics, University of California, Berkeley,* [c]*afeller@berkeley.edu*
[4]*Department of Statistics and Applied Probability, University of California, Santa Barbara,* [d]*afranks@pstat.ucsb.edu*
[5]*Goldman School of Public Policy, University of California, Berkeley,* [e]*stevenraphael@berkeley.edu*

Gun violence is a critical public safety concern in the United States. In 2006, California implemented a unique firearm monitoring program, the Armed and Prohibited Persons System (APPS), to address gun violence in the state. The APPS program first identifies those firearm owners who become prohibited from owning one, due to federal or state law, then confiscates their firearms. Our goal is to assess the effect of APPS on California murder rates using annual, state-level crime data across the U.S. for the years before and after the introduction of the program. To do so, we adapt a nonparametric Bayesian approach, multitask Gaussian processes (MTGPs), to the panel data setting. MTGPs allow for flexible and parsimonious panel data models that nest many existing approaches and allow for direct control over both dependence across time and dependence across units as well as natural uncertainty quantification. We extend this approach to incorporate non-Normal outcomes, auxiliary covariates, and multiple outcome series, which are all important in our application. We also show that this approach has attractive Frequentist properties, including a representation as a weighting estimator with separate weights over units and time periods. Applying this approach, we find that the increased monitoring and enforcement from the APPS program substantially decreased homicides in California. We also find that the effect on murder is driven entirely by declines in gun-related murder with no measurable effect on non-gun murder. Estimated cost per murder avoided are substantially lower than conventional estimates of the value of a statistical life, suggesting a very high benefit-cost ratio for this enforcement effort.

REFERENCES

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Amer. Statist. Assoc.* **105** 493–505. MR2759929 https://doi.org/10.1198/jasa.2009.ap08746

ALAA, A. M. and VAN DER SCHAAR, M. (2017). Bayesian inference of individualized treatment effects using multi-task Gaussian processes. In *Advances in Neural Information Processing Systems* 3424–3432.

ÁLVAREZ, M. (2017). Multi-output Gaussian processes. Gaussian Process Summer School 2017.

ANEJA, A., DONOHUE III, J. J. and ZHANG, A. (2011). The impact of right-to-carry laws and the NRC report: Lessons for the empirical evaluation of law and policy. *Amer. Law Econ. Rev.* **13** 565–631.

ANTONELLI, J. and BECK, B. (2021). Heterogeneous causal effects of neighborhood policing in New York City with staggered adoption of the policy.

ARKHANGELSKY, D. and IMBENS, G. W. (2021). Double-robust identification for causal panel data models. Technical report, National Bureau of Economic Research.

ARKHANGELSKY, D., ATHEY, S., HIRSHBERG, D. A., IMBENS, G. W. and WAGER, S. (2021). Synthetic difference-in-differences. *Amer. Econ. Rev.* **111** 4088–4118.

ATHEY, S., BAYATI, M., DOUDCHENKO, N., IMBENS, G. and KHOSRAVI, K. (2021). Matrix completion methods for causal panel data models. *J. Amer. Statist. Assoc.* **116** 1716–1730. MR4353709 https://doi.org/10.1080/01621459.2021.1891924

Key words and phrases. Gun violence, Gaussian processes, panel data, causal inference.

BARRON, J. (2022). Could one county's success with 'red flag' orders be a model. *New York Times*.

BARTOS, B. J. and KUBRIN, C. E. (2018). Can we downsize our prisons and jails without compromising public safety? Findings from California's Prop 47. *Criminol. Public Policy* **17** 693–715.

BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2021). The augmented synthetic control method. *J. Amer. Statist. Assoc.* **116** 1789–1803. MR4353714 https://doi.org/10.1080/01621459.2021.1929245

BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2022). Synthetic controls with staggered adoption. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 351–381. MR4412990

BEN-MICHAEL, E., ARBOUR, D., FELLER, A., FRANKS, A. and RAPHAEL, S. (2023). Supplement to "Estimating the effects of a California gun control program with multitask Gaussian processes." https://doi.org/10.1214/22-AOAS1654SUPPA, https://doi.org/10.1214/22-AOAS1654SUPPB

BONILLA, E. V., CHAI, K. M. and WILLIAMS, C. (2008). Multi-task Gaussian process prediction. In *Advances in Neural Information Processing Systems* 153–160.

BRANSON, Z., RISCHARD, M., BORNN, L. and MIRATRIX, L. W. (2019). A nonparametric Bayesian methodology for regression discontinuity designs. *J. Statist. Plann. Inference* **202** 14–30. MR3926761 https://doi.org/10.1016/j.jspi.2019.01.003

BRODERSEN, K. H., GALLUSSER, F., KOEHLER, J., REMY, N. and SCOTT, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Ann. Appl. Stat.* **9** 247–274. MR3341115 https://doi.org/10.1214/14-AOAS788

CALIFORNIA DEPARTMENT OF JUSTICE (2015). Legislative report number one—calendar year 2014: Armed and prohibited persons system. Sacramento, CA.

CALIFORNIA DEPARTMENT OF JUSTICE (2016). Legislative report number two—calendar year 2015: Armed and prohibited persons system. Sacramento, CA.

CALIFORNIA DEPARTMENT OF JUSTICE (2017). Legislative report number three—calendar year 2016: Armed and prohibited persons system. Sacramento, CA.

CALIFORNIA DEPARTMENT OF JUSTICE (2018). Legislative report number one—armed and prohibited persons system SB140 legislative report number four—calendar Year 2017 (revised 3/8/2018). Sacramento, CA.

CALIFORNIA DEPARTMENT OF JUSTICE (2019). Legislative report number one—armed and prohibited persons system SB140 legislative report number five—calendar year 2018. Sacramento, CA.

CALIFORNIA DEPARTMENT OF JUSTICE (2022). Attorney general bonta files motion for preliminary injunction against gun retailers to halt sales of illegal ghost gun kits.

CARLSON, D. (2020). Estimating a counter-factual with uncertainty through Gaussian process projection.

CASTILLO-CARNIGLIA, A., KAGAWA, R. M. C., CERDÁ, M., CRIFASI, C. K., VERNICK, J. S., WEBSTER, D. W. and WINTEMUTE, G. J. (2019). California's comprehensive background check and misdemeanor violence prohibition policies and firearm mortality. *Ann. Epidemiol.* **30** 50–56. https://doi.org/10.1016/j.annepidem.2018.10.001

CHEN, Y., GARNETT, R., MONTGOMERY, J. and PRATI, A. (2022). A multi-task Gaussian process model for difference in differences with parallel(ish) trends.

CHENG, L.-F., DUMITRASCU, B., DARNELL, G., CHIVERS, C., DRAUGELIS, M., LI, K. and ENGELHARDT, B. E. (2020). Sparse multi-output Gaussian processes for online medical time series prediction. *BMC Med. Inform. Decis. Mak.* **20** 1–23.

CHERNOZHUKOV, V., WÜTHRICH, K. and ZHU, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *J. Amer. Statist. Assoc.* **116** 1849–1864. MR4353718 https://doi.org/10.1080/01621459.2021.1920957

CHRISTOPHER, B. (2019). How California Got Tough on Guns. Cal Matters, Available at https://calmatters.org/explainers/california-gun-laws-policy-explained/ accessed on October 11, 2019.

COOK, P. J. and DONOHUE, J. J. (2017). Saving lives by regulating guns: Evidence for policy. *Science* **358** 1259–1261. https://doi.org/10.1126/science.aar3067

DING, P. and LI, F. (2018). Causal inference: A missing data perspective. *Statist. Sci.* **33** 214–237. MR3797711 https://doi.org/10.1214/18-STS645

DOMÍNGUEZ, P. and RAPHAEL, S. (2015). The role of the cost-of-crime literature in bridging the gap between social science research and policy making: Potentials and limitations. *Criminol. Public Policy* **14** 589–632.

DONOHUE, J. J., ANEJA, A. and WEBER, K. D. (2019). Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis. *J. Empir. Leg. Stud.* **16** 198–247.

DONOHUE, J. and BOULOUTA, T. (2019). The assault weapon ban saved lives. *Stanford Law School Legal Aggregate*. October 15, 2019; Available at https://stanford.io/2MWNsrV.

DOUDCHENKO, N. and IMBENS, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report. National Bureau of Economic Research.

FERMAN, B. and PINTO, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quant. Econ.* **12** 1197–1221. MR4399595 https://doi.org/10.3982/qe1596

FIEDLER, C., SCHERER, C. W. and TRIMPE, S. (2021). Practical and rigorous uncertainty bounds for Gaussian process regression. In *Proceedings of the AAAI Conference on Artificial Intelligence* **35** 7439–7447.

FLAXMAN, S., WILSON, A., NEILL, D., NICKISCH, H. and SMOLA, A. (2015). Fast Kronecker inference in Gaussian processes with non-Gaussian likelihoods. In *International Conference on Machine Learning* 607–616. PMLR.

FRANKS, A. M., D'AMOUR, A. and FELLER, A. (2020). Flexible sensitivity analysis for observational studies without observable implications. *J. Amer. Statist. Assoc.* **115** 1730–1746. MR4189753 https://doi.org/10.1080/01621459.2019.1604369

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3235677

GIFFORDS LAW CENTER TO PREVENT GUN VIOLENCE (2019). https://lawcenter.giffords.org/gun-laws/policy-areas/who-can-have-a-gun/categories-of-prohibited-people/. Accessed on October 10, 2019.

GIFFORDS LAW CENTER TO PREVENT GUN VIOLENCE (2022). Extreme risk protection orders.

GOBILLON, L. and MAGNAC, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Rev. Econ. Stat.* **98** 535–551.

GOLDSTICK, J. E., CUNNINGHAM, R. M. and CARTER, P. M. (2022). Current causes of death in children and adolescents in the United States. *N. Engl. J. Med.* **386** 1955–1956. https://doi.org/10.1056/NEJMc2201761

GOOVAERTS, P. et al. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford Univ. Press on Demand.

GREEN, E. L. and FERNANDEZ, M. (2018). Trump wants to arm teachers. These schools already do. *New York Times*.

GUSTAFSON, P. (2010). Bayesian inference for partially identified models. *Int. J. Biostat.* **6** Art. 17, 20. MR2602560 https://doi.org/10.2202/1557-4679.1206

HAZLETT, C. and XU, Y. (2018). Trajectory balancing: A general reweighting approach to causal inference with time-series cross-sectional data. Available at SSRN 3214231.

HEATON, P. (2010). *Hidden in Plain Sight*: *What Cost-of-Crime Research Can Tell Us About Investing in Police* **279**. Rand Corporation.

HEMENWAY, D. (2017). *Private Guns*, *Public Health*. Univ. Michigan Press.

HENSMAN, J., MATTHEWS, A. and GHAHRAMANI, Z. (2015). Scalable variational Gaussian process classification.

HUANG, B., CHEN, C. and LIU, J. (2019). GPMatch: A Bayesian doubly robust approach to causal inference with Gaussian process covariance function as a matching tool. arXiv preprint arXiv:1901.10359.

HUANG, B., ZHANG, K. and SCHÖLKOPF, B. (2015). Identification of time-dependent causal model: A Gaussian process treatment. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

IMAI, K. and KIM, I. S. (2019). On the use of two-way fixed effects regression models for causal inference with panel data.

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics*, *Social*, *and Biomedical Sciences*: *An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

JYLÄNKI, P., VANHATALO, J. and VEHTARI, A. (2011). Robust Gaussian process regression with a Student-t likelihood. *J. Mach. Learn. Res.* **12** 3227–3257. MR2877599

KAGAWA, R. M. C., CASTILLO-CARNIGLIA, A., VERNICK, J. S., WEBSTER, D., CRIFASI, C., RUDOLPH, K. E., CERDÁ, M., SHEV, A. and WINTEMUTE, G. J. (2018). Repeal of comprehensive background check policies and firearm homicide and suicide. *Epidemiology* **29** 494–502. https://doi.org/10.1097/EDE.0000000000000838

KANAGAWA, M., HENNIG, P., SEJDINOVIC, D. and SRIPERUMBUDUR, B. K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. arXiv preprint arXiv:1807.02582.

KAPLAN, J. (2019). Jacob Kaplan's concatenated files: Uniform crime reporting (UCR) program data: Supplementary homicide reports, 1976–2017. Inter-university Consortium for Political and Social Research [distributor], Ann Arbor, MI, 2019-07-15. Available at https://doi.org/10.3886/E100699V7.

KARCH, J. D., BRANDMAIER, A. M. and VOELKLE, M. C. (2020). Gaussian process panel modeling—machine learning inspired analysis of longitudinal panel data. *Front. Psychol.* **11** 351.

KIM, S., LEE, C. and GUPTA, S. (2020). Bayesian synthetic control methods. *J. Mark. Res.* **57** 831–852.

LI, M. and KONTAR, R. (2020). On negative transfer and structure of latent functions in multi-output Gaussian processes. arXiv preprint arXiv:2004.02382.

LIU, L., WANG, Y. and XU, Y. (2020). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. Available at SSRN 3555463.

LOFSTROM, M., BIRD, M. and MARTIN, B. (2016). *California's Historic Corrections Reforms*. Public Policy Institute of California Sacramento.

LOFSTROM, M. and RAPHAEL, S. (2016). Incarceration and crime: Evidence from California's public safety realignment reform. *Ann. Am. Acad. Polit. Soc. Sci.* **664** 196–220.

LOTT, J. R. JR and MUSTARD, D. B. (1997). Crime, deterrence, and right-to-carry concealed handguns. *J. Leg. Stud.* **26** 1–68.

MCCOURT, A. D., CRIFASI, C. K., STUART, E. A., VERNICK, J. S., KAGAWA, R. M. C., WINTEMUTE, G. J. and WEBSTER, D. W. (2020). Purchaser licensing, point-of-sale background check laws, and firearm homicide and suicide in 4 US states, 1985–2017. *Amer. J. Publ. Health* **110** 1546–1552. https://doi.org/10.2105/AJPH.2020.305822

MENCHETTI, F. and BOJINOV, I. (2020). Estimating causal effects in the presence of partial interference using multivariate Bayesian structural time series models.

MIRATRIX, L. (2020). Using simulation to analyze interrupted time series designs. arXiv preprint arXiv:2002.05746.

MODI, C. and SELJAK, U. (2019). Generative learning of counterfactual for synthetic control applications in econometrics. arXiv preprint arXiv:1910.07178.

NAISH-GUZMAN, A. and HOLDEN, S. (2008). The generalized FITC approximation. In *Advances in Neural Information Processing Systems* 1057–1064.

NATIONAL RESEARCH COUNCIL (2005). *Firearms and Violence: A Critical Review*. National Academies Press, Washington, DC.

OGANISIAN, A. and ROY, J. A. (2021). A practical introduction to Bayesian estimation of causal effects: Parametric and nonparametric approaches. *Stat. Med.* **40** 518–551. MR4194598 https://doi.org/10.1002/sim.8761

PANG, X., LIU, L. and XU, Y. (2020). A Bayesian alternative to synthetic control for comparative case studies. Available at SSRN.

PEAR, V. A., WINTEMUTE, G. J., JEWELL, N. P. and AHERN, J. (2022). Firearm violence following the implementation of California's gun violence restraining order law. *JAMA Netw. Open* **5** e224216. https://doi.org/10.1001/jamanetworkopen.2022.4216

PETEK, G. (2019). The 2019–2020 budget: Analysis of the governor's criminal justice proposals. Sacramento, CA.

PINKNEY, S. (2021). An improved and extended Bayesian synthetic control. arXiv preprint arXiv:2103.16244.

RASMUSSEN, C. E. (2004). Gaussian processes in machine learning. In *Advanced Lectures on Machine Learning* 63–71. Springer.

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

RAY, K. and VAN DER VAART, A. (2020). Semiparametric Bayesian causal inference. *Ann. Statist.* **48** 2999–3020. MR4152632 https://doi.org/10.1214/19-AOS1919

REN, B., WU, X., BRAUN, D., PILLAI, N. and DOMINICI, F. (2021). Bayesian modeling for exposure response curve via Gaussian processes: Causal effects of exposure to air pollution on health outcomes. arXiv preprint arXiv:2105.03454.

RICHARDSON, T. S., EVANS, R. J. and ROBINS, J. M. (2011). Transparent parametrizations of models for potential outcomes. In *Bayesian Statistics* **9** 569–610. Oxford Univ. Press, Oxford. With discussions by Stephen E. Fienberg, Paul Gustafson, Fabrizia Mealli and Fan Li. MR3204019 https://doi.org/10.1093/acprof:oso/9780199694587.003.0019

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

RUBIN, D. B. (1973). Matching to remove bias in observational studies. *Biometrics* 159–183.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

RUDOLPH, K. E., STUART, E. A., VERNICK, J. S. and WEBSTER, D. W. (2015). Association between Connecticut's permit-to-purchase handgun law and homicides. *Amer. J. Publ. Health* **105** e49–e54.

SAMARTSIDIS, P., SEAMAN, S. R., PRESANIS, A. M., HICKMAN, M. and DE ANGELIS, D. (2019). Assessing the causal effect of binary interventions from observational panel data with few treated units. *Statist. Sci.* **34** 486–503. MR4017525 https://doi.org/10.1214/19-STS713

SAMARTSIDIS, P., SEAMAN, S. R., MONTAGNA, S., CHARLETT, A., HICKMAN, M. and DE ANGELIS, D. (2020). A Bayesian multivariate factor analysis model for evaluating an intervention by using observational time series data on multiple outcomes. *J. Roy. Statist. Soc. Ser. A* **183** 1437–1459. MR4157820

SCHULAM, P. and SARIA, S. (2017). Reliable decision support using counterfactual models. In *Advances in Neural Information Processing Systems* 1697–1708.

SHEPPARD, D. (1999). Promising strategies to reduce gun violence. US Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

SIEGEL, M., XUAN, Z., ROSS, C. S., GALEA, S., KALESAN, B., FLEEGLER, E. and GOSS, K. A. (2017). Easiness of legal access to concealed firearm permits and homicide rates in the United States. *Amer. J. Publ. Health* **107** 1923–1929. https://doi.org/10.2105/AJPH.2017.304057

SOLIN, A. and SÄRKKÄ, S. (2020). Hilbert space methods for reduced-rank Gaussian process regression. *Stat. Comput.* **30** 419–446. MR4064629 https://doi.org/10.1007/s11222-019-09886-w

STAN DEVELOPMENT TEAM (2021). Stan modeling language users guide and reference manual. Version 2.26.

SWANSON, J. W., NORKO, M. A., LIN, H.-J., ALANIS-HIRSCH, K., FRISMAN, L. K., BARANOSKI, M. V., EASTER, M. M., ROBERTSON, A. G., SWARTZ, M. S. et al. (2017). Implementation and effectiveness of Connecticut's risk-based gun removal law: Does it prevent suicides. *Law Contemp. Probl.* **80** 179.

TUOMAALA, E. (2019). The Bayesian synthetic control: Improved counterfactual estimation in the social sciences through probabilistic modeling. arXiv preprint arXiv:1910.06106.

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

WEBSTER, D. W. and WINTEMUTE, G. J. (2015). Effects of policies designed to keep firearms from high-risk individuals. *Annu. Rev. Public Health* **36** 21–37. https://doi.org/10.1146/annurev-publhealth-031914-122516

WILSON, A. G., DANN, C. and NICKISCH, H. (2015). Thoughts on massively scalable Gaussian processes. arXiv preprint arXiv:1511.01870.

WINTEMUTE, G. J., BECKETT, L., KASS, P. H., TANCREDI, D., STUDDERT, D., PIERCE, G., BRAGA, A. A., WRIGHT, M. A. and CERDÁ, M. (2017). Evaluation of California's armed and prohibited persons system: Study protocol for a cluster-randomised trial. *Inj. Prev.* **23** 358. https://doi.org/10.1136/injuryprev-2016-042194

WINTEMUTE, G. J., PEAR, V. A., SCHLEIMER, J. P., PALLIN, R., SOHL, S., KRAVITZ-WIRTZ, N. and TOMSICH, E. A. (2019). Extreme risk protection orders intended to prevent mass shootings: A case series. *Ann. Intern. Med.* **171** 655–658. https://doi.org/10.7326/M19-2162

WITTY, S., TAKATSU, K., JENSEN, D. and MANSINGHKA, V. (2020). Causal inference using Gaussian processes with structured latent confounders. In *Proceedings of the* 37*th International Conference on Machine Learning* (H. Daumé III and A. Singh, eds.). *Proceedings of Machine Learning Research* **119** 10313–10323. PMLR.

XU, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Polit. Anal.* **25** 57–76.

ZHU, X. (2005). *Semi-Supervised Learning with Graphs*. Carnegie Mellon Univ.

ZIMRING, F. E. (1972). The medium is the message: Firearm caliber as a determinant of death from assault. *J. Leg. Stud.* **1** 97–123.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1017–1037
https://doi.org/10.1214/22-AOAS1656

ROBUST JOINT MODELLING OF LEFT-CENSORED LONGITUDINAL DATA AND SURVIVAL DATA WITH APPLICATION TO HIV VACCINE STUDIES

BY TINGTING YU[1,a], LANG WU[2,b], JIN QIU[3,c] AND PETER B. GILBERT[4,d]

[1]*Harvard Pilgrim Health Care Institute and Harvard Medical School,* [a]*tingting_yu@harvardpilgrim.org*
[2]*Department of Statistics, University of British Columbia,* [b]*lang@stat.ubc.ca*
[3]*Department of Statistics, Zhejiang University of Finance and Economics,* [c]*qiujin@zufe.edu.cn*
[4]*Department of Biostatistics, University of Washington,* [d]*pgilbert@scharp.org*

In jointly modelling longitudinal and survival data, the longitudinal data may be complex in the sense that they may contain outliers and may be left censored. Motivated from an HIV vaccine study, we propose a robust method for joint models of longitudinal and survival data, where the outliers in longitudinal data are addressed using a multivariate t-distribution for b-outliers and using an M-estimator for e-outliers. We also propose a computationally efficient method for approximate likelihood inference. The proposed method is evaluated by simulation studies. Based on the proposed models and method, we analyze the HIV vaccine data and find a strong association between longitudinal biomarkers and the risk of HIV infection.

REFERENCES

BARRETT, J., DIGGLE, P., HENDERSON, R. and TAYLOR-ROBINSON, D. (2015). Joint modelling of repeated measurements and time-to-event outcomes: Flexible model specification and exact likelihood inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 131–148. MR3299402 https://doi.org/10.1111/rssb.12060

CANTONI, E. and RONCHETTI, E. (2001). Robust inference for generalized linear models. *J. Amer. Statist. Assoc.* **96** 1022–1030. MR1947250 https://doi.org/10.1198/016214501753209004

COPT, S. and VICTORIA-FESER, M.-P. (2006). High-breakdown inference for mixed linear models. *J. Amer. Statist. Assoc.* **101** 292–300. MR2268046 https://doi.org/10.1198/016214505000000772

ELASHOFF, R. M., LI, G. and LI, N. (2015). *Joint Modeling of Longitudinal and Time-to-Event Data*. CRC Press, Boca Raton.

FLYNN, N., FORTHAL, D., HARRO, C., JUDSON, F., MAYER, K., PARA, M., GILBERT, P. and THE RGP120 HIV VACCINE STUDY GROUP (2005). Placebo-controlled phase 3 trial of recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J. Infect. Dis.* **191** 654–65.

GILL, P. S. (2000). A robust mixed linear model analysis for longitudinal data. *Stat. Med.* **19** 975–987.

HSIEH, F., TSENG, Y.-K. and WANG, J.-L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics* **62** 1037–1043. MR2297674 https://doi.org/10.1111/j.1541-0420.2006.00570.x

HUGHES, J. P. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics* **55** 625–629.

KOLLER, M. (2016). robustlmm: An R package for robust estimation of linear mixed-effects models. *J. Stat. Softw.* **75** 1–24.

KONG, S. and NAN, B. (2016). Semiparametric approach to regression with a covariate subject to a detection limit. *Biometrika* **103** 161–174. MR3465828 https://doi.org/10.1093/biomet/asv055

LANGE, K. L., LITTLE, R. J. A. and TAYLOR, J. M. G. (1989). Robust statistical modeling using the t distribution. *J. Amer. Statist. Assoc.* **84** 881–896. MR1134486

LEE, Y. and NELDER, J. A. (1996). Hierarchical generalized linear models. *J. Roy. Statist. Soc. Ser. B* **58** 619–678. MR1410182

LEE, Y., NELDER, J. A. and PAWITAN, Y. (2018). *Generalized Linear Models with Random Effects*: *Unified Analysis via H-Likelihood* **153**. CRC Press, Boca Raton.

LUCAS, A. (1997). Robustness of the Student t based M-estimator. *Comm. Statist. Theory Methods* **26** 1165–1182. MR1450228 https://doi.org/10.1080/03610929708831974

QIN, G., ZHANG, J., ZHU, Z. and FUNG, W. (2016). Robust estimation of partially linear models for longitudinal data with dropouts and measurement error. *Stat. Med.* **35** 5401–5416. MR3573066 https://doi.org/10.1002/sim.7062

Key words and phrases. Biomarker, outliers, robust joint model, h-likelihood, left censoring.

RIZOPOULOS, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data*: *With Applications in R*. CRC Press, Boca Raton.

SINHA, S. K. (2004). Robust analysis of generalized linear mixed models. *J. Amer. Statist. Assoc.* **99** 451–460. MR2062830 https://doi.org/10.1198/016214504000000340

TAYLOR, J. M. G., PARK, Y., ANKERST, D. P., PROUST-LIMA, C., WILLIAMS, S., KESTIN, L., BAE, K., PICKLES, T. and SANDLER, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* **69** 206–213. MR3058067 https://doi.org/10.1111/j.1541-0420.2012.01823.x

WATERNAUX, C., LAIRD, N. M. and WARE, J. H. (1989). Methods for analysis of longitudinal data: Blood-lead concentrations and cognitive development. *J. Amer. Statist. Assoc.* **84** 33–41.

WU, L. (2009). *Mixed Effects Models for Complex Data*. CRC Press, Boca Raton.

WU, L. and QIU, J. (2011). Approximate bounded influence estimation for longitudinal data with outliers and measurement errors. *J. Statist. Plann. Inference* **141** 2321–2330. MR2775211 https://doi.org/10.1016/j.jspi.2011.01.021

WU, L. and YU, T. (2016). Joint modeling of longitudinal and survival data. In *Wiley StatsRef*: *Statistics Reference Online* 1–9. American Cancer Society.

YU, T., WU, L. and GILBERT, P. B. (2018). A joint model for mixed and truncated longitudinal data and survival data, with application to HIV vaccine studies. *Biostatistics* **19** 374–390. MR3815177 https://doi.org/10.1093/biostatistics/kxx047

YU, T., WU, L., QIU, J. and GILBERT, P. B. (2023). Supplement to “Robust joint modelling of left-censored longitudinal data and survival data with application to HIV vaccine studies.” https://doi.org/10.1214/22-AOAS1656SUPPA, https://doi.org/10.1214/22-AOAS1656SUPPB

ZHENG, X., FUNG, W. K. and ZHU, Z. (2013). Robust estimation in joint mean-covariance regression model for longitudinal data. *Ann. Inst. Statist. Math.* **65** 617–638. MR3094949 https://doi.org/10.1007/s10463-012-0383-8

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1038–1063
https://doi.org/10.1214/22-AOAS1657

BAYESIAN DECISION THEORY FOR TREE-BASED ADAPTIVE SCREENING TESTS WITH AN APPLICATION TO YOUTH DELINQUENCY

BY CHELSEA KRANTSEVICH[1,a], P. RICHARD HAHN[1,b], YI ZHENG[1,c] AND CHARLES KATZ[2,d]

[1]*School of Mathematical and Statistical Sciences, Arizona State University,* [a]*chelsea.krantsevich@asu.edu,* [b]*prhahn@asu.edu,* [c]*yi.isabel.zheng@asu.edu*

[2]*School of Criminology and Criminal Justice, Arizona State University,* [d]*charles.katz@asu.edu*

Crime prevention strategies based on early intervention depend on accurate risk assessment instruments for identifying high-risk youth. It is important in this context that the instruments be convenient to administer, which means, in particular, that they should also be reasonably brief; adaptive screening tests are useful for this purpose. Adaptive tests constructed using classification and regression trees are becoming a popular alternative to traditional item response theory (IRT) approaches for adaptive testing. However, tree-based adaptive tests lack a principled criterion for terminating the test. This paper develops a Bayesian decision theory framework for measuring the trade-off between brevity and accuracy when considering tree-based adaptive screening tests of different lengths. We also present a novel method for designing tree-based adaptive tests, motivated by this framework. The framework and associated adaptive test method are demonstrated through an application to youth delinquency risk assessment in Honduras; it is shown that an adaptive test requiring a subject to answer fewer than 10 questions can identify high-risk youth nearly as accurately as an unabridged survey containing 173 items.

REFERENCES

ABT, T. and WINSHIP, C. (2016). What Works in Reducing Community Violence: A Meta-Review and Field Study for the Northern Triangle. Technical Report, Democracy International, 7600 Wisconsin Avenue, Suite 1010 Bethesda, MD 20814. Available at https://www.usaid.gov/sites/default/files/USAID-2016-What-Works-in-Reducing-Community-Violence-Final-Report.pdf.

ALMOND, R. G. and MISLEVY, R. J. (1998). Graphical models and computerized adaptive testing. *ETS Res. Rep. Ser.* **1998** i–24.

ARTHUR, M. W., HAWKINS, J. D., POLLARD, J. A., CATALANO, R. F. and BAGLIONI, A. J. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors. The Communities That Care Youth Survey. *Eval. Rev.* **26** 575–601. https://doi.org/10.1177/0193841X0202600601

ARTHUR, M. W., BRINEY, J. S., HAWKINS, J. D., ABBOTT, R. D., BROOKE-WEISS, B. L. and CATALANO, R. F. (2007). Measuring risk and protection in communities using the Communities That Care Youth Survey. *Eval. Program Plann.* **30** 197–211. https://doi.org/10.1016/j.evalprogplan.2007.01.009

BASHIR, A., CARVALHO, C. M., HAHN, P. R. and JONES, M. B. (2019). Post-processing posteriors over precision matrices to produce sparse graph estimates. *Bayesian Anal.* **14** 1075–1090. MR4044846 https://doi.org/10.1214/18-BA1139

BERK-SELIGSON, S., ORCÉS, D., PIZZOLITTO, G., SELIGSON, M. A. and WILSON, C. (2014). Impact Evaluation: Honduras Country Report. Technical Report, The Latin American Public Opinion Project (LAPOP), Vanderbilt Univ., Nashville, TN.

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32. https://doi.org/10.1023/a:1010933404324

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees. Wadsworth Statistics/Probability Series*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

Key words and phrases. Classification trees, computerized adaptive diagnostics, computerized adaptive testing, Bayesian decision theory, risk factors, and risk assessment.

CHANG, H.-H. (2004). Understanding computerized adaptive testing: From Robbins-Monro to Lord and beyond. In *The SAGE Handbook of Quantitative Methodology for the Social Sciences* 118–135. SAGE Publications, Inc., Thousand Oaks, CA.

CHANG, H.-H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika* **80** 1–20. MR3325843 https://doi.org/10.1007/s11336-014-9401-5

CHAWLA, N. V., BOWYER, K. W., HALL, L. O. and KEGELMEYER, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *J. Artificial Intelligence Res.* **16** 321–357.

CHOULDECHOVA, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* **5** 153–163. https://doi.org/10.1089/big.2016.0047

CHOULDECHOVA, A. and LUM, K. (2020). The present and future risk of AI in pre-trial risk assessments. Technical Report, Safety & Justice Challenge.

CHOULDECHOVA, A., BENAVIDES-PRADO, D., FIALKO, O. and VAITHIANATHAN, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Proceedings of the* 1*st Conference on Fairness*, *Accountability and Transparency* **81** 134–148.

DE AYALA, R. J. (2009). *The Theory and Practice of Item Response Theory*. *Methodology in the Social Sciences*. Guilford, New York.

DELGADO-GOMEZ, D., BACA-GARCIA, E., AGUADO, D., COURTET, P. and LOPEZ-CASTROMAN, J. (2016). Computerized adaptive test vs. decision trees: Development of a support decision system to identify suicidal behavior. *J. Affective Disorders* **206** 204–209. https://doi.org/10.1016/j.jad.2016.07.032

ECKHOUSE, L., LUM, K., CONTI-COOK, C. and CICCOLINI, J. (2018). Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior* **46** 185–209.

EMBRETSON, S. E. and REISE, S. P. (2000). *Item Response Theory for Psychologists*. *Multivariate Applications Books Series*. Lawrence Erlbaum Associates, Mahwah, NJ, US.

FREY, A. and SEITZ, N.-N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation* **35** 89–94.

GABRY, J., SIMPSON, D., VEHTARI, A., BETANCOURT, M. and GELMAN, A. (2019). Visualization in Bayesian workflow. *J. Roy. Statist. Soc. Ser. A* **182** 389–402. MR3902665 https://doi.org/10.1111/rssa.12378

GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–807. MR1422404

GIBBONS, R. D. and WANG, J. (2019). Personal communication.

GIBBONS, R. D., HOOKER, G., FINKELMAN, M. D., WEISS, D. J., PILKONIS, P. A., FRANK, E., MOORE, T. and KUPFER, D. J. (2013). The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD). *J. Clin. Psychiatry* **74** 669–674.

GIBBONS, R. D., WEISS, D. J., FRANK, E. and KUPFER, D. (2016). Computerized adaptive diagnosis and testing of mental health disorders. *Annu. Rev. Clin. Psychol.* **12** 83–104.

HAHN, P. R. and CARVALHO, C. M. (2015). Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective. *J. Amer. Statist. Assoc.* **110** 435–448. MR3338514 https://doi.org/10.1080/01621459.2014.993077

HAMBLETON, R. K., SWAMINATHAN, H. and ROGERS, H. J. (1991). *Fundamentals of Item Response Theory*. Sage, Thousand Oaks.

HAWKINS, J. D., CATALANO, R. F. and MILLER, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychol. Bull.* **112** 64–105. https://doi.org/10.1037/0033-2909.112.1.64

HE, J., YALOV, S. and HAHN, P. R. (2019). XBART: Accelerated Bayesian Additive Regression Trees. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* **89** 1130–1138.

HENNIGAN, K. M., MAXSON, C. L., SLOANE, D. C., KOLNICK, K. A. and VINDEL, F. (2014). Identifying high-risk youth for secondary gang prevention. *Journal of Crime and Justice* **37** 104–128.

HENNIGAN, K. M., KOLNICK, K. A., VINDEL, F. and MAXSON, C. L. (2015). Targeting youth at risk for gang involvement: Validation of a gang risk assessment to support individualized secondary prevention. *Child. Youth Serv. Rev.* **56** 86–96.

HIGGINSON, A., BENIER, K., SHENDEROVICH, Y., BEDFORD, L., MAZEROLLE, L. and MURRAY, J. (2018). Factors associated with youth gang membership in low- and middle-income countries: A systematic review. *Campbell Systematic Reviews* **14** 1–128.

HOWELL, J. C. and EGLEY JR., A. (2005). Moving risk factors into developmental theories of gang membership. *Youth Violence and Juvenile Justice* **3** 334–354.

JOHNDROW, J. E. and LUM, K. (2019). An algorithm for removing sensitive information: Application to race-independent recidivism prediction. *Ann. Appl. Stat.* **13** 189–220. MR3937426 https://doi.org/10.1214/18-AOAS1201

KATZ, C. M. and FOX, A. M. (2010). Risk and protective factors associated with gang-involved youth in Trinidad and Tobago. *Rev. Panam. Salud. Publica* **27** 187–202. https://doi.org/10.1590/s1020-49892010000300006

KATZ, C. M., CHEON, H., HEDBERG, E. C. and DECKER, S. H. (2021). Impact of family-based secondary prevention programming on risk, resilience, and delinquency: A 6-month follow up within a randomized control trial in Honduras. *Justice Q.* 1–26.

KRANTSEVICH, C., HAHN, P. R., ZHENG, Y. and KATZ, C. (2023). Supplement to "Bayesian decision theory for tree-based adaptive screening tests with an application to youth delinquency." https://doi.org/10.1214/22-AOAS1657SUPPA, https://doi.org/10.1214/22-AOAS1657SUPPB

LOH, W.-Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery* **1** 14–23.

MAGUIRE, E. R., WELLS, W. and KATZ, C. M. (2011). Measuring community risk and protective factors for adolescent problem behaviors: Evidence from a developing nation. *Journal of Research in Crime and Delinquency* **48** 594–620.

MEYER, P. J. (2019). U.S. Strategy for Engagement in Central America: Policy Issues for Congress. CRS Report No. R44812, Congressional Research Service. Available at https://crsreports.congress.gov/product/pdf/R/R44812.

MICHEL, P., BAUMSTARCK, K., LOUNDOU, A., GHATTAS, B., AUQUIER, P. and BOYER, L. (2018). Computerized adaptive testing with decision regression trees: An alternative to item response theory for quality of life measurement in multiple sclerosis. *Patient Preference and Adherence* **12** 1043–1053.

MILBORROW, S. (2021). rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'. R package version 3.1.0. Available at https://CRAN.R-project.org/package=rpart.plot.

MURRAY, J. S. (2021). Log-linear Bayesian additive regression trees for multinomial logistic and count regression models. *J. Amer. Statist. Assoc.* **116** 756–769. MR4270022 https://doi.org/10.1080/01621459.2020.1813587

MURRAY, J. S., DUNSON, D. B., CARIN, L. and LUCAS, J. E. (2013). Bayesian Gaussian copula factor models for mixed data. *J. Amer. Statist. Assoc.* **108** 656–665. MR3174649 https://doi.org/10.1080/01621459.2012.762328

MURRAY, J., SHENDEROVICH, Y., GARDNER, F., MIKTON, C., DERZON, J. H., LIU, J. and EISNER, M. (2018). Risk factors for antisocial behavior in low- and middle-income countries: A systematic review of longitudinal studies. *Crime and Justice* **47** 255–364.

PAAP, M. C. S., KROEZE, K. A., GLAS, C. A. W., TERWEE, C. B., VAN DER PALEN, J. and VELDKAMP, B. P. (2017). Measuring patient-reported outcomes adaptively: Multidimensionality matters! *Appl. Psychol. Meas.* **42** 327–342.

PARMIGIANI, G. and INOUE, L. Y. T. (2009). *Decision Theory: Principles and Approaches. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2604978 https://doi.org/10.1002/9780470746684

PUELZ, D., HAHN, P. R. and CARVALHO, C. M. (2017). Variable selection in seemingly unrelated regressions with random predictors. *Bayesian Anal.* **12** 969–989. MR3724975 https://doi.org/10.1214/17-BA1053

RUDNER, L. M. (2010). Demystifying the GMAT: Computer adaptive testing. *Graduate Management Admission Council: Deans Digest*.

SANDS, W. A., WATERS, B. K. and MCBRIDE, J. R. (1997). *Computerized Adaptive Testing: From Inquiry to Operation*. American Psychological Association.

UNODC (2018). UNODC Statistics. Available at https://dataunodc.un.org/.

VAN DER LINDEN, W. J. (2008). Some new developments in adaptive testing technology. *Zeitschrift Für Psychologie / Journal of Psychology* **216** 3–11.

VAN DER LINDEN, W. J. and HAMBLETON, R. K. (1997). Item response theory: Brief history, common models, and extensions. In *Handbook of Modern Item Response Theory* 1–28. Springer, New York.

WAINER, H. (2000). *Computerized Adaptive Testing: A Primer*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

WANG, C. and CHANG, H.-H. (2011). Item selection in multidimensional computerized adaptive testing—gaining information from different angles. *Psychometrika* **76** 363–384. MR2823005 https://doi.org/10.1007/s11336-011-9215-7

WANG, C., CHANG, H.-H. and BOUGHTON, K. A. (2012). Deriving stopping rules for multidimensional computerized adaptive testing. *Appl. Psychol. Meas.* **37** 99–122.

WANG, M. and HAHN, P. R. (2021). Accelerated Bayesian additive regression trees for fast multi-class classification. Preprint.

WEBB, V. J., NUÑO, L. E. and KATZ, C. (2016). Influence of Risk and Protective Factors on School-aged Youth Involvement with Gangs, Guns, and Delinquency: Findings from the El Salvador Youth Survey. Technical Report, Center for Violence Prevention and Community Safety, Arizona State Univ.

WEERMAN, F. M., MAXSON, C. L., ESBENSEN, F.-A., ALDRIDGE, J., MEDINA, J. and VAN GEMERT, F. (2009). Eurogang Program Manual. Technical Report, Univ. Missouri at St Louis, St Louis, MO.

WOODY, S., CARVALHO, C. M. and MURRAY, J. S. (2021). Model interpretation through lower-dimensional posterior summarization. *J. Comput. Graph. Statist.* **30** 144–161. MR4235972 https://doi.org/10.1080/10618600.2020.1796684

Yao, L., Pommerich, M. and Segall, D. O. (2014). Using multidimensional CAT to administer a short, yet precise, screening test. *Appl. Psychol. Meas.* **38** 614–631.

Zheng, Y., Cheon, H. and Katz, C. M. (2020). Using machine learning methods to develop a short tree-based adaptive classification test: Case study with a high dimensional item pool and imbalanced data. *Appl. Psychol. Meas.* **44** 499–514.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1064–1085
https://doi.org/10.1214/22-AOAS1658

BAYESIAN COX REGRESSION FOR LARGE-SCALE INFERENCE WITH APPLICATIONS TO ELECTRONIC HEALTH RECORDS

BY ALEXANDER WOLFGANG JUNG[1,a] AND MORITZ GERSTUNG[2,b]

[1]*European Molecular Biology Laboratory, European Bioinformatics Institute EMBL-EBI,* [a]*alexwjung@ebi.ac.uk*
[2]*Division of AI in Oncology, German Cancer Research Centre DKFZ,* [b]*moritz.gerstung@dkfz.de*

The Cox model is an indispensable tool for time-to-event analysis, particularly in biomedical research. However, medicine is undergoing a profound transformation, generating data at an unprecedented scale, which opens new frontiers to study and understand diseases. With the wealth of data collected, new challenges for statistical inference arise, as datasets are often high dimensional, exhibit an increasing number of measurements at irregularly spaced time points, and are simply too large to fit in memory. Many current implementations for time-to-event analysis are ill-suited for these problems, as inference is computationally demanding and requires access to the full data at once. Here, we propose a Bayesian version for the counting process representation of Cox's partial likelihood for efficient inference on large-scale datasets with millions of data points and thousands of time-dependent covariates. Through the combination of stochastic variational inference and a reweighting of the log-likelihood, we obtain an approximation for the posterior distribution that factorizes over subsamples of the data, enabling the analysis in big data settings. Crucially, the method produces viable uncertainty estimates for large-scale and high-dimensional datasets. We show the utility of our method through a simulation study and an application to myocardial infarction in the UK Biobank, where we characterize the multivariate effects of risk factors and replicate results from individual studies. Our framework extends the Cox model to new data sources, like biobanks and EHR, the combination of which can provide new insights into our understanding of diseases.

REFERENCES

ALVARES, D., LÁZARO, E., GÓMEZ-RUBIO, V. and ARMERO, C. (2021). Bayesian survival analysis with BUGS. *Stat. Med.* **40** 2975–3020. MR4255788 https://doi.org/10.1002/sim.8933

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes. Springer Series in Statistics*. Springer, New York. MR1198884 https://doi.org/10.1007/978-1-4612-4348-9

ANDERSEN, P. K., POHAR PERME, M., VAN HOUWELINGEN, H. C., COOK, R. J., JOLY, P., MARTINUSSEN, T., TAYLOR, J. M. G., ABRAHAMOWICZ, M. and THERNEAU, T. M. (2021). Analysis of time-to-event for observational studies: Guidance to the use of intensity models. *Stat. Med.* **40** 185–211. MR4194578 https://doi.org/10.1002/sim.8757

BREHENY, P. and HUANG, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **5** 232–253. MR2810396 https://doi.org/10.1214/10-AOAS388

CLIFT, A. K., COUPLAND, C. A. C., KEOGH, R. H., DIAZ-ORDAZ, K., WILLIAMSON, E., HARRISON, E. M., HAYWARD, A., HEMINGWAY, H., HORBY, P. et al. (2020). Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: National derivation and validation cohort study. *BMJ Clin. Res. Ed.* **371** m3731. https://doi.org/10.1136/bmj.m3731

COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

Key words and phrases. Survival analysis, time-dependent covariates, high-dimensional data, variational inference, batch learning.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276. MR0400509 https://doi.org/10.1093/biomet/62.2.269

DAWBER, T. R., MEADORS, G. F. and MOORE, F. E. (1951). Epidemiological approaches to heart disease: The Framingham Study. *Am. J. Public Health Nation's Health* **41** 279–281. https://doi.org/10.2105/ajph.41.3.279

EGEBERG, A., HANSEN, P. R., GISLASON, G. H. and THYSSEN, J. P. (2016). Assessment of the risk of cardiovascular disease in patients with rosacea. *J. Am. Acad. Dermatol.* **75** 336–339. https://doi.org/10.1016/j.jaad.2016.02.1158

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.

HANS, C. (2009). Bayesian lasso regression. *Biometrika* **96** 835–845. MR2564494 https://doi.org/10.1093/biomet/asp047

HIPPISLEY-COX, J. and COUPLAND, C. (2021). Predicting the risk of prostate cancer in asymptomatic men: A cohort study to develop and validate a novel algorithm. *Br. J. Gen. Pract.* **71** e364–e371. https://doi.org/10.3399/bjgp20X714137

HIPPISLEY-COX, J., COUPLAND, C. and BRINDLE, P. (2017). Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study. *BMJ* **357** j2099. https://doi.org/10.1136/bmj.j2099

HJORT, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294. MR1062708 https://doi.org/10.1214/aos/1176347749

IBRAHIM, J. G., CHEN, M.-H. and SINHA, D. (2001). *Bayesian Survival Analysis. Springer Series in Statistics*. Springer, New York. MR1876598 https://doi.org/10.1007/978-1-4757-3447-8

JUNG, A. W. and GERSTUNG, M. (2023). Supplement to "Bayesian Cox regression for large-scale inference with applications to electronic health records." https://doi.org/10.1214/22-AOAS1658SUPPA, https://doi.org/10.1214/22-AOAS1658SUPPB

KALBFLEISCH, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *J. Roy. Statist. Soc. Ser. B* **40** 214–221. MR0517442

KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60** 267–278. MR0326939 https://doi.org/10.1093/biomet/60.2.267

KUCUKELBIR, A., TRAN, D., RANGANATH, R., GELMAN, A. and BLEI, D. M. (2017). Automatic differentiation variational inference. *J. Mach. Learn. Res.* **18** Paper No. 14, 45 pp. MR3634881

KVAMME, H., BORGAN, Ø. and SCHEEL, I. (2019). Time-to-event prediction with neural networks and Cox regression. *J. Mach. Learn. Res.* **20** Paper No. 129, 30 pp. MR4002883

LAUD, P. W., DAMIEN, P. and SMITH, A. F. M. (1998). Bayesian nonparametric and covariate analysis of failure time data. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (D. Dey, P. Müller and D. Sinha, eds.). *Lect. Notes Stat.* **133** 213–225. Springer, New York. MR1630083 https://doi.org/10.1007/978-1-4612-1732-9_11

LEWANDOWSKI, D., KUROWICKA, D. and JOE, H. (2009). Generating random correlation matrices based on vines and extended onion method. *J. Multivariate Anal.* **100** 1989–2001. MR2543081 https://doi.org/10.1016/j.jmva.2009.04.008

LI, R., CHANG, C., JUSTESEN, J. M., TANIGAWA, Y., QIAN, J., HASTIE, T., RIVAS, M. A. and TIBSHIRANI, R. (2022). Corrigendum to: Fast Lasso method for large-scale and ultrahigh-dimensional Cox model with applications to UK Biobank. *Biostatistics* **23** 683. MR4409773 https://doi.org/10.1093/biostatistics/kxab019

MILLETT, E. R. C., PETERS, S. A. E. and WOODWARD, M. (2018). Sex differences in risk factors for myocardial infarction: Cohort study of UK Biobank participants. *BMJ* **363** k4247. https://doi.org/10.1136/bmj.k4247

MITTAL, S., MADIGAN, D., BURD, R. S. and SUCHARD, M. A. (2014). High-dimensional, massive sample-size Cox proportional hazards regression for survival analysis. *Biostatistics* **15** 207–221. https://doi.org/10.1093/biostatistics/kxt043

MOHAMED, S., ROSCA, M., FIGURNOV, M. and MNIH, A. (2020). Monte Carlo gradient estimation in machine learning. *J. Mach. Learn. Res.* **21** Paper No. 132, 62 pp. MR4138116

MORTENSEN, M. B. and NORDESTGAARD, B. G. (2020). Elevated LDL cholesterol and increased risk of myocardial infarction and atherosclerotic cardiovascular disease in individuals aged 70–100 years: A contemporary primary prevention cohort. *Lancet* **396** 1644–1652. https://doi.org/10.1016/S0140-6736(20)32233-9

NIKOOIENEJAD, A., WANG, W. and JOHNSON, V. E. (2020). Bayesian variable selection for survival data using inverse moment priors. *Ann. Appl. Stat.* **14** 809–828. MR4117831 https://doi.org/10.1214/20-AOAS1325

PARK, T. and CASELLA, G. (2008). The Bayesian lasso. *J. Amer. Statist. Assoc.* **103** 681–686. MR2524001 https://doi.org/10.1198/016214508000000337

QIOU, Z., RAVISHANKER, N. and DEY, D. K. (1999). Multivariate survival analysis with positive stable frailties. *Biometrics* **55** 637–644. https://doi.org/10.1111/j.0006-341x.1999.00637.x

RANGANATH, R., GERRISH, S. and BLEI, D. (2014). Black box variational inference. In *Artificial Intelligence and Statistics* 814–822. PMLR.

SHAREF, E., STRAWDERMAN, R. L., RUPPERT, D., COWEN, M. and HALASYAMANI, L. (2010). Bayesian adaptive B-spline estimation in proportional hazards frailty models. *Electron. J. Stat.* **4** 606–642. MR2660535 https://doi.org/10.1214/10-EJS566

SHIN, M., BHATTACHARYA, A. and JOHNSON, V. E. (2018). Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statist. Sinica* **28** 1053–1078. MR3791100

SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **39** 1–13. https://doi.org/10.18637/jss.v039.i05

SINHA, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *J. Amer. Statist. Assoc.* **88** 979–983. https://doi.org/10.2307/2290789

SINHA, D., IBRAHIM, J. G. and CHEN, M.-H. (2003). A Bayesian justification of Cox's partial likelihood. *Biometrika* **90** 629–641. MR2006840 https://doi.org/10.1093/biomet/90.3.629

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., DOWNEY, P., ELLIOTT, P., GREEN, J. et al. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12** e1001779. https://doi.org/10.1371/journal.pmed.1001779

SYLVESTRE, M.-P. and ABRAHAMOWICZ, M. (2008). Comparison of algorithms to generate event times conditional on time-dependent covariates. *Stat. Med.* **27** 2618–2634. MR2440055 https://doi.org/10.1002/sim.3092

TARKHAN, A. and SIMON, N. (2020). BigSurvSGD: Big survival data analysis via stochastic gradient descent. Preprint. Available at arXiv:2003.00116.

THERNEAU, T. M. (2021). A package for survival analysis in R.

THERNEAU, T. M. and GRAMBSCH, P. M. (2000). The Cox model. In *Modeling Survival Data: Extending the Cox Model* 39–77. Springer, New York.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

TIBSHIRANI, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* **16** 385–395. https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

WANG, Y., HONG, C., PALMER, N., DI, Q., SCHWARTZ, J., KOHANE, I. and CAI, T. (2021). A fast divide-and-conquer sparse Cox regression. *Biostatistics* **22** 381–401. MR4246896 https://doi.org/10.1093/biostatistics/kxz036

WILLIAMSON, E. J., WALKER, A. J., BHASKARAN, K., BACON, S., BATES, C., MORTON, C. E., CURTIS, H. J., MEHRKAR, A., EVANS, D. et al. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584** 430–436. https://doi.org/10.1038/s41586-020-2521-4

WITTEN, D. M. and TIBSHIRANI, R. (2010). Survival analysis with high-dimensional covariates. *Stat. Methods Med. Res.* **19** 29–51. MR2744491 https://doi.org/10.1177/0962280209105024

YANG, Y. and ZOU, H. (2013). A cocktail algorithm for solving the elastic net penalized Cox's regression in high dimensions. *Stat. Interface* **6** 167–173. MR3066682 https://doi.org/10.4310/SII.2013.v6.n2.a1

YUSUF, S., JOSEPH, P., RANGARAJAN, S., ISLAM, S., MENTE, A., HYSTAD, P., BRAUER, M., KUTTY, V. R., GUPTA, R. et al. (2020). Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): A prospective cohort study. *Lancet* **395** 795–808. https://doi.org/10.1016/S0140-6736(19)32008-2

ZHANG, H. H. and LU, W. (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika* **94** 691–703. MR2410017 https://doi.org/10.1093/biomet/asm037

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1086–1104
https://doi.org/10.1214/22-AOAS1660

CYTOPT: OPTIMAL TRANSPORT WITH DOMAIN ADAPTATION FOR INTERPRETING FLOW CYTOMETRY DATA

BY PAUL FREULON[1,a], JÉRÉMIE BIGOT[1,b] AND BORIS P. HEJBLUM[2,c]

[1]*Université de Bordeaux, Institut de Mathématiques de Bordeaux et CNRS (UMR 5251),* [a]*paul.freulon@math.u-bordeaux.fr,* [b]*jeremie.bigot@math.u-bordeaux.fr*

[2]*Université de Bordeaux, Bordeaux Population Health Research Center Inserm U1219, Inria SISTM,* [c]*boris.hejblum@u-bordeaux.fr*

The automated analysis of flow cytometry measurements is an active research field. We introduce a new algorithm, referred to as `CytOpT`, using regularized optimal transport to directly estimate the different cell population proportions from a biological sample characterized with flow cytometry measurements. We rely on the regularized Wasserstein metric to compare cytometry measurements from different samples, thus accounting for possible misalignment of a given cell population across samples (due to technical variability from the technology of measurements). In this work we rely on a supervised learning technique, based on the Wasserstein metric, that is used to estimate an optimal reweighting of class proportions in a mixture model from a source distribution (with known segmentation into cell sub-populations) to fit a target distribution with unknown segmentation. Due to the high dimensionality of flow cytometry data, we use stochastic algorithms to approximate the regularized Wasserstein metric to solve the optimization problem involved in the estimation of optimal weights representing the cell population proportions in the target distribution. Several flow cytometry data sets are used to illustrate the performances of `CytOpT` that are also compared to those of existing algorithms for automatic gating based on supervised learning.

REFERENCES

AGHAEEPOUR, N., NIKOLIC, R., HOOS, H. H. and BRINKMAN, R. R. (2011). Rapid cell population identification in flow cytometry data. *Cytometry*, *Part A* **79** 6–13.

AGHAEEPOUR, N., FINAK, G., HOOS, H., MOSMANN, T. R., BRINKMAN, R., GOTTARDO, R., SCHEUERMANN, R. H., CONSORTIUM, F., CONSORTIUM, D. et al. (2013). Critical assessment of automated flow cytometry data analysis techniques. *Nat*. *Methods* **10** 228.

ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein gan. Preprint. Available at arXiv:1701.07875.

BALLU, M., BERTHET, Q. and BACH, F. (2020). Stochastic optimization for regularized Wasserstein estimators. Preprint. Available at arXiv:2002.08695.

BERCU, B. and BIGOT, J. (2021). Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures. *Ann*. *Statist*. **49** 968–987. MR4255115 https://doi.org/10.1214/20-aos1987

BLAND, J. and ALTMAN, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327** 307–310.

COMMENGES, D., ALKHASSIM, C., GOTTARDO, R., HEJBLUM, B. P. and THIÉBAUT, R. (2018). Cytometree: A binary tree algorithm for automatic gating in cytometry analysis. *Cytometry*, *Part A* **93** 1132–1140.

CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* 2292–2300.

DEL BARRIO, E., INOUZHE, H., LOUBES, J., MATRÁN, C. and MAYO-ÍSCAR, A. (2019). OptimalFlow: Optimal-transport approach to flow cytometry gating and population matching. Preprint. Available at arXiv:1907.08006.

DOST, B., WU, C., SU, A. and BAFNA, V. (2010). TCLUST: A fast method for clustering genome-scale expression data. *IEEE/ACM Trans*. *Comput*. *Biol*. *Bioinform*. **8** 808–818.

Key words and phrases. Automatic gating, flow cytometry, optimal transport, stochastic optimization.

FEYDY, J., SÉJOURNÉ, T., VIALARD, F., AMARI, S., TROUVÉ, A. and PEYRÉ, G. (2018). Interpolating between optimal transport and MMD using Sinkhorn divergences. Preprint. Available at arXiv:1810.08278.

FLAMARY, R., CUTURI, M., COURTY, N. and RAKOTOMAMONJY, A. (2018). Wasserstein discriminant analysis. *Mach. Learn.* **107** 1923–1945. MR3854402 https://doi.org/10.1007/s10994-018-5717-1

FREULON, P., BIGOT, J. and HEJBLUM, B. P (2023). Supplement to "CytOpT: Optimal transport with domain adaptation for interpreting flow cytometry data." https://doi.org/10.1214/22-AOAS1660SUPPA, https://doi.org/10.1214/22-AOAS1660SUPPB

GE, Y. and SEALFON, S. C. (2012). FlowPeaks: A fast unsupervised clustering for flow cytometry data via K-means and density peak finding. *Bioinformatics* **28** 2052–2058. https://doi.org/10.1093/bioinformatics/bts300

GENEVAY, A., CUTURI, M., PEYRÉ, G. and BACH, F. (2016). Stochastic optimization for large-scale optimal transport. In *Advances in Neural Information Processing Systems* 3440–3448.

GOTTARDO, V. B. R., KLEINSTEIN, S. H., DAVIS, M. M., HAFLER, D. A., QUILL, H., PALUCKA, A. K., POLAND, G. A., PULENDRAN, B., REINHERZ, E. L. et al. (2014). Computational resources for high-dimensional immune analysis from the human immunology project consortium. *Nat. Biotechnol.* **32** 146.

HAHNE, F., KHODABAKHSHI, A. H., BASHASHATI, A., WONG, C.-J., GASCOYNE, R. D., WENG, A. P., SEYFERT-MARGOLIS, V., BOURCIER, K., ASARE, A. et al. (2010). Per-channel basis normalization methods for flow cytometry data. *Cytometry, Part A* **77** 121–131.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

HEJBLUM, B. P., ALKHASSIM, C., GOTTARDO, R., CARON, F. and THIÉBAUT, R. (2019). Sequential Dirichlet process mixtures of multivariate skew t-distributions for model-based clustering of flow cytometry data. *Ann. Appl. Stat.* **13** 638–660. MR3937443 https://doi.org/10.1214/18-AOAS1209

HENEL, G. and SCHMITZ, J. L. (2007). Basic theory and clinical applications of flow cytometry. *Lab. Med.* **38** 428–436.

JANATI, H., CUTURI, M. and GRAMFORT, A. (2018). Wasserstein regularization for sparse multi-task regression. Preprint. Available at arXiv:1805.07833.

LI, J., SEO, B. and LIN, L. (2019). Optimal transport, mean partition, and uncertainty assessment in cluster analysis. *Stat. Anal. Data Min.* **12** 359–377. MR4018510 https://doi.org/10.1002/sam.11418

LI, H., SHAHAM, U., STANTON, K. P., YAO, Y., MONTGOMERY, R. R. and KLUGER, Y. (2017). Gating mass cytometry data by deep learning. *Bioinformatics* **33** 3423–3430. https://doi.org/10.1093/bioinformatics/btx448

LUX, M., BRINKMAN, R. R., CHAUVE, C., LAING, A., LORENC, A., ABELER-DÖRNER, L. and HAMMER, B. (2018). flowLearn: Fast and precise identification and quality checking of cell populations in flow cytometry. *Bioinformatics* **34** 2245–2253. https://doi.org/10.1093/bioinformatics/bty082

MAECKER, H. T. and MCCOY, J. P. (2010). A model for harmonizing flow cytometry in clinical trials. *Nat. Immunol.* **11** 975–978.

PEYRÉ, G., CUTURI, M. et al. (2019). Computational optimal transport. *Found. Trends Mach. Learn.* **11** 355–607.

REDKO, I., COURTY, N., FLAMARY, R. and TUIA, D. (2018). Optimal transport for multi-source domain adaptation under target shift. Preprint. Available at arXiv:1803.04899.

SAEYS, Y., GASSEN, S. V. and LAMBRECHT, B. N. (2016). Computational flow cytometry: Helping to make sense of high-dimensional immunology data. *Nat. Rev., Immunol.* **16** 449.

SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser/Springer, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

SCHIEBINGER, G., SHU, J., TABAKA, M., CLEARY, B., SUBRAMANIAN, V., SOLOMON, A., GOULD, J., LIU, S., LIN, S. et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176** 928–943.

SOLOMON, J., DE GOES, F., PEYRE, G., CUTURI, M., BUTSCHER, A., NGUYEN, A., DU, T. and GUIBAS, L. (2015). Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.* **34** 1–11.

TUNG, J. W., HEYDARI, K., TIROUVANZIAM, R., SAHAF, B., PARKS, D. R., HERZENBERG, L. A. and HERZENBERG, L. A. (2007). Modern flow cytometry: A practical approach. *Clin. Lab. Med.* **27** 453–468.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1105–1126
https://doi.org/10.1214/22-AOAS1661

A TENSOR DECOMPOSITION MODEL FOR LONGITUDINAL MICROBIOME STUDIES

BY SIYUAN MA[a] AND HONGZHE LI[b]

Department of Biostatistics, Epidemiology, and Informatics, University or Pennsylvania,
[a]*siyuan.ma@pennmedicine.upenn.edu,* [b]*hongzhe@pennmedicine.upenn.edu*

Longitudinal microbiome studies can help delineate true biological signals from the high interindividual variability that is common in microbiome data. However, there are few methods available for unsupervised dimension reduction of time course microbial abundance observations. Existing methods do not fully observe the distribution characteristics of such data types, namely, zero inflation, compositionality, and overdispersion. We present a tensor decomposition model and a semiparametric quasi-likelihood estimation method for the decomposition of longitudinal microbiome data by generalizing existing approaches in tensor decomposition of Gaussian data. Optimization is performed through projected gradient descent, additionally allowing interpretability constraints. We show through simulation studies that our method is able to recover low-rank structures from microbiome time-course data better than existing approaches. Lastly, we apply our method to two existing longitudinal microbiome studies to detect global microbial changes associated with dietary and pharmaceutical effects as well as infant birth modes.

REFERENCES

CHEN, E. Z. and LI, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics* **32** 2611–2617.

CHO, I. and BLASER, M. J. (2012). The human microbiome: At the interface of health and disease. *Nat. Rev. Genet.* **13** 260–270.

CONSORTIUM, H. M. P. et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486** 207.

DE LATHAUWER, L. (2006). A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J. Matrix Anal. Appl.* **28** 642–666. MR2262974 https://doi.org/10.1137/040608830

FRANCINO, M. P. (2016). Antibiotics and the human gut microbiome: Dysbioses and accumulation of resistances. *Front. Microbiol.* **6** 1543. https://doi.org/10.3389/fmicb.2015.01543

GEVERS, D., KUGATHASAN, S., DENSON, L. A., VÁZQUEZ-BAEZA, Y., VAN TREUREN, W., REN, B., SCHWAGER, E., KNIGHTS, D., SONG, S. J. et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15** 382–392.

GIONGO, A., GANO, K. A., CRABB, D. B., MUKHERJEE, N., NOVELO, L. L., CASELLA, G., DREW, J. C., ILONEN, J., KNIP, M. et al. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J.* **5** 82–91.

GOWER, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53** 325–338. MR0214224 https://doi.org/10.1093/biomet/53.3-4.325

HITCHCOCK, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.* **6** 164–189.

HOLMES, I., HARRIS, K. and QUINCE, C. (2012). Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS ONE* **7** e30126. https://doi.org/10.1371/journal.pone.0030126

KOSTIC, A. D., XAVIER, R. J. and GEVERS, D. (2014). The microbiome in inflammatory bowel disease: Current status and the future ahead. *Gastroenterology* **146** 1489–1499.

KOSTIC, A. D., GEVERS, D., SILJANDER, H., VATANEN, T., HYÖTYLÄINEN, T., HÄMÄLÄINEN, A.-M., PEET, A., TILLMANN, V., PÖHÖ, P. et al. (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* **17** 260–273.

Key words and phrases. Metagenomics, dimension reduction, quasi-likelihood.

KUCZYNSKI, J., LAUBER, C. L., WALTERS, W. A., PARFREY, L. W., CLEMENTE, J. C., GEVERS, D. and KNIGHT, R. (2012). Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* **13** 47–58.

KVAM, V. M., LIU, P. and SI, Y. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.* **99** 248–256.

LLOYD-PRICE, J., MAHURKAR, A., RAHNAVARD, G., CRABTREE, J., ORVIS, J., HALL, A. B., BRADY, A., CREASY, H. H., MCCRACKEN, C. et al. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550** 61–66.

LLOYD-PRICE, J., ARZE, C., ANANTHAKRISHNAN, A. N., SCHIRMER, M., AVILA-PACHECO, J., POON, T. W., ANDREWS, E., AJAMI, N. J., BONHAM, K. S. et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569** 655–662.

LOVE, M. I., HUBER, W. and ANDERS, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15** 1–21.

MA, S. and LI, H. (2023). Supplement to "A tensor decomposition model for longitudinal microbiome studies." https://doi.org/10.1214/22-AOAS1661SUPP

MA, S., REN, B., MALLICK, H., MOON, Y. S., SCHWAGER, E., MAHARJAN, S., TICKLE, T. L., LU, Y., CARMODY, R. N. et al. (2021). A statistical model for describing and simulating microbial community profiles. *BioRxiv*.

MALLICK, H., MA, S., FRANZOSA, E. A., VATANEN, T., MORGAN, X. C. and HUTTENHOWER, C. (2017). Experimental design and quantitative analysis of microbial community multiomics. *Genome Biol.* **18** 1–16.

MARTINO, C., SHENHAV, L., MAROTZ, C. A., ARMSTRONG, G., MCDONALD, D., VÁZQUEZ-BAEZA, Y., MORTON, J. T., JIANG, L., DOMINGUEZ-BELLO, M. G. et al. (2021). Context-aware dimensionality reduction deconvolutes gut microbial community dynamics. *Nat. Biotechnol.* **39** 165–168.

MCMURDIE, P. J. and HOLMES, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8** e61217.

MCMURDIE, P. J. and HOLMES, S. (2014). Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10** e1003531.

MU, C., HSU, D. and GOLDFARB, D. (2015). Successive rank-one approximations for nearly orthogonally decomposable symmetric tensors. *SIAM J. Matrix Anal. Appl.* **36** 1638–1659. MR3432147 https://doi.org/10.1137/15M1010890

PASOLLI, E., SCHIFFER, L., MANGHI, P., RENSON, A., OBENCHAIN, V., TRUONG, D. T., BEGHINI, F., MALIK, F., RAMOS, M. et al. (2017). Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14** 1023.

SHAO, Y., FORSTER, S. C., TSALIKI, E., VERVIER, K., STRANG, A., SIMPSON, N., KUMAR, N., STARES, M. D., RODGER, A. et al. (2019). Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574** 117–121.

TANES, C., BITTINGER, K., GAO, Y., FRIEDMAN, E. S., NESSEL, L., PALADHI, U. R., CHAU, L., PANFEN, E., FISCHBACH, M. A. et al. (2021). Role of dietary fiber in the recovery of the human gut microbiome and its metabolome. *Cell Host Microbe* **29** 394–407.

TETT, A., HUANG, K. D., ASNICAR, F., FEHLNER-PEACH, H., PASOLLI, E., KARCHER, N., ARMANINI, F., MANGHI, P., BONHAM, K. et al. (2019). The *Prevotella copri* complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe* **26** 666–679.

TSILIMIGRAS, M. C. and FODOR, A. A. (2016). Compositional data analysis of the microbiome: Fundamentals, tools, and challenges. *Ann. Epidemiol.* **26** 330–335.

VANDEPUTTE, D., KATHAGEN, G., D'HOE, K., VIEIRA-SILVA, S., VALLES-COLOMER, M., SABINO, J., WANG, J., TITO, R. Y., DE COMMER, L. et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551** 507–511.

WANG, M., FISCHER, J. and SONG, Y. S. (2019). Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition. *Ann. Appl. Stat.* **13** 1103–1127. MR3963564 https://doi.org/10.1214/18-AOAS1228

WANG, D. D., NGUYEN, L. H., LI, Y., YAN, Y., MA, W., RINOTT, E., IVEY, K. L., SHAI, I., WILLETT, W. C. et al. (2021). The gut microbiome modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. *Nat. Med.* **27** 333–343.

YASSOUR, M., VATANEN, T., SILJANDER, H., HÄMÄLÄINEN, A.-M., HÄRKÖNEN, T., RYHÄNEN, S. J., FRANZOSA, E. A., VLAMAKIS, H., HUTTENHOWER, C. et al. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci. Transl. Med.* **8** 343ra81.

ZHANG, J., WEI, Z. and CHEN, J. (2018). A distance-based approach for testing the mediation effect of the human microbiome. *Bioinformatics* **34** 1875–1883.

ZHANG, X., PEI, Y.-F., ZHANG, L., GUO, B., PENDEGRAFT, A. H., ZHUANG, W. and YI, N. (2018). Negative binomial mixed models for analyzing longitudinal microbiome data. *Front. Microbiol.* **9** 1683.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1127–1151
https://doi.org/10.1214/22-AOAS1662

A ROTATION-BASED FEATURE AND BAYESIAN HIERARCHICAL MODEL FOR THE FORENSIC EVALUATION OF HANDWRITING EVIDENCE IN A CLOSED SET

BY AMY M. CRAWFORD[1,a], DANICA M. OMMEN[2,b] AND ALICIA L. CARRIQUIRY[2,c]

[1]*Berry Consultants, LLC,* [a]*crawford.m.amy@gmail.com*
[2]*CSAFE/Department of Statistics, Iowa State University,* [b]*dmommen@iastate.edu,* [c]*alicia@iastate.edu*

Forensic handwriting examiners are often tasked with identifying the writer of a particular document. Examples of handwriting evidence include ransom notes, forged documents and signatures, and threatening letters. At present, examiners rely on visual inspection of similarities and differences between the questioned document and reference writing samples. Here, we propose a principled modeling approach to compute the posterior predictive probability of writership when the author of the questioned document is part of a closed set of writers. Given a handwritten document, we extract measurements, including rotation angles that are related to the slant of writing, which are the response variables in a multilevel model. We fit the model and test its posterior predictive performance using writing samples from the United States and from Europe. We find that, as long as the questioned document is longer than a sentence or two, it is possible to correctly associate a writer with a document that he or she wrote with high probability. Earlier versions of this work have been well received by the community of forensic document examiners.

REFERENCES

BAUM, L. F. (1900). *The Wonderful Wizard of Oz*. G. M. Hill Co., Chicago and New York. Illustrated by W. W. Denslow.

BERRY, N., TAYLOR, J. and BAEZ-SANTIAGO, F. (2019). handwriter: Handwriting analysis in R. Available at https://CRAN.R-project.org/package=handwriter. R package version 1.0.1.

BRASSEUR, L. (2005). Florence Nightingale's visual rhetoric in the rose diagrams. *Tech. Commun. Q.* **14** 161–182. https://doi.org/10.1207/s15427625tcq1402_3

CHOI, R. Y., COYNER, A. S., KALPATHY-CRAMER, J., CHIANG, M. F. and CAMPBELL, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Transl. Vis. Sci. Technol.* **9** 14. https://doi.org/10.1167/tvst.9.2.14

CRAWFORD, A. M., BERRY, N. S. and CARRIQUIRY, A. L. (2021). A clustering method for graphical handwriting components and statistical writership analysis. *Stat. Anal. Data Min.* **14** 41–60. MR4223699 https://doi.org/10.1002/sam.11488

CRAWFORD, A. M., OMMEN, D. M. and CARRIQUIRY, A. L. (2023). Supplement to "A rotation-based feature and Bayesian hierarchical model for the forensic evaluation of handwriting evidence in a closed set." https://doi.org/10.1214/22-AOAS1662SUPP

CRAWFORD, A., RAY, A. and CARRIQUIRY, A. (2020). A database of handwriting samples for applications in forensic statistics. *Data Brief* **28** 105059. https://doi.org/10.1016/j.dib.2019.105059

EXPERT WORKING GROUP FOR HUMAN FACTORS IN HANDWRITING EXAMINATION (2020). Forensic handwriting examination and human factors: Improving the practice through a systems approach. U.S. Department of Commerce, National Institute of Standards and Technology. NISTIR 8282.

FEDERAL BUREAU OF INVESTIGATION (2009). Getting physical: Our hardcopy forensic experts. https://archives.fbi.gov/archives/news/stories/2009/april/qdu_040909. Accessed November 12, 2021.

FORGY, E. (1965). Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications. *Biometrics* **21** 768–780.

FRANKE, K., SCHOMAKER, L., VEENHUIS, C., VUURPIJL, L., VAN ERP, M. and GUYON, I. (2004). WANDA: A common ground for forensic handwriting examination and writer identification. *ENFHEX News* **1** 23–47.

Key words and phrases. Handwriter, rotation angle, London Letter, IAM data, K-means clustering, forensic identification.

HUBER, R. A. and HEADRICK, A. M. (1999). *Handwriting Identification*: *Facts and Fundamentals*. CRC Press-Taylor & Francis, Boca Raton, FL.

JOHANSSON, S., LEECH, G. and GOODLUCK, H. (1978). Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computer. Department of English, Univ. Oslo.

JOHNSON, M. Q. and OMMEN, D. M. (2022). Handwriting identification using random forests and score-based likelihood ratios. *Stat*. *Anal*. *Data Min*. **15** 357–375. MR4446410 https://doi.org/10.1002/sam.11566

KENT, J. T. and TYLER, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J*. *Appl*. *Stat*. **15** 247–254. https://doi.org/10.1080/02664768800000029

KLEBER, F., FIEL, S., DIEM, M. and SABLATNIG, R. (2013). CVL-DataBase: An off-line database for writer retrieval, writer identification and word spotting. In 2013 12*th International Conference on Document Analysis and Recognition* 560–564. https://doi.org/10.1109/ICDAR.2013.117

KWAN, Q. Y. (1977). Inference of identify of source. Ph.D. Dissertation in Criminology, Univ. California, Berkeley.

LEEDHAM, G. and SRIHARI, S. (2003). A survey of computer methods in forensic handwritten document examination. In *Proceedings of the Eleventh International Graphonomics Society Conference* 278–281. Scottsdale, AZ.

LLOYD, S. P. (1982). Least squares quantization in PCM. *IEEE Trans*. *Inf*. *Theory* **28** 129–137. MR0651807 https://doi.org/10.1109/TIT.1982.1056489

MARTI, U. and BUNKE, H. (2002). The IAM-Database: An English sentence database for offline handwriting recognition. *Int*. *J*. *Doc*. *Anal*. *Recognit*. **5** 39–46.

MILLER, J. J., PATTERSON, R. B., GANTZ, D. T., SAUNDERS, C. P., WALCH, M. A. and BUSCAGLIA, J. (2017). A set of handwriting features for use in automated writer identification. *J*. *Forensic Sci*. **62** 722–734. https://doi.org/10.1111/1556-4029.13345

NATIONAL RESEARCH COUNCIL (2009). *Strengthening Forensic Science in the United States*: *A Path Forward*. National Academies Press, Washington, D.C.

OSBORN, A. S. (1929). *Questioned Documents*, 2nd edn. Boyd Printing Company, New York, NY.

SAUNDERS, C. P., DAVIS, L. J., LAMAS, A. C., MILLER, J. J. and GANTZ, D. T. (2011). Construction and evaluation of classifiers for forensic document analysis. *Ann*. *Appl*. *Stat*. **5** 381–399. MR2810402 https://doi.org/10.1214/10-AOAS379

SRIHARI, S. N., SRINIVASAN, H. and DESAI, K. (2007). Questioned document examination using. *J*. *Forensic Doc*. *Exam*. **18** 1–20.

STATE V. PICKETT (2021). 246 A.3d 279, 466 N.J. Super. 270 (Super. Ct. App. Div. 2021).

UNITED STATES V. JOHNSON (2016). No. 1:15-cr-00565 VEC, (S.D.N.Y. Jun. 7, 2016).

YOUNG, E. (2018a). Lindbergh kidnapping, FBI: Famous cases and criminals: Volume 2. Independently published. Also available online: https://www.fbi.gov/history/famous-cases/lindbergh-kidnapping. Accessed November 12, 2021.

YOUNG, E. (2018b). Weinberger kidnapping, FBI: Famous cases and criminals: Volume 3. Independently published. Also available online: https://www.fbi.gov/history/famous-cases/weinberger-kidnapping. Accessed November 12, 2021.

ZIMMERMANN, M. and BUNKE, H. (2002). Automatic segmentation of the IAM off-line database for handwritten English text. In *Object Recognition Supported by User Interaction for Service Robots* **4** 35–39. IEEE.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1152–1174
https://doi.org/10.1214/22-AOAS1663

KNOCKOFFS WITH SIDE INFORMATION

BY ZHIMEI REN[1,a] AND EMMANUEL CANDÈS[2,b]

[1]*Department of Statistics, University of Chicago,* [a]*zmren@uchicago.edu*
[2]*Department of Statistics, Stanford University,* [b]*candes@stanford.edu*

We consider the problem of assessing the importance of multiple variables or factors from a dataset when side information is available. In principle, using side information can allow the statistician to pay attention to variables with a greater potential which, in turn, may lead to more discoveries. We introduce an adaptive knockoff filter, which generalizes the knockoff procedure (*Ann. Statist.* **43** (2015) 2055–2085; *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** (2018) 551–577), in that it uses both the data at hand and side information to adaptively order the variables under study and focus on those that are most promising. The *adaptive knockoffs* procedure controls the finite-sample false discovery rate (FDR), and we demonstrate its power by comparing it with other structured multiple testing methods. We also apply our methodology to real genetic data in order to find associations between genetic variants and various phenotypes such as Crohn's disease and lipid levels. Here, the adaptive knockoffs method makes more discoveries than reported in previous studies on the same datasets.

REFERENCES

BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876 https://doi.org/10.1214/15-AOS1337

BARBER, R. F. and CANDÈS, E. J. (2019). A knockoff filter for high-dimensional selective inference. *Ann. Statist.* **47** 2504–2537. MR3988764 https://doi.org/10.1214/18-AOS1755

BASU, P., CAI, T. T., DAS, K. and SUN, W. (2018). Weighted false discovery rate control in large-scale multiple testing. *J. Amer. Statist. Assoc.* **113** 1172–1183. MR3862348 https://doi.org/10.1080/01621459.2017.1336443

BATES, S., CANDÈS, E., JANSON, L. and WANG, W. (2021). Metropolized knockoff sampling. *J. Amer. Statist. Assoc.* **116** 1413–1427. MR4309282 https://doi.org/10.1080/01621459.2020.1729163

BENJAMINI, Y. and HELLER, R. (2007). False discovery rates for spatial signals. *J. Amer. Statist. Assoc.* **102** 1272–1281. MR2412549 https://doi.org/10.1198/016214507000000941

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BENJAMINI, Y. and HOCHBERG, Y. (1997). Multiple hypotheses testing with weights. *Scand. J. Stat.* **24** 407–418. MR1481424 https://doi.org/10.1111/1467-9469.00072

BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 https://doi.org/10.1214/aos/1013699998

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

CAI, T. T., SUN, W. and WANG, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 187–234. MR3928141

CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. MR3798878 https://doi.org/10.1111/rssb.12265

CORAM, M. A., CANDILLE, S. I., DUAN, Q., CHAN, K. H. K., LI, Y., KOOPERBERG, C., REINER, A. P. and TANG, H. (2015). Leveraging multi-ethnic evidence for mapping complex traits in minority populations: An empirical Bayes approach. *Am. J. Hum. Genet.* **96** 740–752.

CORAM, M. A., FANG, H., CANDILLE, S. I., ASSIMES, T. L. and TANG, H. (2017). Leveraging multi-ethnic evidence for risk assessment of quantitative traits in minority populations. *Am. J. Hum. Genet.* **101** 218–226.

Key words and phrases. Multiple testing, variable selection, false discovery rate (FDR), knockoff filters, Bayesian two-group models, genome-wide association study (GWAS).

DEZEURE, R., BÜHLMANN, P., MEIER, L. and MEINSHAUSEN, N. (2015). High-dimensional inference: Confidence intervals, p-values and R-software hdi. *Statist. Sci.* **30** 533–558. MR3432840 https://doi.org/10.1214/15-STS527

EDWARDS, D. (2012). *Introduction to Graphical Modelling*, Springer, New York.

EFRON, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics (IMS) Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758 https://doi.org/10.1017/CBO9780511761362

FERKINGSTAD, E., FRIGESSI, A., RUE, H., THORLEIFSSON, G. and KONG, A. (2008). Unsupervised empirical Bayesian multiple testing with external covariates. *Ann. Appl. Stat.* **2** 714–735. MR2524353 https://doi.org/10.1214/08-AOAS158

FRANKE, A., MCGOVERN, D. P., BARRETT, J. C., WANG, K., RADFORD-SMITH, G. L., AHMAD, T., LEES, C. W., BALSCHUN, T., LEE, J. et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42** 1118.

GENOVESE, C. R., ROEDER, K. and WASSERMAN, L. (2006). False discovery control with p-value weighting. *Biometrika* **93** 509–524. MR2261439 https://doi.org/10.1093/biomet/93.3.509

GIMENEZ, J. R., GHORBANI, A. and ZOU, J. (2019). Knockoffs for the mass: New feature importance statistics with false discovery guarantees. In *The 22nd International Conference on Artificial Intelligence and Statistics* 2125–2133. PMLR, Naha, Okinawa, Japan.

GOYETTE, P., BOUCHER, G., MALLON, D., ELLINGHAUS, E., JOSTINS, L., HUANG, H., RIPKE, S., GUSAREVA, E. S., ANNESE, V. et al. (2015). High-density mapping of the MHC identifies a shared role for HLA-DRB1* 01: 03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat. Genet.* **47** 172.

HASTIE, T. (2017). Generalized additive models. In *Statistical Models in S* 249–307. Routledge.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

IGNATIADIS, N. and HUBER, W. (2021). Covariate powered cross-weighted multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 720–751. MR4319999 https://doi.org/10.1111/rssb.12411

IGNATIADIS, N., KLAUS, B., ZAUGG, J. B. and HUBER, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods* **13** 577–580. https://doi.org/10.1038/nmeth.3885

JANSON, L. B. (2017). A Model-Free Approach to High-Dimensional Inference. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Stanford University. MR4239955

LEI, L. and BICKEL, P. J. (2021). An assumption-free exact test for fixed-design linear models with exchangeable errors. *Biometrika* **108** 397–412. MR4259139 https://doi.org/10.1093/biomet/asaa079

LEI, L. and FITHIAN, W. (2016). Power of ordered hypothesis testing. In *International Conference on Machine Learning* 2924–2932.

LEI, L. and FITHIAN, W. (2018). AdaPT: An interactive procedure for multiple testing with side information. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 649–679. MR3849338 https://doi.org/10.1111/rssb.12253

LI, A. and BARBER, R. F. (2019). Multiple testing with the structure-adaptive Benjamini–Hochberg algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 45–74. MR3904779

LIU, Y. and ZHENG, C. (2018). Auto-encoding knockoff generator for FDR controlled variable selection. Preprint. Available at arXiv:1809.10765.

LIU, J. Z., VAN SOMMEREN, S., HUANG, H., NG, S. C., ALBERTS, R., TAKAHASHI, A., RIPKE, S., LEE, J. C., JOSTINS, L. et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47** 979.

LOH, P.-R., KICHAEV, G., GAZAL, S., SCHOECH, A. P. and PRICE, A. L. (2018). Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50** 906–908. https://doi.org/10.1038/s41588-018-0144-6

LYNCH, G., GUO, W., SARKAR, S. K. and FINNER, H. (2017). The control of the false discovery rate in fixed sequence multiple testing. *Electron. J. Stat.* **11** 4649–4673. MR3724971 https://doi.org/10.1214/17-EJS1359

REN, Z. and CANDÈS, E. (2023). Supplement to "Knockoffs with side information." https://doi.org/10.1214/22-AOAS1663SUPPA, https://doi.org/10.1214/22-AOAS1663SUPPB

ROEDER, K. and WASSERMAN, L. (2009). Genome-wide significance levels and weighted hypothesis testing. *Statist. Sci.* **24** 398–413. MR2779334 https://doi.org/10.1214/09-STS289

ROMANO, Y., SESIA, M. and CANDÈS, E. (2020). Deep knockoffs. *J. Amer. Statist. Assoc.* **115** 1861–1872. MR4189763 https://doi.org/10.1080/01621459.2019.1660174

ROSENBLATT, J. D., FINOS, L., WEEDA, W. D., SOLARI, A. and GOEMAN, J. J. (2018). All-resolutions inference for brain imaging. *NeuroImage* **181** 786–796. https://doi.org/10.1016/j.neuroimage.2018.07.060

SABATTI, C., SERVICE, S. K., HARTIKAINEN, A.-L., POUTA, A., RIPATTI, S., BRODSKY, J., JONES, C. G., ZAITLEN, N. A., VARILO, T. et al. (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41** 35.

SESIA, M., SABATTI, C. and CANDÈS, E. J. (2019). Gene hunting with hidden Markov model knockoffs. *Biometrika* **106** 1–18. MR3912377 https://doi.org/10.1093/biomet/asy033

STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 479–498. MR1924302 https://doi.org/10.1111/1467-9868.00346

STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 187–205. MR2035766 https://doi.org/10.1111/j.1467-9868.2004.00439.x

SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116

SUR, P., CHEN, Y. and CANDÈS, E. J. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a *rescaled* chi-square. *Probab. Theory Related Fields* **175** 487–558. MR4009715 https://doi.org/10.1007/s00440-018-00896-9

WOO, C.-W., KRISHNAN, A. and WAGER, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage* **91** 412–419.

WTCCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* **447** 661.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1175–1198
https://doi.org/10.1214/22-AOAS1664

LATENT MULTIVARIATE LOG-GAMMA MODELS FOR HIGH-DIMENSIONAL MULTITYPE RESPONSES WITH APPLICATION TO DAILY FINE PARTICULATE MATTER AND MORTALITY COUNTS

BY ZHIXING XU[a], JONATHAN R. BRADLEY[b] AND DEBAJYOTI SINHA[c]

Department of Statistics, Florida State University, [a]*zhixing.xu2@sanofi.com,* [b]*jrbradley@fsu.edu,* [c]*sinhad@stat.fsu.edu*

Precise estimation of daily fine particulate matter with a diameter ≤ 2.5 microns (PM2.5) and mortality in the U.S. is an important research challenge in public health because high levels of PM2.5 have been linked to several serious health problems, including lung disease, cardiovascular disease, and stroke. This motivates us to develop a joint Bayesian hierarchical model for bivariate spatial data to obtain precise spatial predictions of two types of responses, continuous skewed PM2.5 levels, and discrete mortality counts over U.S. counties. Our novel modeling framework address several challenges in the area of spatial prediction of mortality counts and PM2.5 levels. Specifically, our model allows for spatial variability and dependence of two types of responses, accommodate an unknown nonlinear spatial relationship between mortality and PM2.5 through basis function expansions, improve the precision of predictions at counties with undisclosed/missing observations, and allow for different missing data patterns for mortality and PM2.5. Furthermore, we introduce a new local measure of association for the cross-dependence between mortality and PM2.5 level. To address the burden of Bayesian computation for large databases, we use the dimension reduction tool and the shared conjugate structure between the Weibull distribution, Poisson distribution, and the multivariate log-gamma distribution. We provide a simulation study to illustrate the performance of our method. Our joint spatial model of "multitype responses" (discrete and continuous responses) and associated Bayesian method are used to analyze bivariate spatial data of daily averaged PM2.5 levels in air and mortality counts (due to diseases related to lung, cardiovascular, respiratory, and stroke) from the Centers for Disease Control and Prevention (CDC) database.

REFERENCES

ALDAZ, J. M. (2009). Self-improvement of the inequality between arithmetic and geometric means. *J. Math. Inequal.* **3** 213–216. MR2542299 https://doi.org/10.7153/jmi-03-21

ANDERSON, J. O., THUNDIYIL, J. G. and STOLBACH, A. (2012). Clearing the air: A review of the effects of particulate matter air pollution on human health. *J. Med. Toxicol.* **8** 166–175.

ANSELIN, L. (1995). Local indicators of spatial association. *Geogr. Anal.* **27** 93–115.

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. MR3362184

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

BESAG, J. (1986). On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B* **48** 259–302. MR0876840

BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–59. MR1105822 https://doi.org/10.1007/BF00116466

BRADLEY, J. R. (2021). An approach to incorporate subsampling into a generic Bayesian hierarchical model. *J. Comput. Graph. Statist.* **30** 889–905. MR4356593 https://doi.org/10.1080/10618600.2021.1923518

BRADLEY, J. R., CRESSIE, N. and SHI, T. (2015). Comparing and selecting spatial predictors using local criteria. *TEST* **24** 1–28. MR3314567 https://doi.org/10.1007/s11749-014-0415-1

Key words and phrases. Bayesian hierarchical model, multitype responses, high-dimensional data, Gibbs sampler.

BRADLEY, J. R., CRESSIE, N. and SHI, T. (2016). A comparison of spatial predictors when datasets could be very large. *Stat. Surv.* **10** 100–131. MR3527662 https://doi.org/10.1214/16-SS115

BRADLEY, J. R., HOLAN, S. H. and WIKLE, C. K. (2015). Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. *Ann. Appl. Stat.* **9** 1761–1791. MR3456353 https://doi.org/10.1214/15-AOAS862

BRADLEY, J. R., HOLAN, S. H. and WIKLE, C. K. (2018). Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion). *Bayesian Anal.* **13** 253–310. MR3773410 https://doi.org/10.1214/17-BA1069

BRADLEY, J. R., HOLAN, S. H. and WIKLE, C. K. (2020). Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. *J. Amer. Statist. Assoc.* **115** 2037–2052. MR4189775 https://doi.org/10.1080/01621459.2019.1677471

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2016). Bayesian spatial change of support for count-valued survey data with application to the American community survey. *J. Amer. Statist. Assoc.* **111** 472–487. MR3538680 https://doi.org/10.1080/01621459.2015.1117471

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2017). Regionalization of multiscale spatial processes by using a criterion for spatial aggregation error. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 815–832. MR3641409 https://doi.org/10.1111/rssb.12179

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2020). Hierarchical models for spatial data with errors that are correlated with the latent process. *Statist. Sinica* **30** 81–109. MR4285486 https://doi.org/10.5705/ss.202016.0230

BROOK, R. D., RAJAGOPALAN, S., POPE III, C. A., BROOK, J. R., BHATNAGAR, A., DIEZ-ROUX, A. V., HOLGUIN, F., HONG, Y., LUEPKER, R. V. et al. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation* **121** 2331–2378.

BURNETTA, R., HONG, C., SZYSZKOWICZA, M., FANN, N., HUBBELLD, B., POPE, C. A., APTEF, J. S., BRAUERG, M., COHENH, A. et al. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci. USA* **115** 9592–9597.

CARROLL, R. J., GAIL, M. H. and LUBIN, J. H. (1993). Case-control studies with errors in covariates. *J. Amer. Statist. Assoc.* **88** 185–199. MR1212486

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press/CRC, Boca Raton, FL. MR2243417 https://doi.org/10.1201/9781420010138

CHAKRABORTY, A. and PANARETOS, V. M. (2017). Regression with genuinely functional errors-in-covariates. ArXiv preprint. Available at arXiv:1712.04290.

CHIB, S. and GREENBERG, E. (1995). Understanding the Metropolis-Hastings algorithm. *Amer. Statist.* **49** 327–335.

CHOI, J., FUENTES, M. and REICH, B. J. (2009). Spatial-temporal association between fine particulate matter and daily mortality. *Comput. Statist. Data Anal.* **53** 2989–3000. MR2667605 https://doi.org/10.1016/j.csda.2008.05.018

CHRISTENSEN, W. F. and AMEMIYA, Y. (2002). Latent variable analysis of multivariate spatial data. *J. Amer. Statist. Assoc.* **97** 302–317. MR1947288 https://doi.org/10.1198/016214502753479437

CLARKE, J. S., NEMERGUT, D., SEYEDNASROLLAH, B., TURNER, P. and ZHANG, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecol. Monogr.* **87** 34–56.

CRESSIE, N. and JOHANNESSON, G. (2006). Spatial prediction for massive datasets. In *Mastering the Data Explosion in the Earth and Environmental Sciences: Australian Academy of Science, Elizabeth and Frederick White Conference* 1–11.

CRESSIE, N. and JOHANNESSON, G. (2008). Fixed rank Kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226. MR2412639 https://doi.org/10.1111/j.1467-9868.2007.00633.x

DATTA, A., BANERJEE, S., FINLEY, A. O., HAMM, N. A. S. and SCHAAP, M. (2016). Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *Ann. Appl. Stat.* **10** 1286–1316. MR3553225 https://doi.org/10.1214/16-AOAS931

DE OLIVEIRA, V. (2003). A note on the correlation structure of transformed Gaussian random fields. *Aust. N. Z. J. Stat.* **45** 353–366. MR1999517 https://doi.org/10.1111/1467-842X.00289

DE OLIVEIRA, V. (2013). Hierarchical Poisson models for spatial count data. *J. Multivariate Anal.* **122** 393–408. MR3189330 https://doi.org/10.1016/j.jmva.2013.08.015

DEPAOLI, S., CLIFTON, J. P. and COBB, P. R. (2016). Just another Gibbs sampler (JAGS) flexible software for MCMC implementation. *J. Educ. Behav. Stat.* **41** 628–649.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

DOCKERY, D. W., POPE, C. A., XU, X., SPENGLER, J. D., WARE, J. H., FAY, M., FERRIS JR, G. B. and SPEIZER, F. E. (1993). An association between air pollution and mortality in six US cities. *N. Engl. J. Med.* **329** 1753–1759.

DOMINICI, F., PENG, R. D., BELL, M. L., PHAM, L., MCDERMOTT, A., ZEGER, S. L. and SAMET, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *J. Am. Med. Assoc.* **295** 1127–1134.

FINLEY, A. O., SANG, H., BANERJEE, S. and GELFAND, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Comput. Statist. Data Anal.* **53** 2873–2884. MR2667597 https://doi.org/10.1016/j.csda.2008.09.008

FRANKLIN, M., ZEKA, A. and SCHWARTZ, J. (2007). Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. *J. Expo. Sci. Environ. Epidemiol.* **17** 279–287. https://doi.org/10.1038/sj.jes.7500530

FUENTES, M., SONG, H.-R., GHOSH, S. K., HOLLAND, D. M. and DAVIS, J. M. (2006). Spatial association between speciated fine particles and mortality. *Biometrics* **62** 855–863. MR2247215 https://doi.org/10.1111/j.1541-0420.2006.00526.x

GAO, L., DATTA, A. and BANERJEE, S. (2022). Hierarchical multivariate directed acyclic graph autoregressive models for spatial diseases mapping. *Stat. Med.* **41** 3057–3075. MR4444886 https://doi.org/10.1002/sim.9404

GELFAND, A. E. and SCHLIEP, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spat. Stat.* **18** 86–104. MR3573271 https://doi.org/10.1016/j.spasta.2016.03.006

GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–807. MR1422404

GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

GELMAN, A., CARLIN, J. B., STERN, H. S., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*. CRC Press/CRC, Boca Raton, FL.

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 https://doi.org/10.1198/016214506000001437

GOTWAY, C. A. and YOUNG, L. J. (2002). Combining incompatible spatial data. *J. Amer. Statist. Assoc.* **97** 632–648. MR1951636 https://doi.org/10.1198/016214502760047140

GRIFFITH, D. A. and TIEFELSDORF, M. (2007). Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environ. Plan. A.* **39** 1193–1221.

HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2018). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–425.

HOEK, G., KRISHNAN, R. M., BEELEN, R., PETERS, A., OSTRO, B., BRUNEKREEF, B. and KAUFMAN, J. D. (2013). Long-term air pollution exposure and cardio-respiratory mortality: A review. *Environ. Health* **12** 43.

HU, G. and BRADLEY, J. (2018). A Bayesian spatial-temporal model with latent multivariate log-gamma random effects with application to earthquake magnitudes. *Stat* **7** e179. MR3796093 https://doi.org/10.1002/sta4.179

HUGHES, J. and HARAN, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 139–159. MR3008275 https://doi.org/10.1111/j.1467-9868.2012.01041.x

JORDAN, A., KRÜGER, F. and LERCH, S. (2017). Evaluating probabilistic forecasts with scoringRules. ArXiv preprint. Available at arXiv:1709.04743.

KAMPA, M. and CASTANAS, E. (2008). Human health effects of air pollution. *Environ. Pollut.* **151** 362–367.

KATZFUSS, M. and CRESSIE, N. (2011). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *J. Time Series Anal.* **32** 430–446. MR2841794 https://doi.org/10.1111/j.1467-9892.2011.00732.x

KATZFUSS, M. and GUINNESS, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statist. Sci.* **36** 124–141. MR4194207 https://doi.org/10.1214/19-STS755

KATZFUSS, M., GUINNESS, J., GONG, W. and ZILBER, D. (2020). Vecchia approximations of Gaussian-process predictions. *J. Agric. Biol. Environ. Stat.* **25** 383–414. MR4139037 https://doi.org/10.1007/s13253-020-00401-7

KIM, J. S. and YUM, B.-J. (2008). Selection between Weibull and lognormal distributions: A comparative simulation study. *Comput. Statist. Data Anal.* **53** 477–485. MR2649102 https://doi.org/10.1016/j.csda.2008.08.012

KNUIMAN, M. W., JAMES, A. L., DIVITINI, M. L., RYAN, G., BARTHOLOMEW, H. C. and MUSK, A. W. (1999). Lung function, respiratory symptoms, and mortality: Results from the busselton health study. *Ann. Epidemiol.* **9** 297–306. https://doi.org/10.1016/s1047-2797(98)00066-0

KRAINSKI, E., GÓMEZ-RUBIO, V., BAKKA, H., LENZI, A., CASTRO-CAMILO, D., SIMPSON, D., LINDGREN, F. and RUE, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. CRC Press/CRC, Boca Raton.

KRISTENSEN, K., NIELSEN, A., BERG, C. W., SKAUG, H. and BELL, B. (2016). TMB: Automatic differentiation and Laplace approximation. *J. Stat. Softw.* **60** 1–21.

LADEN, F., NEAS, L. M., DOCKERY, D. W. and SCHWARTZ, J. (2000). Association of fine particulate matter from different sources with daily mortality in six US cities. *Environ. Health Perspect.* **108** 941.

LEITER, U. and GARBE, C. (2008). Epidemiology of melanoma and nonmelanoma skin cancer the role of sunlight. In *Sunlight, Vitamin D and Skin Cancer* 89–103. Springer, New York, NY.

MARTINO, S. and RIEBLER, A. (2019). Integrated Nested Laplace Approximations (INLA). ArXiv preprint. Available at arXiv:1907.01248.

MENG, X.-L. (1994). Posterior predictive p-values. *Ann. Statist.* **22** 1142–1160. MR1311969 https://doi.org/10.1214/aos/1176325622

MENG, Z. and LU, B. (2007). Dust events as a risk factor for daily hospitalization for respiratory and cardiovascular diseases in Minqin, China. *Atmos. Environ.* **41** 7048–7058.

MILLER, K. A., SISCOVICK, D. S., SHEPPARD, L., SHEPHERD, K., SULLIVAN, J. H., ANDERSON, G. L. and KAUFMAN, D. J. (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *N. Engl. J. Med.* **356** 447–458.

MORAN, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika* **37** 17–23. MR0035933 https://doi.org/10.1093/biomet/37.1-2.17

MUGGLIN, A. S., CARLIN, B. P. and GELFAND (2000). Fully model-based approaches for spatially misaligned data. *J. Amer. Statist. Assoc.* **95** 877–887.

NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. MR1994729 https://doi.org/10.1214/aos/1056562461

NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

PERUZZI, M., BANERJEE, S. and FINLEY, A. O. (2022). Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitioned Domains. *J. Amer. Statist. Assoc.* **117** 969–982. MR4436326 https://doi.org/10.1080/01621459.2020.1833889

QUICK, H., WALLER, L. A. and CASPER, M. (2018). A multivariate space-time model for analysing county level heart disease death rates by race and sex. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **67** 291–304. MR3758767 https://doi.org/10.1111/rssc.12215

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

SCHLIEP, E. M. and HOETING, J. A. (2013). Multilevel latent Gaussian process model for mixed discrete and continuous multivariate response data. *J. Agric. Biol. Environ. Stat.* **18** 492–513. MR3142597 https://doi.org/10.1007/s13253-013-0136-z

SCHWARTZ, J. and NEAS, L. M. (2000). Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren. *Epidemiology* **11** 6–10. https://doi.org/10.1097/00001648-200001000-00004

SELLERS, K. F. and RAIM, A. (2016). A flexible zero-inflated model to address data dispersion. *Comput. Statist. Data Anal.* **99** 68–80. MR3473082 https://doi.org/10.1016/j.csda.2016.01.007

SENGUPTA, A. and CRESSIE, N. (2013). Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions. *Spat. Stat.* **4** 14–44.

SHI, T. and CRESSIE, N. (2007). Global statistical analysis of MISR aerosol data: A massive data product from NASA's Terra satellite. *Environmetrics* **18** 665–680. MR2408937 https://doi.org/10.1002/env.864

STEIN, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spat. Stat.* **8** 1–19. MR3326818 https://doi.org/10.1016/j.spasta.2013.06.003

TURNER, M. C., KREWSKI, D., POPE III, C. A., CHEN, Y., GAPSTUR, S. M. and THUN, M. J. (2011). Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *Am. J. Respir. Crit. Care Med.* **184** 1374–1381.

VALAVANIDIS, A., FIOTAKIS, K. and VLACHOGIANNI, T. (2008). Airborne particulate matter and human health: Toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. *J. Environ. Sci. Health, Part C* **26** 339–362.

WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. MR2756194

WIKLE, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 107–118. CRC Press, Boca Raton, FL. MR2730946 https://doi.org/10.1201/9781420072884-c8

WIKLE, C. K. and CRESSIE, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika* **86** 815–829. MR1741979 https://doi.org/10.1093/biomet/86.4.815

WIKLE, C. K. and HOOTEN, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST* **19** 417–451. MR2745992 https://doi.org/10.1007/s11749-010-0209-z

XU, Z., BRADLEY, J. R. and SINHA, D. (2023). Supplement to "Latent multivariate log-gamma models for high-dimensional multitype responses with application to daily fine particulate matter and mortality counts." https://doi.org/10.1214/22-AOAS1664SUPP

XU, M., GUO, Y., ZHANG, Y., WESTERDAHL, D., MO, Y., LIANG, F. and PAN, X. (2014). Spatiotemporal analysis of particulate air pollution and ischemic heart disease mortality in Beijing, China. *Environ. Health* **13** 109.

XU, Q., LI, X., WANG, S., WANG, C., HUANG, F., GAO, Q., WU, L., TAO, L., GUO, J. et al. (2016). Fine particulate air pollution and hospital emergency room visits for respiratory disease in urban areas in Beijing, China, in 2013. *PLoS ONE* **11** e0153099.

ZHANG, L., TANG, W. and BANERJEE, S. (2021). Fixed-domain asymptotics under Vecchia's approximation of spatial process likelihoods. ArXiv preprint. Available at arXiv:2101.08861.

ZHANG, F., LIU, X., ZHOU, L., YU, Y., WANG, L., LU, J., WANG, W. and KRAFFT, T. (2016). Spatiotemporal patterns of particulate matter (PM) and associations between PM and mortality in Shenzhen, China. *BMC Public Health* **16** 215.

ZIGLER, C. M., DOMINICI, F. and WANG, Y. (2012). Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics* **13** 289–302.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1199–1219
https://doi.org/10.1214/22-AOAS1665

IDENTIFICATION OF IMMUNE RESPONSE COMBINATIONS ASSOCIATED WITH HETEROGENEOUS INFECTION RISK IN THE IMMUNE CORRELATES ANALYSIS OF HIV VACCINE STUDIES

BY CHAERYON KANG[1,a] AND YING HUANG[2,b]

[1]*Department of Biostatistics, University of Pittsburgh,* [a]*crkang@pitt.edu*

[2]*Vaccine and Infectious Disease Division and Public Health Sciences Division, Fred Hutchinson Cancer Research Center,* [b]*yhuang@fredhutch.org*

In HIV vaccine/prevention research, probing into the vaccine-induced immune responses that can help to predict the risk of HIV infection provides valuable information for the development of vaccine regimens. Previous correlate analysis of the Thai vaccine trial aided the discovery of interesting immune correlates related to the risk of developing an HIV infection. The present study aimed to identify the combinations of immune responses associated with the heterogeneous infection risk. We explored a "change-plane" via combination of a subset of immune responses that could help separate vaccine recipients into two heterogeneous subgroups in terms of the association between immune responses and the risk of developing infection. Additionally, we developed a new variable selection algorithm through a penalized likelihood approach to investigate a parsimonious marker combination for the change-plane. The resulting marker combinations can serve as candidate correlates of protection and can be used for predicting the protective effect of the vaccine against HIV infection. The application of the proposed statistical approach to the Thai trial has been presented, wherein the marker combinations were explored among several immune responses and antigens.

REFERENCES

CHEN, G., LIU, Y., SHEN, D. and KOSOROK, M. R. (2016). Composite large margin classifiers with latent subclasses for heterogeneous biomedical data. *Stat. Anal. Data Min.* **9** 75–88. MR3511536 https://doi.org/10.1002/sam.11300

CHUNG, H., FLAHERTY, B. P. and SCHAFER, J. L. (2006). Latent class logistic regression: Application to marijuana use and attitudes among high school seniors. *J. Roy. Statist. Soc. Ser. A* **169** 723–743. MR2291341 https://doi.org/10.1111/j.1467-985X.2006.00419.x

COREY, L., GILBERT, P. B., TOMARAS, G. D., HAYNES, B. F., PANTALEO, G. and FAUCI, A. S. (2015). Immune correlates of vaccine protection against HIV-1 acquisition. *Sci. Transl. Med.* **7** 310rv7. https://doi.org/10.1126/scitranslmed.aac7732

CRAVEN, P. and WAHBA, G. (1978). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. With discussion. MR0501537

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

FAN, A., SONG, R. and LU, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. *J. Amer. Statist. Assoc.* **112** 769–778. MR3671769 https://doi.org/10.1080/01621459.2016.1166115

HAYNES, B. F., GILBERT, P. B., MCELRATH, M. J., ZOLLA-PAZNER, S., TOMARAS, G. D., ALAM, S. M., EVANS, D. T., MONTEFIORI, D. C., KARNASUTA, C. et al. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N. Engl. J. Med.* **366** 1275–1286.

HUANG, Y., CHO, J. and FONG, Y. (2021). Threshold-based subgroup testing in logistic regression models in two-phase sampling designs. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **70** 291–311. MR4226669 https://doi.org/10.1111/rssc.12459

INAN, G. and WANG, L. (2017). PGEE: An R package for analysis of longitudinal data with high-dimensional covariates. *R J.* **9** 393–402.

Key words and phrases. Change-plane, immune-correlates analysis, latent subgroup analysis, RV144 trial.

KANG, C. (2011). New statistical learning methods for chemical toxicity data analysis. PhD Thesis, Univ. North Carolina at Chapel Hill.

KANG, C. and HUANG, Y. (2023). Supplement to "Identification of immune response combinations associated with heterogeneous infection risk in the immune correlates analysis of HIV vaccine studies." https://doi.org/10.1214/22-AOAS1665SUPP

KANG, S., LU, W. and SONG, R. (2017). Subgroup detection and sample size calculation with proportional hazards regression for survival data. *Stat. Med.* **36** 4646–4659. MR3731245 https://doi.org/10.1002/sim.7441

KOSOROK, M. R. and SONG, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.* **35** 957–989. MR2341694 https://doi.org/10.1214/009053606000001244

RERKS-NGARM, S., PITISUTTITHUM, P., NITAYAPHAN, S., KAEWKUNGWAL, J., CHIU, J., PARIS, R., PREMSRI, N., NAMWAT, C., DE SOUZA, M. et al. (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* **361** 2209–2220.

ROLLAND, M. and GILBERT, P. (2012). Evaluating immune correlates in HIV type 1 vaccine efficacy trials: What RV144 may provide. *AIDS Res. Hum. Retrovir.* **28** 400–404.

ROLLAND, M., EDLEFSEN, P. T., LARSEN, B. B., TOVANABUTRA, S., SANDERS-BUELL, E., HERTZ, T., CARRICO, C., MENIS, S., MAGARET, C. A. et al. (2012). Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. *Nature* **490** 417–420.

SHEN, J. and HE, X. (2015). Inference for subgroup analysis with a structured logistic-normal mixture model. *J. Amer. Statist. Assoc.* **110** 303–312. MR3338504 https://doi.org/10.1080/01621459.2014.894763

SHEN, J. and QU, A. (2020). Subgroup analysis based on structured mixed-effects models for longitudinal data. *J. Biopharm. Statist.* **30** 607–622. https://doi.org/10.1080/10543406.2020.1730867

SIMON, R. and WANG, S. (2006). Use of genomic signatures in therapeutics development in oncology and other diseases. *Pharmacogenomics J.* **6** 166–173.

SUN, X., BRIEL, M., WALTER, S. D. and GUYATT, G. H. (2010). Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* **340** c117. https://doi.org/10.1136/bmj.c117

UEKI, M. (2009). A note on automatic variable selection using smooth-threshold estimating equations. *Biometrika* **96** 1005–1011. MR2767286 https://doi.org/10.1093/biomet/asp060

WANG, L., ZHOU, J. and QU, A. (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics* **68** 353–360. MR2959601 https://doi.org/10.1111/j.1541-0420.2011.01678.x

WEI, S. and KOSOROK, M. R. (2013). Latent supervised learning. *J. Amer. Statist. Assoc.* **108** 957–970. MR3174676 https://doi.org/10.1080/01621459.2013.789695

WEI, S. and KOSOROK, M. R. (2018). The change-plane Cox model. *Biometrika* **105** 891–903. MR3877872 https://doi.org/10.1093/biomet/asy050

YATES, N. L., DECAMP, A. C., KORBER, B. T., LIAO, H.-X., IRENE, C., PINTER, A., PEACOCK, J., HARRIS, L. J., SAWANT, S. et al. (2018). HIV-1 envelope glycoproteins from diverse clades differentiate antibody responses and durability among vaccinees. *J. Virol.* **92**.

ZOLLA-PAZNER, S., DECAMP, A., GILBERT, P. B., WILLIAMS, C., YATES, N. L., WILLIAMS, W. T., HOWINGTON, R., FONG, Y., MORRIS, D. E. et al. (2014). Vaccine-induced IgG antibodies to V1V2 regions of multiple HIV-1 subtypes correlate with decreased risk of HIV-1 infection. *PLoS ONE* **9** e87572.

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443 https://doi.org/10.1214/009053607000000802

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1220–1238
https://doi.org/10.1214/22-AOAS1666

BAYESIAN ANALYSIS FOR IMBALANCED POSITIVE-UNLABELLED DIAGNOSIS CODES IN ELECTRONIC HEALTH RECORDS

BY RU WANG[1,a], YE LIANG[1,b], ZHUQI MIAO[2,c] AND TIEMING LIU[3,d]

[1]*Department of Statistics, Oklahoma State University,* [a]*ru.wang@okstate.edu,* [b]*ye.liang@okstate.edu*
[2]*School of Business, State University of New York at New Paltz,* [c]*miaoz@newpaltz.edu*
[3]*School of Industrial Engineering and Management, Oklahoma State University,* [d]*tieming.liu@okstate.edu*

With the increasing availability of electronic health records (EHR), significant progress has been made on developing predictive inference and algorithms by health-data analysts and researchers. However, the EHR data are notoriously noisy, due to missing and inaccurate inputs, despite abundant information. One serious problem is that only a small portion of patients in the database has confirmatory diagnoses, while many other patients remain undiagnosed because they did not comply with the recommended examinations. The phenomenon leads to a so-called positive-unlabelled situation, and the labels are extremely imbalanced. In this paper we propose a model-based approach to classify the unlabelled patients by using a Bayesian finite mixture model. We also discuss the label switching issue for the imbalanced data and propose a consensus Monte Carlo approach to address the imbalance issue and improve computational efficiency simultaneously. Simulation studies show that our proposed model-based approach outperforms existing positive-unlabelled learning algorithms. The proposed method is applied on the Cerner EHR for detecting diabetic retinopathy (DR) patients using laboratory measurements. With only 3% confirmatory diagnoses in the EHR database, we estimate the actual DR prevalence to be 25% which coincides with reported findings in the medical literature.

REFERENCES

ANDREWS, J. L., MCNICHOLAS, P. D. and SUBEDI, S. (2011). Model-based classification via mixtures of multivariate t-distributions. *Comput. Statist. Data Anal.* **55** 520–529. MR2736573 https://doi.org/10.1016/j.csda.2010.05.019

AZZALINI, A. and CAPITANIO, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 367–389. MR1983753 https://doi.org/10.1111/1467-9868.00391

BOTSIS, T., HARTVIGSEN, G., CHEN, F. and WENG, C. (2010). Secondary use of EHR: Data quality issues and informatics opportunities. *Summit on Translational Bioinformatics* **2010** 1.

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

CHAWLA, N. V., BOWYER, K. W., HALL, L. O. and KEGELMEYER, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *J. Artificial Intelligence Res.* **16** 321–357.

CIULLA, T. A., AMADOR, A. G. and ZINMAN, B. (2003). Diabetic retinopathy and diabetic macular edema: Pathophysiology, screening, and novel therapies. *Diabetes Care* **26** 2653–2664.

CLAESEN, M., DE SMET, F., SUYKENS, J. A. and DE MOOR, B. (2015). A robust ensemble approach to learn from positive and unlabeled data using SVM base models. *Neurocomputing* **160** 73–84.

DEAN, N., MURPHY, T. B. and DOWNEY, G. (2006). Using unlabelled data to update classification rules with applications in food authenticity studies. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **55** 1–14. MR2224157 https://doi.org/10.1111/j.1467-9876.2005.00526.x

DZIADKOWIEC, O., CALLAHAN, T., OZKAYNAK, M., REEDER, B. and WELTON, J. (2016). Using a data quality framework to clean data extracted from the electronic health record: A case study. *EGEMS* (*Wash DC*) **4** 1201. https://doi.org/10.13063/2327-9214.1201

FONG, D. S., AIELLO, L., GARDNER, T. W., KING, G. L., BLANKENSHIP, G., CAVALLERANO, J. D., FERRIS, F. L. and KLEIN, R. (2004). Retinopathy in diabetes. *Diabetes Care* **27** s84–s87.

Key words and phrases. Classification, multivariate t mixture, consensus Monte Carlo, diabetic retinopathy.

GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

HRIPCSAK, G. and ALBERS, D. J. (2012). Next-generation phenotyping of electronic health records. *J. Am. Med. Inform. Assoc.* **20** 117–121.

HUANG, Y., ENGLEHART, K. B., HUDGINS, B. and CHAN, A. D. (2005). A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. *IEEE Trans. Biomed. Eng.* **52** 1801–1811.

KOBRIN, K. and BARBARA, E. (2007). Overview of epidemiologic studies of diabetic retinopathy. *Ophthalmic Epidemiol.* **14** 179–183.

LANGE, K. L., LITTLE, R. J. A. and TAYLOR, J. M. G. (1989). Robust statistical modeling using the t distribution. *J. Amer. Statist. Assoc.* **84** 881–896. MR1134486

LI, X. and LIU, B. (2003). Learning to classify texts using positive and unlabeled data. In *IJCAI* **3** 587–592.

LI, X.-L., YU, P. S., LIU, B. and NG, S.-K. (2009). Positive unlabeled learning for data stream classification. In *Proceedings of the* 2009 *SIAM International Conference on Data Mining* 259–270. SIAM.

LIU, B., LEE, W. S., YU, P. S. and LI, X. (2002). Partially supervised classification of text documents. In *ICML* **2** 387–394. Citeseer.

LIU, B., DAI, Y., LI, X., LEE, W. S. and PHILIP, S. Y. (2003). Building text classifiers using positive and unlabeled examples. In *ICDM* **3** 179–188. Citeseer.

LO, K. and GOTTARDO, R. (2012). Flexible mixture modeling via the multivariate t distribution with the Box–Cox transformation: An alternative to the skew-t distribution. *Stat. Comput.* **22** 33–52. MR2865054 https://doi.org/10.1007/s11222-010-9204-1

MARTELLA, F., VERMUNT, J. K., BEEKMAN, M., WESTENDORP, R. G. J., SLAGBOOM, P. E. and HOUWING-DUISTERMAAT, J. J. (2011). A mixture model with random-effects components for classifying sibling pairs. *Stat. Med.* **30** 3252–3264. MR2861473 https://doi.org/10.1002/sim.4365

MCNICHOLAS, P. D. (2017). *Mixture Model-Based Classification*. CRC Press, Boca Raton, FL. MR3642443

MIT-CRITICAL-DATA (2016). *Secondary Analysis of Electronic Health Records*. Springer.

MORDELET, F. and VERT, J.-P. (2011). ProDiGe: Prioritization of disease genes with multitask machine learning from positive and unlabeled examples. *BMC Bioinform.* **12** 389. https://doi.org/10.1186/1471-2105-12-389

MORDELET, F. and VERT, J.-P. (2014). A bagging SVM to learn from positive and unlabeled examples. *Pattern Recogn. Lett.* **37** 201–209.

NG, K., STEINHUBL, S. R., DEFILIPPI, C., DEY, S. and STEWART, W. F. (2016). Early detection of heart failure using electronic health records: Practical implications for time before diagnosis, data diversity, data quantity, and data density. *Circulation: Cardiovascular Quality and Outcomes* **9** 649–658.

PIRI, S., DELEN, D., LIU, T. and ZOLBANIN, H. M. (2017). A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble. *Decis. Support Syst.* **101** 12–27.

ROOS, M., MARTINS, T. G., HELD, L. and RUE, H. (2015). Sensitivity analysis for Bayesian hierarchical models. *Bayesian Anal.* **10** 321–349. MR3420885 https://doi.org/10.1214/14-BA909

SALEH, E., MORENO, A., VALLS, A., ROMERO-AROCA, P. and DE LA RIVA-FERNANDEZ, S. (2016). A fuzzy random forest approach for the detection of diabetic retinopathy on electronic health record data. In *CCIA* 169–174.

SCOTT, S. L., BLOCKER, A. W., BONASSI, F. V., CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *Int. J. Manag. Sci. Eng. Manag.* **11** 78–88.

SKEVOFILAKAS, M., ZARKOGIANNI, K., KARAMANOS, B. G. and NIKITA, K. S. (2010). A hybrid decision support system for the risk assessment of retinopathy development as a long term complication of type 1 diabetes mellitus. In 2010 *Annual International Conference of the IEEE Engineering in Medicine and Biology* 6713–6716. IEEE.

STEPHENS, M. (2000). Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *Ann. Statist.* **28** 40–74. MR1762903 https://doi.org/10.1214/aos/1016120364

SUN, Y. and ZHANG, D. (2019). Diagnosis and analysis of diabetic retinopathy based on electronic health records. *IEEE Access*.

TING, D. S. W., CHEUNG, G. C. M. and WONG, T. Y. (2016). Diabetic retinopathy: Global prevalence, major risk factors, screening practices and public health challenges: A review. *Clinical & Experimental Ophthalmology* **44** 260–277.

XU, L., CRAMMER, K. and SCHUURMANS, D. (2006). Robust support vector machine training via convex outlier ablation. In *AAAI* **6** 536–542.

YANG, P., LI, X.-L., MEI, J.-P., KWOH, C.-K. and NG, S.-K. (2012). Positive-unlabeled learning for disease gene identification. *Bioinformatics* **28** 2640–2647.

YAU, J. W., ROGERS, S. L., KAWASAKI, R., LAMOUREUX, E. L., KOWALSKI, J. W., BEK, T., CHEN, S.-J., DEKKER, J. M., FLETCHER, A. et al. (2012). Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* **35** 556–564.

ZAWISTOWSKI, M., SUSSMAN, J. B., HOFER, T. P., BENTLEY, D., HAYWARD, R. A. and WIITALA, W. L. (2017). Corrected ROC analysis for misclassified binary outcomes. *Stat. Med.* **36** 2148–2160. MR3648645 https://doi.org/10.1002/sim.7260

ZHANG, X., SAADDINE, J. B., CHOU, C.-F., COTCH, M. F., CHENG, Y. J., GEISS, L. S., GREGG, E. W., ALBRIGHT, A. L., KLEIN, B. E. et al. (2010). Prevalence of diabetic retinopathy in the United States, 2005–2008. *JAMA* **304** 649–656.

ZULUAGA, M. A., HUSH, D., LEYTON, E. J. D., HOYOS, M. H. and ORKISZ, M. (2011). Learning from only positive and unlabeled data to detect lesions in vascular CT images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 9–16. Springer.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1239–1259
https://doi.org/10.1214/22-AOAS1667

SURROGATE MARKER ASSESSMENT USING MEDIATION AND INSTRUMENTAL VARIABLE ANALYSES IN A CASE-COHORT DESIGN

BY YEN-TSUNG HUANG[a], JIH-CHANG YU[b] AND JUI-HSIANG LIN[c]

Institute of Statistical Science, Academia Sinica, [a]*ythuang@stat.sinica.edu.tw,* [b]*yujihchang@stat.sinica.edu.tw,* [c]*linjuihsiang@stat.sinica.edu.tw*

The identification of surrogate markers for gold standard outcomes in clinical trials enables future cost-effective trials that target the identified markers. Due to resource limitations, these surrogate markers may be collected only for cases and for a subset of the trial cohort, giving rise to what is termed the case-cohort design. Motivated by a COVID-19 vaccine trial, we propose methods of assessing the surrogate markers for a time-to-event outcome in a case-cohort design by using mediation and instrumental variable (IV) analyses. In the mediation analysis we decomposed the vaccine effect on COVID-19 risk into an indirect effect (the effect mediated through the surrogate marker such as neutralizing antibodies) and a direct effect (the effect not mediated by the marker), and we propose that the mediation proportions are surrogacy indices. In the IV analysis we aimed to quantify the causal effect of the surrogate marker on disease risk in the presence of surrogate-disease confounding which is unavoidable even in randomized trials. We employed weighted estimating equations derived from nonparametric maximum likelihood estimators (NPMLEs) under semiparametric probit models for the time-to-disease outcome. We plugged in the weighted NPMLEs to construct estimators for the aforementioned causal effects and surrogacy indices, and we determined the asymptotic properties of the proposed estimators. Finite sample performance was evaluated in numerical simulations. Applying the proposed mediation and IV analyses to a mock COVID-19 vaccine trial data, we found that 84.2% of the vaccine efficacy was mediated by 50% pseudovirus neutralizing antibody and that neutralizing antibodies had significant protective effects for COVID-19 risk.

REFERENCES

AALEN, O. O. (1989). A linear regression model for the analysis of life times. *Stat. Med.* **8** 907–925.

ANGRIST, J. R., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variable. *J. Amer. Statist. Assoc.* **91** 444–472.

BADEN, L. R., EL SAHLY, H. M., KOTLOFF, K., ESSINK, B., FREY, S., NOVAK, R., DIEMERT, D., SPECTOR, S. A., ROUPHAEL, N. et al. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **384** 403–416.

BARON, R. M. and KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration. *J. Pers. Soc. Psychol.* **51** 1173–1182.

BAYART, J. L., MORIMONT, L., CLOSSET, M., WIEERS, G., ROY, T., GERIN, V., ELSEN, M., EUCHER, C., VAN EECKHOUDT, S. et al. (2021). Confounding factors influencing the kinetics and magnitude of serological response following adminstration of BNT162b2. *Microorganisms* **9** 1340.

BENKESER, D., DIAZ, I. and RAN, J. (2021). Inference for natural mediation effects under case-cohort sampling with applications in identifying COVID-19 vaccine correlates of protection.

BOWDEN, R. J. and TURKINGTON, D. A. (1984). *Instrumental Variables*. *Econometric Society Monographs in Quantitative Economics* **8**. Cambridge Univ. Press, Cambridge. MR0798790

BURGESS, S., DANIEL, R., BUTTERWORTH, A. S., THOMPSON, S. G. and EPIC-INTERACT CONSORTIUM (2015). Network Mendelian randomization: Using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int. J. Epidemiol.* **44** 484–495.

Key words and phrases. Causal mediation model, case-cohort study, clinical trials, instrumental variable, surrogate endpoint, survival analysis.

CHEN, Y.-H. and ZUCKER, D. M. (2009). Case-cohort analysis with semiparametric transformation models. *J. Statist. Plann. Inference* **139** 3706–3717. MR2549118 https://doi.org/10.1016/j.jspi.2009.04.023

COWLING, B. J., LIM, W. W., PERERA, R. A. P. M., FANG, V. J., LEUNG, G. M., MALIK PEIRIS, J. S. and TCHETGEN TCHETGEN, E. J. (2019). Influenza hemagglutination-inhibition antibody titer as a mediator of vaccine-induced protection for influenza B. *Clin. Infect. Dis.* **68** 1713–1717.

COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

COX, R. J. and BROKSTAD, K. A. (2020). Not just antibodies: B cells and T cells mediate immunity to COVID-19. *Nat. Rev., Immunol.* **20** 581–582.

U.S. FDA (2020). Emergency use authorization for vaccines explained. Available at https://www.fda.gov/vaccines-blood-biologics/vaccines/emergency-use-authorization-vaccines-explained.

GILBERT, P. B., FONG, Y., BENKESER, D., ANDRIESEN, J., BORATE, B., CARONE, M., CARPP, L. N., DIAZ, I., FAY, M. P. et al. (2021). USG COVID-19 response team / CoVPN vaccine efficacy trial immune correlates statistical analysis plan. Available at https://figshare.com/articles/online_resource/CoVPN_OWS_COVID-19_Vaccine_Efficacy_Trial_Immune_Correlates_SAP/13198595.

GILBERT, P. B., MONTEFIORI, D. C., MCDERMOTT, A., FONG, Y., BENKESER, D., DENG, W., ZHOU, H., HOUCHENS, C. R., MARTINS, K. et al. (2022). Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science* **375** 43–50.

HSIEH, S. M., LIU, M. C., CHEN, Y. H., LEE, W. S., HWANG, S. H., CHENG, S. J., KO, W. C., HWANG, K. P., WANG, N. C. et al. (2021). Safety and immunogenicity of CpG 1018 and aluminium hydroxide-adjuvanted SARS-CoV-2 S-2P protein vaccine MVC-COV1901: interim results of a large-scale, double-blind, randomised, placebo-controlled phase 2 trial in Taiwan. *Lancet Respir. Med.* **9** 1396–1406.

HUANG, Y.-T. (2020). Mendelian randomization using semiparametric linear transformation models. *Stat. Med.* **39** 890–905. MR4075835 https://doi.org/10.1002/sim.8449

HUANG, Y.-T. and CAI, T. (2016). Mediation analysis for survival data using semiparametric probit models. *Biometrics* **72** 563–574. MR3515783 https://doi.org/10.1111/biom.12445

HUANG, Y.-T., YU, J.-C. and LIN, J.-H. (2023). Supplement to "Surrogate marker assessment using mediation and instrumental variable analyses in a case-cohort design." https://doi.org/10.1214/22-AOAS1667SUPPA, https://doi.org/10.1214/22-AOAS1667SUPPB

IMAI, K., KEELE, L. and YAMAMOTO, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statist. Sci.* **25** 51–71. MR2741814 https://doi.org/10.1214/10-STS321

JIN, P., LI, J., PAN, H., WU, Y. and ZHU, F. (2021). Immunological surrogate endpoints of COVID-2019 vaccines: The evidence we have versus the evidence we need. *Signal Transduct. Targeted Ther.* **6** 48.

KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1924807 https://doi.org/10.1002/9781118032985

KEMP, R. and PRASAD, V. (2017). Surrogate endpoints in oncology: When are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Med.* **15** 134.

KJAERSGAARD, M. I. S. and PARNER, E. T. (2016). Instrumental variable method for time-to-event data using a pseudo-observation approach. *Biometrics* **72** 463–472. MR3515773 https://doi.org/10.1111/biom.12451

LANGE, T. and HANSEN, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology* **22** 575–581.

LAWLOR, D. A., HARBORD, R. M., STERNE, J. A. C., TIMPSON, N. and SMITH, G. D. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27** 1133–1163. MR2420151 https://doi.org/10.1002/sim.3034

LI, J., FINE, J. and BROOKHART, A. (2015). Instrumental variable additive hazards models. *Biometrics* **71** 122–130. MR3335356 https://doi.org/10.1111/biom.12244

LIN, S.-H. and VANDERWEELE, T. (2017). Interventional approach for path-specific effects. *J. Causal Inference* **5** Art. No. 20150027. MR4323810 https://doi.org/10.1515/jci-2015-0027

LIN, D. Y. and YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81** 61–71. MR1279656 https://doi.org/10.1093/biomet/81.1.61

LU, W. and TSIATIS, A. A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika* **93** 207–214. MR2277751 https://doi.org/10.1093/biomet/93.1.207

MACKENZIE, T. A., TOSTESON, T. D., MORDEN, N. E., STUKEL, T. A. and O'MALLEY, A. J. (2014). Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. *Health Serv. Outcomes Res. Methodol.* **14** 54–68.

MACKINNON, D. (2008). *Introduction to Statistical Mediation Analysis*. Taylor & Francis, New York.

MARTINUSSEN, T., VANSTEELANDT, S., TCHETGEN TCHETGEN, E. J. and ZUCKER, D. M. (2017). Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. *Biometrics* **73** 1140–1149. MR3744528 https://doi.org/10.1111/biom.12699

McMahan, K., Yu, J., Mercado, N. B., Loos, C., Tostanoski, L. H., Chandrashekar, A., Liu, J., Peter, L., Atyeo, C. et al. (2020). Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* **590** 630–634.

Miller, K., Wang, M., Gralow, J., Dickler, M., Cobleigh, M., Perez, E. A., Shenkier, T., Cella, D. and Davidson, N. E. (2007). Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N. Engl. J. Med.* **357** 2666–2676.

Ni, L., Ye, F., CHeng, M. L., Feng, Y., Deng, Y. Q., Zhao, H., Wei, P., Ge, J., Gou, M. et al. (2020). Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. *Immunity* **52** 971–977.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence* 411–420. Morgan Kaufmann, San Francisco, CA.

Phillips, D. J., Feng, S., White, T., Sayal, H., Aley, P. K., Bibi, S., Dold, C., Fuskova, M., Gilbert, S. C. et al. (2021). Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nat. Med.* **27** 2032–2040.

Plotkin, S. A. and Gilbert, P. B. (2012). Nomencalture for immune correlates of protection after vaccination. *Clin. Infect. Dis.* **54** 1615–1617.

Prentice, R. L. (1986). A case-cohort design for epidemiological cohort studies and disease prevention trials. *Biometrika* **73** 1–11.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat. Med.* **8** 431–440.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* **3** 143–155.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

Tchetgen Tchetgen, E. J. (2011). On causal mediation analysis with a survival outcome. *Int. J. Biostat.* **7** Art. 33, 38 pp. MR2843528 https://doi.org/10.2202/1557-4679.1351

Tchetgen Tchetgen, E. J., Walter, S., Vansteelandt, S., Martinussen, T. and Glymour, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology* **26** 401–410.

VanderWeele, T. J. (2009). Concerning the consistency assumption in causal inference. *Epidemiology* **20** 880–883.

VanderWeele, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology* **22** 582–585.

VanderWeele, T. J. (2013). Surrogate measures and consistent surrogates. *Biometrics* **69** 561–569. MR3106581 https://doi.org/10.1111/biom.12071

VanderWeele, T. J. and Ding, P. (2017). Sensitivity analysis in observational research: Introducing the e-value. *Ann. Intern. Med.* **167** 268–274.

VanderWeele, T. J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Stat. Interface* **2** 457–468. MR2576399 https://doi.org/10.4310/SII.2009.v2.n4.a7

VanderWeele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M. and Kraft, P. (2014). Methodological challenges in Mendelian randomization. *Epidemiology* **25** 427–435.

Vansteelandt, S. and Daniel, R. M. (2017). Interventional effects for mediation analysis with multiple mediators. *Epidemiology* **28** 258–265.

WHO (2021). WHO meeting on correlates of protection: COVID-19 vaccines. Available at https://www.who.int/news-room/events/detail/2021/06/01/default-calendar/covid-19-vaccineswho-meeting-on-correlates-of-protection.

Wolfson, J. and Gilbert, P. (2010). Statistical identifiability and the surrogate endpoint problem, with application to vaccine trials. *Biometrics* **66** 1153–1161. MR2758503 https://doi.org/10.1111/j.1541-0420.2009.01380.x

Wu, K., Werner, A. P., Koch, M., Choi, A., Narayanan, E., Stewart-Jones, G. B. E., Colpitts, T., Bennett, H., Boyoglu-Barnum, S. et al. (2021). Serum neutralizing activity elicited by mRNA-1273 vaccine. *N. Engl. J. Med.* **384** 1468–1470.

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93** 627–640. MR2261447 https://doi.org/10.1093/biomet/93.3.627

Zucker, D. M. (2005). A pseudo-partial likelihood method for semiparametric survival regression with covariate errors. *J. Amer. Statist. Assoc.* **100** 1264–1277. MR2236440 https://doi.org/10.1198/016214505000000538

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1260–1284
https://doi.org/10.1214/22-AOAS1668

DETECTING DISTRIBUTIONAL DIFFERENCES IN LABELED SEQUENCE DATA WITH APPLICATION TO TROPICAL CYCLONE SATELLITE IMAGERY

BY TREY MCNEELY[1,a], GALEN VINCENT[1,b], KIMBERLY M. WOOD[2,d], RAFAEL IZBICKI[3,e] AND ANN B. LEE[1,c]

[1]*Department of Statistics and Data Science, Carnegie Mellon University,* [a]*treymcneely@gmail.com,* [b]*galenbvincent@gmail.com,* [c]*annlee@andrew.cmu.edu*

[2]*Department of Geosciences, Mississippi State University,* [d]*kimberly.wood@msstate.edu*

[3]*Department of Statistics, Federal University of São Carlos,* [e]*rafaelizbicki@gmail.com*

Our goal is to quantify whether, and if so how, spatiotemporal patterns in tropical cyclone (TC) satellite imagery signal an upcoming rapid intensity change event. To address this question, we propose a new nonparametric test of association between a time series of images and a series of binary event labels. We ask whether there is a difference in distribution between (dependent but identically distributed) 24-hour sequences of images preceding an event vs. a nonevent. By rewriting the statistical test as a regression problem, we leverage neural networks to infer modes of structural evolution of TC convection that are representative of the lead-up to rapid intensity change events. Dependencies between nearby sequences are handled by a bootstrap procedure that estimates the marginal distribution of the label series. We prove that type I error control is guaranteed as long as the distribution of the label series is well estimated which is made easier by the extensive historical data for binary TC event labels. We show empirical evidence that our proposed method identifies archetypes of infrared imagery associated with elevated rapid intensification risk, typically marked by deep or deepening core convection over time. Such results provide a foundation for improved forecasts of rapid intensification.

REFERENCES

AMINIKHANGHAHI, S. and COOK, D. J. (2017). A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **51** 339–367. https://doi.org/10.1007/s10115-016-0987-z

BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. MR4060981

BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72. MR1910074 https://doi.org/10.1214/ss/1023798998

CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. MR3798878 https://doi.org/10.1111/rssb.12265

CHAKRAVARTI, P., KUUSELA, M., LEI, J. and WASSERMAN, L. (2021). Model-independent detection of new physics signals using interpretable semi-supervised classifier tests. arXiv preprint arXiv:2102.07679.

DEMARIA, M. and KAPLAN, J. (1999). An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Weather Forecast.* **14** 326–337.

EVANS, C. and G'SELL, M. (2020). Sequential changepoint detection for label shift in classification. arXiv preprint arXiv:2009.08592.

GALL, R., FRANKLIN, J., MARKS, F., RAPPAPORT, E. N. and TOEPFER, F. (2013). The hurricane forecast improvement project. *Bull. Am. Meteorol. Soc.* **94** 329–343.

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

Key words and phrases. Two-sample testing, high-dimensional time series, association studies, remote sensing, functional data, weather forecasting.

GRIMMETT, G. R. and STIRZAKER, D. R. (2020). *Probability and Random Processes*. Oxford Univ. Press, Oxford. Fourth edition [of 0667520]. MR4229142

HOROWITZ, J. L. (2003). Bootstrap methods for Markov processes. *Econometrica* **71** 1049–1082. MR1995823 https://doi.org/10.1111/1468-0262.00439

HOVMÖLLER, E. (1949). The trough-and-ridge diagram. *Tellus* **1** 62–66.

JANOWIAK, J., JOYCE, B. and XIE, P. (2020). NCEP/CPC L3 half hourly 4 km global (60S–60N) merged IR V1. https://doi.org/10.5067/P4HZB9N27EKU

KANDASAMY, K., KRISHNAMURTHY, A., POCZOS, B., WASSERMAN, L. A. and ROBINS, J. M. (2015). Nonparametric von Mises estimators for entropies, divergences and mutual informations. In *NIPS* **15** 397–405.

KAPLAN, J. and DEMARIA, M. (2003). Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Weather Forecast*. **18** 1093–1108.

KAPLAN, J., DEMARIA, M. and KNAFF, J. A. (2010). A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Weather Forecast*. **25** 220–241.

KAPLAN, J., ROZOFF, C. M., DEMARIA, M., SAMPSON, C. R., KOSSIN, J. P., VELDEN, C. S., CIONE, J. J., DUNION, J. P., KNAFF, J. A. et al. (2015). Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Weather Forecast*. **30** 1374–1396.

KATSEVICH, E. and RAMDAS, A. (2020). A theoretical treatment of conditional independence testing under model-x. arXiv preprint arXiv:2005.05506v4.

KIM, I., LEE, A. B. and LEI, J. (2019). Global and local two-sample tests via regression. *Electron. J. Stat*. **13** 5253–5305. MR4043073 https://doi.org/10.1214/19-EJS1648

KIM, I., RAMDAS, A., SINGH, A. and WASSERMAN, L. (2021). Classification accuracy as a proxy for two-sample testing. *Ann. Statist*. **49** 411–434. MR4206684 https://doi.org/10.1214/20-AOS1962

KLOTZBACH, P. J., BOWEN, S. G., PIELKE, R. and BELL, M. (2018). Continental U.S. hurricane landfall frequency and associated damage: Observations and future risks. *Bull. Am. Meteorol. Soc*. **99** 1359–1376. https://doi.org/10.1175/BAMS-D-17-0184.1

KNAFF, J. A. and DEMARIA, R. T. (2017). Forecasting tropical cyclone eye formation and dissipation in infrared imagery. *Weather Forecast*. **32** 2103–2116.

KREISS, J.-P. and PAPARODITIS, E. (2011). Bootstrap methods for dependent data: A review. *J. Korean Statist. Soc*. **40** 357–378. MR2906623 https://doi.org/10.1016/j.jkss.2011.08.009

LANDSEA, C. W. and FRANKLIN, J. L. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Weather Rev*. **141** 3576–3592. https://doi.org/10.1175/MWR-D-12-00254.1

LI, Q. and RACINE, J. S. (2007). *Nonparametric Econometrics*: *Theory and Practice*. Princeton Univ. Press, Princeton, NJ. MR2283034

LUO, C., LOU, J.-G., LIN, Q., FU, Q., DING, R., ZHANG, D. and WANG, Z. (2014). Correlating events with time series for incident diagnosis. In *Proceedings of the* 20*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1583–1592.

MCNEELY, T., LEE, A. B., HAMMERLING, D. and WOOD, K. (2019). Quantifying the spatial structure of tropical cyclone imagery.

MCNEELY, T., LEE, A. B., WOOD, K. M. and HAMMERLING, D. (2020). Unlocking GOES: A statistical framework for quantifying the evolution of convective structure in tropical cyclones. *J. Appl. Meteorol. Climatol*. **59** 1671–1689.

MCNEELY, T., KHOKHLOV, P., DALMASSO, N., WOOD, K. M. and LEE, A. B. (2022). Structural forecasting for short-term tropical cyclone intensity guidance. arXiv preprint arXiv:2206.00067.

MCNEELY, T., VINCENT, G., IZBICKI, R., WOOD, K. M. and LEE, A. B. (2023). Supplement to "Detecting distributional differences in labeled sequence data with application to tropical cyclone satellite imagery." https://doi.org/10.1214/22-AOAS1668SUPPA, https://doi.org/10.1214/22-AOAS1668SUPPB

MOON, K. and HERO, A. (2014). Multivariate f-divergence estimation with confidence. *Adv. Neural Inf. Process. Syst*. **27** 2420–2428.

ROGERS, R. (2010). Convective-scale structure and evolution during a high-resolution simulation of tropical cyclone rapid intensification. *J. Atmos. Sci*. **67** 44–70.

SANABIA, E. R., BARRETT, B. S. and FINE, C. M. (2014). Relationships between tropical cyclone intensity and eyewall structure as determined by radial profiles of inner-core infrared brightness temperature. *Mon. Weather Rev*. **142** 4581–4599. https://doi.org/10.1175/MWR-D-13-00336.1

SCHARWÄCHTER, E. and MÜLLER, E. (2020a). Does terrorism trigger online hate speech? On the association of events and time series. *Ann. Appl. Stat*. **14** 1285–1303. MR4152133 https://doi.org/10.1214/20-AOAS1338

SCHARWÄCHTER, E. and MÜLLER, E. (2020b). Two-sample testing for event impacts in time series. In *Proceedings of the* 2020 *SIAM International Conference on Data Mining* 10–18. SIAM.

SESIA, M., SABATTI, C. and CANDÈS, E. J. (2019). Gene hunting with hidden Markov model knockoffs. *Biometrika* **106** 1–18. MR3912377 https://doi.org/10.1093/biomet/asy033

WOOD, K. M. and RITCHIE, E. A. (2015). A definition for rapid weakening of North Atlantic and eastern North Pacific tropical cyclones. *Geophys. Res. Lett*. **42** 10,091–10,097.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1285–1306
https://doi.org/10.1214/22-AOAS1669

A BAYESIAN ACCELERATED FAILURE TIME MODEL FOR INTERVAL CENSORED THREE-STATE SCREENING OUTCOMES

BY THOMAS KLAUSCH[a], EDDYMURPHY U. AKWIWU[b], MARK A. VAN DE WIEL[c], VEERLE M. H. COUPÉ[d] AND JOHANNES BERKHOF[e]

Amsterdam University Medical Centers, Epidemiology and Data Science, [a]*t.klausch@amsterdamumc.nl,* [b]*e.akwiwu@amsterdamumc.nl,* [c]*mark.vdwiel@amsterdamumc.nl,* [d]*v.coupe@amsterdamumc.nl,* [e]*h.berkhof@amsterdamumc.nl*

Women infected by the human papillomavirus are at an increased risk to develop cervical intraepithelial neoplasia lesions (CIN). CIN are classified into three grades of increasing severity (CIN-1, CIN-2, and CIN-3) and can eventually develop into cervical cancer. The main purpose of screening is detecting CIN-2 and CIN-3 cases which are usually removed surgically. Screening data from the POBASCAM trial involving 1454 HPV-positive women are analyzed with two objectives, estimate: (a) the transition time from HPV diagnosis to CIN-3 and (b) the transition time from CIN-2 to CIN-3. The screening data have two key characteristics. First, the CIN state is monitored in an interval censored sequence of screening times. Second, a woman's progression to CIN-3 is only observed if the woman progresses to, both, CIN-2 and from CIN-2 to CIN-3 in the same screening interval. We propose a Bayesian accelerated failure time model for the two transition times in this three-state model. To deal with the unusual censoring structure of the screening data, we develop a Metropolis-within-Gibbs algorithm with data augmentation from the truncated transition time distributions.

REFERENCES

ALAA, A. M. and VAN DER SCHAAR, M. (2018). A hidden absorbing semi-Markov model for informatively censored temporal data: Learning and inference. *J. Mach. Learn. Res.* **19** 4. MR3862411

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

ASANJARANI, A., LIQUET, B. and NAZARATHY, Y. (2021). Estimation of semi-Markov multi-state models: A comparison of the sojourn times and transition intensities approaches. *Int. J. Biostat.* **18** 243–262. https://doi.org/10.1515/ijb-2020-0083

BORUVKA, A. and COOK, R. J. (2016). Sieve estimation in a Markov illness-death process under dual censoring. *Biostatistics* **17** 350–363. MR3516005 https://doi.org/10.1093/biostatistics/kxv042

DE GRUTTOLA, V. and LAGAKOS, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* **45** 1–11. MR0999438 https://doi.org/10.2307/2532030

DE WREEDE, L. C., FIOCCO, M. and PUTTER, H. (2011). Mstate: An R package for the analysis of competing risks and multi-state models. *J. Stat. Softw.* **38** 1–30.

DIJKSTRA, M. G., VAN ZUMMEREN, M., ROZENDAAL, L., VAN KEMENADE, F. J., HELMERHORST, T. J., SNIJDERS, P. J., MEIJER, C. J. and BERKHOF, J. (2016). Safety of extending screening intervals beyond five years in cervical screening programmes with testing for high risk human papillomavirus: 14 year follow-up of population based randomised cohort in the Netherlands. *Br. Med. J.* **355** 1–8.

FAY, M. P. and SHAW, P. A. (2010). Exact and asymptotic weighted logrank tests for interval censored data: The interval R package. *J. Stat. Softw.* **36** 1–34.

FOUCHER, Y., GIRAL, M., SOULILLOU, J.-P. and DAURES, J.-P. (2007). A semi-Markov model for multistate and interval-censored data with multiple terminal events. Application in renal transplantation. *Stat. Med.* **26** 5381–5393. MR2416834 https://doi.org/10.1002/sim.3100

FOUCHER, Y., GIRAL, M., SOULILLOU, J. P. and DAURES, J. P. (2010). A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. *Stat. Methods Med. Res.* **19** 127–145. MR2649717 https://doi.org/10.1177/0962280208093889

Key words and phrases. Bayesian, survival, AFT, multistate, screening, interval censoring, semi-Markov.

GELMAN, A., HWANG, J. and VEHTARI, A. (2014). Understanding predictive information criteria for Bayesian models. *Stat*. *Comput*. **24** 997–1016. MR3253850 https://doi.org/10.1007/s11222-013-9416-2

GELMAN, A., JAKULIN, A., PITTAU, M. G. and SU, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Ann*. *Appl*. *Stat*. **2** 1360–1383. MR2655663 https://doi.org/10.1214/08-AOAS191

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2013). *Bayesian Data Analysis*, 3rd ed. CRC Press/CRC, Boca Raton, FL.

JACKSON, C. H. (2011). Multi-state models for panel data: The msm package for R. *J. Stat*. *Softw.* **38** 1–28.

JACKSON, C. H. (2016). flexsurv: A platform for parametric survival modeling in R. *J. Stat*. *Softw.* **70** 1–33.

KANG, M. and LAGAKOS, S. W. (2007). Statistical methods for panel data from a semi-Markov process, with application to HPV. *Biostatistics* **8** 252–264.

KLAUSCH, T., AKWIWU, E. U., VAN DE WIEL, M. A., COUPÉ, V. M. H. and BERKHOF, J. (2023a). Supplementary material to 'A Bayesian accelerated failure time model for interval-censored three-state screening outcomes': Additional details on the simulation and POBASCAM data analysis. https://doi.org/10.1214/22-AOAS1669SUPPA

KLAUSCH, T., AKWIWU, E. U., VAN DE WIEL, M. A., COUPÉ, V. M. H. and BERKHOF, J. (2023b). Supplementary material to 'A Bayesian accelerated failure time model for interval-censored three-state screening outcomes': R code and POBASCAM data. https://doi.org/10.1214/22-AOAS1669SUPPB

KOMÁREK, A. and LESAFFRE, E. (2008). Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *J. Amer. Statist*. *Assoc*. **103** 523–533. MR2523990 https://doi.org/10.1198/016214507000000563

LANGE, J. M. and MININ, V. N. (2013). Fitting and interpreting continuous-time latent Markov models for panel data. *Stat*. *Med*. **32** 4581–4595. MR3118377 https://doi.org/10.1002/sim.5861

LANGE, J. M., HUBBARD, R. A., INOUE, L. Y. T. and MININ, V. N. (2015). A joint model for multistate disease processes and random informative observation times, with applications to electronic medical records data. *Biometrics* **71** 90–101. MR3335353 https://doi.org/10.1111/biom.12252

LANGE, J. M., GULATI, R., LEONARDSON, A. S. et al. (2018). Estimating and comparing cancer progression risks under varying surveillance protocols. *Ann*. *Appl*. *Stat*. **12** 1773–1795. MR3852697 https://doi.org/10.1214/17-AOAS1130

LUO, Y., STEPHENS, D. A., VERMA, A. and BUCKERIDGE, D. L. (2021). Bayesian latent multi-state modeling for nonequidistant longitudinal electronic health records. *Biometrics* **77** 78–90. MR4229722 https://doi.org/10.1111/biom.13261

MANDEL, M. (2010). Estimating disease progression using panel data. *Biostatistics* **11** 304–316.

RAFFA, J. D. and DUBIN, J. A. (2015). Multivariate longitudinal data analysis with mixed effects hidden Markov models. *Biometrics* **71** 821–831. MR3402618 https://doi.org/10.1111/biom.12296

RIJKAART, D. C., BERKHOF, J., ROZENDAAL, L., VAN KEMENADE, F. J., BULKMANS, N. W. J., HEIDEMAN, D. A. M., KENTER, G. G., CUZICK, J., SNIJDERS, P. J. F. et al. (2012). Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: Final results of the POBASCAM randomised controlled trial. *Lancet Oncol*. **13** 78–88. https://doi.org/10.1016/S1470-2045(11)70296-0

TITMAN, A. C. and SHARPLES, L. D. (2010). Semi-Markov models with phase-type sojourn distributions. *Biometrics* **66** 742–752. MR2758210 https://doi.org/10.1111/j.1541-0420.2009.01339.x

TURNBULL, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist*. *Soc*. *Ser. B* **38** 290–295. MR0652727

WEI, S. and KRYSCIO, R. J. (2016). Semi-Markov models for interval censored transient cognitive states with back transitions and a competing risk. *Stat*. *Methods Med*. *Res*. **25** 2909–2924. MR3572890 https://doi.org/10.1177/0962280214534412

WILLIAMS, J. P., STORLIE, C. B., THERNEAU, T. M., JACK, C. R. JR. and HANNIG, J. (2020). A Bayesian approach to multistate hidden Markov models: Application to dementia progression. *J. Amer. Statist*. *Assoc*. **115** 16–31. MR4078442 https://doi.org/10.1080/01621459.2019.1594831

WITTE, B. I., BERKHOF, J. and JONKER, M. A. (2017). An EM algorithm for nonparametric estimation of the cumulative incidence function from repeated imperfect test results. *Stat*. *Med*. **36** 3412–3421. MR3689079 https://doi.org/10.1002/sim.7373

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1307–1332
https://doi.org/10.1214/22-AOAS1670

LOCUS: A REGULARIZED BLIND SOURCE SEPARATION METHOD WITH LOW-RANK STRUCTURE FOR INVESTIGATING BRAIN CONNECTIVITY

BY YIKAI WANG[a] AND YING GUO[b]

Department of Biostatistics and Bioinformatics, Emory University, [a]*johnzon.wyk@gmail.com,* [b]*yguo2@emory.edu*

Network-oriented research has been increasingly popular in many scientific areas. In neuroscience research, imaging-based network connectivity measures have become the key for understanding brain organizations, potentially serving as individual neural fingerprints. There are major challenges in analyzing connectivity matrices, including the high dimensionality of brain networks, unknown latent sources underlying the observed connectivity, and the large number of brain connections leading to spurious findings. In this paper we propose a novel blind source separation method with low-rank structure and uniform sparsity (LOCUS) as a fully data-driven decomposition method for network measures. Compared with the existing method that vectorizes connectivity matrices ignoring brain network topology, LOCUS achieves more efficient and accurate source separation for connectivity matrices using low-rank structure. We propose a novel angle-based uniform sparsity regularization that demonstrates better performance than the existing sparsity controls for low-rank tensor methods. We propose a highly efficient iterative node-rotation algorithm that exploits the block multiconvexity of the objective function to solve the nonconvex optimization problem for learning LOCUS. We illustrate the advantage of LOCUS through extensive simulation studies. Application of LOCUS to Philadelphia Neurodevelopmental Cohort neuroimaging study reveals biologically insightful connectivity traits which are not found using the existing method.

REFERENCES

ALLEN, G. (2012). Sparse higher-order principal components analysis. In *Artificial Intelligence and Statistics* 27–36.

AMICO, E. and GOÑI, J. (2018a). The quest for identifiability in human functional connectomes. *Sci*. *Rep*. **8** 8254. https://doi.org/10.1038/s41598-018-25089-1

AMICO, E. and GOÑI, J. (2018b). Mapping hybrid functional-structural connectivity traits in the human connectome. *Netw*. *Neurosci*. **2** 306–322. https://doi.org/10.1162/netn_a_00049

AMICO, E., MARINAZZO, D., DI PERRI, C., HEINE, L., ANNEN, J., MARTIAL, C., DZEMIDZIC, M., KIRSCH, M., BONHOMME, V. et al. (2017). Mapping the functional connectome traits of levels of consciousness. *NeuroImage* **148** 201–211.

BECKMANN, C. F. and SMITH, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans*. *Med*. *Imag*. **23** 137–152.

BECKMANN, C. F. and SMITH, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *NeuroImage* **25** 294–311. https://doi.org/10.1016/j.neuroimage.2004.10.043

BISWAL, B., ZERRIN YETKIN, F., HAUGHTON, V. M. and HYDE, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn*. *Reson*. *Med*. **34** 537–541.

BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat*. *Rev*. *Neurosci*. **10** 186–198. https://doi.org/10.1038/nrn2575

CHEN, K., DONG, H. and CHAN, K. (2013). Reduced rank regression via adaptive nuclear norm penalization. *Biometrika* **100** 901–920. MR3142340 https://doi.org/10.1093/biomet/ast036

CHUNG, M. K. (2018). Statistical challenges of big brain network data. *Statist*. *Probab*. *Lett*. **136** 78–82. MR3806842 https://doi.org/10.1016/j.spl.2018.02.020

Key words and phrases. Blind source separation, low rank, matrix factorization, network connectivity, neuroimaging.

CHURCH, J. A., FAIR, D. A., DOSENBACH, N. U., COHEN, A. L., MIEZIN, F. M., PETERSEN, S. E. and SCHLAGGAR, B. L. (2008). Control networks in paediatric Tourette syndrome show immature and anomalous patterns of functional connectivity. *Brain* **132** 225–238.

CONTRERAS, J. A., GOÑI, J., RISACHER, S. L., AMICO, E., YODER, K., DZEMIDZIC, M., WEST, J. D., MCDONALD, B. C., FARLOW, M. R. et al. (2017). Cognitive complaints in older adults at risk for Alzheimer's disease are associated with altered resting-state networks. *Alzheimer's Dement*. **6** 40–49.

DAVIES, M. (2004). Identifiability issues in noisy ICA. *IEEE Signal Process*. *Lett*. **11** 470–473.

DECO, G., JIRSA, V. K. and MCINTOSH, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat*. *Rev*. *Neurosci*. **12** 43–56. https://doi.org/10.1038/nrn2961

DURANTE, D., DUNSON, D. B. and VOGELSTEIN, J. T. (2017). Nonparametric Bayes modeling of populations of networks. *J*. *Amer*. *Statist*. *Assoc*. **112** 1516–1530. MR3750873 https://doi.org/10.1080/01621459.2016.1219260

EAVANI, H., SATTERTHWAITE, T. D., FILIPOVYCH, R., GUR, R. E., GUR, R. C. and DAVATZIKOS, C. (2015). Identifying sparse connectivity patterns in the brain using resting-state fMRI. *NeuroImage* **105** 286–299.

ERIKSSON, J. and KOIVUNEN, V. (2004). Identifiability, separability, and uniqueness of linear ICA models. *IEEE Signal Process*. *Lett*. **11** 601–604.

FAN, J., GONG, W. and ZHU, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *J*. *Econometrics* **212** 177–202. MR3994013 https://doi.org/10.1016/j.jeconom.2019.04.026

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J*. *Amer*. *Statist*. *Assoc*. **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

FINN, E. S., SHEN, X., SCHEINOST, D., ROSENBERG, M. D., HUANG, J., CHUN, M. M., PAPADEMETRIS, X. and CONSTABLE, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nat*. *Neurosci*. **18** 1664.

FRISTON, K. J. (2011). Functional and effective connectivity: A review. *Brain Connect*. **1** 13–36.

FRISTON, K., FRITH, C., LIDDLE, P. and FRACKOWIAK, R. (1993). Functional connectivity: The principal-component analysis of large (PET) data sets. *J*. *Cereb*. *Blood Flow Metab*. **13** 5–14.

GLASSER, M. F., SOTIROPOULOS, S. N., WILSON, J. A., COALSON, T. S., FISCHL, B., ANDERSSON, J. L., XU, J., JBABDI, S., WEBSTER, M. et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80** 105–124.

GORSKI, J., PFEUFFER, F. and KLAMROTH, K. (2007). Biconvex sets and optimization with biconvex functions: A survey and extensions. *Math*. *Methods Oper*. *Res*. **66** 373–407. MR2357657 https://doi.org/10.1007/s00186-007-0161-1

GUO, Y. (2011). A general probabilistic model for group independent component analysis and its estimation methods. *Biometrics* **67** 1532–1542. MR2872404 https://doi.org/10.1111/j.1541-0420.2011.01601.x

HOFF, G., VAN DEN HEUVEL, M., BENDERS, M. J., KERSBERGEN, K. J. and DE VRIES, L. S. (2013). On development of functional brain connectivity in the young brain. *Front*. *Human Neurosci*. **7** 650.

HYVÄRINEN, A., KARHUNEN, J. and OJA, E. (2001). *Independent Component Analysis* **46**. Wiley, New York.

HYVÄRINEN, A. and OJA, E. (2000). Independent component analysis: Algorithms and applications. *Neural Netw*. **13** 411–430.

INGALHALIKAR, M., SMITH, A., PARKER, D., SATTERTHWAITE, T. D., ELLIOTT, M. A., RUPAREL, K., HAKONARSON, H., GUR, R. E., GUR, R. C. et al. (2014). Sex differences in the structural connectome of the human brain. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **111** 823–828.

KAGAN, A. M., LINNIK, Y. V., LINNIK, U. V. and LINNIK, I. V. (1973). *Characterization Problems in Mathematical Statistics*. Wiley-Interscience.

KEERATIMAHAT, K. and NICHOLS, T. E. (2021). Discussion on "Distributional independent component analysis for diverse neuroimaging modalities" by Ben Wu, Subhadip Pal, Jian Kang, and Ying Guo. *Biometrics*.

KEMMER, P. B., GUO, Y., WANG, Y. and PAGNONI, G. (2015). Network-based characterization of brain functional connectivity in Zen practitioners. *Front*. *Psychol*. **6**.

KEMMER, P. B., WANG, Y., BOWMAN, F. D., MAYBERG, H. and GUO, Y. (2018). Evaluating the strength of structural connectivity underlying brain functional networks. *Brain Connect*. **8** 579–594.

KIM, H.-J., OLLILA, E. and KOIVUNEN, V. (2013). Sparse regularization of tensor decompositions. In 2013 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* 3836–3840. IEEE.

KUNDU, P., BENSON, B. E., ROSEN, D., FRANGOU, S., LEIBENLUFT, E., LUH, W.-M., BANDETTINI, P. A., PINE, D. S. and ERNST, M. (2018). The integration of functional brain activity from adolescence to adulthood. *J*. *Neurosci*. **38** 3559–3570.

KUNDU, S., LUKEMIRE, J., WANG, Y. and GUO, Y. (2019). A novel joint brain network analysis using longitudinal Alzheimer's disease data. *Sci*. *Rep*. **9** 1–18.

LANG, E. W., TOMÉ, A. M., KECK, I. R., GÓRRIZ-SÁEZ, J. and PUNTONET, C. G. (2012). Brain connectivity analysis: A short survey. *Comput*. *Intell*. *Neurosci*. **2012** 8.

LI, X., XU, D., ZHOU, H. and LI, L. (2018). Tucker tensor regression and neuroimaging analysis. *Stat. Biosci.* **10** 520–545.

LUKEMIRE, J., WANG, Y., VERMA, A. and GUO, Y. (2020). HINT: A hierarchical independent component analysis toolbox for investigating brain functional networks using neuroimaging data. *J. Neurosci. Methods* 108726.

MAIRAL, J., BACH, F., PONCE, J. and SAPIRO, G. (2009). Online dictionary learning for sparse coding. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 689–696.

MAYBERG, H. S. (2003). Modulating dysfunctional limbic-cortical circuits in depression: Towards development of brain-based algorithms for diagnosis and optimised treatment. *Br. Med. Bull.* **65** 193–207. https://doi.org/10.1093/bmb/65.1.193

MEJIA, A. F., NEBEL, M. B., WANG, Y., CAFFO, B. S. and GUO, Y. (2020). Template independent component analysis: Targeted and reliable estimation of subject-level brain networks using big data population priors. *J. Amer. Statist. Assoc.* **115** 1151–1177. MR4143456 https://doi.org/10.1080/01621459.2019.1679638

MINKA, T. P. (2000). Automatic choice of dimensionality for PCA. In *NIPS* **13** 598–604.

POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M. et al. (2011). Functional network organization of the human brain. *Neuron* **72** 665–678.

RABUSSEAU, G. and KADRI, H. (2016). Low-rank regression with tensor responses. In *Advances in Neural Information Processing Systems* 1867–1875.

RASKUTTI, G. and YUAN, M. (2015). Convex regularization for high-dimensional tensor regression. Preprint. Available at arXiv:1512.01215 **639**.

REAL, R. and VARGAS, J. M. (1996). The probabilistic basis of Jaccard's index of similarity. *Syst. Biol.* **45** 380–385.

SATTERTHWAITE, T. D., WOLF, D. H., ROALF, D. R., RUPAREL, K., ERUS, G., VANDEKAR, S., GENNATAS, E. D., ELLIOTT, M. A., SMITH, A. et al. (2014a). Linked sex differences in cognition and functional connectivity in youth. *Cereb. Cortex* **25** 2383–2394.

SATTERTHWAITE, T. D., ELLIOTT, M. A., RUPAREL, K., LOUGHEAD, J., PRABHAKARAN, K., CALKINS, M. E., HOPSON, R., JACKSON, C., KEEFE, J. et al. (2014b). Neuroimaging of the Philadelphia neurodevelopmental cohort. *NeuroImage* **86** 544–553.

SATTERTHWAITE, T. D., WOLF, D. H., ROALF, D. R., RUPAREL, K., ERUS, G., VANDEKAR, S., GENNATAS, E. D., ELLIOTT, M. A., SMITH, A. et al. (2015). Linked sex differences in cognition and functional connectivity in youth. *Cereb. Cortex* **25** 2383–2394.

SEELEY, W. W., CRAWFORD, R. K., ZHOU, J., MILLER, B. L. and GREICIUS, M. D. (2009). Neurodegenerative diseases target large-scale human brain networks. *Neuron* **62** 42–52.

SHI, R. and GUO, Y. (2016). Investigating differences in brain functional networks using hierarchical covariate-adjusted independent component analysis. *Ann. Appl. Stat.* **10** 1930–1957. MR3592043 https://doi.org/10.1214/16-AOAS946

SMITH, S. M., FOX, P. T., MILLER, K. L., GLAHN, D. C., FOX, P. M., MACKAY, C. E., FILIPPINI, N., WATKINS, K. E., TORO, R. et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. USA* **106** 13040–13045.

SMITH, S. M., MILLER, K. L., SALIMI-KHORSHIDI, G., WEBSTER, M., BECKMANN, C. F., NICHOLS, T. E., RAMSEY, J. D. and WOOLRICH, M. W. (2011). Network modelling methods for FMRI. *NeuroImage* **54** 875–891.

SOLO, V., POLINE, J.-B., LINDQUIST, M. A., SIMPSON, S. L., BOWMAN, F. D., CHUNG, M. K. and CASSIDY, B. (2018). Connectivity in fMRI: Blind spots and breakthroughs. *IEEE Trans. Med. Imag.* **37** 1537–1550. https://doi.org/10.1109/TMI.2018.2831261

SUN, W. W. and LI, L. (2017). STORE: Sparse tensor response regression and neuroimaging analysis. *J. Mach. Learn. Res.* **18** Paper No. 135, 37. MR3763769

TZOURIO-MAZOYER, N., LANDEAU, B., PAPATHANASSIOU, D., CRIVELLO, F., ETARD, O., DELCROIX, N., MAZOYER, B. and JOLIOT, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15** 273–289.

WANG, Y. and GUO, Y. (2019). A hierarchical independent component analysis model for longitudinal neuroimaging studies. *NeuroImage* **189** 380–400.

WANG, Y. and GUO, Y. (2023). Supplement to "LOCUS: A regularized blind source separation method with low-rank structure for investigating brain connectivity." https://doi.org/10.1214/22-AOAS1670SUPPA, https://doi.org/10.1214/22-AOAS1670SUPPB

WANG, Y., MENG, D. and YUAN, M. (2018). Sparse recovery: From vectors to tensors. *Nat. Sci. Rev.* **5** 756–767.

WANG, W., ZHANG, X. and LI, L. (2019). Common reducing subspace model and network alternation analysis. *Biometrics* **75** 1109–1120. MR4041815 https://doi.org/10.1111/biom.13099

WANG, Y., KANG, J., KEMMER, P. B. and GUO, Y. (2016). An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. *Front*. *Neurosci*. **10**.

WANG, L., DURANTE, D., JUNG, R. E. and DUNSON, D. B. (2017). Bayesian network–response regression. *Bioinformatics* **33** 1859–1866.

WILLIAMS, L. M. (2016). Precision psychiatry: A neural circuit taxonomy for depression and anxiety. *Lancet Psychiatry* **3** 472–480.

WU, G.-R., STRAMAGLIA, S., CHEN, H., LIAO, W. and MARINAZZO, D. (2013). Mapping the voxel-wise effective connectome in resting state fMRI. *PLoS ONE* **8** e73670.

WU, B., PAL, S., KANG, J. and GUO, Y. (2021). Rejoinder to discussions of "Distributional independent component analysis for diverse neuroimaging modalities". *Biometrics*.

XIA, M., WANG, J. and HE, Y. (2013). BrainNet viewer: A network visualization tool for human brain connectomics. *PLoS ONE* **8** e68910.

YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found*. *Comput*. *Math*. **16** 1031–1068. MR3529132 https://doi.org/10.1007/s10208-015-9269-5

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann*. *Statist*. **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

ZHOU, H., LI, L. and ZHU, H. (2013). Tensor regression with applications in neuroimaging data analysis. *J*. *Amer*. *Statist*. *Assoc*. **108** 540–552. MR3174640 https://doi.org/10.1080/01621459.2013.776499

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1333–1352
https://doi.org/10.1214/22-AOAS1672

SIMULATING FLOOD EVENT SETS USING EXTREMAL PRINCIPAL COMPONENTS

BY CHRISTIAN ROHRBECK[1,a] AND DANIEL COOLEY[2,b]

[1]*Department of Mathematical Sciences, University of Bath,* [a]*cr777@bath.ac.uk*
[2]*Department of Statistics, Colorado State University,* [b]*cooleyd@stat.colostate.edu*

Hazard event sets, a collection of synthetic extreme events over a given period, are important for catastrophe modelling. This paper addresses the issue of generating event sets of extreme river flow for northern England and southern Scotland, a region which has been particularly affected by severe flooding over the past 20 years. We start by analysing historical extreme river flow across 45 gauges, using methods from extreme value analysis, including the concept of extremal principal components. Our analysis reveals interesting connections between the extremal dependence structure and the region's topography/climate. We then introduce a framework which is based on modelling the distribution of the extremal principal components in order to generate synthetic events of extreme river flow. The generative framework is dimension-reducing in that it distinctly handles the principal components based on their contribution to describing the nature of extreme river flow across the study region. We also detail a data-driven approach to select the optimal dimension. Synthetic flood events are subsequently generated efficiently by sampling from the fitted distribution. Our results indicate good agreement between the observed and simulated extreme river flow dynamics and, therefore, illustrate the usefulness of our approach to practitioners. For the considered application, we also find that our approach outperforms existing statistical approaches for generating hazard event sets.

REFERENCES

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050. MR3456363 https://doi.org/10.1214/15-AOAS863

BALLANI, F. and SCHLATHER, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika* **98** 633–645. MR2836411 https://doi.org/10.1093/biomet/asr034

BARLOW, A. M., SHERLOCK, C. and TAWN, J. (2020). Inference for extreme values under threshold-based stopping rules. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **69** 765–789. MR4133146

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2108013 https://doi.org/10.1002/0470012382

BHATIA, S., JAIN, A. and HOOI, B. (2021). ExGAN: Adversarial generation of extreme samples. In *Proceedings of the AAAI Conference on Artificial Intelligence* **35** 6750–6758.

BOLDI, M.-O. and DAVISON, A. C. (2007). A mixture model for multivariate extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 217–229. MR2325273 https://doi.org/10.1111/j.1467-9868.2007.00585.x

BOULAGUIEM, Y., ZSCHEISCHLER, J., VIGNOTTO, E., VAN DER WIEL, K. and ENGELKE, S. (2022). Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks. *Environ. Data Sci.* **1** e5.

BRACKEN, C., RAJAGOPALAN, B., CHENG, L., KLEIBER, W. and GANGOPADHYAY, S. (2016). Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain. *Water Resour. Res.* **52** 6643–6655.

CAMICI, S., BROCCA, L., MELONE, F. and MORAMARCO, T. (2014). Impact of climate change on flood frequency using different climate models and downscaling approaches. *J. Hydrol. Eng.* **19** 04014002.

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics*. Springer London, Ltd., London. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

Key words and phrases. Multivariate extreme value theory, nonparametric bootstrapping, principal component analysis, spatial flood risk analysis.

COLES, S., HEFFERNAN, J. E. and TAWN, J. A. (1999). Dependence measures for extreme value analyses. *Extremes* **2** 339–365.

COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53** 377–392. MR1108334

COOLEY, D., DAVIS, R. A. and NAVEAU, P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *J. Multivariate Anal.* **101** 2103–2117. MR2671204 https://doi.org/10.1016/j.jmva.2010.04.007

COOLEY, D., NYCHKA, D. and NAVEAU, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.* **102** 824–840. MR2411647 https://doi.org/10.1198/016214506000000780

COOLEY, D. and THIBAUD, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika* **106** 587–604. MR3992391 https://doi.org/10.1093/biomet/asz028

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980 https://doi.org/10.1214/11-STS376

DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B* **52** 393–425. MR1086795

DE CARVALHO, M. and DAVISON, A. C. (2014). Spectral density ratio models for multivariate extremes. *J. Amer. Statist. Assoc.* **109** 764–776. MR3223748 https://doi.org/10.1080/01621459.2013.872651

DREES, H. and SABOURIN, A. (2021). Principal component analysis for multivariate extremes. *Electron. J. Stat.* **15** 908–943. MR4255291 https://doi.org/10.1214/21-ejs1803

DREVETON, C. and GUILLOU, Y. (2004). Use of a principal components analysis for the generation of daily time series. *J. Appl. Meteorol.* **43** 984–996.

EASTOE, E. F. (2019). Nonstationarity in peaks-over-threshold river flows: A regional random effects model. *Environmetrics* **30** e2560, 18 pp. MR3999533 https://doi.org/10.1002/env.2560

EASTOE, E. F. and TAWN, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 25–45. MR2662232 https://doi.org/10.1111/j.1467-9876.2008.00638.x

ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 871–932. MR4136498

ENGELKE, S. and IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.* **8** 241–270. MR4243547 https://doi.org/10.1146/annurev-statistics-040620-041554

ENVIRONMENT AGENCY (2018). Estimating the economic costs of the winter floods 2015 to 2016. Ref: LIT 10736. Available at https://www.gov.uk/government/publications/floods-of-winter-2015-to-2016-estimating-the-costs.

GROSSI, P. and KUNREUTHER, H. (2005). *Catastrophe Modeling: A New Approach to Managing Risk*. Springer, New York.

HALL, P., WATSON, G. S. and CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–762. MR0919843 https://doi.org/10.1093/biomet/74.4.751

HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 497–546. MR2088289 https://doi.org/10.1111/j.1467-9868.2004.02050.x

HUSER, R. and WADSWORTH, J. L. (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdiscip. Rev.: Comput. Stat.* **14** Paper No. e1537, 28 pp. MR4376783 https://doi.org/10.1002/wics.1537

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR0980699 https://doi.org/10.1016/0167-7152(89)90106-5

INSTITUTE OF HYDROLOGY (GREAT BRITAIN) (1975). *Flood Studies Report*. Natural Environment Research Council, London.

KEEF, C., TAWN, J. A. and LAMB, R. (2013). Estimating the probability of widespread flood events. *Environmetrics* **24** 13–21. MR3042270 https://doi.org/10.1002/env.2190

LARSSON, M. and RESNICK, S. I. (2012). Extremal dependence measure and extremogram: The regularly varying case. *Extremes* **15** 231–256. MR2915582 https://doi.org/10.1007/s10687-011-0135-9

NORTHROP, P. J., ATTALIDES, N. and JONATHAN, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 93–120. MR3611679 https://doi.org/10.1111/rssc.12159

PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. MR0423667

QUINN, N., BATES, P. D., NEAL, J., SMITH, A., WING, O., SAMPSON, C., SMITH, J. and HEFFERNAN, J. E. (2019). The spatial dependence of flood hazard and risk in the United States. *Water Resour. Res.* **55** 1890–1911.

ROHRBECK, C. and COOLEY, D. (2023). Supplement to "Simulating flood event sets using extremal principal components." https://doi.org/10.1214/22-AOAS1672SUPPA, https://doi.org/10.1214/22-AOAS1672SUPPB

ROHRBECK, C. and TAWN, J. A. (2021). Bayesian spatial clustering of extremal behavior for hydrological variables. *J. Comput. Graph. Statist.* **30** 91–105. MR4235967 https://doi.org/10.1080/10618600.2020.1777139

TAWN, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* **75** 397–415. MR0967580 https://doi.org/10.1093/biomet/75.3.397

WADSWORTH, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics* **58** 116–126. MR3463162 https://doi.org/10.1080/00401706.2014.998345

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1353–1374
https://doi.org/10.1214/22-AOAS1673

ESTIMATING GLOBAL AND COUNTRY-SPECIFIC EXCESS MORTALITY DURING THE COVID-19 PANDEMIC

BY VICTORIA KNUTSON[1,a], SERGE ALESHIN-GUENDEL[1,b], ARIEL KARLINSKY[2,3,c], WILLIAM MSEMBURI[4,d] AND JON WAKEFIELD[1,5,e]

[1]*Department of Biostatistics, University of Washington,* [a]*vknuts@uw.edu,* [b]*aleshing@uw.edu*
[2]*Department of Economics, Hebrew University,* [c]*ariel.karlinsky@mail.huji.ac.il*
[3]*The Federmann Center for the Study of Rationality, Hebrew University*
[4]*Division of Data, Analytics and Delivery for Impact, World Health Organization,* [d]*msemburiw@who.int*
[5]*Department of Statistics, University of Washington,* [e]*jonno@uw.edu*

Estimating the true mortality burden of COVID-19 for every country in the world is a difficult, but crucial, public health endeavor. Attributing deaths, direct or indirect, to COVID-19 is problematic. A more attainable target is the "excess deaths," the number of deaths in a particular period, relative to that expected during "normal times," and we develop a model for this endeavor. The excess mortality requires two numbers, the total deaths and the expected deaths, but the former is unavailable for many countries, and so modeling is required for such countries. The expected deaths are based on historic data, and we develop a model for producing estimates of these deaths for all countries. We allow for uncertainty in the modeled expected numbers when calculating the excess. The methods we describe were used to produce the World Health Organization (WHO) excess death estimates. To achieve both interpretability and transparency we developed a relatively simple overdispersed Poisson count framework within which the various data types can be modeled. We use data from countries with national monthly data to build a predictive log-linear regression model with time-varying coefficients for countries without data. For a number of countries, subnational data only are available, and we construct a multinomial model for such data, based on the assumption that the fractions of deaths in subregions remain approximately constant over time. Our inferential approach is Bayesian, with the covariate predictive model being implemented in the fast and accurate `INLA` software. The subnational modeling was carried out using MCMC in `Stan`. Based on our modeling, the point estimate for global excess mortality during 2020–2021 is 14.8 million, with a 95% credible interval of (13.2, 16.6) million.

REFERENCES

ADAIR, T. and LOPEZ, A. D. (2018). Estimating the completeness of death registration: An empirical method. *PLoS ONE* **13** e0197047.

BAKER, S. G. (1994). The multinomial-Poisson transformation. *J. R. Stat. Soc., Ser. D, Stat.* **43** 495–504.

BÜHLMANN, P. and HOTHORN, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statist. Sci.* **22** 477–505. MR2420454 https://doi.org/10.1214/07-STS242

CARPENTER, B., GELMAN, A., HOFFMAN, M., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P.and RIDDELL, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*.

CHECCHI, F. and ROBERTS, L. (2005). HPN network paper 52: Interpreting and using mortality data in humanitarian emergencies: A primer for non-epidemiologists. Technical report, Overseas Development Institute.

DEZEURE, R., BÜHLMANN, P., MEIER, L. and MEINSHAUSEN, N. (2015). High-dimensional inference: Confidence intervals, p-values and R-software `hdi`. *Statist. Sci.* **30** 533–558. MR3432840 https://doi.org/10.1214/15-STS527

THE ECONOMIST and SOLSTAD (2021a). https://www.economist.com/graphic-detail/coronavirus-excess-deaths-estimates.

Key words and phrases. Bayesian inference, global health, Poisson framework, subnational modeling.

THE ECONOMIST and SOLSTAD (2021b). The pandemic's true death toll. https://www.economist.com/graphic-detail/2021/05/13/how-we-estimated-the-true-death-toll-of-the-pandemic.

FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. MR1873328 https://doi.org/10.1214/aos/1013203451

GARCIA, J., TORRES, C., BARBIERI, M., CAMARDA, C. G., CAMBOIS, E., CAPORALI, A., MESLÉ, F., PONIAKINA, S. and ROBINE, J.-M. (2021). Differences in Covid-19 mortality: Implications of imperfect and diverse data collection systems. *Population* **76** 35–72.

GBD 2019 DEMOGRAPHICS COLLABORATORS AND OTHERS (2020). Global, regional, and national age-sex-specific fertility, mortality, and population estimates, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease study 2019. *Lancet*.

GBD (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019. *Lancet* **396** 1204–1222.

GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869. MR1055437

HALE, T., ANGRIST, N., KIRA, B., PETHERICK, A., PHILLIPS, T. and WEBSTER, S. (2020). Variation in government responses to COVID-19. Technical report, Univ. Oxford.

KARLINSKY, A. (2021). International completeness of death registration 2015–2019. *medRxiv*.

KARLINSKY, A. (2022). Estimating national excess mortality from subnational data: Application to Argentina. *Rev. Panam. Salud Pública* **46** e19. https://doi.org/10.26633/RPSP.2022.19

KARLINSKY, A. and KOBAK, D. (2021). Tracking excess mortality across countries during the Covid-19 pandemic with the world mortality dataset. *eLife* **10**. https://doi.org/10.7554/eLife.69336

KELSALL, J. E., ZEGER, S. L. and SAMET, J. M. (1999). Frequency domain log-linear models; air pollution and mortality. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **48** 331–344.

KNUTSON, V., ALESHIN-GUENDEL, S., KARLINSKY, A., MSEMBURI, W. and WAKEFIELD, J. (2023). Supplement to "Estimating global and country-specific excess mortality during the Covid-19 pandemic." https://doi.org/10.1214/22-AOAS1673SUPPA, https://doi.org/10.1214/22-AOAS1673SUPPB

KUNG, S., DOPPEN, M., BLACK, M., HILLS, T. and KEARNS, N. (2020). Reduced mortality in New Zealand during the Covid-19 pandemic. *Lancet*.

LEON, D. A., SHKOLNIKOV, V. M., SMEETH, L., MAGNUS, P., PECHHOLDOVÁ, M. and JARVIS, C. I. (2020). Covid-19: A need for real-time monitoring of weekly excess deaths. *Lancet* **395** e81.

LINK, W. A. (2013). A cautionary note on the discrete uniform prior for the binomial N. *Ecology* **94** 2173–2179.

LUNDE, B. Å. S., KLEPPE, T. S. and SKAUG, H. J. (2020). An information criterion for automatic gradient tree boosting. Preprint. Available at arXiv:2008.05926.

MIKKELSEN, L., PHILLIPS, D. E., ABOU ZAHR, C., SETEL, P. W., SAVIGNY, D., LOZANO, R. and LOPEZ, A. D. (2015). A global assessment of civil registration and vital statistics systems: Monitoring data quality and progress. *Lancet* **386** 1395–1406.

MSEMBURI, W., KARLINSKY, A., KNUTSON, V., ALESHIN-GUENDEL, S., CHATTERJI, S. and WAKEFIELD, J. (2023). The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature* **613** 130–137. https://doi.org/10.1038/s41586-022-05522-2

NÉMETH, L., JDANOV, D. A. and SHKOLNIKOV, V. M. (2021). An open-sourced, web-based application to analyze weekly excess mortality based on the short-term mortality fluctuations data series. *PLoS ONE* **16** e0246663. https://doi.org/10.1371/journal.pone.0246663

PARKS, R. M., BENNETT, J. E., FOREMAN, K. J., TOUMI, R. and EZZATI, M. (2018). National and regional seasonal dynamics of all-cause and cause-specific mortality in the USA from 1980 to 2016. *eLife* **7**. https://doi.org/10.7554/eLife.35500

RIFFE, T., ACOSTA, E., THE COVERAGE-DB TEAM (2021). Data resource profile: COVerAGE-DB: A global demographic database of Covid-19 cases and deaths. *Int. J. Epidemiol.* **50** 390–390f. https://doi.org/10.1093/ije/dyab027

RIVERA, R., ROSENBAUM, J. E. and QUISPE, W. (2020). Excess mortality in the United States during the first three months of the Covid-19 pandemic. *Epidemiol. Infect.* **148** e264. https://doi.org/10.1017/S0950268820002617

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. CRC Press/CRC, Boca Raton, FL. MR2130347 https://doi.org/10.1201/9780203492024

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

RUE, H., RIEBLER, A., SØRBYE, S. H., ILLIAN, J. B., SIMPSON, D. P. and LINDGREN, F. K. (2017). Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application* **4** 395–421.

SIMPSON, D., RUE, H., RIEBLER, A., MARTINS, T. G. and SØRBYE, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statist. Sci.* **32** 1–28. MR3634300 https://doi.org/10.1214/16-STS576

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

UNSD (2021). Demographic Yearbook 2021. Technical report, United Nations Statistics Division.

VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6** Art. 25, 23. MR2349918 https://doi.org/10.2202/1544-6115.1309

WHO (2020). WHO methods and data sources for life Tables 1990–2019. Technical report, Department of Data and Analytics, Division of Data, Analytics and Delivery for Impact, WHO, Geneva.

WOOD, S. N. (2017). *Generalized Additive Models: An Introduction with ßf R. Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3726911

WANG, H., PAULSON, K. R., PEASE, S. A., WATSON, S., COMFORT, H., ZHENG, P., ARAVKIN, A. Y., BISIGNANO, C., BARBER, R. M., ALAM, T., *et al.* (2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–2021. *The Lancet*.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1375–1397
https://doi.org/10.1214/22-AOAS1674

DYNAMIC RISK PREDICTION TRIGGERED BY INTERMEDIATE EVENTS USING SURVIVAL TREE ENSEMBLES

BY YIFEI SUN[1,a], SY HAN CHIOU[2,b], COLIN O. WU[3,c], MEGHAN E. MCGARRY[4,d] AND CHIUNG-YU HUANG[5,e]

[1]*Department of Biostatistics, Columbia University,* [a]*ys3072@cumc.columbia.edu*
[2]*Department of Mathematical Sciences, University of Texas at Dallas,* [b]*schiou@utdallas.edu*
[3]*National Heart, Lung, and Blood Institute, National Institutes of Health,* [c]*wuc@nhlbi.nih.gov*
[4]*Department of Pediatrics, University of California San Francisco,* [d]*Meghan.McGarry@ucsf.edu*
[5]*Department of Epidemiology and Biostatistics, University of California San Francisco,* [e]*ChiungYu.Huang@ucsf.edu*

With the availability of massive amounts of data from electronic health records and registry databases, incorporating time-varying patient information to improve risk prediction has attracted great attention. To exploit the growing amount of predictor information over time, we develop a unified framework for landmark prediction, using survival tree ensembles, where an updated prediction can be performed when new information becomes available. Compared to conventional landmark prediction with fixed landmark times, our methods allow the landmark times to be subject-specific and triggered by an intermediate clinical event. Moreover, the nonparametric approach circumvents the thorny issue of model incompatibility at different landmark times. In our framework, both the longitudinal predictors and the event time outcome are subject to right censoring, and thus existing tree-based approaches cannot be directly applied. To tackle the analytical challenges, we propose a risk-set-based ensemble procedure by averaging martingale estimating equations from individual trees. Extensive simulation studies are conducted to evaluate the performance of our methods. The methods are applied to the Cystic Fibrosis Foundation Patient Registry (CFFPR) data to perform dynamic prediction of lung disease in cystic fibrosis patients and to identify important prognosis factors.

REFERENCES

BREIMAN, L. (1996). Bagging predictors. *Mach. Learn.* **24** 123–140.

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

CIAMPI, A., THIFFAULT, J., NAKACHE, J.-P. and ASSELAIN, B. (1986). Stratification by stepwise regression, correspondence analysis and recursive partition: A comparison of three methods of analysis for survival data with covariates. *Comput. Statist. Data Anal.* **4** 185–204.

DAVIS, R. B. and ANDERSON, J. R. (1989). Exponential survival trees. *Stat. Med.* **8** 947–961.

DOMANSKI, M. J., TIAN, X., WU, C. O., REIS, J. P., DEY, A. K., GU, Y., ZHAO, L., BAE, S., LIU, K. et al. (2020). Time course of LDL cholesterol exposure and cardiovascular disease event risk. *J. Am. Coll. Cardiol.* **76** 1507–1516.

DÖRING, G., HOIBY, N., GROUP, C. S. et al. (2004). Early intervention and prevention of lung disease in cystic fibrosis: A European consensus. *J. Cyst. Fibros.* **3** 67–91.

FERRER, L., PUTTER, H. and PROUST-LIMA, C. (2019). Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment. *Stat. Methods Med. Res.* **28** 3649–3666. MR4003613 https://doi.org/10.1177/0962280218811837

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2001). *The Elements of Statistical Learning*. Springer, New York.

GORDON, L. and OLSHEN, R. A. (1985). Tree-structured survival analysis. *Cancer Treat. Rep.* **69** 1065–1069.

HARUN, S. N., WAINWRIGHT, C., KLEIN, K. and HENNIG, S. (2016). A systematic review of studies examining the rate of lung function decline in patients with cystic fibrosis. *Paediatr. Respir. Rev.* **20** 55–66. https://doi.org/10.1016/j.prrv.2016.03.002

Key words and phrases. Dynamic prediction, landmark analysis, multi-state model, survival tree, time-dependent predictors.

HEAGERTY, P. J., LUMLEY, T. and PEPE, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56** 337–344.

HELTSHE, S. L., KHAN, U., BECKETT, V., BAINES, A., EMERSON, J., SANDERS, D. B., GIBSON, R. L., MORGAN, W. and ROSENFELD, M. (2018). Longitudinal development of initial, chronic and mucoid Pseudomonas aeruginosa infection in young children with cystic fibrosis. *J. Cyst. Fibros.* **17** 341–347. https://doi.org/10.1016/j.jcf.2017.10.008

HOTHORN, T., LAUSEN, B., BENNER, A. and RADESPIEL-TRÖGER, M. (2004). Bagging survival trees. *Stat. Med.* **23** 77–91. https://doi.org/10.1002/sim.1593

HOTHORN, T., BÜHLMANN, P., DUDOIT, S., MOLINARO, A. and VAN DER LAAN, M. J. (2006). Survival ensembles. *Biostatistics* **7** 355–373.

ISHWARAN, H., KOGALUR, U. B., BLACKSTONE, E. H. and LAUER, M. S. (2008). Random survival forests. *Ann. Appl. Stat.* **2** 841–860. MR2516796 https://doi.org/10.1214/08-AOAS169

JEWELL, N. P. and NIELSEN, J. P. (1993). A framework for consistent prediction rules based on markers. *Biometrika* **80** 153–164. MR1225221 https://doi.org/10.1093/biomet/80.1.153

KAMATA, H., ASAKURA, T., SUZUKI, S., NAMKOONG, H., YAGI, K., FUNATSU, Y., OKAMORI, S., UNO, S., UWAMINO, Y. et al. (2017). Impact of chronic Pseudomonas aeruginosa infection on health-related quality of life in Mycobacterium avium complex lung disease. *BMC Polm. Med.* **17** 198. https://doi.org/10.1186/s12890-017-0544-x

KNAPP, E. A., FINK, A. K., GOSS, C. H., SEWALL, A., OSTRENGA, J., DOWD, C., ELBERT, A., PETREN, K. M. and MARSHALL, B. C. (2016). The cystic fibrosis foundation patient registry. Design and methods of a national observational disease registry. *Ann. Amer. Thorac. Soc.* **13** 1173–1179.

LEBLANC, M. and CROWLEY, J. (1992). Relative risk trees for censored survival data. *Biometrics* **48** 411–425.

LEBLANC, M. and CROWLEY, J. (1993). Survival trees by goodness of split. *J. Amer. Statist. Assoc.* **88** 457–467. MR1224370

LIN, Y. and JEON, Y. (2006). Random forests and adaptive nearest neighbors. *J. Amer. Statist. Assoc.* **101** 578–590. MR2256176 https://doi.org/10.1198/016214505000001230

MANNINO, D. M., REICHERT, M. M. and DAVIS, K. J. (2006). Lung function decline and outcomes in an adult population. *Am. J. Respir. Crit. Care Med.* **173** 985–990.

MAZIARZ, M., HEAGERTY, P., CAI, T. and ZHENG, Y. (2017). On longitudinal prediction with time-to-event outcome: Comparison of modeling options. *Biometrics* **73** 83–93. MR3632354 https://doi.org/10.1111/biom.12562

MCGARRY, M. E., NEUHAUS, J. M., NIELSON, D. W. and LY, N. P. (2019). Regional variations in longitudinal pulmonary function: A comparison of Hispanic and non-Hispanic subjects with cystic fibrosis in the United States. *Pediatr. Pulmonol.* **54** 1382–1390. https://doi.org/10.1002/ppul.24377

MCGARRY, M. E., HUANG, C.-Y., NIELSON, D. W. and LY, N. P. (2021). Early acquisition and conversion of Pseudomonas aeruginosa in Hispanic youth with cystic fibrosis in the United States. *J. Cyst. Fibros.* **20** 424–431. https://doi.org/10.1016/j.jcf.2020.10.002

MCINTOSH, M. W. and PEPE, M. S. (2002). Combining several screening tests: Optimality of the risk score. *Biometrics* **58** 657–664. MR1926119 https://doi.org/10.1111/j.0006-341X.2002.00657.x

MOLINARO, A. M., DUDOIT, S. and VAN DER LAAN, M. J. (2004). Tree-based multivariate regression and density estimation with right-censored data. *J. Multivariate Anal.* **90** 154–177. MR2064940 https://doi.org/10.1016/j.jmva.2004.02.003

MORGAN, W. J., VANDEVANTER, D. R., PASTA, D. J., FOREMAN, A. J., WAGENER, J. S., KONSTAN, M. W., MORGAN, W., KONSTAN, M., LIOU, T. et al. (2016). Forced expiratory volume in 1 second variability helps identify patients with cystic fibrosis at risk of greater loss of lung function. *The Journal of Pediatrics* **169** 116–121.

PARAST, L., CHENG, S.-C. and CAI, T. (2012). Landmark prediction of long-term survival incorporating short-term event time information. *J. Amer. Statist. Assoc.* **107** 1492–1501. MR3036410 https://doi.org/10.1080/01621459.2012.721281

PROUST-LIMA, C., DARTIGUES, J.-F. and JACQMIN-GADDA, H. (2016). Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: A latent process and latent class approach. *Stat. Med.* **35** 382–398. MR3455508 https://doi.org/10.1002/sim.6731

RIZOPOULOS, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67** 819–829. MR2829256 https://doi.org/10.1111/j.1541-0420.2010.01546.x

RIZOPOULOS, D., MOLENBERGHS, G. and LESAFFRE, E. M. E. H. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biom. J.* **59** 1261–1276. MR3731215 https://doi.org/10.1002/bimj.201600238

SEGAL, M. R. (1988). Regression trees for censored data. *Biometrics* **44** 35–47.

STEINGRIMSSON, J. A., DIAO, L. and STRAWDERMAN, R. L. (2019). Censoring unbiased regression trees and ensembles. *J. Amer. Statist. Assoc.* **114** 370–383. MR3941261 https://doi.org/10.1080/01621459.2017.1407775

STEINGRIMSSON, J. A., DIAO, L., MOLINARO, A. M. and STRAWDERMAN, R. L. (2016). Doubly robust survival trees. *Stat. Med.* **35** 3595–3612. MR3537225 https://doi.org/10.1002/sim.6949

SUN, Y., CHIOU, S. H., WU, C. O., MCGARRY, M. E. and HUANG, C.-Y. (2023). Supplement to "Dynamic risk prediction triggered by intermediate events using survival tree ensembles." https://doi.org/10.1214/22-AOAS1674SUPPA, https://doi.org/10.1214/22-AOAS1674SUPPB

SWEETING, M. J., BARRETT, J. K., THOMPSON, S. G. and WOOD, A. M. (2017). The use of repeated blood pressure measures for cardiovascular risk prediction: A comparison of statistical models in the ARIC study. *Stat. Med.* **36** 4514–4528. MR3731236 https://doi.org/10.1002/sim.7144

TANNER, K. T., SHARPLES, L. D., DANIEL, R. M. and KEOGH, R. H. (2021). Dynamic survival prediction combining landmarking with a machine learning ensemble: Methodology and empirical comparison. *J. Roy. Statist. Soc. Ser. A* **184** 3–30. MR4204910 https://doi.org/10.1111/rssa.12611

TAYLOR, J. M. G., PARK, Y., ANKERST, D. P., PROUST-LIMA, C., WILLIAMS, S., KESTIN, L., BAE, K., PICKLES, T. and SANDLER, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* **69** 206–213. MR3058067 https://doi.org/10.1111/j.1541-0420.2012.01823.x

TWALA, B., JONES, M. C. and HAND, D. J. (2008). Good methods for coping with missing data in decision trees. *Pattern Recogn. Lett.* **29** 950–956.

VAN HOUWELINGEN, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scand. J. Stat.* **34** 70–85. MR2325243 https://doi.org/10.1111/j.1467-9469.2006.00529.x

VAN HOUWELINGEN, H. C. and PUTTER, H. (2008). Dynamic predicting by landmarking as an alternative for multi-state modeling: An application to acute lymphoid leukemia data. *Lifetime Data Anal.* **14** 447–463. MR2464769 https://doi.org/10.1007/s10985-008-9099-8

VAN HOUWELINGEN, H. C. and PUTTER, H. (2012). *Dynamic Prediction in Clinical Survival Analysis. Monographs on Statistics and Applied Probability* **123**. CRC Press, Boca Raton, FL. MR3058205

WANG, J., LUO, S. and LI, L. (2017). Dynamic prediction for multiple repeated measures and event time data: An application to Parkinson's disease. *Ann. Appl. Stat.* **11** 1787–1809. MR3709578 https://doi.org/10.1214/17-AOAS1059

WRIGHT, M. N. and ZIEGLER, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77** 1–17. https://doi.org/10.18637/jss.v077.i01

ZHANG, H. (1995). Splitting criteria in survival trees. *Stat. Model.* **104** 305–313.

ZHANG, H., HOLFORD, T. and BRACKEN, M. B. (1996). A tree-based method of analysis for prospective studies. *Stat. Med.* **15** 37–49. https://doi.org/10.1002/(SICI)1097-0258(19960115)15:1<37::AID-SIM144>3.0.CO;2-0

ZHENG, Y. and HEAGERTY, P. J. (2005). Partly conditional survival models for longitudinal data. *Biometrics* **61** 379–391. MR2140909 https://doi.org/10.1111/j.1541-0420.2005.00323.x

ZHU, R. and KOSOROK, M. R. (2012). Recursively imputed survival trees. *J. Amer. Statist. Assoc.* **107** 331–340. MR2949363 https://doi.org/10.1080/01621459.2011.637468

ZHU, Y., LI, L. and HUANG, X. (2019). Landmark linear transformation model for dynamic prediction with application to a longitudinal cohort study of chronic disease. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **68** 771–791. MR3937473

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1398–1418
https://doi.org/10.1214/22-AOAS1675

MIXED-FREQUENCY EXTREME VALUE REGRESSION: ESTIMATING THE EFFECT OF MESOSCALE CONVECTIVE SYSTEMS ON EXTREME RAINFALL INTENSITY

BY DEBBIE J. DUPUIS[1,a] AND LUCA TRAPIN[2,b]

[1]*Department of Decision Sciences, HEC Montréal,* [a]*debbie.dupuis@hec.ca*
[2]*Department of Statistics, University of Bologna,* [b]*luca.trapin@unibo.it*

Understanding and modeling the determinants of extreme hourly rainfall intensity is of utmost importance for the management of flash-flood risk. Increasing evidence shows that mesoscale convective systems (MCS) are the principal driver of extreme rainfall intensity in the United States. We use extreme value statistics to investigate the relationship between MCS activity and extreme hourly rainfall intensity in Greater St. Louis, an area particularly vulnerable to flash floods. Using a block maxima approach with monthly blocks, we find that the impact of MCS activity on monthly maxima is not homogeneous within the month/block. To appropriately capture this relationship, we develop a mixed-frequency extreme value regression framework accommodating a covariate sampled at a frequency higher than that of the extreme observation.

REFERENCES

AHMADALIPOUR, A. and MORADKHANI, H. (2019). A data-driven analysis of flash flood hazard, fatalities, and damages over the CONUS during 1996–2017. *J. Hydrol.* **578**. https://doi.org/10.1016/j.jhydrol.2019.124106

ALMUKHTAR, S., MIGLIOZZI, B., SCHWARTZ, J. and WILLIAMS, J. (2019). The great flood of 2019: A complete picture of a slow-motion disaster. Available at https://www.nytimes.com/interactive/2019/09/11/us/midwest-flooding.html.

ASADI, P., ENGELKE, S. and DAVISON, A. C. (2018). Optimal regionalization of extreme value distributions for flood estimation. *J. Hydrol.* **556** 182–193. https://doi.org/10.1016/j.jhydrol.2017.10.051

ASHLEY, S. T. and ASHLEY, W. S. (2008). Flood fatalities in the United States. *J. Appl. Meteorol. Climatol.* **47** 805–818.

BARBERO, R., FOWLER, H. J., LENDERINK, G. and BLENKINSOP, S. (2017). Is the intensification of precipitation extremes with global warming better detected at hourly than daily resolutions? *Geophys. Res. Lett.* **44** 974–983.

BARBERO, R., FOWLER, H. J., BLENKINSOP, S., WESTRA, S., MORON, V., LEWIS, E., CHAN, S., LENDERINK, G., KENDON, E. et al. (2019). A synthesis of hourly and daily precipitation extremes in different climatic regions. *Weather Clim. Extrem.* **26**. https://doi.org/10.1016/j.wace.2019.100219

BEE, M., DUPUIS, D. J. and TRAPIN, L. (2019). Realized peaks over threshold: A time-varying extreme value approach with high-frequency based measures. *J. Financ. Econom.* **19** 254–283.

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics*. Wiley, Chichester. With contributions from Daniel De Waal and Chris Ferro. MR2108013 https://doi.org/10.1002/0470012382

CALLAU PODUJE, A. C. and HABERLANDT, U. (2017). Short time step continuous rainfall modeling and simulation of extreme events. *J. Hydrol.* **552** 182–197.

CARBONE, R. E. and TUTTLE, J. D. (2008). Rainfall occurrence in the U.S. warm season: The diurnal cycle. *J. Climate* **21** 4132–4146.

CARREAU, J. and TOULEMONDE, G. (2020). Extra-parametrized extreme value copula: Extension to a spatial framework. *Spat. Stat.* **40** 100410, 20. MR4181142 https://doi.org/10.1016/j.spasta.2020.100410

CHANDLER, R. E. and BATE, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika* **94** 167–183. MR2367830 https://doi.org/10.1093/biomet/asm015

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics*. Springer London, Ltd., London. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

Key words and phrases. Extreme value regression, mesoscale convective systems, mixed-frequency data, rainfall intensity.

DAI, A. and TRENBERTH, K. (2004). The diurnal cycle and its depiction in the community climate system model. *J. Climate* **17** 930–951. https://doi.org/10.1175/1520-0442(2004)017<0930:TDCAID>2.0.CO;2

DAVIES, R. (2014a). Federal aid for Missouri September floods. Available at https://floodlist.com/dealing-with-floods/federal-aid-missouri-september-floods.

DAVIES, R. (2014b). Storms hit central USA—Floods in Missouri, Illinois and Iowa. Available at https://floodlist.com/america/usa/storms-hit-central-usa-floods-missouri-illinois-iowa.

DAVIES, R. (2017). USA—Flash floods in New Orleans, Kansas City and Las Vegas. Available at https://floodlist.com/america/usa/usa-flash-floods-in-new-orleans-kansas-city-and-las-vegas.

EINMAHL, J. H. J. and HE, Y. (2022). Extreme value estimation for heterogeneous data. *J. Bus. Econom. Statist.* https://doi.org/10.1080/07350015.2021.2008408

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events: For Insurance and Finance. Applications of Mathematics (New York)* **33**. Springer, Berlin. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

EVANS, J. and WESTRA, S. (2012). Investigating the mechanisms of diurnal rainfall variability using a regional climate model. *J. Climate* **25** 7232–7247. https://doi.org/10.1175/JCLI-D-11-00616.1

FENG, Z., LEUNG, L. R., HAGOS, S., HOUZE, R. A., BURLEYSON, C. D. and BALAGURU, K. (2016). More frequent intense and long-lived storms dominate the springtime trend in central US rainfall. *Nat. Commun.* **7** 13429. https://doi.org/10.1038/ncomms13429

FENG, Z., SONG, F., SAKAGUCHI, K. and LEUNG, L. R. (2021). Evaluation of mesoscale convective systems in climate simulations: Methodological development and results from MPAS-CAM over the United States. *J. Climate* **34** 2611–2633.

FLOODLIST NEWS (2018). USA—Deadly flash floods in Kentucky and Missouri. Available at https://floodlist.com/america/usa/floods-storm-gordon-kentucky-missouri-september-2018.

FOWLER, H. J., WASKO, C. and PREIN, A. F. (2021). Intensification of short-duration rainfall extremes and implications for flood risk: Current state of the art and future directions. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **379** 20190541. https://doi.org/10.1098/rsta.2019.0541

GHYSELS, E. and QIAN, H. (2019). Estimating MIDAS regressions via OLS with polynomial parameter profiling. *Econom. Stat.* **9** 1–16. MR3907670 https://doi.org/10.1016/j.ecosta.2018.02.001

GOURLEY, J., HONG, Y., FLAMIG, Z., LI, L. and WANG, J. (2017). The FLASH project: Improving the tools for flash flood monitoring and prediction across the United States. *Bull. Am. Meteorol. Soc.* **98** 361–372.

KENDON, E. J., BLENKINSOP, S. and FOWLER, H. J. (2018). When will we detect changes in short-duration precipitation extremes? *J. Climate* **31** 2945–2964.

KOUTSOYIANNIS, D., KOZONIS, D. and MANETAS, A. (1998). A mathematical framework for studying rainfall intensity–duration–frequency relationships. *J. Hydrol.* **206** 118–135.

KUNKEL, K. E., KARL, T. R., BROOKS, H., KOSSIN, J., LAWRIMORE, J. H., ARNDT, D., BOSART, L., CHANGNON, D., CUTTER, S. L. et al. (2013). Monitoring and understanding trends in extreme storms: State of knowledge. *Bull. Am. Meteorol. Soc.* **94** 499–514.

LARSON, L. W. (1996). The great USA flood of 1993. Available at https://www.nwrfc.noaa.gov/floods/papers/oh_2/great.htm.

LEADBETTER, M. R. (1973/74). On extreme values in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **28** 289–303. MR0362465 https://doi.org/10.1007/BF00532947

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes. Springer Series in Statistics*. Springer, New York–Berlin. MR0691492

LI, J. (2017). Hourly station-based precipitation characteristics over the Tibetan Plateau. *Int. J. Climatol.* https://doi.org/10.1002/joc.5281

MASSON-DELMOTTE, V., ZHAI, P., PIRANI, A., CONNORS, S. L., PÉAN, C., BERGER, S., CAUD, N., CHEN, Y., GOLDFARB, L., GOMIS, M. I., HUANG, M., LEITZELL, K., LONNOY, E., MATTHEWS, J. B. R., MAYCOCK, T. K., WATERFIELD, T., YELEKÇI, O., YU, R. and ZHOU B. (eds.) (2021). IPCC, 2021: Summary for policymakers. In: *Climate Change* 2021: *The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

MORON, V., BARBERO, R., EVANS, J., WESTRA, S. and FOWLER, H. (2019). Weather types and hourly to multi-day rainfall characteristics in tropical Australia. *J. Climate* **32**. https://doi.org/10.1175/JCLI-D-18-0384.1

MOSELEY, C., HOHENEGGER, C., BERG, P. and HAERTER, J. O. (2016). Intensification of convective extremes driven by cloud-cloud interaction. *Nat. Geosci.* **9**. https://doi.org/10.1038/NGEO2789

NESBITT, S. W., CIFELLI, R. and RUTLEDGE, S. A. (2006). Storm morphology and rainfall characteristics of TRMM precipitation features. *Mon. Weather Rev.* **134** 2702–2721. https://doi.org/10.1175/Mwr3200.1

O'GORMAN, P. A. (2012). Sensitivity of tropical precipitation extremes to climate change. *Nat. Geosci.* **5** 697–700.

OUARDA, T., YOUSEF, L. and CHARRON, C. (2019). Non-stationary intensity–duration–frequency curves integrating information concerning teleconnections and climate change. *Int. J. Climatol.* **39**. https://doi.org/10.1002/joc.5953

PIELKE JR, R. A. and DOWNTON, M. W. (2000). Precipitation and damaging floods: Trends in the United States, 1932–97. *J. Climate* **13** 3625–3637.

PLOSHAY, J. and LAU, N. (2010). Simulation of the diurnal cycle in tropical rainfall and circulation during boreal summer with a high-resolution GCM. *Mon. Weather Rev.* **138** 3434–3453. https://doi.org/10.1175/2010MWR3291.1

PREIN, A. F., LIU, C., IKEDA, K., TRIER, S. B., RASMUSSEN, R. M., HOLLAND, G. J. and CLARK, M. P. (2017). Increased rainfall volume from future convective storms in the US. *Nat. Clim. Change* **7** 880.

PREIN, A. F., LIU, C., IKEDA, K., BULLOCK, R., RASMUSSEN, R. M., HOLLAND, G. J. and CLARK, M. (2020). Simulating North American mesoscale convective systems with a convection-permitting climate model. *Clim. Dyn.* **55** 95–110.

SAHARIA, M., KIRSTETTER, P., VERGARA, H., GOURLEY, J., HONG, Y. and GIROUD, M. (2017). Mapping flash flood severity in the United States. *J. Hydrometeorol.* **18** 397–411.

SEBILLE, Q., FOUGÈRES, A.-L. and MERCADIER, C. (2017). Modeling extreme rainfall: A comparative study of spatial extreme value models. *Spat. Stat.* **21** 187–208. MR3692184 https://doi.org/10.1016/j.spasta.2017.06.009

NATIONAL WEATHER SERVICE (2022). Flash floods and floods...the awesome power! Last consulted April 12, 2022. Available at https://www.weather.gov/pbz/floods.

STEVENSON, S. N. and SCHUMACHER, R. S. (2014). A 10-year survey of extreme rainfall events in the central and eastern United States using gridded multisensor precipitation analyses. *Mon. Weather Rev.* **142** 3147–3162. https://doi.org/10.1175/MWR-D-13-00345.1

TRENBERTH, K. E., DAI, A., RASMUSSEN, R. M. and PARSONS, D. B. (2003). The changing character of precipitation. *Bull. Am. Meteorol. Soc.* **84** 1205–1217.

TYE, M. R. and COOLEY, D. (2015). A spatial model to examine rainfall extremes in Colorado's front range. *J. Hydrol.* **530** 15–23. https://doi.org/10.1016/j.jhydrol.2015.09.023

WESTRA, S., FOWLER, H., EVANS, J., ALEXANDER, L., BERG, P., JOHNSON, F., KENDON, E., LENDERINK, G. and ROBERTS, N. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Rev. Geophys.* **52** 522–555.

ZHU, L., LIU, X. and LUND, R. (2019). A likelihood for correlated extreme series. *Environmetrics* **30** e2546, 15. MR3948465 https://doi.org/10.1002/env.2546

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1419–1443
https://doi.org/10.1214/22-AOAS1676

LAGGED COUPLINGS DIAGNOSE MARKOV CHAIN MONTE CARLO PHYLOGENETIC INFERENCE

BY LUKE J. KELLY[1,a], ROBIN J. RYDER[1,b] AND GRÉGOIRE CLARTÉ[2,c]

[1]*CEREMADE, CNRS, UMR 7534, Université Paris-Dauphine,* [a]*kelly@ceremade.dauphine.fr,* [b]*ryder@ceremade.dauphine.fr*
[2]*Department of Computer Science, University of Helsinki,* [c]*gregoire.clarte@helsinki.fi*

Phylogenetic inference is an intractable statistical problem on a complex space. Markov chain Monte Carlo methods are the primary tool for Bayesian phylogenetic inference, but it is challenging to construct efficient schemes to explore the associated posterior distribution or assess their performance. Existing approaches are unable to diagnose mixing or convergence of Markov schemes jointly across all components of a phylogenetic model. Lagged couplings of Markov chain Monte Carlo algorithms have recently been developed on simpler spaces to diagnose convergence and construct unbiased estimators. We describe a contractive coupling of Markov chains targeting a posterior distribution over a space of phylogenetic trees with branch lengths, scalar parameters and latent variables. We use these couplings to assess mixing and convergence of Markov chains jointly across all components of the phylogenetic model on trees with up to 200 leaves. Samples from our coupled chains may also be used to construct unbiased estimators.

REFERENCES

ALDOUS, D. J. (2000). Mixing time for a Markov chain on cladograms. *Combin. Probab. Comput.* **9** 191–204. MR1774749 https://doi.org/10.1017/S096354830000417X

ALI, R. H., BARK, M., MIRÓ, J., MUHAMMAD, S. A., SJÖSTRAND, J., ZUBAIR, S. M., ABBAS, R. M. and ARVESTAD, L. (2017). VMCMC: A graphical and statistical analysis tool for Markov chain Monte Carlo traces. *BMC Bioinform.* **18** 1–8.

ATKINS, R. and MCDIARMID, C. (2019). Extremal distances for subtree transfer operations in binary trees. *Ann. Comb.* **23** 1–26. MR3921333 https://doi.org/10.1007/s00026-018-0410-4

AYRES, D. L., CUMMINGS, M. P., BAELE, G., DARLING, A. E., LEWIS, P. O., SWOFFORD, D. L., HUELSENBECK, J. P., LEMEY, P., RAMBAUT, A. et al. (2019). BEAGLE 3: Improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *Syst. Biol.* **68** 1052–1061. https://doi.org/10.1093/sysbio/syz020

BAELE, G., LEMEY, P., RAMBAUT, A. and SUCHARD, M. A. (2017). Adaptive MCMC in Bayesian phylogenetics: An application to analyzing partitioned data in BEAST. *Bioinformatics* **33** 1798–1805. https://doi.org/10.1093/bioinformatics/btx088

BASTIDE, P., HO, L. S. T., BAELE, G., LEMEY, P. and SUCHARD, M. A. (2021). Efficient Bayesian inference of general Gaussian models on large phylogenetic trees. *Ann. Appl. Stat.* **15** 971–997. MR4298958 https://doi.org/10.1214/20-aoas1419

BEIKO, R. G., KEITH, J. M., HARLOW, T. J. and RAGAN, M. A. (2006). Searching for convergence in phylogenetic Markov chain Monte Carlo. *Syst. Biol.* **55** 553–565.

BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759

BISWAS, N., JACOB, P. E. and VANETTI, P. (2019). Estimating convergence of Markov chains with L-lag couplings. In *NeurIPS* 7389–7399.

BISWAS, N., BHATTACHARYA, A., JACOB, P. E. and JOHNDROW, J. E. (2022). Coupling-based convergence assessment of some Gibbs samplers for high-dimensional Bayesian regression with shrinkage priors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 973–996. MR4460582 https://doi.org/10.1111/rssb.12495

BOU-RABEE, N. and EBERLE, A. (2022). Couplings for Andersen dynamics. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** 916–944. MR4421613 https://doi.org/10.1214/21-aihp1197

Key words and phrases. Markov chain Monte Carlo methods, couplings, Bayesian phylogenetic inference.

BOUCHARD-CÔTÉ, A., CHERN, K., CUBRANIC, D., HOSSEINI, S., HUME, J., LEPUR, M., OUYANG, Z. and SGARBI, G. (2021). Blang: Bayesian declarative modelling of general data structures and inference via algorithms based on distribution continua. Available at arXiv:1912.10396.

BOUCKAERT, R., VAUGHAN, T. G., BARIDO-SOTTANI, J., DUCHÊNE, S., FOURMENT, M., GAVRYUSHKINA, A., HELED, J., JONES, G., KÜHNERT, D. et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**.

BROMHAM, L., DUCHÊNE, S., HUA, X., RITCHIE, A. M., DUCHÊNE, D. A. and HO, S. Y. W. (2018). Bayesian molecular dating: Opening up the black box. *Biol. Rev. Camb. Philos. Soc.* **93** 1165–1191.

BROWN, D. G. and OWEN, M. (2019). Mean and variance of phylogenetic trees. *Syst. Biol.* **69** 139–154.

BRYANT, D. and MOULTON, V. (1999). A polynomial time algorithm for constructing the refined Buneman tree. *Appl. Math. Lett.* **12** 51–56. MR1749763 https://doi.org/10.1016/S0893-9659(98)00148-7

BRYANT, D. and MOULTON, V. (2004). Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21** 255–265.

COWLES, M. K. and CARLIN, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.* **91** 883–904. MR1395755 https://doi.org/10.2307/2291683

COWLES, M. K., ROBERTS, G. O. and ROSENTHAL, J. S. (1999). Possible biases induced by MCMC convergence diagnostics. *J. Stat. Comput. Simul.* **64** 87–104. MR1741840 https://doi.org/10.1080/00949659908811968

CRAIU, R. V. and MENG, X.-L. (2022). Double happiness: Enhancing the coupled gains of L-lag coupling via control variates. *Statist. Sinica* **32** 1745–1766. MR4478177 https://doi.org/10.5705/ss.202020.0461

DELLICOUR, S., GILL, M. S., FARIA, N. R., RAMBAUT, A., PYBUS, O. G., SUCHARD, M. A. and LEMEY, P. (2021). Relax, keep walking - a practical guide to continuous phylogeographic inference with BEAST. *Mol. Biol. Evol.* **38** 3486–3493. https://doi.org/10.1093/molbev/msab031

DINH, V. and MATSEN, F. A. IV (2017). The shape of the one-dimensional phylogenetic likelihood function. *Ann. Appl. Probab.* **27** 1646–1677. MR3678481 https://doi.org/10.1214/16-AAP1240

DINH, V., BILGE, A., ZHANG, C. and MATSEN IV, F. A. (2017). Probabilistic path Hamiltonian Monte Carlo. In *Proceedings of the* 34*th International Conference on Machine Learning* (D. Precup and Y. W. Teh, eds.). *Proceedings of Machine Learning Research* **70** 1009–1018.

DOUGLAS, J., ZHANG, R. and BOUCKAERT, R. (2021). Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. *PLoS Comput. Biol.* **17** 1–30.

DRUMMOND, A. J., NICHOLLS, G. K., RODRIGO, A. G. and SOLOMON, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161** 1307–1320.

DRUMMOND, A. J., HO, S. Y. W., PHILLIPS, M. J. and RAMBAUT, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**.

FABRETI, L. G. and HÖHNA, S. (2022). Convergence assessment for Bayesian phylogenetic analysis using MCMC simulation. *Methods Ecol. Evol.* **13** 77–90.

FOURMENT, M., MAGEE, A. F., WHIDDEN, C., BILGE, A., MATSEN IV, F. A. and MININ, V. N. (2019). 19 dubious ways to compute the marginal likelihood of a phylogenetic tree topology. *Syst. Biol.* **69** 209–220.

GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

GEWEKE, J. (2004). Getting it right: Joint distribution tests of posterior simulators. *J. Amer. Statist. Assoc.* **99** 799–804. MR2090912 https://doi.org/10.1198/016214504000001132

GEYER, C. J. (2011). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 3–48. CRC Press, Boca Raton, FL. MR2858443

GEYER, C. J. and MØLLER, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Stat.* **21** 359–373. MR1310082

GRAY, R. D., BRYANT, D. and GREENHILL, S. J. (2010). On the shape and fabric of human history. *Philos. Trans. R. Soc. B* **365** 3923–3933.

GRAY, R. D., DRUMMOND, A. J. and GREENHILL, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323** 479–483.

GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810 https://doi.org/10.1093/biomet/82.4.711

GREENHILL, S. J., BLUST, R. and GRAY, R. D. (2008). The austronesian basic vocabulary database: From bioinformatics to lexomics. *Evol. Bioinform.* **4** 271–283.

HARRINGTON, S. M., WISHINGRAD, V. and THOMSON, R. C. (2020). Properties of Markov chain Monte Carlo performance across many empirical alignments. *Mol. Biol. Evol.*

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97

HENG, J. and JACOB, P. E. (2019). Unbiased Hamiltonian Monte Carlo with couplings. *Biometrika* **106** 287–302. MR3949304 https://doi.org/10.1093/biomet/asy074

HOFFMANN, K., BOUCKAERT, R., GREENHILL, S. J. and KÜHNERT, D. (2021). Bayesian phylogenetic analysis of linguistic data using BEAST. *J. Lang. Evol.*

HÖHNA, S., DEFOIN-PLATEL, M. and DRUMMOND, A. J. (2008). Clock-constrained tree proposal operators in Bayesian phylogenetic inference. In *Int. Conf. Bioinform. Biomed. Eng.* 1–7.

HÖHNA, S. and DRUMMOND, A. J. (2012). Guided tree topology proposals for Bayesian phylogenetic inference. *Syst. Biol.* **61** 1–11. https://doi.org/10.1093/sysbio/syr074

JACOB, P. E., LINDSTEN, F. and SCHÖN, T. B. (2020). Smoothing with couplings of conditional particle filters. *J. Amer. Statist. Assoc.* **115** 721–729. MR4107675 https://doi.org/10.1080/01621459.2018.1548856

JACOB, P. E., O'LEARY, J. and ATCHADÉ, Y. F. (2020). Unbiased Markov chain Monte Carlo methods with couplings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 543–600. MR4112777 https://doi.org/10.1111/rssb.12336

JI, X., FISHER, A. A., SU, S., THORNE, J. L., POTTER, B., LEMEY, P., BAELE, G. and SUCHARD, M. A. (2021). Scalable Bayesian divergence time estimation with ratio transformations. Available at arXiv:2110.13298.

JOHNSON, V. E. (1998). A coupling-regeneration scheme for diagnosing convergence in Markov chain Monte Carlo algorithms. *J. Amer. Statist. Assoc.* **93** 238–248. MR1614640 https://doi.org/10.2307/2669620

KELLY, L. J. and NICHOLLS, G. K. (2017). Lateral transfer in stochastic Dollo models. *Ann. Appl. Stat.* **11** 1146–1168. MR3693561 https://doi.org/10.1214/17-AOAS1040

KELLY, L. J., RYDER, R. J. and CLARTÉ, G. (2023). Supplement to "Lagged couplings diagnose Markov chain Monte Carlo phylogenetic inference." https://doi.org/10.1214/22-AOAS1676SUPPA, https://doi.org/10.1214/22-AOAS1676SUPPB

KIM, J., ROSENBERG, N. A. and PALACIOS, J. A. (2020). Distance metrics for ranked evolutionary trees. *Proc. Natl. Acad. Sci. USA* **117** 28876–28886. https://doi.org/10.1073/pnas.1922851117

KOSKELA, J. (2022). Zig-Zag sampling for discrete structures and nonreversible phylogenetic MCMC. *J. Comput. Graph. Statist.* 1–11.

LAKNER, C., VAN DER MARK, P., HUELSENBECK, J. P., LARGET, B. and RONQUIST, F. (2008). Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. *Syst. Biol.* **57** 86–103. https://doi.org/10.1080/10635150801886156

LANFEAR, R., HUA, X. and WARREN, D. L. (2016). Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. *Genome Biol. Evol.* **8** 2319–2332.

LINDVALL, T. (1992). *Lectures on the Coupling Method. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York. MR1180522

MAGEE, A. F., KARCHER, M. D., IV, F. A. M. and MININ, V. N. (2021). How trustworthy is your tree? Bayesian phylogenetic effective sample size through the lens of Monte Carlo error. Available at arXiv:2109.07629.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

MEYER, X. (2021). Adaptive tree proposals for Bayesian phylogenetic inference. *Syst. Biol.* **70** 1015–1032. https://doi.org/10.1093/sysbio/syab004

MIDDLETON, L., DELIGIANNIDIS, G., DOUCET, A. and JACOB, P. E. (2020). Unbiased Markov chain Monte Carlo for intractable target distributions. *Electron. J. Stat.* **14** 2842–2891. MR4132645 https://doi.org/10.1214/20-EJS1727

MOSSEL, E. and VIGODA, E. (2006). Limitations of Markov chain Monte Carlo algorithms for Bayesian inference of phylogeny. *Ann. Appl. Probab.* **16** 2215–2234. MR2288719 https://doi.org/10.1214/105051600000000538

MÜLLER, N. F. and BOUCKAERT, R. R. (2020). Adaptive Metropolis-coupled MCMC for BEAST 2. *PeerJ* **8** e9473. https://doi.org/10.7717/peerj.9473

NASCIMENTO, F. F., DOS REIS, M. and YANG, Z. (2017). A biologist's guide to Bayesian phylogenetic analysis. *Nat. Ecol. Evol.* **1** 1446–1454.

NICHOLLS, G. K. and GRAY, R. D. (2008). Dated ancenstral trees from binary trait data and their application to the diversification of languages. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 545–566. MR2420414 https://doi.org/10.1111/j.1467-9868.2007.00648.x

NICHOLLS, G. K., RYDER, R. J. and WELCH, D. (2013). TraitLab: A MatLab Package for Fitting and Simulating Binary Trait-Like Data.

NYLANDER, J. A. A., WILGENBUSCH, J. C., WARREN, D. L. and SWOFFORD, D. L. (2008). AWTY (are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* **24** 581–583.

Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252. MR1611693 https://doi.org/10.1002/(SICI)1098-2418(199608/09)9:1/2<223::AID-RSA14>3.3.CO;2-R

Rambaut, A., Drummond, A. J., Xie, D., Baele, G. and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst*. *Biol*. **67** 901–904.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab*. *Surv*. **1** 20–71. MR2095565 https://doi.org/10.1214/154957804100000024

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. et al. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst*. *Biol*. **61** 539–542.

Ronquist, F., Huelsenbeck, J. P., Teslenko, M., Zhang, C. and Nylander, J. A. A. (2020). MrBayes version 3.2 Manual: Tutorials and Model Summaries. Accessed 1 June 2022.

Ryder, R. J. and Nicholls, G. K. (2011). Missing data in a stochastic Dollo model for binary trait data, and its application to the dating of Proto-Indo-European. *J*. *R*. *Stat*. *Soc*. *Ser*. *C*. *Appl*. *Stat*. **60** 71–92. MR2758570 https://doi.org/10.1111/j.1467-9876.2010.00743.x

Ryder, R. J., Clarté, G., Hairault, A., Lawless, C. and Robert, C. P. (2020). Comment on article by Jacob, O'Leary and Atchadé. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *B* **82** 590.

Shepherd, D. A. and Klaere, S. (2018). How well does your phylogenetic model fit your data? *Syst*. *Biol*. **68** 157–167.

Smith, M. R. (2021). Robust analysis of phylogenetic tree space. *Syst*. *Biol*.

Spade, D. A., Herbei, R. and Kubatko, L. S. (2014). A note on the relaxation time of two Markov chains on rooted phylogenetic tree spaces. *Statist*. *Probab*. *Lett*. **84** 247–252. MR3131282 https://doi.org/10.1016/j.spl.2013.09.017

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J. and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. **4** vey016. https://doi.org/10.1093/ve/vey016

Syed, S., Bouchard-Côté, A., Deligiannidis, G. and Doucet, A. (2022). Non-reversible parallel tempering: A scalable highly parallel MCMC scheme. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **84** 321–350. MR4412989

Ju, N., Biswas, N., Jacob, P. E., Mena, G., O'Leary, J. and Pompe, E. (2020). Comment on article by Tancredi, Steorts and Liseo. *Bayesian Anal*. **15** 670–672.

Team, R. C. (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

The MathWorks, Inc. (2021). Matlab Release 2021a, Natick, MA, United States.

Trippe, B., Nguyen, T. D. and Broderick, T. (2021). Optimal transport couplings of Gibbs samplers on partitions for unbiased estimation. In *Third Symposium on Advances in Approximate Bayesian Inference*.

Vats, D., Flegal, J. M. and Jones, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. MR3949306 https://doi.org/10.1093/biomet/asz002

Vats, D. and Knudson, C. (2021). Revisiting the Gelman-Rubin diagnostic. *Statist*. *Sci*. **36** 518–529. MR4323050 https://doi.org/10.1214/20-sts812

Wang, L., Bouchard-Côté, A. and Doucet, A. (2015). Bayesian phylogenetic inference using a combinatorial sequential Monte Carlo method. *J*. *Amer*. *Statist*. *Assoc*. **110** 1362–1374. MR3449032 https://doi.org/10.1080/01621459.2015.1054487

Wang, G., O'Leary, J. and Jacob, P. (2021). Maximal couplings of the Metropolis-Hastings algorithm. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* (A. Banerjee and K. Fukumizu, eds.). *Proceedings of Machine Learning Research* **130** 1225–1233. PMLR.

Wang, L., Wang, S. and Bouchard-Côté, A. (2019). An annealed sequential Monte Carlo method for Bayesian phylogenetics. *Syst*. *Biol*. **69** 155–183.

Warren, D. L., Geneva, A. J. and Lanfear, R. (2017). RWTY (R we there yet): An R package for examining convergence of Bayesian phylogenetic analyses. *Mol*. *Biol*. *Evol*. **34** 1016–1020. Version 1.0.2.

Whidden, C. and Matsen IV, F. A. (2015). Quantifying MCMC exploration of phylogenetic tree space. *Syst*. *Biol*. **64** 472–491.

Whidden, C., Claywell, B. C., Fisher, T., Magee, A. F., Fourment, M. and Matsen IV, F. A. (2020). Systematic exploration of the high likelihood set of phylogenetic tree topologies. *Syst*. *Biol*. **69** 280–293.

Wickham, H. (2016). *Ggplot*2: *Elegant Graphics for Data Analysis*. Springer, New York.

Willis, A. (2019). Confidence sets for phylogenetic trees. *J*. *Amer*. *Statist*. *Assoc*. **114** 235–244. MR3941251 https://doi.org/10.1080/01621459.2017.1395342

Willis, A. and Bell, R. (2018). Uncertainty in phylogenetic tree estimates. *J*. *Comput*. *Graph*. *Statist*. **27** 542–552. MR3863756 https://doi.org/10.1080/10618600.2017.1391697

Zhang, C. and Matsen IV, F. A. (2019). Variational Bayesian phylogenetic inference. In *ICLR*.

Zhang, Z., Nishimura, A., Bastide, P., Ji, X., Payne, R. P., Goulder, P., Lemey, P. and Suchard, M. A. (2021). Large-scale inference of correlation among mixed-type biological traits with phylogenetic multivariate probit models. *Ann. Appl. Stat.* **15** 230–251. MR4255276 https://doi.org/10.1214/20-aoas1394

Zhao, T., Wang, Z., Cumberworth, A., Gsponer, J., de Freitas, N. and Bouchard-Côté, A. (2016). Bayesian analysis of continuous time Markov chains with application to phylogenetic modelling. *Bayesian Anal.* **11** 1203–1237. MR3577377 https://doi.org/10.1214/15-BA982

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1444–1468
https://doi.org/10.1214/22-AOAS1677

CO-CLUSTERING OF SPATIALLY RESOLVED TRANSCRIPTOMIC DATA

BY ANDREA SOTTOSANTI[a] AND DAVIDE RISSO[b]

Department of Statistical Sciences, University of Padova, [a]*andrea.sottosanti@unipd.it,* [b]*davide.risso@unipd.it*

Spatial transcriptomics is a groundbreaking technology that allows the measurement of the activity of thousands of genes in a tissue sample and maps where the activity occurs. This technology has enabled the study of the spatial variation of the genes across the tissue. Comprehending gene functions and interactions in different areas of the tissue is of great scientific interest, as it might lead to a deeper understanding of several key biological mechanisms, such as cell-cell communication or tumor-microenvironment interaction. To do so, one can group cells of the same type and genes that exhibit similar expression patterns. However, adequate statistical tools that exploit the previously unavailable spatial information to more coherently group cells and genes are still lacking.

In this work we introduce SPARTACO, a new statistical model that clusters the spatial expression profiles of the genes according to a partition of the tissue. This is accomplished by performing a co-clustering, that is, inferring the latent block structure of the data and inducing two types of clustering: of the genes, using their expression across the tissue, and of the image areas, using the gene expression in the *spots* where the RNA is collected. Our proposed methodology is validated with a series of simulation experiments, and its usefulness in responding to specific biological questions is illustrated with an application to a human brain tissue sample processed with the 10X-Visium protocol.

REFERENCES

ALLEN, G. I. and TIBSHIRANI, R. (2010). Transposable regularized covariance models with an application to missing data imputation. *Ann. Appl. Stat.* **4** 764–790. MR2758420 https://doi.org/10.1214/09-AOAS314

ANDERLUCCI, L. and VIROLI, C. (2015). Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data. *Ann. Appl. Stat.* **9** 777–800. MR3371335 https://doi.org/10.1214/15-AOAS816

BARBIERATO, M., BORRI, M., FACCI, L., ZUSSO, M., SKAPER, S. D. and GIUSTI, P. (2017). Expression and differential responsiveness of central nervous system glial cell populations to the acute phase protein serum amyloid a. *Sci. Rep.* **7** 1–14.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 289–300. MR1325392

BIERNACKI, C., CELEUX, G. and GOVAERT, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 719–725.

BOUVEYRON, C., BOZZI, L., JACQUES, J. and JOLLOIS, F.-X. (2018). The functional latent block model for the co-clustering of electricity consumption curves. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **67** 897–915. MR3832256 https://doi.org/10.1111/rssc.12260

BOUVEYRON, C., CELEUX, G., MURPHY, T. B. and RAFTERY, A. E. (2019). *Model-Based Clustering and Classification for Data Science: With Applications in R. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, Cambridge. MR3967046 https://doi.org/10.1017/9781108644181

BYRD, R. H., LU, P., NOCEDAL, J. and ZHU, C. Y. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16** 1190–1208. MR1346301 https://doi.org/10.1137/0916069

CAPONERA, A., DENTI, F., RIGON, T., SOTTOSANTI, A. and GELFAND, A. (2017). Hierarchical spatio-temporal modeling of resting state fMRI data. In *START UP RESEARCH* 111–130. Springer, Berlin.

CASA, A., BOUVEYRON, C., EROSHEVA, E. and MENARDI, G. (2021). Co-clustering of time-dependent data via the shape invariant model. *J. Classification* **38** 626–649. MR4355226 https://doi.org/10.1007/s00357-021-09402-8

Key words and phrases. Co-clustering, EM algorithm, genomics, human dorsolateral prefrontal cortex, integrated completed log-likelihood, model-based clustering, spatial transcriptomics, 10X-Visium.

CELEUX, G. and GOVAERT, G. (1992). A classification EM algorithm for clustering and two stochastic versions. *Comput*. *Statist*. *Data Anal*. **14** 315–332. MR1192205 https://doi.org/10.1016/0167-9473(92)90042-E

CHEN, K. H., BOETTIGER, A. N., MOFFITT, J. R., WANG, S. and ZHUANG, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348** aaa6090. https://doi.org/10.1126/science.aaa6090

CRESSIE, N. A. C. (2015). *Statistics for Spatial Data*, Revised ed. *Wiley Classics Library*. Wiley, New York. MR3559472

DELATTRE, M., LAVIELLE, M. and POURSAT, M.-A. (2014). A note on BIC in mixed-effects models. *Electron*. *J*. *Stat*. **8** 456–475. MR3200764 https://doi.org/10.1214/14-EJS890

DE LA CRUZ-MESÍA, R. and MARSHALL, G. (2006). Non-linear random effects models with continuous time autoregressive errors: A Bayesian approach. *Stat*. *Med*. **25** 1471–1484. MR2227289 https://doi.org/10.1002/sim.2290

DRIES, R., ZHU, Q., DONG, R., ENG, C.-H. L., LI, H., LIU, K. et al. (2021). Giotto: A toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol*. **22** 1–31.

EDSGÄRD, D., JOHNSSON, P. and SANDBERG, R. (2018). Identification of spatial expression trends in single-cell gene expression data. *Nat*. *Methods* **15** 339–342. https://doi.org/10.1038/nmeth.4634

EFRON, B. (2009). Are a set of microarrays independent of each other? *Ann*. *Appl*. *Stat*. **3** 922–942. MR2750220 https://doi.org/10.1214/09-AOAS236

EISENBERG, E. and LEVANON, E. Y. (2003). Human housekeeping genes are compact. *Trends Genet*. **19** 362–365.

GOVAERT, G. and NADIF, M. (2008). Block clustering with Bernoulli mixture models: Comparison of different approaches. *Comput*. *Statist*. *Data Anal*. **52** 3233–3245. MR2424788 https://doi.org/10.1016/j.csda.2007.09.007

GOVAERT, G. and NADIF, M. (2010). Latent block model for contingency table. *Comm*. *Statist*. *Theory Methods* **39** 416–425. MR2745285 https://doi.org/10.1080/03610920903140197

GOVAERT, G. and NADIF, M. (2013). *Co-Clustering*: *Models*, *Algorithms and Applications*. Wiley, New York.

GUPTA, A. K. and NAGAR, D. K. (2000). *Matrix Variate Distributions*. *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics* **104**. CRC Press/CRC, Boca Raton, FL. MR1738933

HODGE, R. D., BAKKEN, T. E., MILLER, J. A., SMITH, K. A., BARKAN, E. R., GRAYBUCK, L. T. et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573** 61–68.

KERIBIN, C., BRAULT, V., CELEUX, G. and GOVAERT, G. (2015). Estimation and selection for the latent block model on categorical data. *Stat*. *Comput*. **25** 1201–1216. MR3401881 https://doi.org/10.1007/s11222-014-9472-2

LUBECK, E., COSKUN, A. F., ZHIYENTAYEV, T., AHMAD, M. and CAI, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. *Nat*. *Methods* **11** 360–361.

MARX, V. (2021). Method of the year 2020: Spatially resolved transcriptomics. *Nat*. *Methods* **18** 9–14.

MAYNARD, K. R., COLLADO-TORRES, L., WEBER, L. M., UYTINGCO, C., BARRY, B. K., WILLIAMS, S. R. et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat*. *Neurosci*.. https://doi.org/10.1038/s41593-020-00787-0

MORAN, G. E., ROČKOVÁ, V. and GEORGE, E. I. (2021). Spike-and-slab Lasso biclustering. *Ann*. *Appl*. *Stat*. **15** 148–173. MR4255269 https://doi.org/10.1214/20-aoas1385

MURUA, A. and QUINTANA, F. A. (2021). Biclustering via semiparametric Bayesian inference. *Bayesian Anal*. 1–27. https://doi.org/10.1214/21-BA1284

NOBILE, A. and FEARNSIDE, A. T. (2007). Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Stat*. *Comput*. **17** 147–162. MR2380643 https://doi.org/10.1007/s11222-006-9014-7

PARDO, B., SPANGLER, A., WEBER, L. M., HICKS, S. C., JAFFE, A. E., MARTINOWICH, K. et al. (2021). SpatialLIBD: An R/Bioconductor package to visualize spatially-resolved transcriptomics data. BioRxiv. https://doi.org/10.1101/2021.04.29.440149

RAO, N., CLARK, S. and HABERN, O. (2020). Bridging genomics and tissue pathology. *Genet*. *Eng*. *Biotechnol*. *News* **40** 50–51. https://doi.org/10.1089/gen.40.02.16

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

RIGHELLI, D., WEBER, L. M., CROWELL, H. L., PARDO, B., COLLADO-TORRES, L., GHAZANFAR, S. et al. (2022). SpatialExperiment: Infrastructure for spatially resolved transcriptomics data in R using bioconductor. *Bioinformatics* **38** 3128–3131. https://doi.org/10.1126/science.aaw1219

RODRIQUES, S. G., STICKELS, R. R., GOEVA, A., MARTIN, C. A., MURRAY, E., VANDERBURG, C. R. et al. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363** 1463–1467. https://doi.org/10.1126/science.aaw1219

SMYTH, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat*. *Appl*. *Genet*. *Mol*. *Biol*. **3** Art. 3, 29. MR2101454 https://doi.org/10.2202/1544-6115.1027

SOTTOSANTI, A. and RISSO, D. (2023). Supplement to “Co-clustering of spatially resolved transcriptomic data.” https://doi.org/10.1214/22-AOAS1677SUPPA, https://doi.org/10.1214/22-AOAS1677SUPPB

STARZYK, R. M., ROSENOW, C., FRYE, J., LEISMANN, M., RODZINSKI, E., PUTNEY, S. and TUOMANEN, E. I. (2000). Cerebral cell adhesion molecule: A novel leukocyte adhesion determinant on blood-brain barrier capillary endothelium. *J*. *Infect*. *Dis*. **181** 181–187.

SUN, S., ZHU, J. and ZHOU, X. (2020). Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat*. *Methods* **17** 193–200.

SVENSSON, V., TEICHMANN, S. A. and STEGLE, O. (2018). SpatialDE: Identification of spatially variable genes. *Nat*. *Methods* **15** 343–346.

TAN, K. M. and WITTEN, D. M. (2014). Sparse biclustering of transposable data. *J*. *Comput*. *Graph*. *Statist*. **23** 985–1008. MR3270707 https://doi.org/10.1080/10618600.2013.852554

TOWNES, F. W., HICKS, S. C., ARYEE, M. J. and IRIZARRY, R. A. (2019). Feature selection and dimension reduction for single-cell RNA-seq based on a multinomial model. *Genome Biol*. **20** 1–16.

VAN LAARHOVEN, P. J. M. and AARTS, E. H. L. (1987). *Simulated Annealing*: *Theory and Applications*. *Mathematics and Its Applications* **37**. Reidel, Dordrecht. MR0904050 https://doi.org/10.1007/978-94-015-7744-1

WITTEN, D. M. and TIBSHIRANI, R. (2009). Covariance-regularized regression and classification for high dimensional problems. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **71** 615–636. MR2749910 https://doi.org/10.1111/j.1467-9868.2009.00699.x

WITTEN, D. M. and TIBSHIRANI, R. (2010). A framework for feature selection in clustering. *J*. *Amer*. *Statist*. *Assoc*. **105** 713–726. MR2724855 https://doi.org/10.1198/jasa.2010.tm09415

WYSE, J. and FRIEL, N. (2012). Block clustering with collapsed latent block models. *Stat*. *Comput*. **22** 415–428. MR2865026 https://doi.org/10.1007/s11222-011-9233-4

ZHAO, E., STONE, M. R., REN, X., GUENTHOER, J., SMYTHE, K. S., PULLIAM, T. et al. (2021). Spatial transcriptomics at subspot resolution with BayesSpace. *Nat*. *Biotechnol*. 1–10.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1469–1490
https://doi.org/10.1214/22-AOAS1678

BALANCING WEIGHTS FOR REGION-LEVEL ANALYSIS: THE EFFECT OF MEDICAID EXPANSION ON THE UNINSURANCE RATE AMONG STATES THAT DID NOT EXPAND MEDICAID

BY MAX RUBINSTEIN[1,a], AMELIA HAVILAND[1,b] AND DAVID CHOI[2,c]

[1]*Heinz College and Department of Statistics & Data Science, Carnegie Mellon University,* [a]*mrubinst@andrew.cmu.edu,* [b]*amelia@andrew.cmu.edu*

[2]*Heinz College, Carnegie Mellon University,* [c]*davidch@andrew.cmu.edu*

We predict the average effect of Medicaid expansion on the nonelderly adult uninsurance rate among states that did not expand Medicaid in 2014, as if they had expanded their Medicaid eligibility requirements. Using American Community Survey data aggregated to the region level, we estimate this effect by reweighting the expansion regions to approximately match the covariate distribution of the nonexpansion regions. Existing methods to estimate balancing weights often assume that the covariates are measured without error and do not account for dependencies in the outcome model. Our covariates have random noise that is uncorrelated with the outcome errors, and our assumed outcome model contains state-level random effects. To correct for measurement error induced bias, we propose generating weights on a linear approximation to the true covariates, extending an idea from the measurement error literature known as "regression calibration." This requires auxiliary data to estimate the measurement error variability. We also propose an objective function to reduce the variance of our estimator when the outcome model errors are homoskedastic and equicorrelated within states. We then estimate that Medicaid expansion would have caused a -2.33 (-3.54, -1.11) percentage point change in the adult uninsurance rate among states that did not expand Medicaid.

REFERENCES

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Amer. Statist. Assoc.* **105** 493–505. MR2759929 https://doi.org/10.1198/jasa.2009.ap08746

ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2015). Comparative politics and the synthetic control method. *Amer. J. Polit. Sci.* **59** 495–510.

BEN-MICHAEL, E., FELLER, A. and HARTMAN, E. (2021). Multilevel calibration weighting for survey data. ArXiv preprint. Available at arXiv:2102.09052.

BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2021). The augmented synthetic control method. *J. Amer. Statist. Assoc.* **116** 1789–1803. MR4353714 https://doi.org/10.1080/01621459.2021.1929245

BONVINI, M. and KENNEDY, E. H. (2021). Sensitivity analysis via the proportion of unmeasured confounding. *J. Amer. Statist. Assoc.* 1–11.

BOTOSARU, I. and FERMAN, B. (2019). On the role of covariates in the synthetic control method. *Econom. J.* **22** 117–130. MR4021116 https://doi.org/10.1093/ectj/utz001

CAMERON, A. C. and MILLER, D. L. (2015). A practitioner's guide to cluster-robust inference. *J. Hum. Resour.* **50** 317–372.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press/CRC, Boca Raton, FL. MR2243417 https://doi.org/10.1201/9781420010138

COURTEMANCHE, C., MARTON, J., UKERT, B., YELOWITZ, A. and ZAPATA, D. (2017). Early impacts of the affordable care act on health insurance coverage in Medicaid expansion and non-expansion states. *J. Policy Anal. Manage.* **36** 178–210. https://doi.org/10.1002/pam.21961

Key words and phrases. Medicaid expansion, balancing weights, measurement error, hierarchical data.

DAW, J. R. and HATFIELD, L. A. (2018). Matching and regression to the mean in difference-in-differences analysis. *Health Serv. Res.* **53** 4138–4156.

DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.* **87** 376–382. MR1173804

DEVILLE, J.-C., SÄRNDAL, C.-E. and SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *J. Amer. Statist. Assoc.* **88** 1013–1020.

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596. MR0615434

FREAN, M., GRUBER, J. and SOMMERS, B. D. (2017). Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the affordable care act. *J. Health Econ.* **53** 72–86. https://doi.org/10.1016/j.jhealeco.2017.02.004

GLESER, L. J. (1992). The importance of assessing measurement reliability in multivariate regression. *J. Amer. Statist. Assoc.* **87** 696–707. MR1185191

GREIFER, N. (2021). optweight: Targeted Stable Balancing Weights Using Optimization. R package version 0.2.5.9000.

HABERMAN, S. J. (1984). Adjustment by minimum discriminant information. *Ann. Statist.* **12** 971–988. MR0751286 https://doi.org/10.1214/aos/1176346715

HAINMUELLER, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20** 25–46.

HUQUE, M. H., BONDELL, H. D. and RYAN, L. (2014). On the impact of covariate measurement error on spatial regression modelling. *Environmetrics* **25** 560–570. MR3295549 https://doi.org/10.1002/env.2305

ILLENBERGER, N. A., SMALL, D. S. and SHAW, P. A. (2020). Impact of regression to the mean on the synthetic control method: Bias and sensitivity analysis. *Epidemiology* **31** 815–822.

IMAI, K. and RATKOVIC, M. (2014). Covariate balancing propensity score. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 243–263. MR3153941 https://doi.org/10.1111/rssb.12027

IMBENS, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.

KAESTNER, R., GARRETT, B., CHEN, J., GANGOPADHYAYA, A. and FLEMING, C. (2017). Effects of ACA Medicaid expansions on health insurance coverage and labor supply. *J. Policy Anal. Manage.* **36** 608–642.

KEELE, L., BEN-MICHAEL, E., FELLER, A., KELZ, R. and MIRATRIX, L. (2020). Hospital quality risk standardization via approximate balancing weights. ArXiv preprint. Available at arXiv:2007.09056.

KLINE, P. (2011). Oaxaca-blinder as a reweighting estimator. *Am. Econ. Rev.* **101** 532–37.

LADHANIA, R., HAVILAND, A. M., VENKAT, A., TELANG, R. and PINES, J. M. (2021). The effect of Medicaid expansion on the nature of new enrollees' emergency department use. *Med. Care Res. Rev.* **78** 24–35. https://doi.org/10.1177/1077558719848270

MILLER, S., JOHNSON, N. and WHERRY, L. R. (2021). Medicaid and mortality: New evidence from linked survey and administrative data. *Q. J. Econ.* **136** 1783–1829.

RUBIN, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* **100** 322–331. MR2166071 https://doi.org/10.1198/016214504000001880

RUBINSTEIN, M., HAVILAND, A. and CHOI, D. (2023). Supplement to "Balancing weights for region-level analysis: The effect of Medicaid expansion on the uninsurance rate among states that did not expand Medicaid." https://doi.org/10.1214/22-AOAS1678SUPPA, https://doi.org/10.1214/22-AOAS1678SUPPB

SÄRNDAL, C.-E. and LUNDSTRÖM, S. (2005). *Estimation in Surveys with Nonresponse*. *Wiley Series in Survey Methodology*. Wiley, Chichester. MR2163886 https://doi.org/10.1002/0470011351

SOMMERS, B., KRONICK, R., FINEGOLD, K., PO, R., SCHWARTZ, K. and GLIED, S. (2012). Understanding participation rates in Medicaid: Implications for the affordable care act. *Publ. Health* **93** 67–74.

TEAM, R. C. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

WANG, Y. and ZUBIZARRETA, J. R. (2020). Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika* **107** 93–105. MR4064142 https://doi.org/10.1093/biomet/asz050

WICKHAM, H., AVERICK, M., BRYAN, J., CHANG, W., MCGOWAN, L. D., FRANÇOIS, R., GROLEMUND, G., HAYES, A., HENRY, L. et al. (2019). Welcome to the tidyverse. *J. Open Sour. Softw.* **4** 1686.

ZHANG, Z., KIM, H. J., LONJON, G., ZHU, Y. et al. (2019). Balance diagnostics after propensity score matching. *Ann. Transl. Med.* **7**.

ZUBIZARRETA, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *J. Amer. Statist. Assoc.* **110** 910–922. MR3420672 https://doi.org/10.1080/01621459.2015.1023805

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1491–1520
https://doi.org/10.1214/22-AOAS1679

SPATIOTEMPORAL LOCAL INTERPOLATION OF GLOBAL OCEAN HEAT TRANSPORT USING ARGO FLOATS: A DEBIASED LATENT GAUSSIAN PROCESS APPROACH

BY BEOMJO PARK[1,a], MIKAEL KUUSELA[1,b], DONATA GIGLIO[2,c] AND ALISON GRAY[3,d]

[1]*Department of Statistics and Data Science, Carnegie Mellon University,* [a]*beomjo@stat.cmu.edu,* [b]*mkuusela@andrew.cmu.edu*

[2]*Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder,* [c]*donata.giglio@colorado.edu*

[3]*School of Oceanography, University of Washington,* [d]*argray@uw.edu*

The world ocean plays a key role in redistributing heat in the climate system and hence in regulating Earth's climate. Yet statistical analysis of ocean heat transport suffers from partially incomplete large-scale data intertwined with complex spatiotemporal dynamics as well as from potential model misspecification. We present a comprehensive spatiotemporal statistical framework tailored to interpolating the global ocean heat transport using in situ Argo profiling float measurements. We formalize the statistical challenges using latent local Gaussian process regression accompanied by a two-stage fitting procedure. We introduce an approximate expectation-maximization algorithm to jointly estimate both the mean field and the covariance parameters, and refine the potentially underspecified mean field model with a debiasing procedure. This approach provides data-driven global ocean heat transport fields that vary in both space and time and can provide insights into crucial dynamical phenomena, such as El Niño & La Niña, as well as the global climatological mean heat transport field which by itself is of scientific interest. The proposed framework and the Argo-based estimates are thoroughly validated with state-of-the-art multimission satellite products and shown to yield realistic subsurface ocean heat transport estimates.

REFERENCES

ARGO (2020). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). https://doi.org/10.17882/42182

ARIAS-CASTRO, E., SALMON, J. and WILLETT, R. (2012). Oracle inequalities and minimax rates for non-local means and related adaptive kernel-based methods. *SIAM J. Imaging Sci.* **5** 944–992. MR3022184 https://doi.org/10.1137/110859403

BANERJEE, S., GELFAND, A. E. and SIRMANS, C. F. (2003). Directional rates of change under spatial process models. *J. Amer. Statist. Assoc.* **98** 946–954. MR2041483 https://doi.org/10.1198/C16214503000000909

BARKER, P. M. and MCDOUGALL, T. J. (2020). Two interpolation methods using multiply-rotated piecewise cubic Hermite interpolating polynomials. *J. Atmos. Ocean. Technol.* **37** 605–619.

BAYARRI, M. J., BERGER, J. O., CAFEO, J., GARCIA-DONATO, G., LIU, F., PALOMO, J., PARTHASARATHY, R. J., PAULO, R., SACKS, J. et al. (2007). Computer model validation with functional output. *Ann. Statist.* **35** 1874–1906. MR2363956 https://doi.org/10.1214/009053607000000163

BEHRENS, E., FERNANDEZ, D. and SUTTON, P. (2019). Meridional oceanic heat transport influences marine heatwaves in the Tasman Sea on interannual to decadal timescales. *Front. Mar. Sci.* **6** 228. https://doi.org/10.3389/fmars.2019.00228

BOLIN, D. and WALLIN, J. (2020). Multivariate type G Matérn stochastic partial differential equation random fields. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 215–239. MR4060983

BRYDEN, H. L. and IMAWAKI, S. (2001). Chapter 6.1—Ocean heat transport. In *International Geophysics* (G. Siedler, J. Church and J. Gould, eds.). *Ocean Circulation and Climate* **77** 455–474. Academic Press, San Diego. https://doi.org/10.1016/S0074-6142(01)80134-0

Key words and phrases. Latent Gaussian process regression, local kriging, approximate EM, model misspecification, physical oceanography.

BRYNJARSDÓTTIR, J. and O'HAGAN, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Probl.* **30** 114007, 24 pp. MR3274591 https://doi.org/10.1088/0266-5611/30/11/114007

CHARNIGO, R., HALL, B. and SRINIVASAN, C. (2011). A generalized C_p criterion for derivative estimation. *Technometrics* **53** 238–253. MR2857702 https://doi.org/10.1198/TECH.2011.09147

COLIN DE VERDIÈRE, A., MEUNIER, T. and OLLITRAULT, M. (2019). Meridional overturning and heat transport from Argo floats displacements and the planetary geostrophic method (PGM): Application to the subpolar North Atlantic. *J. Geophys. Res., Oceans* **124** 6270–6285. https://doi.org/10.1029/2018JC014565

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1239641 https://doi.org/10.1002/9781119115151

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

CSISZÁR, I. and TUSNÁDY, G. (1984). Information geometry and alternating minimization procedures. *Statist. Decisions* **1** 205–237. MR0785210

DAI, W., TONG, T. and GENTON, M. G. (2016). Optimal estimation of derivatives in nonparametric regression. *J. Mach. Learn. Res.* **17** Paper No. 164, 25 pp. MR3555052

DE BRABANTER, K., DE BRABANTER, J., DE MOOR, B. and GIJBELS, I. (2013). Derivative estimation with local polynomial fitting. *J. Mach. Learn. Res.* **14** 281–301. MR3033332

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

DONG, S., BARINGER, M., GONI, G. and GARZOLI, S. (2011). Importance of the assimilation of Argo float measurements on the meridional overturning circulation in the South Atlantic. *Geophys. Res. Lett.* **38** L18603. https://doi.org/10.1029/2011GL048982

FAN, J. and GIJBELS, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B* **57** 371–394. MR1323345 https://doi.org/10.1111/j.2517-6161.1995.tb02034.x

FAN, J., GASSER, T., GIJBELS, I., BROCKMANN, M. and ENGEL, J. (1997). Local polynomial regression: Optimal kernels and asymptotic minimax efficiency. *Ann. Inst. Statist. Math.* **49** 79–99. MR1450693 https://doi.org/10.1023/A:1003162622169

FORGET, G. and FERREIRA, D. (2019). Global ocean heat transport dominated by heat export from the tropical Pacific. *Nat. Geosci.* **12** 351–354. https://doi.org/10.1038/s41561-019-0333-7

FORGET, G. and PONTE, R. M. (2015). The partition of regional sea level variability. *Prog. Oceanogr.* **137** 173–195. https://doi.org/10.1016/j.pocean.2015.06.002

FRITSCH, F. N. and CARLSON, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.* **17** 238–246. MR0567271 https://doi.org/10.1137/0717021

FUENTES, M. (2002). Interpolation of nonstationary air pollution processes: A spatial spectral approach. *Stat. Model.* **2** 281–298. MR1951586 https://doi.org/10.1191/1471082x02st034oa

GANACHAUD, A. and WUNSCH, C. (2000). Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature* **408** 453–457. https://doi.org/10.1038/35044048

GIGLIO, D., ROEMMICH, D. and CORNUELLE, B. (2013). Understanding the annual cycle in global steric height. *Geophys. Res. Lett.* **40** 4349–4354. https://doi.org/10.1002/grl.50774

GLANTZ, M. H. and RAMIREZ, I. J. (2020). Reviewing the Oceanic Niño Index (ONI) to enhance societal readiness for El Niño's impacts. *Int. J. Disaster Risk Sci.* **11** 394–403. https://doi.org/10.1007/s13753-020-00275-w

GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. MR2325275 https://doi.org/10.1111/j.1467-9868.2007.00587.x

GOOD, S. A., MARTIN, M. J. and RAYNER, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res., Oceans* **118** 6704–6716. https://doi.org/10.1002/2013JC009067

GRAY, A. R. and RISER, S. C. (2014). A global analysis of Sverdrup balance using absolute geostrophic velocities from Argo. *J. Phys. Oceanogr.* **44** 1213–1229. https://doi.org/10.1175/JPO-D-12-0206.1

GRAY, A. R. and RISER, S. C. (2015). A method for multiscale optimal analysis with application to Argo data. *J. Geophys. Res., Oceans* **120** 4340–4356. https://doi.org/10.1002/2014JC010208

GUERRIER, S., KAREMERA, M., ORSO, S. and VICTORIA-FESER, M.-P. (2020). Asymptotically optimal bias reduction for parametric models. Preprint. Available at arXiv:2002.08757.

HAAS, T. C. (1990). Kriging and automated variogram modeling within a moving window. *Atmos. Environ., A Gen. Top.* **24** 1759–1769. https://doi.org/10.1016/0960-1686(90)90508-K

HAAS, T. C. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *J. Amer. Statist. Assoc.* **90** 1189–1199. https://doi.org/10.2307/2291511

HALLIN, M., LU, Z. and TRAN, L. T. (2004). Local linear spatial regression. *Ann. Statist.* **32** 2469–2500. MR2153992 https://doi.org/10.1214/009053604000000850

HIGDON, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environ. Ecol. Stat.* **5** 173–190. https://doi.org/10.1023/A:1009666805688

JAYNE, S., ROEMMICH, D., ZILBERMAN, N., RISER, S., JOHNSON, K., JOHNSON, G. and PIOTROWICZ, S. (2017). The Argo program: Present and future. *Oceanography* **30** 18–28. https://doi.org/10.5670/oceanog.2017.213

KAWAI, Y., HOSODA, S., UEHARA, K. and SUGA, T. (2021). Heat and salinity transport between the permanent pycnocline and the mixed layer due to the obduction process evaluated from a gridded Argo dataset. *J. Oceanogr.* **77** 75–92. https://doi.org/10.1007/s10872-020-00559-1

KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398 https://doi.org/10.1111/1467-9868.00294

KUK, A. Y. C. (1995). Asymptotically unbiased estimation in generalized linear models with random effects. *J. Roy. Statist. Soc. Ser. B* **57** 395–407. MR1323346

KUUSELA, M. and STEIN, M. L. (2018a). Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **474** 20180400. https://doi.org/10.1098/rspa.2018.0400

KUUSELA, M. and STEIN, M. L. (2018b). Supplementary material from "Locally stationary spatio-temporal interpolation of Argo profiling float data". https://doi.org/10.6084/m9.figshare.c.4310771.v3

LAGERLOEF, G. S. E., MITCHUM, G. T., LUKAS, R. and NIILER, P. P. (1999). Tropical Pacific near-surface currents estimated from altimeter, wind, and drifter data. *J. Geophys. Res.* **104** 23313–23326. https://doi.org/10.1029/1999JC900197

LEBEDEV, K., YOSHINARI, H., MAXIMENKO, N. A. and HACKER, P. W. (2007). YoMaHa'07: Velocity data assessed from trajectories of Argo floats at parking level and at the sea surface. IPRC Technical Note No. **4** 16.

LI, X. and YUAN, D. (2020). An assessment of the CMIP5 models in simulating the Argo geostrophic meridional transport in the North Pacific Ocean. *J. Oceanol. Limnol.* **38** 1445–1463. https://doi.org/10.1007/s00343-020-0002-0

LIU, Y. and DE BRABANTER, K. (2018). Derivative estimation in random design. In *Advances in Neural Information Processing Systems* **31** (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) 3445–3454.

MACDONALD, A. M. and BARINGER, M. (2013). Ocean heat transport. In *International Geophysics* **103** 759–785. https://doi.org/10.1016/B978-0-12-391851-2.00029-5

MCDOUGALL, T. and BARKER, P. M. (2011). Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox. *SCOR/IAPSO WG* **127** 1–28.

MCMONIGAL, K., GUNN, K. L., BEAL, L. M., ELIPOT, S. and WILLIS, J. K. (2022). Reduction in meridional heat export contributes to recent Indian Ocean warming. *J. Phys. Oceanogr.* **52** 329–345. https://doi.org/10.1175/JPO-D-21-0085.1

MCPHADEN, M. J., SANTOSO, A. and CAI, W. (2020). *El Niño Southern Oscillation in a Changing Climate*. American Geophysical Union (AGU). https://doi.org/10.1002/9781119548164

MERCHANT, C. J., EMBURY, O., BULGIN, C. E., BLOCK, T., CORLETT, G. K., FIEDLER, E., GOOD, S. A., MITTAZ, J., RAYNER, N. A. et al. (2019). Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Sci. Data* **6** 223. https://doi.org/10.1038/s41597-019-0236-x

MEYSSIGNAC, B., BOYER, T., ZHAO, Z., HAKUBA, M. Z., LANDERER, F. W., STAMMER, D., KÖHL, A., KATO, S., L'ECUYER, T. et al. (2019). Measuring global ocean heat content to estimate the Earth energy imbalance. *Front. Mar. Sci.* **6** 432. https://doi.org/10.3389/fmars.2019.00432

MOGEN, S. C., LOVENDUSKI, N. S., DALLMANN, A. R., GREGOR, L., SUTTON, A. J., BOGRAD, S. J., QUIROS, N. C., DI LORENZO, E., HAZEN, E. L. et al. (2022). Ocean biogeochemical signatures of the North Pacific blob. *Geophys. Res. Lett.* **49** e2021GL096938. https://doi.org/10.1029/2021GL096938

NEAL, R. M. and HINTON, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models* (M. I. Jordan, ed.). *NATO ASI Series* 355–368. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-011-5014-9_12

NOCEDAL, J. (1980). Updating quasi-Newton matrices with limited storage. *Math. Comp.* **35** 773–782. MR0572855 https://doi.org/10.2307/2006193

NYCHKA, D., WIKLE, C. and ROYLE, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Stat. Model.* **2** 315–331. MR1951588 https://doi.org/10.1191/1471082x02st037oa

OLLITRAULT, M. and RANNOU, J.-P. (2013). ANDRO: An Argo-based deep displacement dataset. *J. Atmos. Ocean. Technol.* **30** 759–788. https://doi.org/10.1175/JTECH-D-12-00073.1

PACIOREK, C. J. and SCHERVISH, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17** 483–506. MR2240939 https://doi.org/10.1002/env.785

PARK, B., KUUSELA, M., GIGLIO, D. and GRAY, A. (2023a). Supplement to "Spatiotemporal local interpolation of global ocean heat transport using Argo floats: A debiased latent Gaussian process approach." https://doi.org/10.1214/22-AOAS1679SUPPA

PARK, B., KUUSELA, M., GIGLIO, D. and GRAY, A. (2023b). Code for "Spatio-temporal local interpolation of global ocean heat transport using Argo floats: A debiased latent Gaussian process approach." https://doi.org/10.1214/22-AOAS1679SUPPB

RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

REN, H.-L., WANG, R., ZHAI, P., DING, Y. and LU, B. (2017). Upper-ocean dynamical features and prediction of the super El Niño in 2015/16: A comparison with the cases in 1982/83 and 1997/98. *J. Meteorol. Res.* **31** 278–294. https://doi.org/10.1007/s13351-017-6194-3

RIDGWAY, K. R., DUNN, J. R. and WILKIN, J. L. (2002). Ocean interpolation by four-dimensional weighted least squares—Application to the waters around Australasia. *J. Atmos. Ocean. Technol.* **19** 1357–1375. https://doi.org/10.1175/1520-0426(2002)019<1357:OIBFDW>2.0.CO;2

RIO, M. H. and SANTOLERI, R. (2018). Improved global surface currents from the merging of altimetry and sea surface temperature data. *Remote Sens. Environ.* **216** 770–785. https://doi.org/10.1016/j.rse.2018.06.003

RIO, M.-H., MULET, S., ETIENNE, H., PICOT, N. and DIBARBOURE, G. (2018). New CNES-CLS18 mean dynamic topography of the global ocean from altimetry, gravity and in-situ data. In *OSTST* 22.

RISER, S. C., FREELAND, H. J., ROEMMICH, D., WIJFFELS, S., TROISI, A., BELBÉOCH, M., GILBERT, D., XU, J., POULIQUEN, S. et al. (2016). Fifteen years of ocean observations with the global Argo array. *Nat. Clim. Change* **6** 145–153. https://doi.org/10.1038/nclimate2872

ROEMMICH, D. and GILSON, J. (2009). The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo program. *Prog. Oceanogr.* **82** 81–100. https://doi.org/10.1016/j.pocean.2009.03.004

ROEMMICH, D., BOEBEL, O., FREELAND, H. J., KING, B. A., LE TRAON, P.-Y., MOLINARI, R., OWENS, W. B., RISER, S., SEND, U. et al. (1998). *On the Design and Implementation of Argo*: *A Global Array of Profiling Floats*. *ICPO Publication Series* **21**. GODAE International Project Office, Melbourne, Vic.

ROULSTON, M. S. and SMITH, L. A. (2002). Evaluating probabilistic forecasts using information theory. *Mon. Weather Rev.* **130** 1653–1660. https://doi.org/10.1175/1520-0493(2002)130<1653:EPFUIT>2.0.CO;2

RUDNICK, D. L. (2016). Ocean research enabled by underwater gliders. *Annu. Rev. Mar. Sci.* **8** 519–541. https://doi.org/10.1146/annurev-marine-122414-033913

RUDNICK, D. L., DAVIS, R. E. and SHERMAN, J. T. (2016). Spray underwater glider operations. *J. Atmos. Ocean. Technol.* **33** 1113–1122. https://doi.org/10.1175/JTECH-D-15-0252.1

RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370. MR1311979 https://doi.org/10.1214/aos/1176325632

SCOTT, R. B., ARBIC, B. K., CHASSIGNET, E. P., COWARD, A. C., MALTRUD, M., MERRYFIELD, W. J., SRINIVASAN, A. and VARGHESE, A. (2010). Total kinetic energy in four global eddying ocean circulation models and over 5000 current meter records. *Ocean Model.* **32** 157–169. https://doi.org/10.1016/j.ocemod.2010.01.005

SHERMAN, J., DAVIS, R. E., OWENS, W. B. and VALDES, J. (2001). The autonomous underwater glider "Spray". *IEEE J. Oceanic Eng.* **26** 437–446. https://doi.org/10.1109/48.972076

SIBSON, R. (1981). A brief description of natural neighbor interpolation. In *Interpolating Multivariate Data* Chapter 2 21–36. Wiley, New York.

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. *Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

STEIN, M. L. (2013). Statistical properties of covariance tapers. *J. Comput. Graph. Statist.* **22** 866–885. MR3173747 https://doi.org/10.1080/10618600.2012.719844

STEIN, M. L. (2020). Some statistical issues in climate science. *Statist. Sci.* **35** 31–41. MR4071356 https://doi.org/10.1214/19-STS730

STOCKER, T. F. (2013). Chapter 1—The ocean as a component of the climate system. In *International Geophysics* (G. Siedler, S. M. Griffies, J. Gould and J. A. Church, eds.). *Ocean Circulation and Climate* **103** 3–30. Academic Press, San Diego. https://doi.org/10.1016/B978-0-12-391851-2.00001-5

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360. MR0594650 https://doi.org/10.1214/aos/1176345206

SUN, B., LIU, C. and WANG, F. (2019). Global meridional Eddy heat transport inferred from Argo and altimetry observations. *Sci. Rep.* **9** 1345. https://doi.org/10.1038/s41598-018-38069-2

TABURET, G., SANCHEZ-ROMAN, A., BALLAROTTA, M., PUJOL, M.-I., LEGEAIS, J.-F., FOURNIER, F., FAUGERE, Y. and DIBARBOURE, G. (2019). DUACS DT2018: 25 years of reprocessed sea level altimetry products. *Ocean Sci.* **15** 1207–1224. https://doi.org/10.5194/os-15-1207-2019

TALLEY, L. D., PICKARD, G. L., EMERY, W. J. and SWIFT, J. H. (2011). *Descriptive Physical Oceanography*: *An Introduction*, 6th ed. Academic Press, San Diego.

TRENBERTH, K. E. and CARON, J. M. (2001). Estimates of meridional atmosphere and ocean heat transports. *J. Climate* **14** 3433–3443. https://doi.org/10.1175/1520-0442(2001)014<3433:EOMAAO>2.0.CO;2

TRENBERTH, K. E. and SOLOMON, A. (1994). The global heat balance: Heat transports in the atmosphere and ocean. *Clim. Dyn.* **10** 107–134. https://doi.org/10.1007/BF00210625

TRENBERTH, K. E., FASULLO, J. T., SCHUCKMANN, K. V. and CHENG, L. (2016). Insights into Earth's energy imbalance from multiple sources. *J. Climate* **29** 7495–7505. https://doi.org/10.1175/JCLI-D-16-0339.1

VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B* **50** 297–312. MR0964183 https://doi.org/10.1111/j.2517-6161.1988.tb01729.x

WANG, W. W. and LIN, L. (2015). Derivative estimation based on difference sequence via locally weighted least squares regression. *J. Mach. Learn. Res.* **16** 2617–2641. MR3450519

WILLIS, J. K. and FU, L.-L. (2008). Combining altimeter and subsurface float data to estimate the time-averaged circulation in the upper ocean. *J. Geophys. Res., Oceans* **113** C12017. https://doi.org/10.1029/2007JC004690

WONG, A. P. S., WIJFFELS, S. E., RISER, S. C., POULIQUEN, S., HOSODA, S., ROEMMICH, D., GILSON, J., JOHNSON, G. C., MARTINI, K. et al. (2020). Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. *Front. Mar. Sci.* **7** 700. https://doi.org/10.3389/fmars.2020.00700

YARGER, D., STOEV, S. and HSING, T. (2022). A functional-data approach to the Argo data. *Ann. Appl. Stat.* **16** 216–246. MR4400508 https://doi.org/10.1214/21-aoas1477

ZHENG, Y. and GIESE, B. S. (2009). Ocean heat transport in simple ocean data assimilation: Structure and mechanisms. *J. Geophys. Res., Oceans* **114** C11009. https://doi.org/10.1029/2008JC005190

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1521–1542
https://doi.org/10.1214/22-AOAS1680

LATENT VARIABLE MODELS FOR MULTIVARIATE DYADIC DATA WITH ZERO INFLATION: ANALYSIS OF INTERGENERATIONAL EXCHANGES OF FAMILY SUPPORT

BY JOUNI KUHA[1,a], SILIANG ZHANG[2,c] AND FIONA STEELE[1,b]

[1]*Department of Statistics, London School of Economics and Political Science,* [a]*j.kuha@lse.ac.uk,* [b]*f.a.steele@lse.ac.uk*
[2]*Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University,* [c]*slzhang@fem.ecnu.edu.cn*

Understanding the help and support that is exchanged between family members of different generations is of increasing importance, with research questions in sociology and social policy focusing on both predictors of the levels of help given and received, and on reciprocity between them. We propose general latent variable models for analysing such data, when helping tendencies in each direction are measured by multiple binary indicators of specific types of help. The model combines two continuous latent variables, which represent the helping tendencies, with two binary latent class variables which allow for high proportions of responses where no help of any kind is given or received. This defines a multivariate version of a zero-inflation model. The main part of the models is estimated using MCMC methods, with a bespoke data augmentation algorithm. We apply the models to analyse exchanges of help between adult individuals and their noncoresident parents, using survey data from the UK Household Longitudinal Study.

REFERENCES

BAKK, Z. and KUHA, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika* **83** 871–892. MR3875886 https://doi.org/10.1007/s11336-017-9592-7

BARTHOLOMEW, D., KNOTT, M. and MOUSTAKI, I. (2011). *Latent Variable Models and Factor Analysis*: *A Unified Approach*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2849614 https://doi.org/10.1002/9781119970583

BURT, R. S. (1973). Confirmatory factor-analytic structures and the theory construction process. *Sociol*. *Methods Res*. **2** 131–190.

BURT, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociol*. *Methods Res*. **5** 3–52.

CHAN, T. W. (2008). The structure of intergenerational exchange in the UK. University of Oxford. Working Paper Series 2008-05.

CHAN, T. W. and ERMISCH, J. (2015). Residential proximity of parents and their adult offspring in the United Kingdom, 2009–2010. *Popul*. *Stud*. **69** 355–372.

CHENG, Y.-P., BIRDITT, K. S., ZARIT, S. H. and FINGERMAN, K. L. (2015). Young adults' provision of support to middle-aged parents. *J*. *Gerontol*., *Ser*. *B* **70** 407–416.

CRAGG, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* **39** 829–844.

DAGUM, L. and MENON, R. (1998). OpenMP: An industry-standard API for shared-memory programming. *IEEE Comput*. *Sci*. *Eng*. **5** 46–55.

DAVEY, A. and EGGEBEEN, D. J. (1998). Patterns of intergenerational exchange and mental health. *J*. *Gerontol*. *Psychol*. *Sci*. **53B** P86–P95.

DE AYALA, R. J. (2009). *The Theory and Practice of Item Response Theory*. The Guilford Press, New York.

FINGERMAN, K. L., KIM, K., DAVIS, E. M., FURSTENBERG, F. F. JR, BIRDITT, K. S. and ZARIT, S. H. (2015). "I'll give you the world": Socioeconomic differences in parental support of adult children. *J*. *Marriage Fam*. **77** 844–865.

Key words and phrases. Item response theory models, latent class analysis, mixture models, nonequivalence of measurement, two-step estimation.

FINKELMAN, M. D., GREEN, J. G., GRUBER, M. J. and ZASLAVSKY, A. M. (2011). A zero- and K-inflated mixture model for health questionnaire data. *Stat. Med.* **30** 1028–1043. MR2767820 https://doi.org/10.1002/sim.4217

GILKS, W. R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *J. Appl. Stat.* **41** 337–348.

GIN, B., SIM, N., SKRONDAL, A. and RABE-HESKETH, S. (2020). A dyadic IRT model. *Psychometrika* **85** 815–836. MR4169498 https://doi.org/10.1007/s11336-020-09718-1

GRUNDY, E. (2005). Reciprocity in relationships: Socio-economic and health influences on intergenerational exchanges between third age parents and their adult children in Great Britain. *Br. J. Sociol.* **56** 233–255.

HENRETTA, J., VAN VOORHIS, M. and SOLDO, B. (2018). Cohort differences in parental financial help to adult children. *Demography* **55** 1567–1582.

HOGAN, D. P., EGGEBEEN, D. J. and CLOGG, C. C. (1993). The structure of intergenerational exchanges in American families. *Am. J. Sociol.* **98** 1428–1458.

HUANG, H.-Y. (2016). Mixture random-effect IRT models for controlling extreme response style on rating scales. *Front. Psychol.* **7** 1706. https://doi.org/10.3389/fpsyg.2016.01706

KALMIJN, M. (2014). Adult intergenerational relationships. In *The Wiley Blackwell Companion to the Sociology of Families* (J. Treas, J. Scott and M. Rochards, eds.) 385–403. Wiley, Chichester.

KANKARAŠ, M., VERMUNT, J. K. and MOORS, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociol. Methods Res.* **40** 279–310. MR2767836 https://doi.org/10.1177/0049124111405301

KENNY, D. A. and LAVOIE, L. (1984). The social relations model. *Adv. Exp. Soc. Psychol.* **18** 141–182.

KNIES, G., ed. (2018) *Understanding Society: The UK Household Longitudinal Study Waves* 1-8. *User Guide*. Institute for Social and Economic Research, Univ. Essex, Colchester.

KUHA, J., ZHANG, S. and STEELE, F. (2023). Supplement to "Latent variable models for multivariate dyadic data with zero inflation: Analysis of intergenerational exchanges of family support." https://doi.org/10.1214/22-AOAS1680SUPPA, https://doi.org/10.1214/22-AOAS1680SUPPB

KÜNEMUND, H., MOTEL-LINGEBIEL, A. and KOHLI, M. (2005). Do intergenerational transfers from elderly parents increase social inquality among their middle-aged children? Evidence from the German ageing survey. *J. Gerontol., Psychol. Sci.* **60B** S30–S36.

LAMBERT, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34** 1–14.

LESTHAEGHE, R. (2014). The second demographic transition: A concise overview of its development. *Proc. Natl. Acad. Sci. USA* **111** 18112–18115.

MAGNUS, B. E. and LIU, Y. (2017). A zero-inflated Box-Cox normal unipolar item response model for measuring constructs of psychopathology. *J. Educ. Behav. Stat.* **42** 531–558.

MAGNUS, B. E. and THISSEN, D. (2017). Item response modeling of multivariate count data with zero inflation, maximum inflation, and heaping. *J. Educ. Behav. Stat.* **42** 531–558.

MASON, A. and LEE, R. (2018). Intergenerational transfers and the older population. In *Future Directions for the Demography of Aging: Proceedings of a Workshop* (M. Hayward and M. Majmundar, eds.) The National Academies Press, Washington DC.

MILLSAP, R. E. (2011). *Statistical Approaches to Measurement Invariance*. Routledge, New York.

MIN, Y. and AGRESTI, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Stat. Model.* **5** 1–19. MR2133525 https://doi.org/10.1191/1471082X05st084oa

MULLAHY, J. (1986). Specification and testing of some modified count data models. *J. Econometrics* **33** 341–365. MR0867980 https://doi.org/10.1016/0304-4076(86)90002-3

MUTHÉN, B. and ASPAROUHOV, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addict. Behav.* **31** 1050–1066.

MUTHÉN, L. K. and MUTHÉN, B. O. (2010). *Mplus User's Guide*, 6th ed. Muthén & Muthén, Los Angeles, CA.

PICKARD, L. (2015). A growing care gap? The supply of unpaid care for older people by their adult children in England to 2032. *Ageing Soc.* **35** 96–123.

SILVERSTEIN, M., CONROY, S. J., WANG, H., GIARRUSSO, R. and BENGTSON, V. L. (2002). Reciprocity in parent-child relations over the adult life course. *J. Gerontol., Psychol. Sci.* **57B** S3–S13.

SKRONDAL, A. and RABE-HESKETH, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Interdisciplinary Statistics*. CRC Press/CRC, Boca Raton, FL. MR2059021 https://doi.org/10.1201/9780203489437

SNIJDERS, T. A. B. and KENNY, D. A. (1999). The social relations model for family data: A multilevel approach. *Pers. Relatsh.* **6** 471–486.

STEELE, F. and GRUNDY, E. (2021). Random effects dynamic panel models for unequally spaced multivariate categorical repeated measures: An application to child-parent exchanges of support. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **70** 3–23. MR4204935 https://doi.org/10.1111/rssc.12446

R CORE TEAM (2020). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

TOBIN, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26** 24–36. MR0090462 https://doi.org/10.2307/1907382

VAN DER LINDEN, W., ed. (2016) *Handbook of Item Response Theory*. CRC Press/CRC, New York.

WALL, M. M., GUO, J. and AMEMIYA, Y. (2012). Mixture factor analysis for approximating a nonnormally distributed continuous latent factor with continuous and dichotomous observed variables. *Multivar. Behav. Res*. **47** 276–313.

WALL, M. M., PARK, J. Y. and MOUSTAKI, I. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Appl*. *Psychol*. *Meas*. **39** 583–597.

XUE, Q.-L. and BANDEEN-ROCHE, K. (2002). Combining complete multivariate outcomes with incomplete covariate information: A latent class approach. *Biometrics* **58** 110–120. MR1891049 https://doi.org/10.1111/j.0006-341X.2002.00110.x

University of Essex, Institute for Social and Economic Research, NatCen Social Research, and Kantar Public (2018). Understanding Society: Waves 1–8, 2009–2017 and Harmonised BHPS: Waves 1–18, 1991–2009. [data collection], 11th edition ed. UK Data Service. SN 6614. https://doi.org/10.5255/UKDA-SN-6614-12

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1543–1573
https://doi.org/10.1214/22-AOAS1681

A BAYESIAN PANEL VECTOR AUTOREGRESSION TO ANALYZE THE IMPACT OF CLIMATE SHOCKS ON HIGH-INCOME ECONOMIES

BY FLORIAN HUBER[1,a], TAMÁS KRISZTIN[2,1,b] AND MICHAEL PFARRHOFER[3,c]

[1]*Department of Economics, University of Salzburg,* [a]*florian.huber@plus.ac.at*
[2]*Integrated Biosphere Futures Group, International Institute for Applied Systems Analysis (IIASA),* [b]*krisztin@iiasa.ac.at*
[3]*Department of Economics, University of Vienna,* [c]*michael.pfarrhofer@univie.ac.at*

In this paper we assess the impact of climate shocks on futures markets for agricultural commodities and a set of macroeconomic quantities for multiple high-income economies. To capture relations among countries, markets, and climate shocks, this paper proposes parsimonious methods to estimate high-dimensional panel vector autoregressions. We assume that coefficients associated with domestic lagged endogenous variables arise from a Gaussian mixture model while further parsimony is achieved using suitable global-local shrinkage priors on several regions of the parameter space. Our results point toward pronounced global reactions of key macroeconomic quantities to climate shocks. Moreover, the empirical findings highlight substantial linkages between regionally located shocks and global commodity markets.

REFERENCES

AGUILAR, O. and WEST, M. (2000). Bayesian dynamic factor models and portfolio allocation. *J. Bus. Econom. Statist.* **18** 338–357.

AKRAM, Q. F. (2009). Commodity prices, interest rates and the dollar. *Energy Econ.* **31** 838–851. https://doi.org/10.1016/j.eneco.2009.05.016

ALESSANDRI, P. and MUMTAZ, H. (2021). The macroeconomic cost of climate volatility. Preprint. Available at arXiv:2108.01617.

ALLENBY, G. M., ARORA, N. and GINTER, J. L. (1998). On the heterogeneity of demand. *J. Mark. Res.* **35** 384–389.

AMARE, M., JENSEN, N. D., SHIFERAW, B. and CISSÉ, J. D. (2018). Rainfall shocks and agricultural productivity: Implication for rural household consumption. *Agric. Syst.* **166** 79–89. https://doi.org/10.1016/j.agsy.2018.07.014

BAFFES, J. and HANIOTIS, T. (2010). Placing the recent commodity boom into perspective. In *Food Prices and Rural Poverty* 40–70. World Bank, Washington DC.

BAI, J. and NG, S. (2007). Determining the number of primitive shocks in factor models. *J. Bus. Econom. Statist.* **25** 52–60. MR2338870 https://doi.org/10.1198/073500106000000413

BAKER, J., HAVLIK, P., BEACH, R., LECLÈRE, D., SCHMID, E., VALIN, H., COLE, J., CREASON, J., OHREL, S. et al. (2018). Evaluating the effects of climate change on US agricultural systems: Sensitivity to regional impact and trade expansion scenarios. *Environ. Res. Lett.* **13** 1–48.

BALKOVIČ, J., VAN DER VELDE, M., SKALSKÝ, R., XIONG, W., FOLBERTH, C., KHABAROV, N., SMIRNOV, A., MUELLER, N. D. and OBERSTEINER, M. (2014). Global wheat production potentials and management flexibility under the representative concentration pathways. *Glob. Planet. Change* **122** 107–121. https://doi.org/10.1016/j.gloplacha.2014.08.010

BAUMEISTER, C. and PEERSMAN, G. (2013). The role of time-varying price elasticities in accounting for volatility changes in the crude oil market. *J. Appl. Econometrics* **28** 1087–1109. MR3137716 https://doi.org/10.1002/jae.2283

BEGUERÍA, S., VICENTE-SERRANO, S. M., REIG, F. and LATORRE, B. (2014). Standardized Precipitation Evapotranspiration Index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *Int. J. Climatol.* **34** 3001–3023.

BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *J. Amer. Statist. Assoc.* **110** 1479–1490. MR3449048 https://doi.org/10.1080/01621459.2014.960967

Key words and phrases. Climate change impacts, commodity markets, food security, hierarchical modeling, factor stochastic volatility models, second keyword.

BURKE, M. and TANUTAMA, V. (2019). Climatic constraints on aggregate economic output. *NBER Working Paper* **25779**.

CANOVA, F. and CICCARELLI, M. (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *J. Econometrics* **120** 327–359. MR2058892 https://doi.org/10.1016/S0304-4076(03)00216-1

CANOVA, F. and CICCARELLI, M. (2009). Estimating multicountry VAR models. *Internat. Econom. Rev.* **50** 929–959. MR2542805 https://doi.org/10.1111/j.1468-2354.2009.00554.x

CANOVA, F. and CICCARELLI, M. (2013). Panel vector autoregressive models: A survey. In *VAR Models in Macroeconomics—New Developments and Applications*: *Essays in Honor of Christopher A. Sims*. *Adv. Econom.* **32** 205–246. Emerald Group Publ, Ltd., Bingley. MR3496837 https://doi.org/10.1108/S0731-905320130000031006

CARRIERO, A., CLARK, T. E. and MARCELLINO, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *J. Econometrics* **212** 137–154. MR3994011 https://doi.org/10.1016/j.jeconom.2019.04.024

CASHIN, P., MOHADDES, K. and RAISSI, M. (2017). Fair weather or foul? The macroeconomic effects of El Niño. *J. Int. Econ.* **106** 37–54.

CHAN, J., EISENSTAT, E. and YU, X. (2022). Large Bayesian VARs with factor stochastic volatility: Identification, order invariance and structural analysis. arXiv:2207.03988.

CRESPO CUARESMA, J., FELDKIRCHER, M. and HUBER, F. (2016). Forecasting with global vector autoregressive models: A Bayesian approach. *J. Appl. Econometrics* **31** 1371–1391. MR3580905 https://doi.org/10.1002/jae.2504

DE NICOLA, F., DE PACE, P. and HERNANDEZ, M. A. (2016). Co-movement of major energy, agricultural, and food commodity price returns: A time-series assessment. *Energy Econ.* **57** 28–41. https://doi.org/10.1016/j.eneco.2016.04.012

DEES, S., DI MAURO, F., PESARAN, M. H. and SMITH, L. V. (2007). Exploring the international linkages of the Euro area: A global VAR analysis. *J. Appl. Econometrics* **22** 1–38. MR2359200 https://doi.org/10.1002/jae.932

DOAN, T. R., LITTERMAN, B. R. and SIMS, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Rev.* **3** 1–100.

DRUDI, F., MOENCH, E., HOLTHAUSEN, C., WEBER, P.-F., FERRUCCI, G., SETZER, R., NINO, V. D., BARBIERO, F., FACCIA, D. et al. (2021). Climate change and monetary policy in the euro area.

ELLER, M., HUBER, F. and SCHUBERTH, H. (2020). How important are global factors for understanding the dynamics of international capital flows? *J. Int. Money Financ.* **109** 102221.

ENDERS, W. and HOLT, M. T. (2014). *The Evolving Relationships Between Agricultural and Energy Commodity Prices*: *A Shifting-Mean Vector Autoregressive Analysis*. Univ. Chicago Press, Chicago.

FAO (2017). *The Future of Food and Agriculture*: *Trends and Challenges*. FAO, Rome.

FELDKIRCHER, M. and HUBER, F. (2016). The international transmission of US shocks?evidence from Bayesian global vector autoregressions. *Eur. Econ. Rev.* **81** 167–188.

FELDKIRCHER, M., HUBER, F., KOOP, G. and PFARRHOFER, M. (2022). Approximate Bayesian inference and forecasting in huge-dimensional multicountry VARs. *Internat. Econom. Rev.* forthcoming. https://doi.org/10.1111/iere.12577

FRÜHWIRTH-SCHNATTER, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *J. Amer. Statist. Assoc.* **96** 194–209. MR1952732 https://doi.org/10.1198/016214501750333063

FRÜHWIRTH-SCHNATTER, S. (2006). *Finite Mixture and Markov Switching Models*. *Springer Series in Statistics*. Springer, New York. MR2265601

FRÜHWIRTH-SCHNATTER, S. (2011). Dealing with label switching under model uncertainty. In *Mixtures*: *Estimation and Applications*. *Wiley Ser. Probab. Stat.* 213–239. Wiley, Chichester. MR2883354 https://doi.org/10.1002/9781119995678.ch10

FRÜHWIRTH-SCHNATTER, S. and KAUFMANN, S. (2008). Model-based clustering of multiple time series. *J. Bus. Econom. Statist.* **26** 78–89. MR2422063 https://doi.org/10.1198/073500107000000106

FRÜHWIRTH-SCHNATTER, S., TÜCHLER, R. and OTTER, T. (2004). Bayesian analysis of the heterogeneity model. *J. Bus. Econom. Statist.* **22** 2–15. MR2028204 https://doi.org/10.1198/073500103288619331

GARCIA, P., IRWIN, S. H. and SMITH, A. (2015). Futures market failure? *Am. J. Agric. Econ.* **97** 40–64. https://doi.org/10.1093/ajae/aau067

GAUPP, F., PFLUG, G., HOCHRAINER-STIGLER, S., HALL, J. and DADSON, S. (2017). Dependency of crop production between global breadbaskets: A copula approach for the assessment of global and regional risk pools. *Risk Anal.* **37** 2212–2228. https://doi.org/10.1111/risa.12761

GEORGIADIS, G. (2015). Examining asymmetries in the transmission of monetary policy in the euro area: Evidence from a mixed cross-section global VAR model. *Eur. Econ. Rev.* **75** 195–215.

GILBERT, C. L. (2010). How to understand high food prices. *J. Agric. Econ.* **61** 398–425. https://doi.org/10.1111/j.1477-9552.2010.00248.x

GRIFFIN, J. E. and BROWN, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Anal*. **5** 171–188. MR2596440 https://doi.org/10.1214/10-BA507

GUERRERO, S., HERNÁNDEZ-DEL VALLE, G. and JUÁREZ-TORRES, M. (2017). Using a functional approach to test trending volatility in the price of Mexican and international agricultural products. *J. Agric. Econ*. **48** 3–13. https://doi.org/10.1111/agec.12290

HARARI, M. and FERRARA, E. L. (2018). Conflict, climate and cells: A disaggregated analysis. *Rev. Econ. Stat*. **100** 594–608.

HARRI, A., NALLEY, L. and HUDSON, D. (2009). The relationship between oil, exchange rates, and commodity prices. *Journal of Agricultural and Applied Economics* **41** 501–510. https://doi.org/10.1017/S1074070800002959

HAUZENBERGER, N., HUBER, F., KOOP, G. and ONORANTE, L. (2021). Fast and flexible Bayesian inference in time-varying parameter regression models. *J. Bus. Econom. Statist*. 1–15.

HAVLÍK, P., SCHNEIDER, U. A., SCHMID, E., BÖTTCHER, H., FRITZ, S., SKALSKÝ, R., AOKI, K., CARA, S. D., KINDERMANN, G. et al. (2011). Global land-use implications of first and second generation biofuel targets. *Energy Policy* **39** 5690–5702. https://doi.org/10.1016/j.enpol.2010.03.030

HEADEY, D. (2011). Rethinking the global food crisis: The role of trade shocks. *Food Policy* **36** 136–146. https://doi.org/10.1016/j.foodpol.2010.10.003

HIRSCH, C., KRISZTIN, T. and SEE, L. (2020). Water resources as determinants for foreign direct investments in land-a gravity analysis of foreign land acquisitions. *Ecol. Econ*. **170** 106516.

HUANG, H., VON LAMPE, M. and VAN TONGEREN, F. (2011). Climate change and trade in agriculture. *Food Policy* **36** S9–S13. https://doi.org/10.1016/j.foodpol.2010.10.008

HUBER, F. (2016). Density forecasting using Bayesian global vector autoregressions with stochastic volatility. *Int. J. Forecast*. **32** 818–837.

HUBER, F. and FELDKIRCHER, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. *J. Bus. Econom. Statist*. **37** 27–39. MR3910223 https://doi.org/10.1080/07350015.2016.1256217

HUBER, F., KRISZTIN, T. and PFARRHOFER, M. (2023). Supplement to “A Bayesian panel vector autoregression to analyze the impact of climate shocks on high-income economies.” https://doi.org/10.1214/22-AOAS1681SUPPA, https://doi.org/10.1214/22-AOAS1681SUPPB

HUBER, F., KRISZTIN, T. and PIRIBAUER, P. (2017). Forecasting global equity indices using large Bayesian vars. *Bull. Econ. Res*. **69** 288–308. MR3680268 https://doi.org/10.1111/boer.12094

IFPRI (2008). *High Food Prices: The What, Who, and How of Proposed Policy Actions*. IFPRI, Washington DC.

ISHWARAN, H., JAMES, L. F. and SUN, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *J. Amer. Statist. Assoc*. **96** 1316–1332. MR1946579 https://doi.org/10.1198/016214501753382255

JANSSENS, C., HAVLÍK, P., KRISZTIN, T., BAKER, J., FRANK, S., HASEGAWA, T., LECLÈRE, D., OHREL, S., RAGNAUTH, S. et al. (2020). Global hunger and climate change adaptation through international trade. *Nat. Clim. Change* **10** 829–835.

JAROCIŃSKI, M. (2010). Responses to monetary policy shocks in the East and the West of Europe: A comparison. *J. Appl. Econometrics* **25** 833–868. MR2756988 https://doi.org/10.1002/jae.1082

JEBABLI, I., AROURI, M. and TEULON, F. (2014). On the effects of world stock market and oil price shocks on food prices: An empirical investigation based on TVP-VAR models with stochastic volatility. *Energy Econ*. **45** 66–98. https://doi.org/10.1016/j.eneco.2014.06.008

KASTNER, G. (2019a). Sparse Bayesian time-varying covariance estimation in many dimensions. *J. Econometrics* **210** 98–115. MR3944765 https://doi.org/10.1016/j.jeconom.2018.11.007

KASTNER, G. (2019b). `factorstochvol`: Bayesian Estimation of (Sparse) Latent Factor Stochastic Volatility Models. R-package version 0.9.

KASTNER, G. and HUBER, F. (2020). Sparse Bayesian vector autoregressions in huge dimensions. *J. Forecast*. **39** 1142–1165. MR4161021 https://doi.org/10.1002/for.2680

KIM, H. S., MATTHES, C. and PHAN, T. (2021). Extreme Weather and the Macroeconomy. Available at SSRN 3918533.

KOOP, G. and KOROBILIS, D. (2016). Model uncertainty in panel vector autoregressive models. *Eur. Econ. Rev*. **81** 115–131.

KOOP, G. and KOROBILIS, D. (2018). Forecasting with high-dimensional panel VARs. *Essex Finance Centre Working Papers* **31**.

KOROBILIS, D. (2016). Prior selection for panel vector autoregressions. *Comput. Statist. Data Anal*. **101** 110–120. MR3504839 https://doi.org/10.1016/j.csda.2016.02.011

LENG, G. and HALL, J. (2019). Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future. *Science of the Total Environment* **654** 811–821. https://doi.org/10.1016/j.scitotenv.2018.10.434

LENK, P. J. and DESARBO, W. S. (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika* **65** 93–119.

LUCOTTE, Y. (2016). Co-movements between crude oil and food prices: A post-commodity boom perspective. *Econom. Lett.* **147** 142–147. MR3552172 https://doi.org/10.1016/j.econlet.2016.08.032

MALSINER-WALLI, G., FRÜHWIRTH-SCHNATTER, S. and GRÜN, B. (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Stat. Comput.* **26** 303–324. MR3439375 https://doi.org/10.1007/s11222-014-9500-2

MINOT, N. (2014). Food price volatility in sub-Saharan Africa: Has it really increased? *Food Policy* **45** 45–56. https://doi.org/10.1016/j.foodpol.2013.12.008

MIRANDA-AGRIPPINO, S. and REY, H. (2020). U.S. monetary policy and the global financial cycle. *Rev. Econ. Stud.* **87** 2754–2776. MR4170663 https://doi.org/10.1093/restud/rdaa019

NAZLIOGLU, S. (2011). World oil and agricultural commodity prices: Evidence from nonlinear causality. *Energy Policy* **39** 2935–2943. https://doi.org/10.1016/j.enpol.2011.03.001

NAZLIOGLU, S. and SOYTAS, U. (2011). World oil prices and agricultural commodity prices: Evidence from an emerging market. *Energy Economics* **33** 488–496. https://doi.org/10.1016/j.eneco.2010.11.012

NAZLIOGLU, S. and SOYTAS, U. (2012). Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. *Energy Economics* **34** 1098–1104. https://doi.org/10.1016/j.eneco.2011.09.008

PARK, T. and CASELLA, G. (2008). The Bayesian lasso. *J. Amer. Statist. Assoc.* **103** 681–686. MR2524001 https://doi.org/10.1198/016214508000000337

PESARAN, M. H., SCHUERMANN, T. and WEINER, S. M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *J. Bus. Econom. Statist.* **22** 129–181. MR2041460 https://doi.org/10.1198/073500104000000019

PITT, M. K. and SHEPHARD, N. (1999). Time-varying covariances: A factor stochastic volatility approach. In *Bayesian Statistics*, 6 (*Alcoceber*, 1998) 547–570. Oxford Univ. Press, New York. MR1724873

PUMA, M. J., BOSE, S., CHON, S. Y. and COOK, B. I. (2015). Assessing the evolving fragility of the global food system. *Environmental Research Letters* **10**. https://doi.org/10.1088/1748-9326/10/2/024007

ROBERTS, M. J. and SCHLENKER, W. (2013). Identifying supply and demand elasticities of agricultural commodities: Implications for the US ethanol mandate. *American Economic Review* **103** 2265–2295.

ROUSSEAU, J. and MENGERSEN, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 689–710. MR2867454 https://doi.org/10.1111/j.1467-9868.2011.00781.x

SAGHAIAN, S. H. (2010). The impact of the oil sector on commodity prices: Correlation or causation? *Journal of Agricultural and Applied Economics* **42** 477–485. https://doi.org/10.1017/S1074070800003667

SANDSTRÖM, V., VALIN, H., KRISZTIN, T., HAVLÍK, P., HERRERO, M. and KASTNER, T. (2018). The role of trade in the greenhouse gas footprints of EU diets. *Global Food Security* **19** 48–55. https://doi.org/10.1016/j.gfs.2018.08.007

SERRA, T., ZILBERMAN, D., GIL, J. M. and GOODWIN, B. K. (2011). Nonlinearities in the U.S. corn-ethanol-oil-gasoline price system. *Agricultural Economics* **42** 35–45. https://doi.org/10.1111/j.1574-0862.2010.00464.x

URBAN, D. W., ROBERTS, M. J., SCHLENKER, W. and LOBELL, D. B. (2015). The effects of extremely wet planting conditions on maize and soybean yields. *Climatic Change* **130** 247–260.

VAN HUELLEN, S. (2018). How financial investment distorts food prices: Evidence from U.S. grain markets. *Agricultural Economics* **49** 171–181. https://doi.org/10.1111/agec.12406

VAN DER VELDE, M., TUBIELLO, F. N., VRIELING, A. and BOURAOUI, F. (2012). Impacts of extreme weather on wheat and maize in France: Evaluating regional crop simulations against observed data. *Climatic Change* **113** 751–765. https://doi.org/10.1007/s10584-011-0368-2

YAU, C. and HOLMES, C. (2011). Hierarchical Bayesian nonparametric mixture models for clustering with variable relevance determination. *Bayesian Anal.* **6** 329–351. MR2806247 https://doi.org/10.1214/11-BA612

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1574–1591
https://doi.org/10.1214/22-AOAS1682

ESTIMATION AND INFERENCE FOR EXPOSURE EFFECTS WITH LATENCY IN THE COX PROPORTIONAL HAZARDS MODEL IN THE PRESENCE OF EXPOSURE MEASUREMENT ERROR

BY SARAH B. PESKOE[1,a], NING ZHANG[2,b], DONNA SPIEGELMAN[3,d] AND MOLIN WANG[2,c]

[1]*Department of Biostatistics and Bioinformatics, Duke University School of Medicine,* [a]*sarah.peskoe@duke.edu*
[2]*Department of Epidemiology, Harvard T.H. Chan School of Public Health,* [b]*ning.zhang@channing.harvard.edu,* [c]*stmow@channing.harvard.edu*
[3]*Center for Methods in Implementation and Prevention Science, Yale School of Public Health,* [d]*donna.spiegelman@yale.edu*

Researchers are often interested in estimating the effects of time-varying exposures on health outcomes. The latency period, defined as the critical period of susceptibility, can be an important component of exposure effect assessment. Although it is widely known that many environmental, nutritional, and other exposure measurements are prone to error and are also likely to act only during a critical time window of susceptibility, no one has yet considered the impact of this on the estimation of latency parameters in survival models. In this paper we derived methods for point and interval estimation for the latency parameter and the regression coefficients in rare disease situations. Under a linear measurement model, although the estimated hazard ratios are biased, as has been previously demonstrated, we show that the latency parameter is approximately unbiased. Simulations and an illustrative example investigating the prospective association between $PM_{2.5}$ and lung cancer incidence in the Nurses' Health Study are included to evaluate the performance of our method.

REFERENCES

BARUL, C., FAYOSSÉ, A., CARTON, M., PILORGET, C., WORONOFF, A.-S., STÜCKER, I. and LUCE, D. (2017). Occupational exposure to chlorinated solvents and risk of head and neck cancer in men: A population-based case-control study in France. *Environ. Health* **16** 1–12.

BELENCHIA, A. M., JOHNSON, S. A., ELLERSIECK, M. R., ROSENFELD, C. S. and PETERSON, C. A. (2017). *J. Endocrinol.* **234** 301–313. https://doi.org/10.1530/JOE-17-0015

BROWN, K. W., SARNAT, J. A., SUH, H. H., COULL, B. A., SPENGLER, J. D. and KOUTRAKIS, P. (2008). Ambient site, home outdoor and home indoor particulate concentrations as proxies of personal exposures. *J. Environ. Monit.* **10** 1041–1051.

BROWN, K. W., SARNAT, J. A., SUH, H. H., COULL, B. A. and KOUTRAKIS, P. (2009). Factors influencing relationships between personal and ambient concentrations of gaseous and particulate pollutants. *Sci. Total Environ.* **407** 3754–3765.

CARROLL, R. J. and WAND, M. P. (1991). Semiparametric estimation in logistic measurement error models. *J. Roy. Statist. Soc. Ser. B* **53** 573–585. MR1125715

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press/CRC, Boca Raton, FL. MR2243417 https://doi.org/10.1201/9781420010138

CARROLL, A. J., CARNETHON, M. R., LIU, K., JACOBS JR., D. R., COLANGELO, L. A., STEWART, J. C., CARR, J. J., WIDOME, R., AUER, R. et al. (2017). Interaction between smoking and depressive symptoms with subclinical heart disease in the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Health Psychology* **36** 101.

CHEN, L. H., KNUTSEN, S. F., SHAVLIK, D., BEESON, W. L., PETERSEN, F., GHAMSARY, M. and ABBEY, D. (2005). The association between fatal coronary heart disease and ambient particulate air pollution: Are females at greater risk? *Environ. Health Perspect.* **113** 1723–1729.

Key words and phrases. Cox proportional hazard model, cumulative average exposure, exposure metrics, latency parameter, measurement error, time-varying exposure.

CHIUVE, S. E., REXRODE, K. M., SPIEGELMAN, D., LOGROSCINO, G., MANSON, J. E. and RIMM, E. B. (2008). Primary prevention of stroke by healthy lifestyle. *Circulation* **118** 947–954. https://doi.org/10.1161/CIRCULATIONAHA.108.781062

FINKELSTEIN, M. M. (1991). Use of "time windows" to investigate lung cancer latency intervals at an Ontario steel plant. *Am. J. Ind. Med*. **19** 229–235.

GILLIES, M., RICHARDSON, D. B., CARDIS, E., DANIELS, R. D., O'HAGAN, J. A., HAYLOCK, R., LAURIER, D., LEURAUD, K., MOISSONNIER, M. et al. (2017). Mortality from circulatory diseases and other non-cancer outcomes among nuclear workers in France, the United Kingdom and the United States (INWORKS). *Radiat*. *Res*. **188** 276–290.

GIOVANNUCCI, E., LIU, Y., STAMPFER, M. J. and WILLETT, W. C. (2006). A prospective study of calcium intake and incident and fatal prostate cancer. *Cancer Epidemiol*. *Biomark*. *Prev.* **15** 203–210.

HART, J. E., LIAO, X., HONG, B., PUETT, R. C., YANOSKY, J. D., SUH, H., KIOUMOURTZOGLOU, M.-A., SPIEGELMAN, D. and LADEN, F. (2015). The association of long-term exposure to PM2. 5 on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. *Environ*. *Health* **14** 1–9.

HAUPTMANN, M., WELLMANN, J., LUBIN, J. H., ROSENBERG, P. S. and KREIENBROCK, L. (2000). Analysis of exposure-time–response relationships using a spline weight function. *Biometrics* **56** 1105–1108. MR1815589 https://doi.org/10.1111/j.0006-341X.2000.01105.x

HUNTER, D. J., SPIEGELMAN, D., ADAMI, H.-O., BEESON, L., VAN DEN BRANDT, P. A., FOLSOM, A. R., FRASER, G. E., GOLDBOHM, R. A., GRAHAM, S. et al. (1996). Cohort studies of fat intake and the risk of breast cancer—a pooled analysis. *N. Engl. J. Med*. **334** 356–361.

KIOUMOURTZOGLOU, M.-A., SPIEGELMAN, D., SZPIRO, A. A., SHEPPARD, L., KAUFMAN, J. D., YANOSKY, J. D., WILLIAMS, R., LADEN, F., HONG, B. et al. (2014). Exposure measurement error in PM 2.5 health effects studies: A pooled analysis of eight personal exposure validation studies. *Environ*. *Health* **13** 1–11.

KOUTRAKIS, P., SUH, H. H., SARNAT, J. A., BROWN, K. W., COULL, B. A. and SCHWARTZ, J. (2005). Characterization of particulate and gas exposures of sensitive subpopulations living in Baltimore and Boston. *Res*. *Rep*. *Health Eff*. *Inst*. **131** 1–65.

KUHA, J. (1994). Corrections for exposure measurement error in logistic regression models with an application to nutritional data. *Stat*. *Med*. **13** 1135–1148.

LANGHOLZ, B., THOMAS, D., XIANG, A. and STRAM, D. (1999). Latency analysis in epidemiologic studies of occupational exposures: Application to the Colorado Plateau uranium miners cohort. *Am*. *J*. *Ind*. *Med*. **35** 246–256.

LIAO, X., ZUCKER, D. M., LI, Y. and SPIEGELMAN, D. (2011). Survival analysis with error-prone time-varying covariates: A risk set calibration approach. *Biometrics* **67** 50–58. MR2898816 https://doi.org/10.1111/j.1541-0420.2010.01423.x

LIAO, X., ZHOU, X., WANG, M., HART, J. E., LADEN, F. and SPIEGELMAN, D. (2018). Survival analysis with functions of mismeasured covariate histories: The case of chronic air pollution exposure in relation to mortality in the nurses' health study. *J*. *R*. *Stat*. *Soc*. *Ser*. *C*. *Appl*. *Stat*. **67** 307–327. MR3758768 https://doi.org/10.1111/rssc.12229

LIN, K. J., MITCHELL, A. A., YAU, W.-P., LOUIK, C. and HERNÁNDEZ-DÍAZ, S. (2012). Maternal exposure to amoxicillin and the risk of oral clefts. *Epidemiology* **23** 699.

LIU, L. S., BOX, M., KALMAN, D., KAUFMAN, J., KOENIG, J., LARSON, T., LUMLEY, T., SHEPPARD, L. and WALLACE, L. (2003). Exposure assessment of particulate matter for susceptible populations in Seattle. *Environ*. *Health Perspect*. **111** 909–918.

LOGAN, R. and SPIEGELMAN, D. (2012). The SAS% BLINPLUS Macro.

MENG, Q. Y., TURPIN, B. J., KORN, L., WEISEL, C. P., MORANDI, M., COLOME, S., ZHANG, J., STOCK, T., SPEKTOR, D. et al. (2005). Influence of ambient (outdoor) sources on residential indoor and personal PM 2.5 concentrations: Analyses of RIOPA data. *J*. *Expo*. *Sci*. *Environ*. *Epidemiol*. **15** 17–28.

MICHELS, K., SONG, M., WILLETT, W. C. and ROSNER, B. (2020). Latency estimation for chronic disease risk: A damped exponential weighting model. *Eur*. *J*. *Epidemiol*. **35** 807–819.

PACIOREK, C. J., YANOSKY, J. D., PUETT, R. C., LADEN, F. and SUH, H. H. (2009). Practical large-scale spatio-temporal modeling of particulate matter concentrations. *Ann*. *Appl*. *Stat*. **3** 370–397. MR2668712 https://doi.org/10.1214/08-AOAS204

PESKOE, S. B., SPIEGELMAN, D. and WANG, M. (2019). There is no impact of exposure measurement error on latency estimation in linear models. *Stat*. *Med*. **38** 1245–1261. MR3920609 https://doi.org/10.1002/sim.8038

PRENTICE, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69** 331–342. MR0671971 https://doi.org/10.1093/biomet/69.2.331

PUETT, R. C., SCHWARTZ, J., HART, J. E., YANOSKY, J. D., SPEIZER, F. E., SUH, H., PACIOREK, C. J., NEAS, L. M. and LADEN, F. (2008). Chronic particulate exposure, mortality, and coronary heart disease in the nurses' health study. *Am. J. Epidemiol.* **168** 1161–1168.

PUETT, R. C., HART, J. E., YANOSKY, J. D., PACIOREK, C., SCHWARTZ, J., SUH, H., SPEIZER, F. E. and LADEN, F. (2009). Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *Environ. Health Perspect.* **117** 1697–1701.

PUETT, R. C., HART, J. E., YANOSKY, J. D., SPIEGELMAN, D., WANG, M., FISHER, J. A., HONG, B. and LADEN, F. (2014). Particulate matter air pollution exposure, distance to road, and incident lung cancer in the nurses' health study cohort. *Environ. Health Perspect.* **122** 926–932.

RICHARDSON, D. B., MACLEHOSE, R. F., LANGHOLZ, B. and COLE, S. R. (2011a). Hierarchical latency models for dose-time-response associations. *Am. J. Epidemiol.* **173** 695–702. https://doi.org/10.1093/aje/kwq387

RICHARDSON, D. B., COLE, S. R., CHU, H. and LANGHOLZ, B. (2011b). Lagging exposure information in cumulative exposure-response analyses. *Am. J. Epidemiol.* **174** 1416–1422.

ROSNER, B., SPIEGELMAN, D. and WILLETT, W. C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error. *Am. J. Epidemiol.* **132** 734–745.

ROSNER, B., SPIEGELMAN, D. and WILLETT, W. C. (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *Am. J. Epidemiol.* **136** 1400–1413.

ROSNER, B., WILLETT, W. and SPIEGELMAN, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat. Med.* **8** 1051–1069.

ROTHMAN, K. J., GREENLAND, S., LASH, T. L. et al. (2008). *Modern Epidemiology* **3**. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia.

SARNAT, J. A., KOUTRAKIS, P. and SUH, H. H. (2000). Assessing the relationship between personal particulate and gaseous exposures of senior citizens living in Baltimore, MD. *J. Air Waste Manage. Assoc.* **50** 1184–1198.

SARNAT, S. E., COULL, B. A., SCHWARTZ, J., GOLD, D. R. and SUH, H. H. (2006). Factors affecting the association between ambient concentrations and personal exposures to particles and gases. *Environ. Health Perspect.* **114** 649–654.

SCHWARTZ, J., COULL, B., LADEN, F. and RYAN, L. (2008). The effect of dose and timing of dose on the association between airborne particles and survival. *Environ. Health Perspect.* **116** 64–69.

SPIEGELMAN, D., MCDERMOTT, A. and ROSNER, B. (1997). Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am. J. Clin. Nutr.* **65** 1179S–1186S. https://doi.org/10.1093/ajcn/65.4.1179S

SYLVESTRE, M.-P. and ABRAHAMOWICZ, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Stat. Med.* **28** 3437–3453. MR2744373 https://doi.org/10.1002/sim.3701

THOMAS, D. C. (2009). *Statistical Methods in Environmental Epidemiology*. Oxford Univ. Press, Oxford. MR3154379

THOMAS, D., STRAM, D. and DWYER, J. (1993). Exposure measurement error: Influence on exposure-disease. Relationships and methods of correction. *Annu. Rev. Public Health* **14** 69–93. https://doi.org/10.1146/annurev.pu.14.050193.000441

WANG, C., LIU, H. and GAO, S. (2017). A penalized Cox proportional hazards model with multiple time-varying exposures. *Ann. Appl. Stat.* **11** 185–201. MR3634320 https://doi.org/10.1214/16-AOAS999

WANG, C. Y., HSU, L., FENG, Z. D. and PRENTICE, R. L. (1997). Regression calibration in failure time regression. *Biometrics* **53** 131–145. MR1450183 https://doi.org/10.2307/2533103

WANG, M., LIAO, X., LADEN, F. and SPIEGELMAN, D. (2016). Quantifying risk over the life course—latency, age-related susceptibility, and other time-varying exposure metrics. *Stat. Med.* **35** 2283–2295. MR3513513 https://doi.org/10.1002/sim.6864

WEUVE, J., PUETT, R. C., SCHWARTZ, J., YANOSKY, J. D., LADEN, F. and GRODSTEIN, F. (2012). Exposure to particulate air pollution and cognitive decline in older women. *Arch. Intern. Med.* **172** 219–227.

WILLIAMS, R., SUGGS, J., REA, A., LEOVIC, K., VETTE, A., CROGHAN, C., SHELDON, L., RODES, C., THORNBURG, J. et al. (2003a). The Research Triangle Park particulate matter panel study: PM mass concentration relationships. *Atmos. Environ.* **37** 5349–5363.

WILLIAMS, R., SUGGS, J., REA, A., SHELDON, L., RODES, C. and THORNBURG, J. (2003b). The Research Triangle Park particulate matter panel study: Modeling ambient source contribution to personal and residential PM mass concentrations. *Atmos. Environ.* **37** 5365–5378.

YANOSKY, J. D., PACIOREK, C. J., SCHWARTZ, J., LADEN, F., PUETT, R. and SUH, H. H. (2008). Spatio-temporal modeling of chronic PM10 exposure for the Nurses' Health Study. *Atmos. Environ.* **42** 4047–4062.

YANOSKY, J. D., PACIOREK, C. J., LADEN, F., HART, J. E., PUETT, R. C., LIAO, D. and SUH, H. H. (2014). Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors. *Environ. Health* **13** 1–15.

ZANOBETTI, A., WAND, M. P., SCHWARTZ, J. and RYAN, L. M. (2000). Generalized additive distributed lag models: Quantifying mortality displacement. *Biostatistics* **1** 279–292.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1592–1614
https://doi.org/10.1214/22-AOAS1684

RANDOMIZATION INFERENCE FOR CLUSTER-RANDOMIZED TEST-NEGATIVE DESIGNS WITH APPLICATION TO DENGUE STUDIES: UNBIASED ESTIMATION, PARTIAL COMPLIANCE, AND STEPPED-WEDGE DESIGN

BY BINGKAI WANG[1,a], SUZANNE M. DUFAULT[2,c], DYLAN S. SMALL[1,b] AND NICHOLAS P. JEWELL[3,d]

[1]*Department of Statistics and Data Science, The Wharton School, University of Pennsylvania,* [a]*bingkai@wharton.upenn.edu,* [b]*dsmall@wharton.upenn.edu*

[2]*Division of Pulmonary and Critical Care Medicine, University of California, San Francisco,* [c]*suzanne.dufault@ucsf.edu*

[3]*Department of Medical Statistics, London School of Hygiene & Tropical Medicine,* [d]*Nicholas.Jewell@lshtm.ac.uk*

In 2019, the World Health Organization identified dengue as one of the top 10 global health threats. For the control of dengue, the Applying *Wolbachia* to Eliminate Dengue (AWED) study group conducted a cluster-randomized trial in Yogyakarta, Indonesia, and used a novel design, called the cluster-randomized test-negative design (CR-TND). This design can yield valid statistical inference with data collected by a passive surveillance system and thus has the advantage of cost-efficiency compared to traditional cluster-randomized trials. We investigate the statistical assumptions and properties of CR-TND under a randomization inference framework, which is known to be robust for small-sample problems. We find that, when the differential healthcare-seeking behavior comparing intervention and control varies across clusters (in contrast to the setting of Dufault and Jewell (*Stat. Med.* **39** (2020a) 1429–1439) where the differential healthcare-seeking behavior is constant across clusters), current analysis methods for CR-TND can be biased and have inflated type I error. We propose the log-contrast estimator that can eliminate such bias and improve precision by adjusting for covariates. Furthermore, we extend our methods to handle partial intervention compliance and a stepped-wedge design, both of which appear frequently in cluster-randomized trials. Finally, we demonstrate our results by simulation studies and reanalysis of the AWED study.

REFERENCES

ANDERS, K. L., CUTCHER, Z., KLEINSCHMIDT, I., DONNELLY, C. A., FERGUSON, N. M., INDRIANI, C., RYAN, P. A., O'NEILL, S. L., JEWELL, N. P. et al. (2018a). Cluster-randomized test-negative design trials: A novel and efficient method to assess the efficacy of community-level Dengue interventions. *Am. J. Epidemiol.* **187** 2021–2028.

ANDERS, K. L., INDRIANI, C., AHMAD, R. A., TANTOWIJOYO, W., ARGUNI, E., ANDARI, B., JEWELL, N. P., RANCES, E., O'NEILL, S. L. et al. (2018b). The AWED trial (applying Wolbachia to eliminate Dengue) to assess the efficacy of Wolbachia-infected mosquito deployments to reduce Dengue incidence in Yogyakarta, Indonesia: Study protocol for a cluster randomised controlled trial. *Trials* **19** 1–16.

ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.

ARONOW, P. M., GREEN, D. P. and LEE, D. K. K. (2014). Sharp bounds on the variance in randomized experiments. *Ann. Statist.* **42** 850–871. MR3210989 https://doi.org/10.1214/13-AOS1200

BRESLOW, N. E. and CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88** 9–25.

CATTARINO, L., RODRIGUEZ-BARRAQUER, I., IMAI, N., CUMMINGS, D. A. T. and FERGUSON, N. M. (2020). Mapping global variation in Dengue transmission intensity. *Sci. Transl. Med.* **12**. https://doi.org/10.1126/scitranslmed.aax4144

Key words and phrases. Case control, healthcare-seeking behavior, partial compliance, stepped-wedge design.

CHUA, H., FENG, S., LEWNARD, J. A., SULLIVAN, S. G., BLYTH, C. C., LIPSITCH, M. and COWLING, B. J. (2020). The use of test-negative controls to monitor vaccine effectiveness: A systematic review of methodology. *Epidemiology* **31** 43.

CLARKE, P. S. and WINDMEIJER, F. (2012). Instrumental variable estimators for binary outcomes. *J. Amer. Statist. Assoc.* **107** 1638–1652. MR3036422 https://doi.org/10.1080/01621459.2012.734171

DING, P., FELLER, A. and MIRATRIX, L. (2016). Randomization inference for treatment effect variation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 655–671. MR3506797 https://doi.org/10.1111/rssb.12124

DUFAULT, S. M. and JEWELL, N. P. (2020a). Analysis of counts for cluster randomized trials: Negative controls and test-negative designs. *Stat. Med.* **39** 1429–1439. MR4098500 https://doi.org/10.1002/sim.8488

DUTRA, H. L. C., ROCHA, M. N., DIAS, F. B. S., MANSUR, S. B., CARAGATA, E. P. and MOREIRA, L. A. (2016). Wolbachia blocks currently circulating Zika virus isolates in Brazilian Aedes aegypti mosquitoes. *Cell Host Microbe* **19** 771–774.

ENDO, A., FUNK, S. and KUCHARSKI, A. J. (2020). Bias correction methods for test-negative designs in the presence of misclassification. *Epidemiol. Infect.* **148**.

HABER, M., AN, Q., FOPPA, I. M., SHAY, D. K., FERDINANDS, J. M. and ORENSTEIN, W. A. (2015). A probability model for evaluating the bias and precision of influenza vaccine effectiveness estimates from case-control studies. *Epidemiol. Infect.* **143** 1417–1426.

HODGES, J. L. JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Stat.* **34** 598–611. MR0152070 https://doi.org/10.1214/aoms/1177704172

HUSSEY, M. A. and HUGHES, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials* **28** 182–191.

JACKSON, M. L. and NELSON, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. *Vaccine* **31** 2165–2168.

JEWELL, N. P., DUFAULT, S., CUTCHER, Z., SIMMONS, C. P. and ANDERS, K. L. (2019). Analysis of cluster-randomized test-negative designs: Cluster-level methods. *Biostatistics* **20** 332–346. MR3922137 https://doi.org/10.1093/biostatistics/kxy005

JI, X., FINK, G., ROBYN, P. J. and SMALL, D. S. (2017). Randomization inference for stepped-wedge cluster-randomized trials: An application to community-based health insurance. *Ann. Appl. Stat.* **11** 1–20. MR3634312 https://doi.org/10.1214/16-AOAS969

JOHNSON, K. N. (2015). The impact of Wolbachia on virus infection in mosquitoes. *Viruses* **7** 5705–5717.

LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

LI, X. and DING, P. (2020). Rerandomization and regression adjustment. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 241–268. MR4060984

LI, F., HUGHES, J. P., HEMMING, K., TALJAARD, M., MELNICK, E. R. and HEAGERTY, P. J. (2021). Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat. Methods Med. Res.* **30** 612–639. MR4236826 https://doi.org/10.1177/0962280220932962

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430 https://doi.org/10.1093/biomet/73.1.13

LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420 https://doi.org/10.1214/12-AOAS583

MCNEISH, D. and STAPLETON, L. M. (2016). Modeling clustered data with very few clusters. *Multivar. Behav. Res.* **51** 495–518.

MURRAY, D. M., VARNELL, S. P. and BLITSTEIN, J. L. (2004). Design and analysis of group-randomized trials: A review of recent methodological developments. *Am. J. Publ. Health* **94** 423–432.

ORENSTEIN, E. W., DE SERRES, G., HABER, M. J., SHAY, D. K., BRIDGES, C. B., GARGIULLO, P. and ORENSTEIN, W. A. (2007). Methodologic issues regarding the use of three observational study designs to assess influenza vaccine effectiveness. *Int. J. Epidemiol.* **36** 623–631.

RAINEY, S. M., SHAH, P., KOHL, A. and DIETRICH, I. (2014). Understanding the Wolbachia-mediated inhibition of arboviruses in mosquitoes: Progress and challenges. *J Gen Virol* **95** 517–530. https://doi.org/10.1099/vir.0.057422-0

ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138 https://doi.org/10.1007/978-1-4757-3692-2

ROTH, J. and SANT'ANNA, P. H. (2021). Efficient estimation for staggered rollout designs. ArXiv preprint. Available at arXiv:2102.01291.

SMALL, D. S., TEN HAVE, T. R. and ROSENBAUM, P. R. (2008). Randomization inference in a group-randomized trial of treatments for depression: Covariate adjustment, noncompliance, and quantile effects. *J. Amer. Statist. Assoc.* **103** 271–279. MR2420232 https://doi.org/10.1198/016214507000000897

SU, F. and DING, P. (2021). Model-assisted analyses of cluster-randomized experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 994–1015. MR4349125 https://doi.org/10.1111/rssb.12468

SULLIVAN, S. G., TCHETGEN TCHETGEN, E. J. and COWLING, B. J. (2016). Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. *Am. J. Epidemiol.* **184** 345–353.

UTARINI, A., INDRIANI, C., AHMAD, R. A., TANTOWIJOYO, W., ARGUNI, E., ANSARI, M. R., SUPRIYATI, E., WARDANA, D. S., MEITIKA, Y. et al. (2021). Efficacy of Wolbachia-infected mosquito deployments for the control of Dengue. *N. Engl. J. Med.* **384** 2177–2186. https://doi.org/10.1056/NEJMoa2030243

WALKER, T. J. P. H., JOHNSON, P. H., MOREIRA, L. A., ITURBE-ORMAETXE, I., FRENTIU, F. D., MCMENIMAN, C. J., LEONG, Y. S., DONG, Y., AXFORD, J. et al. (2011). The wMel Wolbachia strain blocks Dengue and invades caged Aedes aegypti populations. *Nature* **476** 450–453.

WANG, B., HARHAY, M. O., SMALL, D. S., MORRIS, T. P. and LI, F. (2021). On the robustness and precision of mixed-model analysis of covariance in cluster-randomized trials. ArXiv preprint. Available at arXiv:2112.00832.

WANG, B., DUFAULT, S. M., SMALL, D. S. and JEWELL, N. P. (2023). Supplement to "Randomization inference for cluster-randomized test-negative Designs with application to Dengue studies: Unbiased estimation, partial compliance, and stepped-wedge design." https://doi.org/10.1214/22-AOAS1684SUPP

WESTREICH, D. and HUDGENS, M. G. (2016). Invited commentary: Beware the test-negative design. *Am. J. Epidemiol.* **184** 354–356.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1615–1640
https://doi.org/10.1214/22-AOAS1685

VARIATIONAL BAYESIAN ANALYSIS OF NONHOMOGENEOUS HIDDEN MARKOV MODELS WITH LONG AND ULTRALONG SEQUENCES

BY XINYUAN CHEN[1,a], YIWEI LI[2,b], XIANGNAN FENG[3,c], AND JOSEPH T. CHANG[4,d]

[1]*Department of Mathematics and Statistics, Mississippi State University,* [a]*xchen@math.msstate.edu*

[2]*Department of Marketing and International Business, Lingnan University,* [b]*victor.li@ln.edu.hk*

[3]*Department of Statistics and Data Science, Fudan University,* [c]*fengxiangnan@fudan.edu.cn*

[4]*Department of Statistics and Data Science, Yale University,* [d]*joseph.chang@yale.edu*

Nonhomogeneous hidden Markov models (NHMMs) are useful in modeling sequential and autocorrelated data. Bayesian approaches, particularly Markov chain Monte Carlo (MCMC) methods, are principal statistical inference tools for NHMMs. However, MCMC sampling is computationally demanding, especially for long observation sequences. We develop a variational Bayes (VB) method for NHMMs, which utilizes a structured variational family of Gaussian distributions with factorized covariance matrices to approximate target posteriors, combining a forward-backward algorithm and stochastic gradient ascent in estimation. To improve efficiency and handle ultralong sequences, we further propose a subsequence VB (SVB) method that works on subsamples. The SVB method exploits the memory decay property of NHMMs and uses buffers to control for bias caused by breaking sequential dependence from subsampling. We highlight that the local nonhomogeneity of NHMMs substantially affects the required buffer lengths and propose the use of local Lyapunov exponents that characterize local memory decay rates of NHMMs and adaptively determine buffer lengths. Our methods are validated in simulation studies and in modeling ultralong sequences of customers' telecom records to uncover the relationship between their mobile Internet usage behaviors and conventional telecommunication behaviors.

REFERENCES

ABARBANEL, H. D. I., BROWN, R. and KENNEL, M. B. (1992). Local Lyapunov exponents computed from observed data. *J. Nonlinear Sci.* **2** 343–365. MR1186765 https://doi.org/10.1007/BF01208929

AICHER, C., MA, Y.-A., FOTI, N. J. and FOX, E. B. (2019). Stochastic gradient MCMC for state space models. *SIAM J. Math. Data Sci.* **1** 555–587. MR4010763 https://doi.org/10.1137/18M1214780

ALTMAN, R. M. (2007). Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *J. Amer. Statist. Assoc.* **102** 201–210. MR2345538 https://doi.org/10.1198/016214506000001086

ANDREWS, M., LUO, X., FANG, Z. and GHOSE, A. (2016). Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Mark. Sci.* **35** 218–233. https://doi.org/10.1287/mksc.2015.0905

ANSARI, A., LI, Y. and ZHANG, J. Z. (2018). Probabilistic topic model for hybrid recommender systems: A stochastic variational Bayesian approach. *Mark. Sci.* **37** 987–1008.

ARNOLD, L. (1998). *Random Dynamical Systems. Springer Monographs in Mathematics*. Springer, Berlin. MR1723992 https://doi.org/10.1007/978-3-662-12878-7

ASCARZA, E., NETZER, O. and HARDIE, B. G. S. (2018). Some customers would rather leave without saying goodbye. *Mark. Sci.* **37** 54–77. https://doi.org/10.1287/mksc.2017.1057

BAUM, L. E., PETRIE, T., SOULES, G. and WEISS, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.* **41** 164–171. MR0287613 https://doi.org/10.1214/aoms/1177697196

BLEI, D. M. and JORDAN, M. I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Anal.* **1** 121–143. MR2227367 https://doi.org/10.1214/06-BA104

Key words and phrases. Nonhomogeneous hidden Markov model, variational Bayesian inference, local Lyapunov exponents, mobile Internet usage.

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

BOULDING, W., STAELIN, R., EHRET, M. and JOHNSTON, W. J. (2005). A customer relationship management roadmap: What is known, potential pitfalls, and where to go. *J. Mark.* **69** 155–166.

BRAUN, M. and MCAULIFFE, J. (2010). Variational inference for large-scale models of discrete choice. *J. Amer. Statist. Assoc.* **105** 324–335. MR2757203 https://doi.org/10.1198/jasa.2009.tm08030

CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. *Springer Series in Statistics*. Springer, New York. MR2159833

CHEN, X., LI, Y., CHANG, J. and FENG, X. (2023). Supplement to "Variational Bayesian analysis of nonhomogeneous hidden Markov models with long and ultralong sequences." https://doi.org/10.1214/22-AOAS1685SUPPA, https://doi.org/10.1214/22-AOAS1685SUPPB

COLLET, P. and LEONARDI, F. (2014). Loss of memory of hidden Markov models and Lyapunov exponents. *Ann. Appl. Probab.* **24** 422–446. MR3161652 https://doi.org/10.1214/13-AAP929

FONG, N. M., FANG, Z. and LUO, X. (2015). Geo-conquesting: Competitive locational targeting of mobile promotions. *J. Mark. Res.* **52** 726–735. https://doi.org/10.1509/jmr.14.0229

FOTI, N., XU, J., LAIRD, D. and FOX, E. (2014). Stochastic variational inference for hidden Markov models. In *Advances in Neural Information Processing Systems* 3599–3607.

FRÜHWIRTH-SCHNATTER, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *J. Amer. Statist. Assoc.* **96** 194–209. MR1952732 https://doi.org/10.1198/016214501750333063

GENTZKOW, M., KELLY, B. and TADDY, M. (2019). Text as data. *J. Econ. Lit.* **57** 535–574.

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 https://doi.org/10.1198/016214506000001437

HEAPS, S. E., BOYS, R. J. and FARROW, M. (2015). Bayesian modelling of rainfall data by using non-homogeneous hidden Markov models and latent Gaussian variables. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64** 543–568. MR3325463 https://doi.org/10.1111/rssc.12094

HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14** 1303–1347. MR3081926

HOLSCLAW, T., GREENE, A. M., ROBERTSON, A. W. and SMYTH, P. (2017). Bayesian nonhomogeneous Markov models via Pólya-gamma data augmentation with applications to rainfall modeling. *Ann. Appl. Stat.* **11** 393–426. MR3634329 https://doi.org/10.1214/16-AOAS1009

HUGHES, J. P., GUTTORP, P. and CHARLES, S. P. (1999). A non-homogeneous hidden Markov model for precipitation occurrence. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **48** 15–30.

IP, E., ZHANG, Q., REJESKI, J., HARRIS, T. and KRITCHEVSKY, S. (2013). Partially ordered mixed hidden Markov model for the disablement process of older adults. *J. Amer. Statist. Assoc.* **108** 370–384. MR3174627 https://doi.org/10.1080/01621459.2013.770307

JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233. https://doi.org/10.1023/A:1007665907178

KANG, K., CAI, J., SONG, X. and ZHU, H. (2019). Bayesian hidden Markov models for delineating the pathology of Alzheimer's disease. *Stat. Methods Med. Res.* **28** 2112–2124. MR3977095 https://doi.org/10.1177/0962280217748675

KANI, A., DESARBO, W. S. and FONG, D. K. H. (2018). A factorial hidden Markov model for the analysis of temporal change in choice models. *Cust. Needs Solut.* **5** 162–177. https://doi.org/10.1007/s40547-018-0088-0

KINGMA, D. P. and BA, J. (2015). Adam: A method for stochastic optimization. In 3*rd International Conference on Learning Representations*, *ICLR* 2015, *San Diego*, *CA*, *USA*, *May* 7–9, 2015, *Conference Track Proceedings*.

KINGMA, D. P. and WELLING, M. (2014). Auto-encoding variational Bayes. In *Proceedings of the* 2*nd International Conference on Learning Representations* (*ICLR*).

KOTLER, P. and KELLER, K. L. (2016). *Marketing Management*. Pearson Italia Spa.

KUCUKELBIR, A., TRAN, D., RANGANATH, R., GELMAN, A. and BLEI, D. M. (2017). Automatic differentiation variational inference. *J. Mach. Learn. Res.* **18** Paper No. 14, 45 pp. MR3634881

LE GLAND, F. and MEVEL, L. (2000a). Basic properties of the projective product with application to products of column-allowable nonnegative matrices. *Math. Control Signals Systems* **13** 41–62. MR1742139 https://doi.org/10.1007/PL00009860

LE GLAND, F. and MEVEL, L. (2000b). Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Systems* **13** 63–93. MR1742140 https://doi.org/10.1007/PL00009861

LUO, X., ANDREWS, M., FANG, Z. and PHANG, C. W. (2014). Mobile targeting. *Manage. Sci.* **60** 1738–1756. https://doi.org/10.1287/mnsc.2013.1836

MA, Y., FOTI, N. J. and FOX, E. B. (2017). Stochastic gradient MCMC methods for hidden Markov models. In *Proceedings of the* 34*th International Conference on Machine Learning* **70** 2265–2274. JMLR.org.

MA, L., SUN, B. and KEKRE, S. (2015). The squeaky wheel gets the grease—An empirical analysis of customer voice and firm intervention on Twitter. *Mark. Sci.* **34** 627–645.

MCCULLAGH, P. (1980). Regression models for ordinal data. *J. Roy. Statist. Soc. Ser. B* **42** 109–142. MR0583347

MELIGKOTSIDOU, L. and DELLAPORTAS, P. (2011). Forecasting with non-homogeneous hidden Markov models. *Stat. Comput.* **21** 439–449. MR2806620 https://doi.org/10.1007/s11222-010-9180-5

MONTOYA, R., NETZER, O. and JEDIDI, K. (2010). Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Mark. Sci.* **29** 909–924. https://doi.org/10.1287/mksc.1100.0570

NEMIROVSKI, A., JUDITSKY, A., LAN, G. and SHAPIRO, A. (2009). Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19** 1574–1609. MR2486041 https://doi.org/10.1137/070704277

NETZER, O., LATTIN, J. M. and SRINIVASAN, V. (2008). A hidden Markov model of customer relationship dynamics. *Mark. Sci.* **27** 185–204. https://doi.org/10.1287/mksc.1070.0294

ONG, V. M.-H., NOTT, D. J. and SMITH, M. S. (2018). Gaussian variational approximation with a factor covariance structure. *J. Comput. Graph. Statist.* **27** 465–478. MR3863750 https://doi.org/10.1080/10618600.2017.1390472

PADILLA, N., MONTOYA, R. and NETZER, O. (2020). Heterogeneity in HMMs: Allowing for heterogeneity in the number of states. Working Paper.

POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. MR3174712 https://doi.org/10.1080/01621459.2013.829001

RANGANATH, R., GERRISH, S. and BLEI, D. (2014). Black box variational inference. In *Artificial Intelligence and Statistics* 814–822.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

SONG, X., XIA, Y. and ZHU, H. (2017). Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics* **73** 313–323. MR3632377 https://doi.org/10.1111/biom.12536

SPEZIA, L. (2006). Bayesian analysis of non-homogeneous hidden Markov models. *J. Stat. Comput. Simul.* **76** 713–725. MR2253194 https://doi.org/10.1080/10629360500108798

YE, X. (2018). Stochastic dynamics: Markov chains, random transformations and applications. Ph.D. thesis, Univ. Washington.

ZEILER, M. D. (2012). ADADELTA: An adaptive learning rate method. arXiv E-prints.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1641–1662
https://doi.org/10.1214/22-AOAS1687

HOW CLOSE AND HOW MUCH? LINKING HEALTH OUTCOMES TO BUILT ENVIRONMENT SPATIAL DISTRIBUTIONS

BY ADAM T. PETERSON[1,a], VERONICA J. BERROCAL[2], EMMA V. SANCHEZ-VAZNAUGH[3] AND BRISA N. SÁNCHEZ[4,b]

[1]*Department of Biostatistics, University of Michigan,* [a]*atpvyc@umich.edu*
[2]*Department of Statistics, University of California, Irvine*
[3]*Department of Public Health, San Francisco State University*
[4]*Department of Biostatistics and Epidemiology, Drexel University,* [b]*bns48@drexel.edu*

Built environment features (BEFs) refer to aspects of the human constructed environment which may, in turn, support or restrict health related behaviors and thus impact health. In this paper we are interested in understanding whether the spatial distribution and quantity of fast-food restaurants (FFRs) influence the risk of obesity in schoolchildren. To achieve this goal, we propose a two-stage Bayesian hierarchical modeling framework. In the first stage, examining the position of FFRs relative to that of some reference locations—in our case, schools—we model the distances of FFRs from these reference locations as realizations of inhomogenous Poisson processes (IPP). With the goal of identifying representative spatial patterns of exposure to FFRs, we model the intensity functions of the IPPs using a Bayesian nonparametric model, specifying a nested Dirichlet process prior. The second-stage model relates exposure patterns to obesity. We offer two different approaches to carry out the second stage; they differ in how they accommodate uncertainty in the exposure patterns. In the first approach, the odds of obesity at the school level is regressed on cluster indicators, each representing a major pattern of exposure to FFRs. In the second, we employ Bayesian kernel machine regression to relate the odds of obesity to the multivariate vector reporting the degree of similarity of a given school to all other schools. Our analysis on the influence of patterns of FFR occurrence on obesity among Californian schoolchildren has indicated that, in 2010, among schools that are consistently assigned to a cluster, there is a lower odds of obesity among ninth graders who attend schools with most distant FFR occurrences in a one-mile radius, as compared to others.

REFERENCES

AUCHINCLOSS, A. H., MOORE, K. A., MOORE, L. V. and ROUX, A. V. D. (2012). Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. *Health Place* **18** 1341–1347.

AUSTIN, S. B., MELLY, S. J., SANCHEZ, B. N., PATEL, A., BUKA, S. and GORTMAKER, S. L. (2005). Clustering of fast-food restaurants around schools: A novel application of spatial statistics to the study of food environments. *Am. J. Publ. Health* **95** 1575–1581.

BAEK, J., SÁNCHEZ, B. N., BERROCAL, V. J. and SANCHEZ-VAZNAUGH, E. V. (2016). Distributed lag models: Examining associations between the built environment and health. *Epidemiology* **27** 116.

BESSER, L. M., RODRIGUEZ, D. A., MCDONALD, N., KUKULL, W. A., FITZPATRICK, A. L., RAPP, S. R. and SEEMAN, T. (2018). Neighborhood built environment and cognition in non-demented older adults: The multiethnic study of atherosclerosis. *Soc. Sci. Med.* **200** 27–35. https://doi.org/10.1016/j.socscimed.2018.01.007

BOBB, J. F., VALERI, L., CLAUS HENN, B., CHRISTIANI, D. C., WRIGHT, R. O., MAZUMDAR, M., GODLESKI, J. J. and COULL, B. A. (2015). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics* **16** 493–508. MR3365442 https://doi.org/10.1093/biostatistics/kxu058

Key words and phrases. Food environment, child obesity, inhomogenous Poisson process, density estimation, nested Dirichlet process.

BOJORQUEZ, I. and OJEDA-REVAH, L. (2018). Urban public parks and mental health in adult women: Mediating and moderating factors. *Int. J. Soc. Psychiatry* **64** 637–646. https://doi.org/10.1177/0020764018795198

CAMERLENGHI, F., DUNSON, D. B., LIJOI, A., PRÜNSTER, I. and RODRÍGUEZ, A. (2019). Latent nested nonparametric priors (with discussion). *Bayesian Anal.* **14** 1303–1356. With discussions and a rejoinder. MR4044854 https://doi.org/10.1214/19-BA1169

CARPENTER, B., GELMAN, A., HOFFMAN, M., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M. A., GUO, J., LI, P. et al. (2016). Stan: A probabilistic programming language. *J. Stat. Softw.* **20** 1–37.

CHIANG, S., GUINDANI, M., YEH, H. J., DEWAR, S., HANEEF, Z., STERN, J. M. and VANNUCCI, M. (2017). A hierarchical Bayesian model for the identification of PET markers associated to the prediction of surgical outcome after anterior temporal lobe resection. *Front. Neurosci.* **11** 669.

CURRIE, J., DELLAVIGNA, S., MORETTI, E. and PATHANIA, V. (2010). The effect of fast food restaurants on obesity and weight gain. *Amer. Econ. J.: Econ. Policy* **2** 32–63.

DAVIS, B. and CARPENTER, C. (2009). Proximity of fast-food restaurants to schools and adolescent obesity. *Am. J. Publ. Health* **99** 505–510.

DIEBOLT, J. and ROBERT, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *J. Roy. Statist. Soc. Ser. B* **56** 363–375. MR1281940

DIGGLE, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd ed. *Monographs on Statistics and Applied Probability* **128**. CRC Press, Boca Raton, FL. MR3113855

EVENSON, K. R., JONES, S. A., HOLLIDAY, K. M., COHEN, D. A. and MCKENZIE, T. L. (2016). Park characteristics, use, and physical activity: A review of studies using SOPARC (system for observing play and recreation in communities). *Prev. Med.* **86** 153–166.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3235677

GOODRICH, B., GABRY, J., ALI, I. and BRILLEMAN, S. (2020). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.19.3.

GRAZIANI, R., GUINDANI, M. and THALL, P. F. (2015). Bayesian nonparametric estimation of targeted agent effects on biomarker change to predict clinical outcome. *Biometrics* **71** 188–197. MR3335363 https://doi.org/10.1111/biom.12250

HARTIGAN, J. A. (1975). *Clustering Algorithms*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0405726

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. MR1952729 https://doi.org/10.1198/016214501750332758

MACEACHERN, S. N. (2000). Dependent dirichlet processes. Unpublished Manuscript, Dept. Statistics, The Ohio State Univ., 1–40.

MACEACHERN, S. N. and SHEN, X. (1999). Variable selection and function estimation in additive nonparametric regression using a data-based prior: Comment. *J. Amer. Statist. Assoc.* **94** 799–802.

MCGUIRE, S. (2012). Institute of medicine (IOM) early childhood obesity prevention policies. Washington, DC: The National Academies Press; 2011. *Adv. Nutrition* **3** 56–57. https://doi.org/10.3945/an.111.001347

MILLER, J. W. and HARRISON, M. T. (2013). A simple example of Dirichlet process mixture inconsistency for the number of components. In *Advances in Neural Information Processing Systems* 199–206.

MILLER, J. W. and HARRISON, M. T. (2014). Inconsistency of Pitman–Yor process mixtures for the number of components. *J. Mach. Learn. Res.* **15** 3333–3370. MR3277163

MILLER, J. W. and HARRISON, M. T. (2018). Mixture models with a prior on the number of components. *J. Amer. Statist. Assoc.* **113** 340–356. MR3803469 https://doi.org/10.1080/01621459.2016.1255636

NGUYEN, H. D. and MCLACHLAN, G. (2019). On approximations via convolution-defined mixture models. *Comm. Statist. Theory Methods* **48** 3945–3955. MR3976714 https://doi.org/10.1080/03610926.2018.1487069

NYLUND-GIBSON, K., GRIMM, R. P. and MASYN, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Struct. Equ. Model.* **26** 967–985. MR4022809 https://doi.org/10.1080/10705511.2019.1590146

PAPASTAMOULIS, P. (2016). label.switching: An R package for dealing with the label switching problem in MCMC outputs. *J. Stat. Softw.* **69** 1–24. https://doi.org/10.18637/jss.v069.c01

PETERSON, A. (2020). bendr: Built environment nested Dirichlet processes in R. R package version 0.1.0-alpha.

PETERSON, A. T., BERROCAL, V. J., SANCHEZ-VAZNAUGH, E. V. and SÁNCHEZ, B. N. (2023). Supplement to “How close and how much? Linking health outcomes to built environment spatial distributions.” https://doi.org/10.1214/22-AOAS1687SUPP

R CORE TEAM (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RAFTERY, A. E. and LEWIS, S. M. (1995). The number of iterations, convergence diagnostics and generic Metropolis algorithms. In *Practical Markov Chain Monte Carlo* **7** 763–773.

REN, L., DU, L., CARIN, L. and DUNSON, D. B. (2011). Logistic stick-breaking process. *J. Mach. Learn. Res.* **12** 203–239. MR2773552

RODRÍGUEZ, A., DUNSON, D. B. and GELFAND, A. E. (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.* **103** 1131–1144. MR2528831 https://doi.org/10.1198/016214508000000553

RODRÍGUEZ, C. E. and WALKER, S. G. (2014). Label switching in Bayesian mixture models: Deterministic relabeling strategies. *J. Comput. Graph. Statist.* **23** 25–45. MR3173759 https://doi.org/10.1080/10618600.2012.735624

ROOF, K. and OLERU, N. (2008). Public health: Seattle and King County's push for the built environment. *J. Environ. Health* **71** 24–27.

SACKS, G., SWINBURN, B. and XUEREB, G. (2012). Population-based approaches to childhood obesity prevention.

SÁNCHEZ, B. N., SANCHEZ-VAZNAUGH, E. V., USCILKA, A., BAEK, J. and ZHANG, L. (2012). Differential associations between the food environment near schools and childhood overweight across race/ethnicity, gender, and grade. *Am. J. Epidemiol.* **175** 1284–1293.

SANCHEZ-VAZNAUGH, E. V., WEVERKA, A., MATSUZAKI, M. and SÁNCHEZ, B. N. (2019). Changes in fast food outlet availability near schools: Unequal patterns by income, race/ethnicity, and urbanicity. *Am. J. Prev. Med.* **57** 338–345. https://doi.org/10.1016/j.amepre.2019.04.023

SKINNER, A. C., RAVANBAKHT, S. N., SKELTON, J. A., PERRIN, E. M. and ARMSTRONG, S. C. (2018). Prevalence of obesity and severe obesity in US children, 1999–2016. *Pediatrics* **141**. https://doi.org/10.1542/peds.2017-3459

STAN DEVELOPMENT TEAM (2020). RStan: The R interface to Stan. R package version 2.19.3.

STEPHENS, M. (2000). Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 795–809. MR1796293 https://doi.org/10.1111/1467-9868.00265

VALERI, L., MAZUMDAR, M. M., BOBB, J. F., CLAUS HENN, B., RODRIGUES, E., SHARIF, O. I., KILE, M. L., QUAMRUZZAMAN, Q., AFROZ, S. et al. (2017). The joint effect of prenatal exposure to metal mixtures on neurodevelopmental outcomes at 20–40 months of age: Evidence from rural Bangladesh. *Environ. Health Perspect.* **125** 067015.

VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. MR3647105 https://doi.org/10.1007/s11222-016-9696-4

WADE, S. (2015). mcclust.ext: Point estimation and credible balls for Bayesian cluster analysis. R package version 1.0.

WADE, S. and GHAHRAMANI, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Anal.* **13** 559–626. With discussion and a reply by the authors. MR3807860 https://doi.org/10.1214/17-BA1073

WALL, M. M. and LIU, X. (2009). Spatial latent class analysis model for spatially distributed multivariate binary data. *Comput. Statist. Data Anal.* **53** 3057–3069. MR2667610 https://doi.org/10.1016/j.csda.2008.07.037

WALL, M. M., LARSON, N. I., FORSYTH, A., VAN RIPER, D. C., GRAHAM, D. J., STORY, M. T. and NEUMARK-SZTAINER, D. (2012). Patterns of obesogenic neighborhood features and adolescent weight: A comparison of statistical approaches. *Am. J. Prev. Med.* **42** e65–e75.

WALLS, D. (2013). National establishment time-series (NETS) database: 2012 database description. Walls & Associates, Oakland.

XIAO, S., KOTTAS, A. and SANSÓ, B. (2015). Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences. *Ann. Appl. Stat.* **9** 353–382. MR3341119 https://doi.org/10.1214/14-AOAS796

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303 https://doi.org/10.1198/016214504000000241

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1663–1680
https://doi.org/10.1214/22-AOAS1688

TRUNCATED RANK-BASED TESTS FOR TWO-PART MODELS WITH EXCESSIVE ZEROS AND APPLICATIONS TO MICROBIOME DATA

BY WANJIE WANG[1,a], ERIC CHEN[2,b] AND HONGZHE LI[2,c]

[1]*Department of Statistics and Applied Probability, National University of Singapore,* [a]*staww@nus.edu.sg*
[2]*Division of Biostatistics, University of Pennsylvania,* [b]*chvlyl@gmail.com,* [c]*hongzhe@upenn.edu*

High-throughput sequencing technology allows us to test the compositional difference of bacteria in different populations. One important feature of human microbiome data is that it often includes a large number of zeros. Such data can be treated as being generated from a two-part model that includes a zero-point mass. Motivated by analysis of such nonnegative data with excessive zeros, we introduce several truncated rank-based two-group and multigroup tests, including a truncated rank-based Wilcoxon rank-sum test for two-group comparison and two truncated Kruskal–Wallis tests for multigroup comparisons. We show, both analytically through asymptotic relative efficiency analysis and by simulations, that the proposed tests have higher power than the standard rank-based tests in typical microbiome data settings, especially when the proportion of zeros in the data is high. The tests can also be applied to repeated measurements of compositional data via simple within-subject permutations. In a simple before-and-after treatment experiment, the within-subject permutation is similar to the paired rank test. However, the proposed tests handle the excessive zeros which leads to a better power. We apply the tests to compare the microbiome compositions of healthy children and pediatric Crohn's disease patients and to assess the treatment effects on microbiome compositions. We identify several bacterial genera that are missed by the standard rank-based tests.

REFERENCES

GEVERS, D., KUGATHASAN, S., DENSON, L. A. et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15** 382–392.

HALLSTROM, A. P. (2010). A modified Wilcoxon test for non-negative distributions with a clump of zeros. *Stat. Med.* **29** 391–400. MR2750556 https://doi.org/10.1002/sim.3785

HUSON, D. H., AUCH, A. F., QI, J. and SCHUSTER, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* **17** 377–386.

LACHENBRUCH, P. (1976). Analysis of data with clumping at zero. *Biom. J.* **18** 351–356.

LACHENBRUCH, P. (2001). Comparison of two-part models with competitors. *Stat. Med.* **20** 1215–1234.

LACHENBRUCH, P. A. (2002). Analysis of data with excess zeros. *Stat. Methods Med. Res.* **11** 297–302. https://doi.org/10.1191/0962280202sm289ra

LAU, S. K., WOO, P. C., FUNG, A. M., CHAN, K.-M., WOO, G. K. and YUEN, K.-Y. (2004). Anaerobic, non-sporulating, Gram-positive bacilli bacteraemia characterized by 16S rRNA gene sequencing. *J. Med. Microbiol.* **53** 1247–1253. https://doi.org/10.1099/jmm.0.45803-0

LEWIS, J., CHEN, E. Z., BALDASSANO, R. N., OTLEY, A. R., GRIFFITHS, A. M., LEE, D., BITTINGER, K., BAILEY, A., FRIEDMAN, E. S. et al. (2015). Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe* **18** 489–500.

MACHIELS, K., JOOSSENS, M., SABINO, J., DE PRETER, V., ARIJS, I., EECKHAUT, V., BALLET, V., CLAES, K., VAN IMMERSEEL, F. et al. (2014). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* **63** 1275–1283.

MANICHANH, C., BORRUEL, N., CASELLAS, F. and GUARNER, F. (2012). The gut microbiota in IBD. *Nat. Rev. Gastroenterol. Hepatol.* **9** 599–608. https://doi.org/10.1038/nrgastro.2012.152

Key words and phrases. Asymptotic relative efficiency, differential abundance analysis, two-part model, truncation.

QIN, J., LI, R., RAES, J., ARUMUGAM, M., BURGDORF, K. S., MANICHANH, C., NIELSEN, T., PONS, N., LEVENEZ, F. et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464** 59–65.

QIN, J., LI, Y., CAI, Z., LI, S., ZHU, J., ZHANG, F., LIANG, S., ZHANG, W., GUAN, Y. et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490** 55–60.

SARTOR, R. and MAZMANIAN, S. (2012). Intestinal microbes in inflammatory bowel diseases. *Am. J. Gastroenterol*. **1** 15–21.

SEGATA, N., WALDRON, L., BALLARINI, A., NARASIMHAN, V., JOUSSON, O. and HUTTENHOWER, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9** 811–814.

TURNBAUGH, P. J., LEY, R. E., MAHOWALD, M. A., MAGRINI, V., MARDIS, E. R. and GORDON, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444** 1027–1031. https://doi.org/10.1038/nature05414

TURNBAUGH, P. J., LEY, R. E., HAMADY, M., FRASER-LIGGETT, C. M., KNIGHT, R. and GORDON, J. I. (2007). The human microbiome project. *Nature* **449** 804–810.

WAGNER, B. D., ROBERTSON, C. E. and HARRIS, J. K. (2011). Application of two-part statistics for comparison of sequence variant counts. *PLoS ONE* **6** e20296. https://doi.org/10.1371/journal.pone.0020296

WANG, W., CHEN, E. and LI, H. (2023). Supplement to “Truncated rank-based tests for two-part models with excessive zeros and applications to microbiome data.” https://doi.org/10.1214/22-AOAS1688SUPP

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1681–1700
https://doi.org/10.1214/22-AOAS1689

distinct: A NOVEL APPROACH TO DIFFERENTIAL DISTRIBUTION ANALYSES

BY SIMONE TIBERI[1,2,a], HELENA L. CROWELL[2,b], PANTELIS SAMARTSIDIS[3,d], LUKAS M. WEBER[4,e] AND MARK D. ROBINSON[2,c]

[1]*Department of Statistical Sciences, University of Bologna,* [a]*Simone.Tiberi@unibo.it*
[2]*Department of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich,* [b]*Helena.Crowell@uzh.ch,* [c]*mark.robinson@mls.uzh.ch*
[3]*MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge,* [d]*pantelis.samartsidis@mrc-bsu.cam.ac.uk*
[4]*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health,* [e]*lukas.weber@jhu.edu*

We present *distinct*, a general method for differential analysis of full distributions that is well suited to applications on single-cell data, such as single-cell RNA sequencing and high-dimensional flow or mass cytometry data. High-throughput single-cell data reveal an unprecedented view of cell identity and allow complex variations between conditions to be discovered; nonetheless, most methods for differential expression target differences in the mean and struggle to identify changes where the mean is only marginally affected. *distinct* is based on a hierarchical nonparametric permutation approach and, by comparing empirical cumulative distribution functions, identifies both differential patterns involving changes in the mean as well as more subtle variations that do not involve the mean. We performed extensive benchmarks across both simulated and experimental datasets from single-cell RNA sequencing and mass cytometry data, where *distinct* shows favourable performance, identifies more differential patterns than competitors, and displays good control of false positive and false discovery rates. *distinct* is available as a Bioconductor R package.

REFERENCES

AMEZQUITA, R. A., LUN, A. T., BECHT, E., CAREY, V. J., CARPP, L. N., GEISTLINGER, L., MARINI, F., RUE-ALBRECHT, K., RISSO, D. et al. (2020). Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17** 137–145.

AZODI, C. B., ZAPPIA, L., OSHLACK, A. and MCCARTHY, D. J. (2021). splatPop: Simulating population scale single-cell RNA sequencing data. *Genome Biol*. **22** 1–16.

BACHER, R., CHU, L.-F., LENG, N., GASCH, A. P., THOMSON, J. A., STEWART, R. M., NEWTON, M. and KENDZIORSKI, C. (2017). SCnorm: Robust normalization of single-cell RNA-seq data. *Nat. Methods* **14** 584–586.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BODENMILLER, B., ZUNDER, E. R., FINCK, R., CHEN, T. J., SAVIG, E. S., BRUGGNER, R. V., SIMONDS, E. F., BENDALL, S. C., SACHS, K. et al. (2012). Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat. Biotechnol*. **30** 858–867.

CROWELL, H. L. (2020). muscData: Multi-sample multi-group scRNA-seq data. R package version 1.1.2.

CROWELL, H. L., SONESON, C., GERMAIN, P.-L., CALINI, D., COLLIN, L., RAPOSO, C., MALHOTRA, D. and ROBINSON, M. D. (2020). *muscat* detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun*. **11** 1–12.

CSARDI, G. and NEPUSZ, T. (2006). The igraph software package for complex network research. *Int. J. Complex Syst*. **1695** 1–9.

DOHERTY, M. R., CHEON, H., JUNK, D. J., VINAYAK, S., VARADAN, V., TELLI, M. L., FORD, J. M., STARK, G. R. and JACKSON, M. W. (2017). Interferon-beta represses cancer stem cell properties in triple-negative breast cancer. *Proc. Natl. Acad. Sci. USA* **114** 13792–13797.

Key words and phrases. Differential distribution, differential analyses, differential state, high-throughput single-cell data, single-cell RNA-seq, single-cell flow and mass cytometry, permutation tests.

ELING, N., RICHARD, A. C., RICHARDSON, S., MARIONI, J. C. and VALLEJOS, C. A. (2018). Correcting the mean-variance dependency for differential variability testing using single-cell RNA sequencing data. *Cell Syst*. **7** 284–294.

FINAK, G., MCDAVID, A., YAJIMA, M., DENG, J., GERSUK, V., SHALEK, A. K., SLICHTER, C. K., MILLER, H. W., MCELRATH, M. J. et al. (2015). MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol*. **16** 1–13.

HAFEMEISTER, C. and SATIJA, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. **20** 1–15.

KANG, H. M., SUBRAMANIAM, M., TARG, S., NGUYEN, M., MALISKOVA, L., MCCARTHY, E., WAN, E., WONG, S., BYRNES, L. et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol*. **36** 89.

KHARCHENKO, P. V., SILBERSTEIN, L. and SCADDEN, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11** 740–742.

KORTHAUER, K. D., CHU, L.-F., NEWTON, M. A., LI, Y., THOMSON, J., STEWART, R. and KENDZIORSKI, C. (2016). A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol*. **17** 222.

LOVE, M. I., HUBER, W. and ANDERS, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15** 550. https://doi.org/10.1186/s13059-014-0550-8

LÜTGE, A., ZYPRYCH-WALCZAK, J., KUNZMANN, U. B., CROWELL, H. L., CALINI, D., MALHOTRA, D., SONESON, C. and ROBINSON, M. D. (2021). CellMixS: Quantifying and visualizing batch effects in single-cell RNA-seq data. *Life Sci. Alliance* **4** e202001004.

MASSEY JR, F. J. (1951). The Kolmogorov–Smirnov test for goodness of fit. *J. Amer. Statist. Assoc*. **46** 68–78.

MCCARTHY, D. J., CAMPBELL, K. R., LUN, A. T. and WILLS, Q. F. (2017). Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33** 1179–1186.

NOWICKA, M., KRIEG, C., CROWELL, H. L., WEBER, L. M., HARTMANN, F. J., GUGLIETTA, S., BECHER, B., LEVESQUE, M. P. and ROBINSON, M. D. (2017). CyTOF workflow: Differential discovery in high-throughput high-dimensional cytometry datasets. *F1000Res*. **6** 748. https://doi.org/10.12688/f1000research.11622.3

PHIPSON, B. and SMYTH, G. K. (2010). Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol*. **9** Art. 39, 14 pp. MR2746025 https://doi.org/10.2202/1544-6115.1585

QIN, X.-Q., TAO, N., DERGAY, A., MOY, P., FAWELL, S., DAVIS, A., WILSON, J. M. and BARSOUM, J. (1998). Interferon-β gene therapy inhibits tumor formation and causes regression of established tumors in immune-deficient mice. *Proc. Natl. Acad. Sci. USA* **95** 14411–14416.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. and SMYTH, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43** e47–e47.

ROBINSON, M. D., MCCARTHY, D. J. and SMYTH, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26** 139–140.

SONESON, C. and ROBINSON, M. D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15** 255–261. https://doi.org/10.1038/nmeth.4612

SQUAIR, J. W., GAUTIER, M., KATHE, C., ANDERSON, M. A., JAMES, N. D., HUTSON, T. H., HUDELLE, R., QAISER, T., MATSON, K. J. et al. (2021). Confronting false discoveries in single-cell differential expression. *BioRxiv*.

TIBERI, S., CROWELL, H. L., SAMARTSIDIS, P. and WEBER, L. M. (2023). Supplement to "*distinct*: A novel approach to differential distribution analyses." https://doi.org/10.1214/22-AOAS1689SUPPA, https://doi.org/10.1214/22-AOAS1689SUPPB

TUNG, P.-Y., BLISCHAK, J. D., HSIAO, C. J., KNOWLES, D. A., BURNETT, J. E., PRITCHARD, J. K. and GILAD, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep*. **7** 39921.

UHLÉN, M., FAGERBERG, L., HALLSTRÖM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, Å., KAMPF, C., SJÖSTEDT, E. et al. (2015). Tissue-based map of the human proteome. *Science* **347** 1260419.

VALLEJOS, C. A., MARIONI, J. C. and RICHARDSON, S. (2015). BASiCS: Bayesian analysis of single-cell sequencing data. *PLoS Comput. Biol*. **11** e1004333. https://doi.org/10.1371/journal.pcbi.1004333

VALLEJOS, C. A., RICHARDSON, S. and MARIONI, J. C. (2016). Beyond comparisons of means: Understanding changes in gene expression at the single-cell level. *Genome Biol*. **17** 1–14.

WANG, T., LI, B., NELSON, C. E. and NABAVI, S. (2019). Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. *BMC Bioinform*. **20** 40.

WEBER, L. M. and SONESON, C. (2019). HDCytoData: Collection of high-dimensional cytometry benchmark datasets in Bioconductor object formats. *F1000Res*. **8** 1459. https://doi.org/10.12688/f1000research.20210.2

WEBER, L. M., NOWICKA, M., SONESON, C. and ROBINSON, M. D. (2019). diffcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering. *Commun*. *Biol*. **2** 1–11.

YIP, S. H., WANG, P., KOCHER, J.-P. A., SHAM, P. C. and WANG, J. (2017). Linnorm: Improved statistical analysis for single cell RNA-seq expression data. *Nucleic Acids Res*. **45** e179–e179.

ZAPPIA, L., PHIPSON, B. and OSHLACK, A. (2017). Splatter: Simulation of single-cell RNA sequencing data. *Genome Biol*. **18** 1–15.

ZHANG, Y., ZHENG, L., ZHANG, L., HU, X., REN, X. and ZHANG, Z. (2019). Deep single-cell RNA sequencing data of individual T cells from treatment-naive colorectal cancer patients. *Sci*. *Data* **6** 1–15.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1701–1721
https://doi.org/10.1214/22-AOAS1691

THE RISK OF MATERNAL COMPLICATIONS AFTER CESAREAN DELIVERY: NEAR-FAR MATCHING FOR INSTRUMENTAL VARIABLES STUDY DESIGNS WITH LARGE OBSERVATIONAL DATASETS

BY RUOQI YU[1,a], RACHEL KELZ[2,b], SCOTT LORCH[3,d] AND LUKE J. KEELE[2,c]

[1]*Department of Statistics, University of California, Davis,* [a]*ruoqi.yu.ry@gmail.com*
[2]*Department of Surgery, University of Pennsylvania,* [b]*rachel.kelz@pennmedicine.upenn.edu,* [c]*luke.keele@gmail.com*
[3]*Department of Pediatrics, The Children's Hospital of Philadelphia,* [d]*LORCH@email.chop.edu*

Cesarean delivery is used when there are problems with the placenta or umbilical cord, for twin pregnancies, and breech births. However, research has found that Cesarean delivery increases the risk of maternal complications like blood transfusions and admission to the intensive care unit. Here, using an instrumental variables study design to reduce bias from unobserved confounders, we study whether Cesarean delivery increases the risk of maternal complications. We use a variant of matching—near-far matching—to render our study design more plausible. In a near-far match the investigator seeks to strengthen the effect of the instrument on the exposure while balancing observable characteristics between groups of subjects with low and high values of the instrument. Extant near-far matching methods are computationally intensive for large data sets, and computing time can be very lengthy. To reduce the computational complexity of near-far matching in large observational studies, we apply an iterative form of Glover's algorithm for a doubly convex bipartite graph to determine an optimal reverse caliper for the instrument which reduces the number of candidate matches and allows for an optimal match in a large but much sparser graph. We also incorporate a variety of balance constraints, including exact matching, fine and near-fine balance, and covariate balance prioritization. We illustrate this new matching method using medical claims data from Pennsylvania, New York, and Florida. In our application we match on physician's preferences for delivery via Cesarean section which is the instrument in our study. We compare the computing time from our match to extant methods, and we find that we can reduce the computational time required for the match by more than 11 hours. If our matched sample came from a paired randomized experiment, we could conclude that Cesarean delivery elevates the risk of maternal complications and increases the time spent in the hospital. Sensitivity analysis shows that the estimates for complications could be the result of a minor amount of confounding due to an unobserved covariate. The effects on the length of stay outcome, however, are more insensitive to hidden confounders.

REFERENCES

ABADIE, A. and GARDEAZABAL, J. (2003). The economic costs of conflict: A case study of the Basque country. *Am. Econ. Rev.* **93** 112–132.

ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.

ARORA, S. and BARAK, B. (2009). *Computational Complexity: A Modern Approach*. Cambridge Univ. Press, Cambridge. MR2500087 https://doi.org/10.1017/CBO9780511804090

ASCH, D. A., NICHOLSON, S., SRINIVAS, S., HERRIN, J. and EPSTEIN, A. J. (2009). Evaluating obstetrical residency programs using patient outcomes. *JAMA* **302** 1277–1283.

BAIOCCHI, M., CHENG, J. and SMALL, D. S. (2014). Instrumental variable methods for causal inference. *Stat. Med.* **33** 2297–2340. MR3257582 https://doi.org/10.1002/sim.6128

Key words and phrases. Instrumental variables, matching, causal inference.

BAIOCCHI, M., SMALL, D. S., LORCH, S. and ROSENBAUM, P. R. (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *J. Amer. Statist. Assoc.* **105** 1285–1296. MR2796550 https://doi.org/10.1198/jasa.2010.ap09490

BAIOCCHI, M., SMALL, D. S., YANG, L., POLSKY, D. and GROENEVELD, P. W. (2012). Near/far matching: A study design approach to instrumental variables. *Health Serv. Outcomes Res. Methodol.* **12** 237–253.

BERTSEKAS, D. P. (1998). *Network Optimization: Continuous and Discrete Models*. Athena Scientific, Belmont, MA.

BRANSON, Z. and KEELE, L. (2020). Evaluating a key instrumental variable assumption using randomization tests. *Am. J. Epidemiol.* **189** 1412–1420. https://doi.org/10.1093/aje/kwaa089

BROOKHART, M. A. and SCHNEEWEISS, S. (2007). Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. *Int. J. Biostat.* **3** Art. 14, 25. MR2383610 https://doi.org/10.2202/1557-4679.1072

BROOKHART, M. A., WANG, P., SOLOMON, D. H. and SCHNEEWEISS, S. (2006). Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* **17** 268.

BROOKS, J. M., CHRISCHILLES, E. A., SCOTT, S. D. and CHEN-HARDEE, S. S. (2003). Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. *Health Serv. Res.* **38** 1385–1402.

CLARK, S. L., FRYE, D. R., MEYERS, J. A., BELFORT, M. A., DILDY, G. A., KOFFORD, S., ENGLEBRIGHT, J. and PERLIN, J. A. (2010). Reduction in elective delivery at <39 weeks of gestation: Comparative effectiveness of 3 approaches to change and the impact on neonatal intensive care admission and stillbirth. *Am. J. Obstet. Gynecol.* **203** 449.e1–449.e6. https://doi.org/10.1016/j.ajog.2010.05.036

COCHRAN, W. G. and RUBIN, D. B. (1973). Controlling bias in observational studies: A review. *Sankhya, Ser. A* **35** 417–446.

CURTIN, S. C., GREGORY, K. D., KORST, L. M. and UDDIN, S. F. (2015). Maternal morbidity for vaginal and cesarean deliveries, according to previous cesarean history: New data from the birth certificate, 2013. National Vital Statistics Reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System **64** 1–13.

DAVIES, N. M. (2015). Commentary: An even clearer portrait of bias in observational studies? *Epidemiology* **26** 505.

DAVIES, N. M., SMITH, G. D., WINDMEIJER, F. and MARTIN, R. M. (2013). Cox-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: An instrumental variable analysis. *Epidemiology* **24** 352–362.

DAVIES, N. M., THOMAS, K. H., TAYLOR, A. E., TAYLOR, G. M., MARTIN, R. M., MUNAFÒ, M. R. and WINDMEIJER, F. (2017). How to compare instrumental variable and conventional regression analyses using negative controls and bias plots. *Int. J. Epidemiol.* **46** 2067–2077.

ERTEFAIE, A., SMALL, D. S. and ROSENBAUM, P. R. (2018). Quantitative evaluation of the trade-off of strengthened instruments and sample size in observational studies. *J. Amer. Statist. Assoc.* **113** 1122–1134. MR3862344 https://doi.org/10.1080/01621459.2017.1305275

FOGARTY, C. B., LEE, K., KELZ, R. R. and KEELE, L. J. (2021). Biased encouragements and heterogeneous effects in an instrumental variable study of emergency general surgical outcomes. *J. Amer. Statist. Assoc.* **116** 1625–1636. MR4353701 https://doi.org/10.1080/01621459.2020.1863220

GLOVER, F. (1967). Maximum matching in a convex bipartite graph. *Nav. Res. Logist. Q.* **14** 313–316.

HERNÁN, M. A. and ROBINS, J. M. (2006). Instruments for causal inference: An epidemiologists dream. *Epidemiology* **17** 360–372.

HIRAI, A. H., SAPPENFIELD, W. M., GHANDOUR, R. M., DONAHUE, S., LEE, V. and LU, M. C. (2018). The collaborative improvement and innovation network (CoIIN) to reduce infant mortality: An outcome evaluation from the US South, 2011 to 2014. *Am. J. Publ. Health* **108** 815–821.

HODGES, J. L. JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Stat.* **34** 598–611. MR0152070 https://doi.org/10.1214/aoms/1177704172

JACKSON, J. W. and SWANSON, S. A. (2015). Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology* **26** 498–504.

KANG, H., PECK, L. and KEELE, L. (2018). Inference for instrumental variables: A randomization inference approach. *J. Roy. Statist. Soc. Ser. A* **181** 1231–1254. MR3876390 https://doi.org/10.1111/rssa.12353

KANG, H., KREUELS, B., MAY, J. and SMALL, D. S. (2016). Full matching approach to instrumental variables estimation with application to the effect of malaria on stunting. *Ann. Appl. Stat.* **10** 335–364. MR3480499 https://doi.org/10.1214/15-AOAS894

KEELE, L. and MORGAN, J. W. (2016). How strong is strong enough? Strengthening instruments through matching and weak instrument tests. *Ann. Appl. Stat.* **10** 1086–1106. MR3528373 https://doi.org/10.1214/16-AOAS932

KEELE, L. J., SHAROKY, C. E., SELLERS, M. M., WIRTALLA, C. J. and KELZ, R. R. (2018). An instrumental variables design for the effect of emergency general surgery. *Epidemiol. Methods* **7**.

KEELE, L. J., ZHAO, Q., KELZ, R. R. and SMALL, D. S. (2019). Falsification tests for instrumental variable desings with an application to tendency to operate. *Med. Care* **57** 167–171.

KEELE, L., HARRIS, S., PIMENTEL, S. D. and GRIEVE, R. (2020). Stronger instruments and refined covariate balance in an observational study of the effectiveness of prompt admission to intensive care units. *J. Roy. Statist. Soc. Ser. A* **183** 1501–1521. MR4157823

KENNEDY, E. H., SJÖLANDER, A. and SMALL, D. S. (2015). Semiparametric causal inference in matched cohort studies. *Biometrika* **102** 739–746. MR3394290 https://doi.org/10.1093/biomet/asv025

KORB, D., GOFFINET, F., SECO, A., CHEVRET, S., DENEUX-THARAUX, C., GROUP, E. S. et al. (2019). Risk of severe maternal morbidity associated with cesarean delivery and the role of maternal age: A population-based propensity score analysis. *CMAJ, Can. Med. Assoc. J.* **191** E352–E360.

KORTE, B. and VYGEN, J. (2012). *Combinatorial Optimization: Theory and Algorithms*, 5th ed. *Algorithms and Combinatorics* **21**. Springer, Heidelberg. MR2850465 https://doi.org/10.1007/978-3-642-24488-9

LAVENDER, T., HOFMEYR, G. J., NEILSON, J. P., KINGDON, C. and GYTE, G. M. L. (2012). Caesarean section for non-medical reasons at term. *Cochrane Database Syst. Rev.* **3** CD004660. https://doi.org/10.1002/14651858.CD004660.pub3

LEONARD, S. A., MAIN, E. K. and CARMICHAEL, S. L. (2019). The contribution of maternal characteristics and cesarean delivery to an increasing trend of severe maternal morbidity. *BMC Pregnancy Childbirth* **19** 1–9.

LIU, S., LISTON, R. M., JOSEPH, K., HEAMAN, M., SAUVE, R., KRAMER, M. S. et al. (2007). Maternal mortality and severe morbidity associated with low-risk planned cesarean delivery versus planned vaginal delivery at term. *CMAJ, Can. Med. Assoc. J.* **176** 455–460.

LORCH, S. A., BAIOCCHI, M., AHLBERG, C. E. and SMALL, D. S. (2012). The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics* **130** 270–278.

LU, B., ZANUTTO, E., HORNIK, R. and ROSENBAUM, P. R. (2001). Matching with doses in an observational study of a media campaign against drug abuse. *J. Amer. Statist. Assoc.* **96** 1245–1253. MR1973668 https://doi.org/10.1198/016214501753381896

MAIN, E. K., MOORE, D., FARRELL, B., SCHIMMEL, L. D., ALTMAN, R. J., ABRAHAMS, C., BLISS, M. C., POLIVY, L. and STERLING, J. (2006). Is there a useful cesarean birth measure? Assessment of the nulliparous term singleton vertex cesarean birth rate as a tool for obstetric quality improvement. *Am. J. Obstet. Gynecol.* **194** 1644–1651.

MCCLELLAN, M., MCNEIL, B. J. and NEWHOUSE, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: Analysis using instrumental variables. *JAMA* **272** 859–866.

OSHIRO, B. T., KOWALEWSKI, L., SAPPENFIELD, W., ALTER, C. C., BETTEGOWDA, V. R., RUSSELL, R., CURRAN, J., REEVES, L., KACICA, M. et al. (2013). A multistate quality improvement program to decrease elective deliveries before 39 weeks of gestation. *Obstet. Gynecol.* **121** 1025–1031.

ROSENBAUM, P. R. (1989). Optimal matching for observational studies. *J. Amer. Statist. Assoc.* **84** 1024–1032.

ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138 https://doi.org/10.1007/978-1-4757-3692-2

ROSENBAUM, P. R. (2010). *Design of Observational Studies*. *Springer Series in Statistics*. Springer, New York. MR2561612 https://doi.org/10.1007/978-1-4419-1213-8

ROSENBAUM, P. R. (2012). Optimal matching of an optimally chosen subset in observational studies. *J. Comput. Graph. Statist.* **21** 57–71. MR2913356 https://doi.org/10.1198/jcgs.2011.09219

RUBIN, D. B. (1980). Bias reduction using Mahalanobis-metric matching. *Biometrics* **36** 293–298.

SHETTY, K. D., VOGT, W. B. and BHATTACHARYA, J. (2009). Hormone replacement therapy and cardiovascular health in the United States. *Med. Care* 600–605.

SMALL, D. S. and ROSENBAUM, P. R. (2008). War and wages: The strength of instrumental variables and their sensitivity to unobserved biases. *J. Amer. Statist. Assoc.* **103** 924–933. MR2528819 https://doi.org/10.1198/016214507000001247

SMALL, D. S., TAN, Z., RAMSAHAI, R. R., LORCH, S. A. and BROOKHART, M. A. (2017). Instrumental variable estimation with a stochastic monotonicity assumption. *Statist. Sci.* **32** 561–579. MR3730522 https://doi.org/10.1214/17-STS623

STOCK, J. H. and TREBBI, F. (2003). Retrospectives: Who invented instrumental variable regression? *J. Econ. Perspect.* **17** 177–194.

STOCK, J. H. and YOGO, M. (2005). Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models* 80–108. Cambridge Univ. Press, Cambridge. MR2232140 https://doi.org/10.1017/CBO9780511614491.006

SWANSON, S. A. and HERNÁN, M. A. (2014). Think globally, act globally: An epidemiologist's perspective on instrumental variable estimation [discussion of MR3264545]. *Statist. Sci.* **29** 371–374. MR3264549 https://doi.org/10.1214/14-STS491

SWANSON, S. A. and HERNÁN, M. A. (2017). The challenging interpretation of instrumental variable estimates under monotonicity. *Int. J. Epidemiol.* **47** 1289–1297.

TAN, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. *J. Amer. Statist. Assoc.* **101** 1607–1618. MR2279483 https://doi.org/10.1198/016214505000001366

VANSTEELANDT, S., BOWDEN, J., BABANEZHAD, M. and GOETGHEBEUR, E. (2011). On instrumental variables estimation of causal odds ratios. *Statist. Sci.* **26** 403–422. MR2917963 https://doi.org/10.1214/11-STS360

WALD, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.* **11** 285–300. MR0002739 https://doi.org/10.1214/aoms/1177731868

WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. MR2768559

YANG, D., SMALL, D. S., SILBER, J. H. and ROSENBAUM, P. R. (2012). Optimal matching with minimal deviation from fine balance in a study of obesity and surgical outcomes. *Biometrics* **68** 628–636. MR2959630 https://doi.org/10.1111/j.1541-0420.2011.01691.x

YANG, F., ZUBIZARRETA, J. R., SMALL, D. S., LORCH, S. and ROSENBAUM, P. R. (2014). Dissonant conclusions when testing the validity of an instrumental variable. *Amer. Statist.* **68** 253–263. MR3280612 https://doi.org/10.1080/00031305.2014.962764

YU, R. and ROSENBAUM, P. R. (2019). Directional penalties for optimal matching in observational studies. *Biometrics* **75** 1380–1390. MR4041838 https://doi.org/10.1111/biom.13098

YU, R., SILBER, J. H. and ROSENBAUM, P. R. (2020). Matching methods for observational studies derived from large administrative databases. *Statist. Sci.* **35** 338–355. MR4148206 https://doi.org/10.1214/19-STS699

YU, R., KELZ, R., LORCH, S. and KEELE, L. J (2023). Supplement to "The risk of maternal complications after Cesarean delivery: Near-far matching for instrumental variables study designs with large observational datasets." https://doi.org/10.1214/22-AOAS1691SUPPA, https://doi.org/10.1214/22-AOAS1691SUPPB

ZHANG, B., HENG, S., MACKAY, E. J. and YE, T. (2021). Bridging preference-based instrumental variable studies and cluster-randomized encouragement experiments: Study design, noncompliance, and average cluster effect ratio. *Biometrics*. https://doi.org/10.1111/biom.13500

ZHAO, Q. (2019). On sensitivity value of pair-matched observational studies. *J. Amer. Statist. Assoc.* **114** 713–722. MR3963174 https://doi.org/10.1080/01621459.2018.1429277

ZUBIZARRETA, J. R., SMALL, D. S., GOYAL, N. K., LORCH, S. and ROSENBAUM, P. R. (2013). Stronger instruments via integer programming in an observational study of late preterm birth outcomes. *Ann. Appl. Stat.* **7** 25–50. MR3086409 https://doi.org/10.1214/12-AOAS582

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1722–1739
https://doi.org/10.1214/22-AOAS1692

INTEGRATING MULTIPLE BUILT ENVIRONMENT DATA SOURCES

BY JUNG YEON WON[1,a], MICHAEL R. ELLIOTT[1,b], EMMA V. SANCHEZ-VAZNAUGH[2,c] AND BRISA N. SÁNCHEZ[3,d]

[1]*Department of Biostatistics, University of Michigan,* [a]*jungywon@umich.edu,* [b]*mrelliot@umich.edu*
[2]*Department of Health Education, San Francisco State University,* [c]*emmav@sfsu.edu*
[3]*Department of Epidemiology and Biostatistics, Drexel University,* [d]*bns48@drexel.edu*

Studies examining the contribution of the built environment to health often rely on commercial data sources to derive exposure measures, such as the number of specific food outlets in study participants' neighborhoods. Data on the location of community amenities (e.g., food outlets) can be collected from multiple sources. However, these commercial listings are known to have ascertainment errors and thus provide conflicting claims about the number and location of amenities. We propose a method that integrates exposure measures from different databases, while accounting for ascertainment errors, and obtains unbiased health effects of latent exposure. We frame the problem of conflicting exposure measures as a problem of two contingency tables with partially known margins, with the entries of the tables modeled using a multinomial distribution. Available estimates of source quality were embedded in a joint model for observed exposure counts, latent exposures, and health outcomes. Simulations show that our modeling framework yields substantially improved inferences regarding the health effects. We used the proposed method to estimate the association between children's body mass index (BMI) and the concentration of food outlets near their schools when both the NETS and Reference USA databases are available.

REFERENCES

ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII—1983*. *Lecture Notes in Math*. **1117** 1–198. Springer, Berlin. MR0883646 https://doi.org/10.1007/BFb0099421

ATHENS, J. K., DUNCAN, D. T. and ELBEL, B. (2016). Proximity to fast-food outlets and supermarkets as predictors of fast-food dining frequency. *Journal of the Academy of Nutrition and Dietetics* **116** 1266–1275.

BARTOLUCCI, F., PANDOLFI, S. and PENNONI, F. (2022). Discrete latent variable models. *Annu. Rev. Stat. Appl.* **9** 425–452. MR4394915 https://doi.org/10.1146/annurev-statistics-040220-091910

CALIFORNIA DEPARTMENT OF EDUCATION (2019). Physical Fitness Testing (PFT). Available at http://www.cde.ca.gov/ta/tg/pf/.

CASPI, C. E. and FRIEBUR, R. (2016). Modified ground-truthing: An accurate and cost-effective food environment validation method for town and rural areas. *Int. J. Behav. Nutr. Phys. Act.* **13** 37. https://doi.org/10.1186/s12966-016-0360-3

DONG, X. L., BERTI-EQUILLE, L. and SRIVASTAVA, D. (2009). Integrating conflicting data: The role of source dependence. *Proc. VLDB Endow.* **2** 550–561.

DORAZIO, R. M., MUKHERJEE, B., ZHANG, L., GHOSH, M., JELKS, H. L. and JORDAN, F. (2008). Modeling unobserved sources of heterogeneity in animal abundance using a Dirichlet process prior. *Biometrics* **64** 635–644. MR2432438 https://doi.org/10.1111/j.1541-0420.2007.00873.x

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

FIENBERG, S. E. (1972a). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika* **59** 591–603. MR0383619 https://doi.org/10.1093/biomet/59.3.591

FIENBERG, S. E. (1972b). The analysis of incomplete multi-way contingency tables. *Biometrics* **28** 177–202.

FLEGAL, K. M., WEI, R., OGDEN, C. L., FREEDMAN, D. S., JOHNSON, C. L. and CURTIN, L. R. (2009). Characterizing extreme values of body mass index–for-age by using the 2000 Centers for Disease Control and Prevention growth charts. *Am. J. Clin. Nutr.* **90** 1314–1320.

Key words and phrases. Built-environment, count exposure, data integration, measurement error, Dirichlet process mixture model, commercial business lists.

Fleischhacker, S. E., Evenson, K. R., Sharkey, J., Pitts, S. B. J. and Rodriguez, D. A. (2013). Validity of secondary retail food outlet data: A systematic review. *Am. J. Prev. Med.* **45** 462–473.

Ghosal, S. (2010). The Dirichlet process, related priors and posterior asymptotics. *Bayesian Nonparametrics* **28** 35.

Grafova, I. B. (2008). Overweight children: Assessing the contribution of the built environment. *Prev. Med.* **47** 304–308.

Hougaard, P., Lee, M.-L. T. and Whitmore, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics* **53** 1225–1238. MR1614370 https://doi.org/10.2307/2533492

Howard, P. H., Fitzpatrick, M. and Fulfrost, B. (2011). Proximity of food retailers to schools and rates of overweight ninth grade students: An ecological study in California. *BMC Public Health* **11** 1–8.

InfoUSA (2012). Infousa Business Listing Description. Available at https://www.Infousa.Com/Product/Business-Lists/.

Joe, H. and Zhu, R. (2005). Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial distribution. *Biom. J.* **47** 219–229. MR2137236 https://doi.org/10.1002/bimj.200410102

Jones, K. K., Zenk, S. N., Tarlov, E., Powell, L. M., Matthews, S. A. and Horoi, I. (2017). A step-by-step approach to improve data quality when using commercial business lists to characterize retail food environments. *BMC Res. Notes* **10** 1–12.

Larson, N. I., Story, M. T. and Nelson, M. C. (2009). Neighborhood environments: Disparities in access to healthy foods in the US. *Am. J. Prev. Med.* **36** 74–81.

Lebel, A., Daepp, M. I., Block, J. P., Walker, R., Lalonde, B., Kestens, Y. and Subramanian, S. (2017). Quantifying the foodscape: A systematic review and meta-analysis of the validity of commercially available business data. *PLoS ONE* **12** e0174417.

Lee, H. (2012). The role of local food availability in explaining obesity risk among young school-aged children. *Social Science & Medicine* **74** 1193–1203.

Levin, B. (1981). A representation for multinomial cumulative distribution functions. *Ann. Statist.* **9** 1123–1126. MR0628769

Liese, A. D., Barnes, T. L., Lamichhane, A. P., Hibbert, J. D., Colabianchi, N. and Lawson, A. B. (2013). Characterizing the food retail environment: Impact of count, type, and geospatial error in 2 secondary data sources. *Journal of Nutrition Education and Behavior* **45** 435–442.

Lucan, S. C., Maroko, A. R., Bumol, J., Torrens, L., Varona, M. and Berke, E. M. (2013). Business list vs ground observation for measuring a food environment: Saving time or waste of time (or worse)? *Journal of the Academy of Nutrition and Dietetics* **113** 1332–1339.

Manrique-Vallier, D. (2016). Bayesian population size estimation using Dirichlet process mixtures. *Biometrics* **72** 1246–1254. MR3591609 https://doi.org/10.1111/biom.12502

Miller, J. W. and Harrison, M. T. (2014). Inconsistency of Pitman–Yor process mixtures for the number of components. *J. Mach. Learn. Res.* **15** 3333–3370. MR3277163

Miller, J. W. and Harrison, M. T. (2018). Mixture models with a prior on the number of components. *J. Amer. Statist. Assoc.* **113** 340–356. MR3803469 https://doi.org/10.1080/01621459.2016.1255636

Must, A. and Anderson, S. (2006). Body mass index in children and adolescents: Considerations for population-based applications. *Int. J. Obes.* **30** 590–594.

Nelder, J. A. and Lee, Y. (1992). Likelihood, quasi-likelihood and pseudolikelihood: Some comparisons. *J. Roy. Statist. Soc. Ser. B* **54** 273–284. MR1157725

NETS (2021). Business Dynamics Research Consortium, National Establishment Time-Series (NETS) Database: Database Description. Available at http://exceptionalgrowth.org.

Pollock, K. H. and Otto, M. C. (1983). Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics* **39** 1035–1049.

Powell, L. M., Han, E., Zenk, S. N., Khan, T., Quinn, C. M., Gibbs, K. P., Pugach, O., Barker, D. C., Resnick, E. A. et al. (2011). Field validation of secondary commercial data sources on the retail food outlet environment in the U.S. *Health Place* **17** 1122–1131.

R Core Team (2021). R: A language and environment for statistical computing.

Richardson, S. and Gilks, W. R. (1993a). A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *Am. J. Epidemiol.* **138** 430–442.

Richardson, S. and Gilks, W. R. (1993b). Conditional independence models for epidemiological studies with covariate measurement error. *Stat. Med.* **12** 1703–1722.

Ross, G. J. and Markwick, D. (2020). dirichletprocess: Build Dirichlet Process Objects for Bayesian Modelling. R package version 0.4.0.

Teh, Y. W. (2010). Dirichlet process. In *In Encyclopedia of Machine Learning* 280–287. Springer, Berlin.

VAN SMEDEN, M., LASH, T. L. and GROENWOLD, R. H. (2020). Reflection on modern methods: Five myths about measurement error in epidemiological research. *Int. J. Epidemiol*. **49** 338–347.

WANG, N., CARROLL, R. and LIANG, K.-Y. (1996). Quasilikelihood estimation in measurement error models with correlated replicates. *Biometrics* 401–411.

WON, J. Y., ELLIOTT, M. R., SANCHEZ-VAZNAUGH, E. V. and SÁNCHEZ, B. N. (2023). Supplement to "Integrating Multiple Built Environment Data Sources." https://doi.org/10.1214/22-AOAS1692SUPPA, https://doi.org/10.1214/22-AOAS1692SUPPB

ZHAO, B., RUBINSTEIN, B. I. P., GEMMELL, J. and HAN, J. (2012). A Bayesian approach to discovering truth from conflicting sources for data integration. *Proc. VLDB Endow.* **5** 550–561.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1740–1763
https://doi.org/10.1214/22-AOAS1693

MARGINALLY CALIBRATED RESPONSE DISTRIBUTIONS FOR END-TO-END LEARNING IN AUTONOMOUS DRIVING

BY CLARA HOFFMANN[a] AND NADJA KLEIN[b]

Chair of Statistics and Data Science, Humboldt-Universität zu Berlin, [a]*clara.c.hoffmann@gmail.com,* [b]*nadja.klein@hu-berlin.de*

End-to-end learners for autonomous driving are deep neural networks that predict the instantaneous steering angle directly from images of the street ahead. These learners must provide reliable uncertainty estimates for their predictions in order to meet safety requirements and to initiate a switch to manual control in areas of high uncertainty. However, end-to-end learners typically only deliver point predictions, since distributional predictions are associated with large increases in training time or additional computational resources during prediction. To address this shortcoming, we investigate efficient and scalable approximate inference for the deep distributional model of Klein, Nott and Smith (*J. Comput. Graph. Statist.* **30** (2021) 467–483) in order to quantify uncertainty for the predictions of end-to-end learners. A special merit of this model, which we refer to as implicit copula neural linear model (IC-NLM), is that it produces densities for the steering angle that are marginally calibrated, that is, the average of the estimated densities equals the empirical distribution of steering angles. To ensure the scalability to large n regimes, we develop efficient estimation based on variational inference as a fast alternative to computationally intensive, exact inference via Hamiltonian Monte Carlo. We demonstrate the accuracy and speed of the variational approach on two end-to-end learners trained for highway driving using the comma2k19 dataset. The IC-NLM is competitive with other established uncertainty quantification methods for end-to-end learning in terms of nonprobabilistic predictive performance and outperforms them in terms of marginal calibration for in-distribution prediction. Our proposed approach also allows the identification of overconfident learners and contributes to the explainability of black-box end-to-end learners by using the predictive densities to understand which steering actions the learner sees as valid.

REFERENCES

ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J. et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

ABBASI, S., HAJABDOLLAHI, M., KARIMI, N. and SAMAVI, S. (2020). Modeling teacher-student techniques in deep neural networks for knowledge distillation. In 2020 *International Conference on Machine Vision and Image Processing* (*MVIP*) 1–6. https://doi.org/10.1109/MVIP49855.2020.9116923

AMINI, A., SOLEIMANY, A., KARAMAN, S. and RUS, D. (2019). Spatial uncertainty sampling for end-to-end control. 1–5. arXiv:1805.04829.

ARNEZ, F., ESPINOZA, H., RADERMACHER, A. and TERRIER, F. (2020). A Comparison of uncertainty estimation approaches in deep learning components for autonomous vehicle applications. arXiv:2006.15172.

BISHOP, C. M. (1994). Mixture density networks. Unpublished.

BLUNDELL, C., CORNEBISE, J., KAVUKCUOGLU, K. and WIERSTRA, D. (2015). Weight uncertainty in neural networks. In *Proceedings of the* 32*nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.). *Proceedings of Machine Learning Research* **37** 1613–1622. PMLR, Lille, France.

BOJARSKI, M., TESTA, D. D., DWORAKOWSKI, D., FIRNER, B., FLEPP, B., GOYAL, P., JACKEL, L. D., MONFORT, M., MULLER, U. et al. (2016). End to end learning for self-driving cars. 1–4,6. arXiv:1604.07316.

Key words and phrases. Calibration, deep neural network, implicit copula, neural linear models, probabilistic forecasting, uncertainty quantification, variational inference.

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751 https://doi.org/10.1093/biomet/asq017

CHEN, C., SEFF, A., KORNHAUSER, A. and XIAO, J. (2015). DeepDriving: Learning affordance for direct perception in autonomous driving. 2722–2730.

CHI, L. and MU, Y. (2017). Learning end-to-end autonomous steering model from spatial and temporal visual cues. In *Proceedings of the Workshop on Visual Analysis in Smart and Connected Communities*. *VSCC '17* 9–16. Association for Computing Machinery, New York. https://doi.org/10.1145/3132734.3132737

CHOLLET, F. et al. (2015). Keras. https://keras.io.

CRAIU, V. R. and SABETI, A. (2012). In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes. *J. Multivariate Anal.* **110** 106–120. MR2927512 https://doi.org/10.1016/j.jmva.2012.03.010

DOSOVITSKIY, A., BEYER, L., KOLESNIKOV, A., WEISSENBORN, D., ZHAI, X., UNTERTHINER, T., DEHGHANI, M., MINDERER, M., HEIGOLD, G. et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations*, *ICLR* 2021.

GAL, Y. and GHAHRAMANI, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.). *Proceedings of Machine Learning Research* **48** 1050–1059. PMLR, New York.

GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007a). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. MR2325275 https://doi.org/10.1111/j.1467-9868.2007.00587.x

GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007b). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. MR2325275 https://doi.org/10.1111/j.1467-9868.2007.00587.x

GNEITING, T. and RANJAN, R. (2013). Combining predictive distributions. *Electron. J. Stat.* **7** 1747–1782. MR3080409 https://doi.org/10.1214/13-EJS823

GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

GUNAWAN, D., KOHN, R. and NOTT, D. (2021). Flexible variational bayes based on a copula of a mixture of normals.

GUO, C., PLEISS, G., SUN, Y. and WEINBERGER, K. Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning* (D. Precup and Y. W. Teh, eds.). *Proceedings of Machine Learning Research* **70** 1321–1330. PMLR.

HE, B., LAKSHMINARAYANAN, B. and TEH, Y. W. (2020). Bayesian deep ensembles via the neural tangent kernel. In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) **33** 1010–1022. Curran Associates, Red Hook, NY.

HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14** 1303–1347. MR3081926

HOFFMANN, C. and KLEIN, N. (2023). Supplement to "Marginally calibrated response distributions for end-to-end learning in autonomous driving." https://doi.org/10.1214/22-AOAS1693SUPPA, https://doi.org/10.1214/22-AOAS1693SUPPB

KENDALL, A. and GAL, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) **30** 5574–5584. Curran Associates, Red Hook, NY.

KINGMA, D. P. and WELLING, M. (2014). Auto-encoding variational Bayes. In *2nd International Conference on Learning Representations* (Y. Bengio and Y. LeCun, eds.).

KLEIN, N. and KNEIB, T. (2016a). Simultaneous inference in structured additive conditional copula regression models: A unifying Bayesian approach. *Stat. Comput.* **26** 841–860. MR3515025 https://doi.org/10.1007/s11222-015-9573-6

KLEIN, N. and KNEIB, T. (2016b). Scale-dependent priors for variance parameters in structured additive distributional regression. *Bayesian Anal.* **11** 1071–1106. MR3545474 https://doi.org/10.1214/15-BA983

KLEIN, N., NOTT, D. J. and SMITH, M. S. (2021). Marginally calibrated deep distributional regression. *J. Comput. Graph. Statist.* **30** 467–483. MR4270517 https://doi.org/10.1080/10618600.2020.1807996

KLEIN, N. and SMITH, M. S. (2019). Implicit copulas from Bayesian regularized regression smoothers. *Bayesian Anal.* **14** 1143–1171. MR4044849 https://doi.org/10.1214/18-BA1138

KLEIN, N., SMITH, M. S. and NOTT, D. J. (2021). Deep distributional time series models and the probabilistic forecasting of intraday electricity prices. arXiv:2010.0184.

KULESHOV, V., FENNER, N. and ERMON, S. (2018). Accurate uncertainties for deep learning using calibrated regression. In *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 2796–2804. PMLR.

LAKSHMINARAYANAN, B., PRITZEL, A. and BLUNDELL, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) **30**. Curran Associates, Red Hook, NY.

LIN, S.-C., ZHANG, Y., HSU, C.-H., SKACH, M., HAQUE, M. E., TANG, L. and MARS, J. (2018). The architectural implications of autonomous driving: Constraints and acceleration. In *Proceedings of the Twenty-Third International Conference on Architectural Support for Programming Languages and Operating Systems*. *ASPLOS '18* 751–766. Association for Computing Machinery, New York.

MARTIN, C. and DUHAIME, D. (2020). keras-mdn-layer. Published with MIT license.

MICHELMORE, R., KWIATKOWSKA, M. and GAL, Y. (2018). Evaluating uncertainty quantification in end-to-end autonomous driving control. 1–5. arXiv:1811.06817.

MILLER, A., FOTI, N., D'AMOUR, A. and ADAMS, R. P. (2017). Reducing reparameterization gradient variance. In *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) **30** 3–10. Curran Associates, Red Hook, NY.

NAIR, V. and HINTON, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the* 27*th International Coference on International Conference on Machine Learning*. *ICML'*10 807–814. Omnipress, Madison, WI.

NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds.) 113–162. CRC Press, Boca Raton, FL. MR2858447

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2197664 https://doi.org/10.1007/s11229-005-3715-x

NOTT, D. J., TAN, S. L., VILLANI, M. and KOHN, R. (2012). Regression density estimation with variational methods and stochastic approximation. *J. Comput. Graph. Statist.* **21** 797–820. MR2970920 https://doi.org/10.1080/10618600.2012.679897

OBER, S. W. and RASMUSSEN, C. E. (2019). Benchmarking the neural linear model for regression. arXiv:1912.08416.

ONG, V. M.-H., NOTT, D. J. and SMITH, M. S. (2018). Gaussian variational approximation with a factor covariance structure. *J. Comput. Graph. Statist.* **27** 465–478. MR3863750 https://doi.org/10.1080/10618600.2017.1390472

ORMEROD, J. T. and WAND, M. P. (2010). Explaining variational approximations. *Amer. Statist.* **64** 140–153. MR2757005 https://doi.org/10.1198/tast.2010.09058

PEARCE, T., BRINTRUP, A., ZAKI, M. and NEELY, A. (2018). High-quality prediction intervals for deep learning: A distribution-free, ensembled approach. In *Proceedings of the* 35*th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 4088. PMLR.

PIIRONEN, J. and VEHTARI, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron. J. Stat.* **11** 5018–5051. MR3738204 https://doi.org/10.1214/17-EJS1337SI

PINSLER, R., GORDON, J., NALISNICK, E. and HERNÁNDEZ-LOBATO, J. M. (2019). Bayesian batch active learning as sparse subset approximation. In *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.) **32** 7–8. Curran Associates, Red Hook, NY.

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

POLSON, N. G. and SOKOLOV, V. (2017). Deep learning: A Bayesian perspective. *Bayesian Anal.* **12** 1275–1304. MR3724986 https://doi.org/10.1214/17-BA1082

RACINE, J. S. (2008). Nonparametric econometrics: A primer. *Found Trends Econom.* **3** 1–88. https://doi.org/10.1561/0800000009

REZENDE, D. J., MOHAMED, S. and WIERSTRA, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the* 31*st International Conference on Machine Learning* (E. P. Xing and T. Jebara, eds.) **32** 1278–1286.

RIQUELME, C., TUCKER, G. and SNOEK, J. (2018). Deep Bayesian bandits showdown: An empirical comparison of Bayesian deep networks for Thompson sampling. 4. arXiv:1802.09127.

RODRIGUES, F. and PEREIRA, F. C. (2020). Beyond expectation: Deep joint mean and quantile regression for spatiotemporal problems. *IEEE Trans. Neural Netw. Learn. Syst.* **31** 5377–5389. https://doi.org/10.1109/TNNLS.2020.2966745

RUMELHART, D., HINTON, G. E. and WILLIAMS, R. J. (1986). Learning representations by back-propagating errors. *Nature* **323** 533–536.

SCHAFER, H., SANTANA, E., HADEN, A. and BIASINI, R. (2018). A commute in data: The comma2k19 dataset. 1. arXiv:1812.05752.

SKLAR, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **8** 229–231. MR0125600

SMITH, M. S. and KLEIN, N. (2021). Bayesian inference for regression copulas. *J. Bus. Econom. Statist.* **39** 712–728. MR4272930 https://doi.org/10.1080/07350015.2020.1721295

SMITH, M. S. and LOAIZA-MAYA, R. (2021). Implicit copula variational inference.

SMITH, M. S., LOAIZA-MAYA, R. and NOTT, D. J. (2020). High-dimensional copula variational approximation through transformation. *J. Comput. Graph. Statist.* **29** 729–743. MR4191239 https://doi.org/10.1080/10618600.2020.1740097

SNOEK, J., RIPPEL, O., SWERSKY, K., KIROS, R., SATISH, N., SUNDARAM, N., PATWARY, M., PRABHAT, M. and ADAMS, R. (2015). Scalable Bayesian optimization using deep neural networks. In *Proceedings of the* 32*nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.). *Proceedings of Machine Learning Research* **37** 2171–2176. PMLR, Lille, France.

SONG, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scand. J. Stat.* **27** 305–320. MR1777506 https://doi.org/10.1111/1467-9469.00191

SONG, P. X.-K., LI, M. and YUAN, Y. (2009). Joint regression analysis of correlated data using Gaussian copulas. *Biometrics* **65** 60–68. MR2665846 https://doi.org/10.1111/j.1541-0420.2008.01058.x

TITSIAS, M. and LÁZARO-GREDILLA, M. (2014). Doubly stochastic variational Bayes for non-conjugate inference. In *Proceedings of the* 31*st International Conference on Machine Learning* (E. P. Xing and T. Jebara, eds.) **32** 1971–1979.

TRAN, M.-N., NGUYEN, N., NOTT, D. and KOHN, R. (2020). Bayesian deep net GLM and GLMM. *J. Comput. Graph. Statist.* **29** 97–113. MR4085867 https://doi.org/10.1080/10618600.2019.1637747

URIA, B., MURRAY, I. and LAROCHELLE, H. (2013). RNADE: The real-valued neural autoregressive density-estimator. In *Advances in Neural Information Processing Systems* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) **26** 2175–2183. Curran Associates, Red Hook, NY.

WILSON, A. G. and GHAHRAMANI, Z. (2010). Copula processes. In *Advances in Neural Information Processing Systems* (J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel and A. Culotta, eds.) **23**. Curran Associates, Red Hook, NY.

XU, H., GAO, Y., YU, F. and DARRELL, T. (2017). End-to-end learning of driving models from large-scale video datasets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*) 1,7–8.

ZEILER, M. D. (2012). ADADELTA: An adaptive learning rate method. 3–4. arXiv:1212.5701.

ZHANG, Z., DALCA, A. V. and SABUNCU, M. R. (2019). Confidence calibration for convolutional neural networks using structured dropout. arXiv:1906.09551.

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1764–1781
https://doi.org/10.1214/22-AOAS1695

LEVERAGING HARDY–WEINBERG DISEQUILIBRIUM FOR ASSOCIATION TESTING IN CASE-CONTROL STUDIES

BY LIN ZHANG[1,a], LISA J. STRUG[1,2,b] AND LEI SUN[1,3,c]

[1]*Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto,* [a]*linzhang@utstat.toronto.edu*
[2]*Department of Computer Science, Faculty of Arts and Science, University of Toronto,* [b]*lisa.strug@utoronto.ca*
[3]*Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto,* [c]*sun@utstat.toronto.edu*

Modern genome-wide association studies (GWAS) remove single nucleotide polymorphisms (SNPs) that are in Hardy–Weinberg disequilibrium (HWD), despite limited rigor for this practice. In a case-control GWAS, although HWD in the control sample is an evidence for genotyping error, a truly associated SNP may be in HWD in the case and/or control populations. We, therefore, develop a new case-control association test that: (i) leverages HWD attributed to true association to increase power, (ii) is robust to HWD caused by genotyping error, and (iii) is easy-to-implement at the genome-wide level. The proposed robust *allele-based* joint test incorporates the *difference* in HWD between the case and control samples into the traditional association measure to gain power. We provide the asymptotic distribution of the proposed test statistic under the null hypothesis. We evaluate its type 1 error control at the genome-wide significance level of 5×10^{-8} in the presence of HWD attributed to factors unrelated to phenotype-genotype association, such as genotyping error. Finally, we demonstrate that the power of the proposed allele-based joint test is higher than the standard association test for a variety of genetic models, through derivations of the noncentrality parameters of the tests, as well as simulation and application studies.

REFERENCES

ANDERSON, C. A., PETTERSSON, F. H., CLARKE, G. M., CARDON, L. R., MORRIS, A. P. and ZONDERVAN, K. T. (2010). Data quality control in genetic case-control association studies. *Nat. Protoc.* **5** 1564–1573. https://doi.org/10.1038/nprot.2010.116

APPLE, R. J., ERLICH, H. A., KLITZ, W., MANOS, M. M., BECKER, T. M. and WHEELER, C. M. (1994). HLA DR–DQ associations with cervical carcinoma show papillomavirus–type specificity. *Nat. Genet.* **6** 157–162.

BUNIELLO, A., MACARTHUR, J. A. L., CEREZO, M., HARRIS, L. W., HAYHURST, J., MALANGONE, C., MCMAHON, A., MORALES, J., MOUNTJOY, E. et al. (2019). The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47** D1005–D1012.

BYCROFT, C., FREEMAN, C., PETKOVA, D., BAND, G., ELLIOTT, L. T., SHARP, K., MOTYER, A., VUKCEVIC, D., DELANEAU, O. et al. (2018). The UK biobank resource with deep phenotyping and genomic data. *Nature* **562** 203–209.

CHEN, C.-F. (1983). Score tests for regression models. *J. Amer. Statist. Assoc.* **78** 158–161.

CHEN, B., CRAIU, R. V., STRUG, L. J. and SUN, L. (2021). The X factor: A robust and powerful approach to X-chromosome-inclusive whole-genome association studies. *Genet. Epidemiol.* **45** 694–709. https://doi.org/10.1002/gepi.22422

DERKACH, A., LAWLESS, J. F. and SUN, L. (2014). Pooled association tests for rare genetic variants: A review and some new results. *Statist. Sci.* **29** 302–321. MR3264544 https://doi.org/10.1214/13-STS456

DEVLIN, B. and ROEDER, K. (1999). Genomic control for association studies. *Biometrics* **55** 997–1004.

DUDBRIDGE, F. and GUSNANTO, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32** 227–234.

FREIDLIN, B., ZHENG, G., LI, Z. and GASTWIRTH, J. L. (2002). Trend tests for case-control studies of genetic markers: Power, sample size and robustness. *Hum. Hered.* **53** 146–152.

Key words and phrases. Case-control studies, association tests, Hardy–Weinberg equilibrium.

GONG, J., WANG, F., XIAO, B., PANJWANI, N., LIN, F., KEENAN, K., AVOLIO, J., ESMAEILI, M., ZHANG, L. et al. (2019). Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. *PLoS Genet*. **15** e1008007.

GONZÁLEZ, J. R., CARRASCO, J. L., DUDBRIDGE, F., ARMENGOL, L., ESTIVILL, X. and MORENO, V. (2008). Maximizing association statistics over genetic models. *Genet*. *Epidemiol*. **32** 246–254.

LI, M. and LI, C. (2008). Assessing departure from Hardy–Weinberg equilibrium in the presence of disease association. *Genet*. *Epidemiol*. **32** 589–599.

LIN, Y.-C., BROOKS, J. D., BULL, S. B., GAGNON, F., GREENWOOD, C. M., HUNG, R. J., LAWLESS, J., PATERSON, A. D., SUN, L. et al. (2020). Statistical power in Covid-19 case-control host genomic study design. *Gen*. *Med*. **12** 1–8.

LIU, Y. and XIE, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Amer. Statist*. *Assoc*. **115** 393–402. MR4078471 https://doi.org/10.1080/01621459.2018.1554485

MAREES, A. T., DE KLUIVER, H., STRINGER, S., VORSPAN, F., CURIS, E., MARIE-CLAIRE, C. and DERKS, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int*. *J*. *Methods Psychiatr. Res*. **27** e1608.

MCCARTHY, M. I., ABECASIS, G. R., CARDON, L. R., GOLDSTEIN, D. B., LITTLE, J., IOANNIDIS, J. P. A. and HIRSCHHORN, J. N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat*. *Rev. Genet*. **9** 356–369. https://doi.org/10.1038/nrg2344

SASIENI, P. D. (1997). From genotypes to genes: Doubling the sample size. *Biometrics* **53** 1253–1261. MR1614374 https://doi.org/10.2307/2533494

SCHAID, D. J. and JACOBSEN, S. J. (1999). Biased tests of association: Comparisons of allele frequencies when departing from Hardy–Weinberg proportions. *Am*. *J*. *Epidemiol*. **149** 706–711. https://doi.org/10.1093/oxfordjournals.aje.a009878

SONG, K. and ELSTON, R. C. (2006). A powerful method of combining measures of association and Hardy-Weinberg disequilibrium for fine-mapping in case-control studies. *Stat*. *Med*. **25** 105–126. MR2222077 https://doi.org/10.1002/sim.2350

SUN, L., ROMMENS, J. M., CORVOL, H., LI, W., LI, X., CHIANG, T. A., LIN, F., DORFMAN, R., BUSSON, P.-F. et al. (2012). Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. *Nat*. *Genet*. **44** 562.

TAYLOR, J. and TIBSHIRANI, R. (2006). A tail strength measure for assessing the overall univariate significance in a dataset. *Biostatistics* **7** 167–181.

TURNER, S., ARMSTRONG, L. L., BRADFORD, Y., CARLSON, C. S., CRAWFORD, D. C., CRENSHAW, A. T., DE ANDRADE, M., DOHENY, K. F., HAINES, J. L. et al. (2011). Quality control procedures for genome-wide association studies. *Curr. Protoc*. *Hum*. *Genet*. **68** 1–19.

WANG, J. and SHETE, S. (2008). A test for genetic association that incorporates information about deviation from Hardy–Weinberg proportions in cases. *Am*. *J*. *Hum*. *Genet*. **83** 53–63. https://doi.org/10.1016/j.ajhg.2008.06.010

WANG, J. and SHETE, S. (2010). Using both cases and controls for testing Hardy–Weinberg proportions in a genetic association study. *Hum*. *Hered*. **69** 212–218.

WASSERSTEIN, R. L. and LAZAR, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose [Editorial]. *Amer. Statist*. **70** 129–133. MR3511040 https://doi.org/10.1080/00031305.2016.1154108

WITTKE-THOMPSON, J. K., PLUZHNIKOV, A. and COX, N. J. (2005). Rational inferences about departures from Hardy–Weinberg equilibrium. *Am*. *J*. *Hum*. *Genet*. **76** 967–986. https://doi.org/10.1086/430507

YU, C., ZHANG, S., ZHOU, C. and SILE, S. (2009). A likelihood ratio test of population Hardy–Weinberg equilibrium for case-control studies. *Genet*. *Epidemiol*. **33** 275–280.

ZHANG, L. (2021). *A General Study of Genetic Association Tests and the Test of Hardy-Weinberg Equilibrium*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—University of Toronto (Canada). MR4360687

ZHANG, L. and SUN, L. (2022a). A generalized robust allele-based genetic association test. *Biometrics* **78** 487–498. MR4450570 https://doi.org/10.1111/biom.13456

ZHANG, L., STRUG, L. J. and SUN, L. (2023). Supplement to "Leveraging Hardy–Weinberg disequilibrium for association testing in case-control studies." https://doi.org/10.1214/22-AOAS1695SUPPA, https://doi.org/10.1214/22-AOAS1695SUPPB

ZHENG, G. and NG, H. K. T. (2008). Genetic model selection in two-phase analysis for case-control association studies. *Biostatistics* **9** 391–399. https://doi.org/10.1093/biostatistics/kxm039

The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1782–1800
https://doi.org/10.1214/22-AOAS1699

BIVARIATE HIERARCHICAL BAYESIAN MODEL FOR COMBINING SUMMARY MEASURES AND THEIR UNCERTAINTIES FROM MULTIPLE SOURCES

BY YUJING YAO[1,a], R. TODD OGDEN[1,b], CHUBING ZENG[1,2,d] AND QIXUAN CHEN[1,c]

[1]*Department of Biostatistics, Mailman School of Public Health, Columbia University,* [a]*yy2725@cumc.columbia.edu,* [b]*to166@cumc.columbia.edu,* [c]*qc2138@cumc.columbia.edu*

[2]*Division of Biostatistics, Department of Preventive Medicine, University of Southern California,* [d]*chubingzeng@gmail.com*

It is often of interest to combine available estimates of a similar quantity from multiple data sources. When the corresponding variances of each estimate are also available, a model should take into account the *uncertainty of the estimates* themselves as well as the *uncertainty in the estimation of variances*. In addition, if there exists a strong association between estimates and their variances, the correlation between these two quantities should also be considered. In this paper we propose a bivariate hierarchical Bayesian model that jointly models the estimates and their estimated variances, assuming a correlation between these two measures. We conduct simulations to explore the performance of the proposed bivariate Bayesian model and compare it to other commonly used methods under different correlation scenarios. The proposed bivariate Bayesian model has a wide range of applications. We illustrate its application in three very different areas: PET brain imaging studies, meta-analysis, and small area estimation.

REFERENCES

BETANCOURT, M. and GIROLAMI, M. (2015). Hamiltonian Monte Carlo for hierarchical models. In *Current Trends in Bayesian Methodology with Applications* 79–101. CRC Press, Boca Raton, FL. MR3644666

BORENSTEIN, M., HEDGES, L. V., HIGGINS, J. P. and ROTHSTEIN, H. R. (2011). *Introduction to Meta-Analysis*. Wiley, New York.

BROWNE, W. J. and DRAPER, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal*. **1** 473–513. MR2221283 https://doi.org/10.1214/06-BA117

CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76** 1–32.

CARSON, R. E. (2005). Tracer kinetic modeling in PET. In *Positron Emission Tomography* 127–159. Springer, New York.

CHANG, B.-H., WATERNAUX, C. and LIPSITZ, S. (2001). Meta-analysis of binary data: Which within study variance estimate to use? *Stat. Med.* **20** 1947–1956.

CHEN, Q., ELLIOTT, M. R., HAZIZA, D., YANG, Y., GHOSH, M., LITTLE, R. J. A., SEDRANSK, J. and THOMPSON, M. (2017). Approaches to improving survey-weighted estimates. *Statist. Sci.* **32** 227–248. MR3648957 https://doi.org/10.1214/17-STS609

CHU, H. and COLE, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: A generalized linear mixed model approach. *J. Clin. Epidemiol.* **59** 1331–1332; author reply 1332–1333. https://doi.org/10.1016/j.jclinepi.2006.06.011

COCHRAN, W. G. (1954). The combination of estimates from different experiments. *Biometrics* **10** 101–129.

CRESPI, C. M. and BOSCARDIN, W. J. (2009). Bayesian model checking for multivariate outcome data. *Comput. Statist. Data Anal.* **53** 3765–3772. MR2749921 https://doi.org/10.1016/j.csda.2009.03.024

DUMOUCHEL, W. (1994). Hierarchical Bayes linear models for meta-analysis. Technical Report 27, National Institute of Statistical Sciences. Available at http://www.niss.org/sites/default/files/pdfs/technicalreports/tr27.pdf.

Key words and phrases. Brain imaging, correlation between measures and uncertainty, estimation of uncertainty, meta-analysis, small area estimation.

EMERSON, J. D., HOAGLIN, D. C. and MOSTELLER, F. (1993). A modified random-effect procedurefor combining risk differences in sets of 2×2 tables from clinical trials. *Stat. Methods Appl.* **2** 269–290.

FABRIZI, E., MONTANARI, G. E. and RANALLI, M. G. (2016). A hierarchical latent class model for predicting disability small area counts from survey data. *J. Roy. Statist. Soc. Ser. A* **179** 103–131. MR3461570 https://doi.org/10.1111/rssa.12112

FARS (2016). Fatality analysis reporting system(FARS): Analytic users manual 1975–2015. National Highway Traffic Safety Administration, Washington, DC.

FAY, R. E. III and HERRIOT, R. A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277. MR0548019

GELMAN, A. and HILL, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge Univ. Press.

GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–807. With comments and a rejoinder by the authors. MR1422404

GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3235677

GHOSH, M. and RAO, J. N. K. (1994). Small area estimation: An appraisal. *Statist. Sci.* **9** 55–93. With comments and a rejoinder by the authors. MR1278679

GLASS, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educ. Res.* **5** 3–8.

GOLDSTEIN, H. (2011). *Multilevel Statistical Models* **922**. Wiley, New York.

GUO, J., RIEBLER, A. and RUE, H. (2017). Bayesian bivariate meta-analysis of diagnostic test studies with interpretable priors. *Stat. Med.* **36** 3039–3058. MR3670407 https://doi.org/10.1002/sim.7313

HIGGINS, J. P. and THOMPSON, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21** 1539–1558.

HIGGINS, J. P. T., THOMPSON, S. G. and SPIEGELHALTER, D. J. (2009). A re-evaluation of random-effects meta-analysis. *J. Roy. Statist. Soc. Ser. A* **172** 137–159. MR2655609 https://doi.org/10.1111/j.1467-985X.2008.00552.x

HOAGLIN, D. C. (2015). We know less than we should about methods of meta-analysis. *Res. Synth. Methods* **6** 287–289.

HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. MR3214779

HWANG, J. T. G., QIU, J. and ZHAO, Z. (2009). Empirical Bayes confidence intervals shrinking both means and variances. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 265–285. MR2655533 https://doi.org/10.1111/j.1467-9868.2008.00681.x

KORKMAZ, S., GOKSULUK, D. and ZARARSIZ, G. (2014). MVN: An R package for assessing multivariate normality. *R J.* **6** 151–162.

MAITI, T., REN, H. and SINHA, S. (2014). Prediction error of small area predictors shrinking both means and variances. *Scand. J. Stat.* **41** 775–790. MR3249428 https://doi.org/10.1111/sjos.12061

MORRIS, E. D., ENDRES, C. J., SCHMIDT, K. C., CHRISTIAN, B. T., MUZIC, R. F. and FISHER, R. E. (2004). Kinetic modeling in positron emission tomography. In *Emission Tomography: The Fundamentals of PET and SPECT* **46** 499–540.

NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

OGDEN, R. T. and TARPEY, T. (2005). Estimation in regression models with externally estimated parameters. *Biostatistics* **7** 115–129.

PAUL, M., RIEBLER, A., BACHMANN, L. M., RUE, H. and HELD, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Stat. Med.* **29** 1325–1339. MR2757228 https://doi.org/10.1002/sim.3858

PFEFFERMANN, D. (2002). Small area estimation-new developments and directions. *Int. Stat. Rev.* **70** 125–143.

POTTER, F. J. (1988). Survey of procedures to control extreme sampling weights. In *Proceedings of the American Statistical Association, Section on Survey Research Methods* 453–458. Amer. Statist. Assoc., Washington, DC.

POTTER, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In *Proceedings of the American Statistical Association, Section on Survey Research Methods* **225230**. Amer. Statist. Assoc., Washington, DC.

PUSTEJOVSKY, J. E. and TIPTON, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prev. Sci.* **23** 425–438. https://doi.org/10.1007/s11121-021-01246-3

RAO, J. N. K. and MOLINA, I. (2015). *Small Area Estimation*, 2nd ed. *Wiley Series in Survey Methodology*. Wiley, Hoboken, NJ. With a foreword by Graham Kalton. MR3380626 https://doi.org/10.1002/9781118735855

REITSMA, J. B., GLAS, A. S., RUTJES, A. W., SCHOLTEN, R. J., BOSSUYT, P. M. and ZWINDERMAN, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J. Clin. Epidemiol.* **58** 982–990.

RUBIN, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12** 1151–1172. MR0760681 https://doi.org/10.1214/aos/1176346785

SMITH, T. C., SPIEGELHALTER, D. J. and THOMAS, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Stat. Med.* **14** 2685–2699.

SUGASAWA, S., KUBOKAWA, T. and RAO, J. N. K. (2019). Hierarchical Bayes small-area estimation with an unknown link function. *Scand. J. Stat.* **46** 885–897. MR3994173 https://doi.org/10.1111/sjos.12376

SUGASAWA, S., TAMAE, H. and KUBOKAWA, T. (2017). Bayesian estimators for small area models shrinking both means and variances. *Scand. J. Stat.* **44** 150–167. MR3619699 https://doi.org/10.1111/sjos.12246

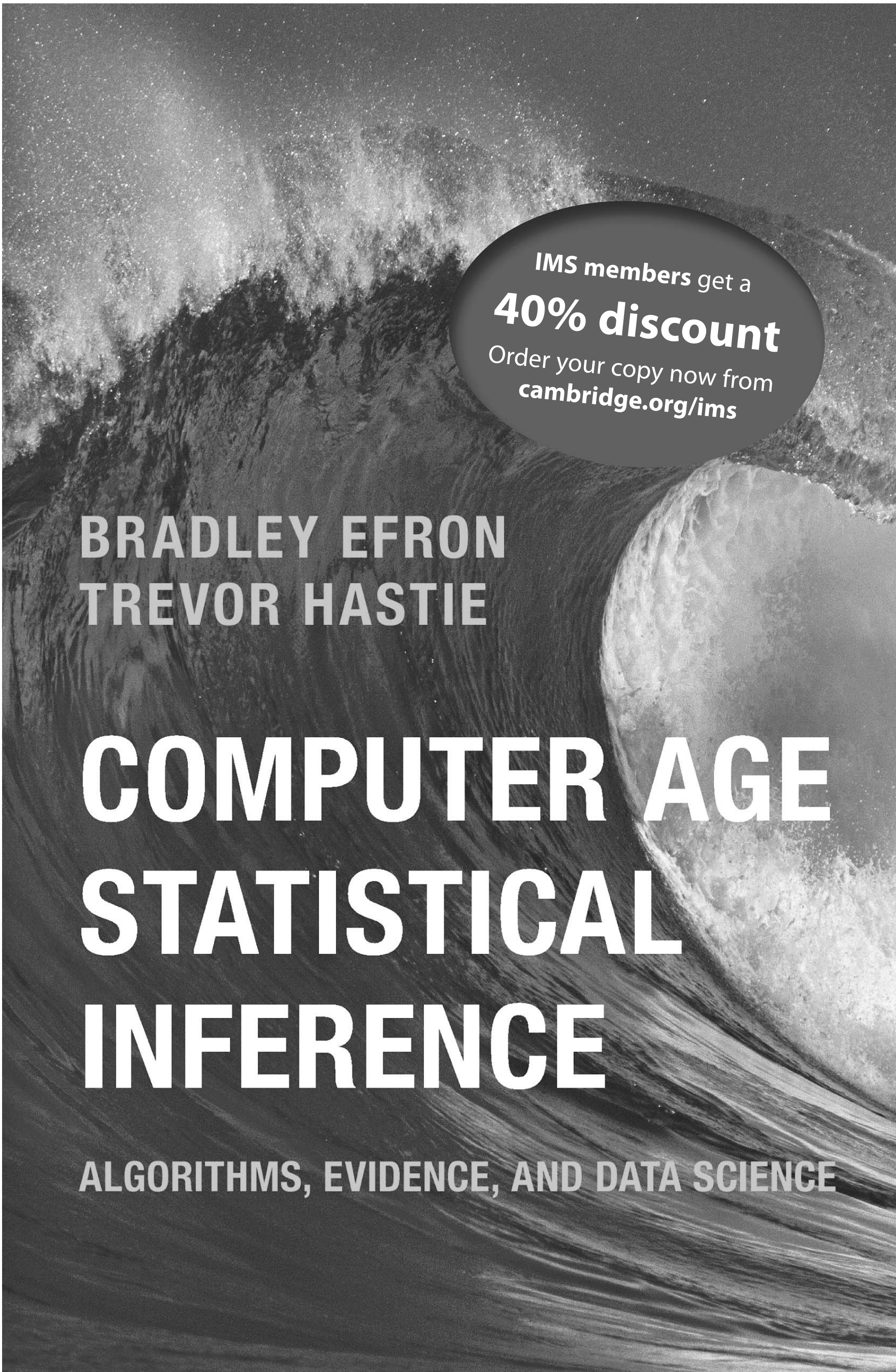
SUTTON, A. J. and ABRAMS, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Stat. Methods Med. Res.* **10** 277–303. https://doi.org/10.1177/096228020101000404

TURNER, R. M., OMAR, R. Z., YANG, M., GOLDSTEIN, H. and THOMPSON, S. G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat. Med.* **19** 3417–3432.

WANG, J. and FULLER, W. A. (2003). The mean squared error of small area predictors constructed with estimated area variances. *J. Amer. Statist. Assoc.* **98** 716–723. MR2011685 https://doi.org/10.1198/016214503000000620

YAO, Y., OGDEN, R. T., ZENG, C. and CHEN, Q. (2023). Supplement to "Bivariate hierarchical bayesian model for combining summary measures and their uncertainties from multiple sources." https://doi.org/10.1214/22-AOAS1699SUPPA, https://doi.org/10.1214/22-AOAS1699SUPPB

YOU, Y. and CHAPMAN, B. (2006). Small area estimation using area level models and estimated sampling variances. *Surv. Methodol.* **32** 97.


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE