STATISTICAL SCIENCE

Volume 34, Number 4 November 2019

Memorial Issue for Lawrence D. Brown

***Statistical Science* [ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 34, Number 4, November 2019. Published quarterly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, OH 44122, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

POSTMASTER: Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike—Suite L2310, Bethesda, MD 20814-3998, USA.

Statistical Science Volume 34, Number 4 (523–691) November 2019

Volume 34 Number 4 November 2019

Memorial Issue for Lawrence D. Brown

Statistical Science
2019, Vol. 34, No. 4, 523–544
https://doi.org/10.1214/18-STS693

Models as Approximations I: Consequences Illustrated with Linear Regression

Andreas Buja, Lawrence Brown, Richard Berk, Edward George, Emil Pitkin, Mikhail Traskin, Kai Zhang and Linda Zhao

Abstract. In the early 1980s, Halbert White inaugurated a "model-robust" form of statistical inference based on the "sandwich estimator" of standard error. This estimator is known to be "heteroskedasticity-consistent," but it is less well known to be "nonlinearity-consistent" as well. Nonlinearity, however, raises fundamental issues because in its presence regressors are not ancillary, hence cannot be treated as fixed. The consequences are deep: (1) population slopes need to be reinterpreted as statistical functionals obtained from OLS fits to largely arbitrary joint x-y distributions; (2) the meaning of slope parameters needs to be rethought; (3) the regressor distribution affects the slope parameters; (4) randomness of the regressors becomes a source of sampling variability in slope estimates of order $1/\sqrt{N}$; (5) inference needs to be based on model-robust standard errors, including sandwich estimators or the x-y bootstrap. In theory, model-robust and model-trusting standard errors can deviate by arbitrary magnitudes either way. In practice, significant deviations between them can be detected with a diagnostic test.

Key words and phrases: Ancillarity of regressors, misspecification, econometrics, sandwich estimator, bootstrap.

REFERENCES

BERK, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Stat.* **37** 51–58. MR0189176

BERK, R. H. (1970). Consistency a posteriori. *Ann. Math. Stat.* **41** 894–906. MR0266356

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122

BERK, R., KRIEGLER, B. and YLVISAKER, D. (2008). Counting the homeless in Los Angeles county. In *Probability and Statis-*

Andreas Buja is the Liem Sioe Liong/First Pacific Company Professor of Statistics, Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA. Lawrence Brown was the Miers Busch Professor of Statistics, Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA. Richard Berk is Professor of Criminology and Statistics, Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA. Edward George is the Universal Furniture Professor of Statistics, Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA. Emil Pitkin is Lecturer and Research Scholar in Statistics, Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA. Mikhail Traskin was Assistant Professor of Statistics, work done while with the Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA. Kai Zhang is Associate Professor, Department of Statistics & Operations Research, University of North Carolina at Chapel Hill, 306 Hanes Hall, CB#3260, Chapel Hill, North Carolina 27599-3260, USA. Linda Zhao is Professor of Statistics, Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA (e-mail: lzhao@wharton.upenn.edu).

tics: *Essays in Honor of David A. Freedman*. *Inst. Math. Stat.* (*IMS*) *Collect*. **2** 127–141. IMS, Beachwood, OH. MR2459951

BERMAN, M. (1988). A theorem of Jacobi and its generalization. *Biometrika* **75** 779–783. MR0995120

BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist*. *Sinica* **7** 1–31. MR1441142

BOX, G. E. P.(1979). Robustness in the strategy of scientific model building. In *Robustness in Statistics*: *Proceedings of a Workshop* (R. L. Launer and G. N. Wilkinson, eds.) Academic Press (Elsevier), Amsterdam.

BÜHLMANN, P. and VAN DE GEER, S. (2015). High-dimensional inference in misspecified linear models. *Electron*. *J*. *Stat*. **9** 1449–1473. MR3367666

BUJA, A., BROWN, L., BERK, R., GEORGE, E., PITKIN, E., TRASKIN, M., ZHANG, K., and ZHAO, L. (2019). Supplement to "Models as Approximations I: Consequences Illustrated with Linear Regression." 10.1214/18-STS693SUPP.

COX, D. R. (1962). Further results on tests of separate families of hypotheses (1962). *J. Roy. Statist*. *Soc*. *Ser. B* **24** 406–424.

COX, D. R. (1995). Discussion of Chatfield (1995). *J. Roy. Statist*. *Soc*. *Ser. A* **158** 455–456.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. CRC Press, London. MR0370837

DAVIES, L. (2014). *Data Analysis and Approximate Models*. *Monographs on Statistics and Applied Probability* **133**. CRC Press, Boca Raton, FL. MR3241514

DE BLASI, P. (2013). Discussion on article "Bayesian inference with misspecified models" by Stephen G. Walker. *J. Statist*. *Plann*. *Inference* **143** 1634–1637. MR3082221

DIGGLE, P. J., HEAGERTY, P. J., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. *Oxford Statistical Science Series* **25**. Oxford Univ. Press, Oxford. MR2049007

DONOHO, D. and MONTANARI, A. (2014). Variance Breakdown of Huber (M)-estimators: $n/p \to m \in (1, \infty)$. Available at arXiv:1503.02106.

EFRON, B. (1982). *The Jackknife*, *the Bootstrap and Other Resampling Plans*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **38**. SIAM, Philadelphia, PA. MR0659849

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903

EICKER, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Ann*. *Math*. *Stat*. **34** 447–456. MR0148177

EL KAROUI, N., BEAN, D., BICKEL, P. and YU, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc*. *Natl*. *Acad*. *Sci*. **110** 14563–14568.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann*. *Statist*. **9** 1218–1228. MR0630104

FREEDMAN, D. A. (2006). On the so-called "Huber sandwich estimator" and "robust standard errors". *Amer. Statist*. **60** 299–302. MR2291297

GELMAN, A. and PARK, D. K. (2009). Splitting a predictor at the upper quarter or third and the lower quarter or third. *Amer. Statist*. **63** 1–8. MR2655696

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237

HALL, A. R. (2005). *Generalized Method of Moments*. *Advanced Texts in Econometrics*. Oxford Univ. Press, Oxford. MR2135106

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0829458

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054. MR0666123

HARTIGAN, J. A. (1983). Asymptotic normality of posterior distributions. In *Bayes Theory*. *Springer Series in Statistics* 107–118. Springer, New York, NY.

HAUSMAN, J. A. (1978). Specification tests in econometrics. *Econometrica* **46** 1251–1271. MR0513692

HINKLEY, D. V. (1977). Jacknifing in unbalanced situations. *Technometrics* **19** 285–292. MR0458734

HOFF, P. and WAKEFIELD, J. (2013). Bayesian sandwich posteriors for pseudo-true parameters. A discussion of "Bayesian inference with misspecified models" by Stephen Walker. *J. Statist*. *Plann*. *Inference* **143** 1638–1642. MR3082222

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc*. *Fifth Berkeley Sympos*. *Math*. *Statist*. *and Probability* (*Berkeley*, *Calif*., 1965/66), *Vol. I*: *Statistics* 221–233. Univ. California Press, Berkeley, CA. MR0216620

HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795

KAUERMANN, G. and CARROLL, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *J. Amer. Statist*. *Assoc*. **96** 1387–1396. MR1946584

KRASKER, W. S. and WELSCH, R. E. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist*. *Assoc*. **77** 595–604. MR0675886

LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann*. *Statist*. **44** 907–927. MR3485948

LEHMANN, E. L. and ROMANO, J. P. (2008). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR1481711

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

LOH, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann*. *Statist*. **45** 866–896. MR3650403

LONG, J. S. and ERVIN, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear model. *Amer. Statist*. **54** 217–224.

MACKINNON, J. and WHITE, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econometrics* **29** 305–325.

MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann*. *Statist*. **21** 255–285. MR1212176

MAMMEN, E. (1996). Empirical process of residuals for high-dimensional linear models. *Ann*. *Statist*. **24** 307–335. MR1389892

MCCARTHY, D., ZHANG, K., BROWN, L. D., BERK, R., BUJA, A., GEORGE, E. I. and ZHAO, L. (2018). Calibrated percentile double bootstrap for robust linear regression inference. *Statist*. *Sinica* **28** 2565–2589. MR3839874

NEWEY, W. K. and WEST, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. MR0890864

O'HAGAN, A. (2013). Bayesian inference with misspecified models: Inference about what?. *J. Statist. Plann. Inference* **143** 1643–1648. MR3082223

OWEN, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton, FL.

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. MR1329181

R Development Core Team (2008). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org.

STIGLER, S. M. (2001). Ancillary history. In *State of the Art in Probability and Statistics: Festschrift for W. R. van Zwet* (M. DeGunst, C. Klaassen and A. van der Vaart, eds.) 555–567.

SZPIRO, A. A., RICE, K. M. and LUMLEY, T. (2010). Model-robust regression and a Bayesian "sandwich" estimator. *Ann. Appl. Stat.* **4** 2099–2113. MR2829948

WALKER, S. G. (2013). Bayesian inference with misspecified models. *J. Statist. Plann. Inference* **143** 1621–1633. MR3082220

WASSERMAN, L. (2011). Low assumptions, high dimensions. *Ration. Mark. Moral.* **2** 201–209.

WEBER, N. C. (1986). The jackknife and heteroskedasticity: Consistent variance estimation for regression models. *Econom. Lett.* **20** 161–163. MR0830028

WHITE, H. (1980a). Using least squares to approximate unknown regression functions. *Internat. Econom. Rev.* **21** 149–170. MR0572464

WHITE, H. (1980b). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838. MR0575027

WHITE, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76** 419–433. MR0624344

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163

WHITE, H. (1994). *Estimation, Inference and Specification Analysis. Econometric Society Monographs* **22**. Cambridge Univ. Press, Cambridge. MR1292251

WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350. MR0868303

Statistical Science
2019, Vol. 34, No. 4, 545–565
https://doi.org/10.1214/18-STS694

Models as Approximations II: A Model-Free Theory of Parametric Regression

Andreas Buja, Lawrence Brown, Arun Kumar Kuchibhotla, Richard Berk, Edward George and Linda Zhao

Abstract. We develop a model-free theory of general types of parametric regression for i.i.d. observations. The theory replaces the parameters of parametric models with statistical functionals, to be called "regression functionals," defined on large nonparametric classes of joint x-y distributions, without assuming a correct model. Parametric models are reduced to heuristics to suggest plausible objective functions. An example of a regression functional is the vector of slopes of linear equations fitted by OLS to largely arbitrary x-y distributions, without assuming a linear model (see Part I). More generally, regression functionals can be defined by minimizing objective functions, solving estimating equations, or with ad hoc constructions. In this framework, it is possible to achieve the following: (1) define a notion of "well-specification" for regression functionals that replaces the notion of correct specification of models, (2) propose a well-specification diagnostic for regression functionals based on reweighting distributions and data, (3) decompose sampling variability of regression functionals into two sources, one due to the conditional response distribution and another due to the regressor distribution interacting with misspecification, both of order $N^{-1/2}$, (4) exhibit plug-in/sandwich estimators of standard error as limit cases of x-y bootstrap estimators, and (5) provide theoretical heuristics to indicate that x-y bootstrap standard errors may generally be preferred over sandwich estimators.

Key words and phrases: Ancillarity of regressors, misspecification, econometrics, sandwich estimator, bootstrap, bagging.

REFERENCES

BERK, R., KRIEGLER, B. and YLVISAKER, D. (2008). Counting the homeless in Los Angeles county. In *Probability and Statistics*: *Essays in Honor of David A. Freedman* (D. Nolan and S. Speed, eds.). *Inst. Math. Stat.* (*IMS*) *Collect.* **2** 127–141. IMS, Beachwood, OH. MR2459951

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122

BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31. MR1441142

BREIMAN, L. (1996). Bagging predictors. *Mach. Learn.* **24** 123–140.

BUJA, A. and STUETZLE, W. (2001,2016). Smoothing effects of bagging: Von Mises expansions of bagged statistical functionals. Available at arXiv:1612.02528.

BUJA, A., BROWN, L., KUCHIBHOTLA, A. K., BERK, R., GEORGE, E. and ZHAO, L. (2019). Supplement to "Models as Approximations II: A Model-Free Theory of Parametric Regression." 10.1214/18-STS694SUPP.

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the*

Andreas Buja is the Liem Sioe Liong/First Pacific Company Professor of Statistics, Lawrence Brown was the Miers Busch Professor of Statistics, Arun Kumar Kuchibhotla is Doctoral Student of Statistics, Richard Berk is Professor of Criminology and Statistics, Edward George is the Universal Furniture Professor of Statistics, Linda Zhao is Professor of Statistics, Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA (e-mail: lzhao@wharton.upenn.edu).

Bootstrap. *Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903

GELMAN, A. and PARK, D. K. (2009). Splitting a predictor at the upper quarter or third and the lower quarter or third. *Amer*. *Statist*. **63** 1–8. MR2655696

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0829458

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054. MR0666123

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

HAUSMAN, J. A. (1978). Specification tests in econometrics. *Econometrica* **46** 1251–1271. MR0513692

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann*. *Math*. *Stat*. **35** 73–101. MR0161415

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc*. *Fifth Berkeley Sympos*. *Math*. *Statist*. *and Probability* (*Berkeley*, *Calif*., 1965/66), *Vol*. *I*: *Statistics* 221–233. Univ. California Press, Berkeley, CA. MR0216620

KUCHIBHOTLA, A. K., BROWN, L. D. and BUJA, A. (2018). Model-free study of ordinary least squares linear regression. Available at arXiv:1809.10538.

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166

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

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann*. *Statist*. **22** 2031–2050. MR1329181

R DEVELOPMENT CORE TEAM (2008). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*,Vienna, Austria. Available at http://www.R-project.org.

RIEDER, H. (1994). *Robust Asymptotic Statistics*. *Springer Series in Statistics*. Springer, New York. MR1284041

TUKEY, J. W. (1962). The future of data analysis. *Ann*. *Math*. *Stat*. **33** 1–67. MR0133937

WHITE, H. (1980). Using least squares to approximate unknown regression functions. *Internat*. *Econom*. *Rev*. **21** 149–170. MR0572464

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163

Statistical Science
2019, Vol. 34, No. 4, 566–568
https://doi.org/10.1214/19-STS722

Discussion of Models as Approximations I & II

Sara van de Geer

Abstract. We discuss the papers "Models as Approximations" I & II, by A. Buja, R. Berk, L. Brown, E. George, E. Pitkin, M. Traskin, L. Zao and K. Zhang (Part I) and A. Buja, L. Brown, A. K. Kuchibhota, R. Berk, E. George and L. Zhao (Part II). We present a summary with some details for the generalized linear model.

Key words and phrases: Misspecification, sandwich formula.

REFERENCES

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. I*: *Statistics* 221–233. Univ. California Press, Berkeley, Calif. MR0216620

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

Sara van de Geer is Full Professor, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland (e-mail: geer@stat.math.ethz.ch).

Statistical Science
2019, Vol. 34, No. 4, 569–571
https://doi.org/10.1214/19-STS723

Comment on Models as Approximations, Parts I and II, by Buja et al.

Jerald F. Lawless

Abstract. I comment on the papers Models as Approximations I and II, by A. Buja, R. Berk, L. Brown, E. George, E. Pitkin, M. Traskin, L. Zhao and K. Zhang.

Key words and phrases: Covariate distributions, misspecification, regression models, transportability.

REFERENCES

BREIMAN, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–231. MR1874152

CHATTERJEE, N., CHEN, Y.-H., MAAS, P. and CARROLL, R. J. (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *J. Amer. Statist. Assoc.* **111** 107–117. MR3494641

COX, D. R. (1990). Role of models in statistical analysis. *Statist. Sci.* **5** 169–174. MR1062575

HAN, P. and LAWLESS, J. F. (2019). Empirical likelihood estimation using auxiliary summary information with different covariate distributions. *Statist. Sinica* **29** 1321–1342. MR3932520

KEIDING, N. and LOUIS, T. A. (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *J. Roy. Statist. Soc. Ser. A* **179** 319–376. MR3461587

SHMUELI, G. (2010). To explain or to predict? *Statist. Sci.* **25** 289–310. MR2791669

Jerald F. Lawless is Distinguished Professor Emeritus, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 (e-mail: jlawless@uwaterloo.ca).

Statistical Science
2019, Vol. 34, No. 4, 572–574
https://doi.org/10.1214/19-STS724

Comment: Models as Approximations

Nikki L. B. Freeman, Xiaotong Jiang, Owen E. Leete, Daniel J. Luckett, Teeranan Pokaprakarn and Michael R. Kosorok

REFERENCES

BOX, G. E. P. (1979). Robustness in the strategy of scientific model building. In *Robustness in Statistics* (R. L. Launer and G. N. Wilkinson, eds.) 201–236. Academic Press, New York.

BOX, G. E. P. and DRAPER, N. R. (1987). *Empirical Model-Building and Response Surfaces*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0861118

CHENG, S. C., WEI, L. J. and YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82** 835–845. MR1380818

KOSOROK, M. R., LEE, B. L. and FINE, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *Ann. Statist.* **32** 1448–1491. MR2089130

LIN, D. Y. and WEI, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84** 1074–1078. MR1134495

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR0727836

ROSSET, S. and TIBSHIRANI, R. J. (2018). From fixed-x to random-x regression: Bias-variance decompositions, covariance penalties, and prediction error estimation. *J. Amer. Statist. Assoc.* 1–14.

SEN, A. and SEN, B. (2014). Testing independence and goodness-of-fit in linear models. *Biometrika* **101** 927–942. MR3286926

STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. MR1439316

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163

Nikki L. B. Freeman is a graduate student, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA (e-mail: nlbf@live.unc.edu). Xiaotong Jiang is a graduate student, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA (e-mail: xiaotong@live.unc.edu). Owen E. Leete is a graduate student, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA (e-mail: oleete@email.unc.edu). Daniel J. Luckett is a postdoctoral research associate, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA (e-mail: luckett@live.unc.edu). Teeranan Pokaprakarn is a graduate student, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA (e-mail: terranan@live.unc.edu). Michael R. Kosorok is the W.R. Kenan, Jr. Distinguished Professor and Chair, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA (e-mail: kosorok@bios.unc.edu).

Statistical Science
2019, Vol. 34, No. 4, 575–579
https://doi.org/10.1214/19-STS725

Discussion of Models as Approximations I & II

Dag Tjøstheim

REFERENCES

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

GONÇALVES, S. and WHITE, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *J. Econometrics* **119** 199–219. MR2041897

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774

HJORT, N. L. and JONES, M. C. (1996). Locally parametric nonparametric density estimation. *Ann. Statist.* **24** 1619–1647. MR1416653

LACAL, V. and TJØSTHEIM, D. (2019). Estimating and testing nonlinear local dependence between two time series. *J. Bus. Econom. Statist.* **37** 648–660. MR4016160

LU, Z., LUNDERVOLD, A., TJØSTHEIM, D. and YAO, Q. (2007). Exploring spatial nonlinearity using additive approximation. *Bernoulli* **13** 447–472. MR2331259

MAMMEN, E., LINTON, O. and NIELSEN, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27** 1443–1490. MR1742496

NORDMAN, D. J. and LAHIRI, S. N. (2014). Convergence rates of empirical block length selectors for block bootstrap. *Bernoulli* **20** 958–978. MR3178523

TJØSTHEIM, D. and HUFTHAMMER, K. O. (2013). Local Gaussian correlation: A new measure of dependence. *J. Econometrics* **172** 33–48. MR2997128

WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838. MR0575027

WHITE, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76** 419–433. MR0624344

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. MR0067450

Dag Tjøstheim is Professor Emeritus, Department of Mathematics, University of Bergen, Norway (e-mail: Dag.Tjostheim@math.uib.no).

Statistical Science
2019, Vol. 34, No. 4, 580–583
https://doi.org/10.1214/19-STS726

Comment: “Models as Approximations I: Consequences Illustrated with Linear Regression” by A. Buja, R. Berk, L. Brown, E. George, E. Pitkin, L. Zhan and K. Zhang

Roderick J. Little

REFERENCES

BIRNBAUM, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.* **57** 269–326. MR0138176

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *J. Roy. Statist. Soc. Ser. A* **143** 383–430. MR0603745

LITTLE, R. J. (2004). To model or not to model? Competing modes of inference for finite population sampling. *J. Amer. Statist. Assoc.* **99** 546–556. MR2109316

LITTLE, R. J. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *Amer. Statist.* **60** 213–223. MR2246754

LITTLE, R. (2011). Calibrated Bayes, for statistics in general, and missing data in particular. *Statist. Sci.* **26** 162–174. MR2858391

LITTLE, R. J. (2012). Calibrated Bayes: An alternative inferential paradigm for official statistics (with discussion and rejoinder). *J. Off. Stat.* **28** 309–337.

LITTLE, R. J. (2013). In praise of simplicity not mathematistry! Ten simple powerful ideas for the statistical scientist. *J. Amer. Statist. Assoc.* **108** 359–369. MR3174626

LITTLE, R. and AN, H. (2004). Robust likelihood-based analysis of multivariate data with missing values. *Statist. Sinica* **14** 949–968. MR2089342

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196

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

RUBIN, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12** 1151–1172. MR0760681

RUBIN, D. B. (2019). Conditional calibration and the sage statistician. *Surv. Methodol.* To appear.

SZPIRO, A. A., RICE, K. M. and LUMLEY, T. (2010). Model-robust regression and a Bayesian "sandwich" estimator. *Ann. Appl. Stat.* **4** 2099–2113. MR2829948

ZHANG, G. and LITTLE, R. (2009). Extensions of the penalized spline of propensity prediction method of imputation. *Biometrics* **65** 911–918. MR2649864

ZHENG, H. and LITTLE, R. J. (2005). Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. *J. Off. Stat.* **21** 1–20.

Roderick J. Little is Distinguished University Professor, Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, Michigan 48109-2029, USA (e-mail: rlittle@umich.edu).

Statistical Science
2019, Vol. 34, No. 4, 584–590
https://doi.org/10.1214/19-STS746

Comment: Models Are Approximations!

Anthony C. Davison, Erwan Koch and Jonathan Koh

Abstract. This discussion focuses on areas of disagreement with the papers, particularly the target of inference and the case for using the robust 'sandwich' variance estimator in the presence of moderate mis-specification. We also suggest that existing procedures may be appreciably more powerful for detecting mis-specification than the authors' RAV statistic, and comment on the use of the pairs bootstrap in balanced situations.

Key words and phrases: Bootstrap, designed experiment, infinitesimal jackknife, model mis-specification, regression diagnostics, sandwich variance estimator.

REFERENCES

COOK, R. D. and WEISBERG, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70** 1–10. MR0742970

DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. MR1478673

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596. MR0615434

FERNHOLZ, L. T. (1983). *Von Mises Calculus for Statistical Functionals*. *Lecture Notes in Statistics* **19**. Springer, New York. MR0713611

FISHER, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

FOX, T., HINKLEY, D. and LARNTZ, K. (1980). Jackknifing in nonlinear regression. *Technometrics* **22** 29–33. MR0559682

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. Wiley, New York. MR0829458

HINKLEY, D. V. (1977). Jacknifing in unbalanced situations. *Technometrics* **19** 285–292. MR0458734

HINKLEY, D. V. and WANG, S. (1991). Efficiency of robust standard errors for regression coefficients. *Comm. Statist. Theory Methods* **20** 1–11. MR1114631

KALBFLEISCH, J. D. (1975). Sufficiency and conditionality. *Biometrika* **62** 251–268. MR0386075

PANARETOS, V. M., KRAUS, D. and MADDOCKS, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Amer. Statist. Assoc.* **105** 670–682. MR2724851

TUKEY, J. W. (1949). One degree of freedom for non-additivity. *Biometrics* **5** 232–242.

Anthony C. Davison is Professor, École Polytechnique Fédérale de Lausanne, EPFL-FSB-MATH-STAT, Station 8, 1015 Lausanne, Switzerland (e-mail: Anthony.Davison@epfl.ch). Erwan Koch is Instructor, École Polytechnique Fédérale de Lausanne, EPFL-FSB-MATH-STAT, Station 8, 1015 Lausanne, Switzerland (e-mail: Erwan.Koch@epfl.ch). Jonathan Koh is PhD candidate, École Polytechnique Fédérale de Lausanne, EPFL-FSB-MATH-STAT, Station 8, 1015 Lausanne, Switzerland (e-mail: Jonathan.Koh@epfl.ch).

Statistical Science
2019, Vol. 34, No. 4, 591–598
https://doi.org/10.1214/19-STS747

Comment: Models as (Deliberate) Approximations

David Whitney, Ali Shojaie and Marco Carone

REFERENCES

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249

CARONE, M., LUEDTKE, A. R. and VAN DER LAAN, M. J. (2019). Toward computerized efficient estimation in infinite-dimensional models. *J. Amer. Statist. Assoc.* **114** 1174–1190. MR4011771

CHAMBAZ, A., NEUVIAL, P. and VAN DER LAAN, M. J. (2012). Estimation of a non-parametric variable importance measure of a continuous exposure. *Electron. J. Stat.* **6** 1059–1099. MR2988439

GRAHAM, B. S. and PINTO, C. C. D. X. (2018). Semiparametrically efficient estimation of the average linear regression function. Technical report, National Bureau of Economic Research.

HATTORI, S. and HENMI, M. (2012). Estimation of treatment effects based on possibly misspecified Cox regression. *Lifetime Data Anal.* **18** 408–433. MR2973773

O'QUIGLEY, J. (2008). *Proportional Hazards Regression. Statistics for Biology and Health*. Springer, New York. MR2400249

SCHEMPER, M., WAKOUNIG, S. and HEINZE, G. (2009). The estimation of average hazard ratios by weighted Cox regression. *Stat. Med.* **28** 2473–2489. MR2751535

STRUTHERS, C. A. and KALBFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73** 363–369. MR0855896

VAN DER LAAN, M. J. and ROBINS, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York.

VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data. Springer Series in Statistics*. Springer, New York. MR2867111

XU, R. and O'QUIGLEY, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics* **1** 423–439.

YU, Z. and VAN DER LAAN, M. J. (2003). Measuring treatment effects using semiparametric models. Technical report, Univ. California, Berkeley, Dept. Biostatistics Working Paper Series.

David Whitney is Teaching Fellow in Statistics, Department of Mathematics, Imperial College London, Huxley Building, South Kensington Campus, Imperial College, London SW7 2AZ, United Kingdom (e-mail: d.whitney@imperial.ac.uk). Ali Shojaie is Associate Professor, Department of Biostatistics, University of Washington, 1705 NE Pacific Street, Seattle, Washington 98195, USA (e-mail: ashojaie@uw.edu). Marco Carone is Assistant Professor and Norman Breslow Endowed Faculty Fellow, Department of Biostatistics, University of Washington, 1705 NE Pacific Street, Seattle, Washington 98195, USA (e-mail: mcarone@uw.edu).

Statistical Science
2019, Vol. 34, No. 4, 599–603
https://doi.org/10.1214/19-STS748

Comment: Statistical Inference from a Predictive Perspective

Alessandro Rinaldo, Ryan J. Tibshirani and Larry Wasserman

Abstract. What is the meaning of a regression parameter? Why is this the de facto standard object of interest for statistical inference? These are delicate issues, especially when the model is misspecified. We argue that focusing on predictive quantities may be a desirable alternative.

Key words and phrases: Regression, prediction, variable importance, effect sizes.

REFERENCES

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

CHEN, J., SONG, L., WAINWRIGHT, M. J. and JORDAN, M. I. (2017). L-Shapley and c-Shapley: Efficient model interpretation for structured data. Preprint. Available at arXiv:1808.02610.

LEI, J., G'SELL, M., RINALDO, A., TIBSHIRANI, R. J. and WASSERMAN, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113** 1094–1111. MR3862342

MARKOVIC, J., XIA, L. and TAYLOR, J. (2017). Unifying approach to selective inference with applications to cross-validation. Preprint. Available at arXiv:1703.06559.

RINALDO, A., WASSERMAN, L. and G'SELL, M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *Ann. Statist.* **47** 3438–3469. MR4025748

SHAH, R. and PETERS, J. (2018). The hardness of conditional independence testing and the generalised covariance measure. Preprint. Available at arXiv:1804.07203.

SHAPLEY, L. S. (1953). A value for n-person games. In *Contributions to the Theory of Games*, *Vol*. 2. *Annals of Mathematics Studies* **28** 307–317. Princeton Univ. Press, Princeton, NJ. MR0053477

VOVK, V., GAMMERMAN, A. and SHAFER, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. MR2161220

WILLIAMSON, B., GILBERT, P., SIMON, N. and CARONE, M. (2017). Nonparametric variable importance assessment using machine learning techniques. Univ. Washington Biostatistics Working Paper Series, Working Paper 422.

Alessandro Rinaldo is Professor, Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: arinaldo@cmu.edu). Ryan J. Tibshirani is Associate Professor, Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: ryantibs@cmu.edu). Larry Wasserman is Professor, Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: larry@cmu.edu).

Statistical Science
2019, Vol. 34, No. 4, 604–605
https://doi.org/10.1214/19-STS756

Discussion: Models as Approximations

Dalia Ghanem and Todd A. Kuffner

REFERENCES

ATHEY, S. and IMBENS, G. W. (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments* **1** 73–140. Elsevier, Amsterdam.

ELLIOTT, G., GHANEM, D. and KRÜGER, F. (2016). Forecasting conditional probabilities of binary outcomes under misspecification. *Rev. Econ. Stat.* **98** 742–755.

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

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

Dalia Ghanem is Assistant Professor, Department of Agricultural and Resource Economics, University of California, Davis, One Shields Avenue, Davis, California 95616, USA (e-mail: dghanem@ucdavis.edu). Todd A. Kuffner is Associate Professor, Department of Mathematics and Statistics, Washington University in St. Louis, 1 Brookings Dr., St. Louis, Missouri 63131, USA (e-mail: kuffner@wustl.edu).

Statistical Science
2019, Vol. 34, No. 4, 606–620
https://doi.org/10.1214/19-STS762

Models as Approximations—Rejoinder

Andreas Buja, Arun Kumar Kuchibhotla, Richard Berk, Edward George, Eric Tchetgen Tchetgen and Linda Zhao

Abstract. We respond to the discussants of our articles emphasizing the importance of inference under misspecification in the context of the reproducibility/replicability crisis. Along the way, we discuss the roles of diagnostics and model building in regression as well as connections between our well-specification framework and semiparametric theory.

Key words and phrases: Well-specification, reproducibility/replicability, proper scoring rules, causal inference, semiparametrics, diagnostics.

REFERENCES

ADAM, D. (2019). Psychology's reproducibility solution fails first test. *Science* **364** 813. 10.1126/science.364.6443.813.

ARONOV, P. M. and MILLER, B. T. (2019). *Foundations of Agnostic Statistics*. Cambridge Univ. Press, Cambridge.

ATHEY, S. and IMBENS, G. (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments* 1 73–140. Elsevier, Amsterdam.

AZRIEL, D., BROWN, L. D., SKLAR, M., BERK, R., BUJA, A. and ZHAO, L. (2016). Semi-supervised linear regression. Available at arXiv:1612.02391.

BERK, R., BUJA, A., BROWN, L., GEORGE, E., KUCHIBHOTLA, A. K., SU, W. and ZHAO, L. (2019). Assumption lean regression. *Amer. Statist.* 10.1080/00031305.2019.1592781.

BERK, R., OLSON, M., BUJA, A. and OUSS, A. (2020). Using recursive partitioning to find and estimate heterogeneous treatment effects in randomized clinical trials. *J. Exp. Criminol.*. To appear. Available at arXiv.org/abs/1807.04164.

BOOS, D. D. (1992). On generalized score tests. *Amer. Statist.* **46** 327–333.

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80** 580–619. MR0803258

BUJA, A., STUETZLE, W. and YI, S. (2005). Loss Functions for Binary Class Probability Estimation and Classification: Structure and Applications. Unpublished manuscript. Available at www-stat.wharton.upenn.edu/~buja.

CANTONI, E. and RONCHETTI, E. (2001). Robust inference for generalized linear models. *J. Amer. Statist. Assoc.* **96** 1022–1030. MR1947250

DAVIES, L. (2014). *Data Analysis and Approximate Models: Model Choice, Location-Scale, Analysis of Variance, Nonparametric Regression and Image Analysis. Monographs on Statistics and Applied Probability* **133**. CRC Press, Boca Raton, FL. MR3241514

ELLIOTT, G., GHANEM , D. and KRÜGER, F. (2016). Forecasting conditional probabilities of binary outcomes under misspecification. *The Review of Economics and Statistics* **98** 742–755.

EMA, FDA (2017). ICH E9(R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials. www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials, www.regulations.gov/docket?D=FDA-2017-D-6113.

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548

GODAMBE, V. P. and THOMPSON, M. E. (1984). Robust estimation through estimating equations. *Biometrika* **71** 115–125. MR0738332

HARTMAN, N. (2014). Who really found the Higgs Boson. https://getpocket.com/explore/item/who-really-found-the-higgs-boson.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. I*: *Statistics* 221–233. Univ. California Press, Berkeley, CA. MR0216620

IOANNIDIS, J. P. A. (2005). Why most published research findings are false. *Chance* **18** 40–47. MR2216666

KOLLER, M. and STAHEL, W. A. (2017). Nonsingular subsampling for regression S estimators with categorical predictors. *Comput. Statist.* **32** 631–646. MR3656977

Andreas Buja is the Liem Sioe Liong/First Pacific Company Professor of Statistics. Arun Kumar Kuchibhotla is Doctoral Student of Statistics. Richard Berk is Professor of Criminology and Statistics. Edward George is the Universal Furniture Professor of Statistics. Eric Tchetgen Tchetgen is the Luddy Family President's Distinguished Professor. Linda Zhao is Professor of Statistics. Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, USA (e-mail: lzhao@wharton.upenn.edu).

KUCHIBHOTLA, A. K., BROWN, L. D. and BUJA, A. (2018a). Model-free study of ordinary least squares linear regression. Available at arXiv:1809.10538.

KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A., GEORGE, E. I. and ZHAO, L. (2018b). A model free perspective for linear regression: Uniform-in-model bounds for post selection inference. Available at arXiv:1802.05801.

LEI, J., G'SELL, M., RINALDO, A., TIBSHIRANI, R. J. and WASSERMAN, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113** 1094–1111. MR3862342

MCCARTHY, D., ZHANG, K., BROWN, L. D., BERK, R., BUJA, A., GEORGE, E. I. and ZHAO, L. (2018). Calibrated percentile double bootstrap for robust linear regression inference. *Statist. Sinica* **28** 2565–2589. MR3839874

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized linear models*. Chapman and Hall, London.

NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382. MR1303237

NEWEY, W. K., HSIEH, F. and ROBINS, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica* **72** 947–962. MR2051442

PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166

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

PITKIN, E., BERK, R., BROWN, L., BUJA, A., GEORGE, E., ZHANG, K. and ZHAO, L. (2013). Improved precision in estimating average treatment effects. Available at arXiv:1311.0291.

SHAH, R. and PETERS, J. (2018). The hardness of conditional independence testing and the generalised covariance measure. Available at arXiv:1804.07203.

SIMMONS, J. P., NELSON, L. D. and SIMONSOHN, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22** 1359–1366.

SZPIRO, A. A., RICE, K. M. and LUMLEY, T. (2010). Model-robust regression and a Bayesian "sandwich" estimator. *Ann. Appl. Stat.* **4** 2099–2113. MR2829948

STEINBERGER, L. and LEEB, H. (2018). Conditional predictive inference for high-dimensional stable algorithms. Available at arXiv:1809.01412v1.

STOKER, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica* **54** 1461–1481. MR0868152

WHITE, H. (1980a). Using least squares to approximate unknown regression functions. *Internat. Econom. Rev.* **21** 149–170. MR0572464

WHITE, H. (1980b). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838. MR0575027

Statistical Science
2019, Vol. 34, No. 4, 621–634
https://doi.org/10.1214/19-STS717

Larry Brown's Contributions to Parametric Inference, Decision Theory and Foundations: A Survey

James O. Berger and Anirban DasGupta

Abstract. This article gives a panoramic survey of the general area of parametric statistical inference, decision theory and foundations of statistics for the period 1965–2010 through the lens of Larry Brown's contributions to varied aspects of this massive area. The article goes over sufficiency, shrinkage estimation, admissibility, minimaxity, complete class theorems, estimated confidence, conditional confidence procedures, Edgeworth and higher order asymptotic expansions, variational Bayes, Stein's SURE, differential inequalities, geometrization of convergence rates, asymptotic equivalence, aspects of empirical process theory, inference after model selection, unified frequentist and Bayesian testing, and Wald's sequential theory. A reasonably comprehensive bibliography is provided.

Key words and phrases: Admissibility, ancillary, asymptotic equivalence, Bayes, conditional confidence, differential inequality, Edgeworth expansions, estimated confidence, minimax, sequential, shrinkage, sufficiency.

REFERENCES

AGRESTI, A. and COULL, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *Amer. Statist.* **52** 119–126. MR1628435

BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Stat.* **25** 423–462. MR0063630

BARANKIN, E. W. and MAITRA, A. P. (1963). Generalization of the Fisher–Darmois–Koopman–Pitman theorem on sufficient statistics. *Sankhya*, *Ser. A* **25** 217–244. MR0171342

BARRON, A. R. (1986). Entropy and the central limit theorem. *Ann. Probab.* **14** 336–342. MR0815975

BERGER, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226. MR0397940

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR0804611

BERGER, J. O., BROWN, L. D. and WOLPERT, R. L. (1994). A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing. *Ann. Statist.* **22** 1787–1807. MR1329168

BERGER, J. O. and STRAWDERMAN, W. E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *Ann. Statist.* **24** 931–951. MR1401831

BERGER, J., BOCK, M. E., BROWN, L. D., CASELLA, G. and GLESER, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5** 763–771. MR0443156

BERK, R., BROWN, L. D. and ZHAO, L. (2009). Statistical inference after model selection. *J. Quant. Criminol.* **26** 217–236.

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122

BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309. MR0630112

BICKEL, P. J. and DOKSUM, K. A. (2016). *Mathematical Statistics—Basic Ideas and Selected Topics*. *Vol.* 2, 2nd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3287337

BLACKWELL, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Stat.* **22** 393–399. MR0043418

BLYTH, C. R. and STILL, H. A. (1983). Binomial confidence intervals. *J. Amer. Statist. Assoc.* **78** 108–116. MR0696854

BOROVKOV, A. A. and SAKHANENKO, A. I. (1980). Estimates for averaged quadratic risk. *Probab. Math. Statist.* **1** 185–195. MR0626310

James O. Berger is Arts and Sciences Professor, Department of Statistical Science, Duke University, Durham, North Carolina, USA (e-mail: berger@stat.duke.edu). Anirban DasGupta is Professor, Department of Statistics, Purdue University, West Lafayette, Indiana, USA (e-mail: dasgupta@purdue.edu).

BROWN, L. (1964). Sufficient statistics in the case of independent random variables. *Ann. Inst. Statist. Math.* **35** 1456–1474. MR0216611

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Inst. Statist. Math.* **37** 1087–1136. MR0216647

BROWN, L. (1967). The conditional level of Student's t test. *Ann. Inst. Statist. Math.* **38** 1068–1071. MR0214210

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Inst. Statist. Math.* **42** 855–903. MR0286209

BROWN, L. D. (1978). A contribution to Kiefer's theory of conditional confidence procedures. *Ann. Statist.* **6** 59–71. MR0471160

BROWN, L. D. (1979). A heuristic method for determining admissibility of estimators—with applications. *Ann. Statist.* **7** 960–994. MR0536501

BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9** 1289–1300. MR0630111

BROWN, L. D. (1982). A proof of the central limit theorem motivated by the Cramér–Rao inequality. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 141–148. North-Holland, Amsterdam. MR0659464

BROWN, L. D. (1983). Comments on "The Robust Bayesian Viewpoint" by Berger, J. O.. In *Robustness of Bayesian Analyses* (J. B. Kadane, ed.) 126–133. Elsevier, Amsterdam.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* (S. S. Gupta, ed.) **9**. IMS, Hayward, CA. MR0882001

BROWN, L. D. (1988). The differential inequality of a statistical estimation problem. In *Statistical Decision Theory and Related Topics, IV, Vol.* 1 (*West Lafayette, Ind.*, 1986) (S. S. Gupta and J. O. Berger, eds.) 299–324. Springer, New York. MR0927109

BROWN, L. D. (1990). The 1985 Wald Memorial Lectures. An ancillarity paradox which appears in multiple linear regression. *Ann. Statist.* **18** 471–538. MR1056325

BROWN, L. D. (1994). Minimaxity, more or less. In *Statistical Decision Theory and Related Topics, V* (*West Lafayette, IN*, 1992) (S. S. Gupta and J. O. Berger, eds.) 1–18. Springer, New York. MR1286291

BROWN, L. D. (1998). *Minimax Theory, in Encyclopedia of Biostatistics* (P. Armitage and T. Colton, eds.). Wiley, New York.

BROWN, L. D. (2000). An essay on statistical decision theory. *J. Amer. Statist. Assoc.* **95** 1277–1281. MR1825275

BROWN, L. D., ed. (2002). An analogy between statistical equivalence and stochastic strong limit theorems. Invited address at the International Congress of Mathematicians, Beijing, August, 2002.

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.* **16** 101–133. MR1861069

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30** 160–201. MR1892660

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2003). Interval estimation in exponential families. *Statist. Sinica* **13** 19–49. MR1963918

BROWN, L. D. and COHEN, A. (1981). Inadmissibility of large classes of sequential tests. *Ann. Statist.* **9** 1239–1247. MR0630106

BROWN, L. D., COHEN, A. and SAMUEL-CAHN, E. (1983). A sharp necessary condition for admissibility of sequential tests—necessary and sufficient conditions for admissibility of SPRTs. *Ann. Statist.* **11** 640–653. MR0696075

BROWN, L. D., COHEN, A. and STRAWDERMAN, W. E. (1979). Monotonicity of Bayes sequential tests. *Ann. Statist.* **7** 1222–1230. MR0550145

BROWN, L. D. and FARRELL, R. H. (1990). A lower bound for the risk in estimating the value of a probability density. *J. Amer. Statist. Assoc.* **85** 1147–1153. MR1134513

BROWN, L. D. and GAJEK, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18** 1578–1594. MR1074424

BROWN, L. D. and HWANG, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics, III, Vol.* 1 (*West Lafayette, Ind.*, 1981) 205–230. Academic Press, New York. MR0705290

BROWN, L. D. and LIU, R. C. (1993). Bounds on the Bayes and minimax risk for signal parameter estimation. *IEEE Trans. Inform. Theory* **39** 1386–1394.

BROWN, L. D. and LOW, M. G. (1991). Information inequality bounds on the minimax risk (with an application to nonparametric regression). *Ann. Statist.* **19** 329–337. MR1091854

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398. MR1425958

BROWN, L. D. and PURVES, R. (1973). Measurable selections of extrema. *Ann. Statist.* **1** 902–912. MR0432846

BROWN, L. D. and ZHAO, L. (2003). Direct asymptotic equivalence of nonparametric regression and the infinite-dimensional location problem. Preprint Univ. Pennsylvania, Philadelphia, PA.

BROWN, L. D., CARTER, A. V., LOW, M. G. and ZHANG, C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *Ann. Statist.* **32** 2074–2097. MR2102503

BROWN, L., DASGUPTA, A., HAFF, L. R. and STRAWDERMAN, W. E. (2006). The heat equation and Stein's identity: Connections, applications. *J. Statist. Plann. Inference* **136** 2254–2278. MR2235057

CLEVENSON, M. L. and ZIDEK, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705. MR0394962

COHEN, A. (1965). Estimates of linear combinations of the parameters in the mean vector of a multivariate distribution. *Ann. Inst. Statist. Math.* **36** 78–87. MR0172399

DASGUPTA, A. (1983). Bayes and minimax estimation in one and multiparameter families. Ph.D. thesis, Indian Statistical Institute.

DASGUPTA, A. (2005). A conversation with Larry Brown. *Statist. Sci.* **20** 193–203. MR2183449

DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability*. *Springer Texts in Statistics*. Springer, New York. MR2664452

DASGUPTA, A. and SINHA, B. K. (1986). Estimation in the multiparameter exponential family: Admissibility and inadmissibility results. *Statist. Decisions* **4** 101–130. MR0848322

DIACONIS, P. and STEIN, C. (1982). Lecture notes on statistical decision theory. Dept. Statistics, Stanford Univ., Stanford, CA.

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437. MR1062717

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974

DYNKIN, E. B. (1961). Necessary and sufficient statistics for a family of probability distributions. In *Select. Transl. Math. Statist. and Probability*, *Vol.* 1 17–40. IMS and Amer. Math. Soc., Providence, RI. MR0116407

EATON, M. L. (1992). A statistical diptych: Admissible inferences—recurrence of symmetric Markov chains. *Ann. Statist.* **20** 1147–1179. MR1186245

FISHER, R. A. (1923). On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond. Ser. A* **222** 309–368.

GHOSH, B. K. (1979). A comparison of some approximate confidence intervals for the binomial parameter. *J. Amer. Statist. Assoc.* **74** 894–900. MR0556485

HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.* **20** 225–241. MR0030730

HSUAN, F. C. (1979). A stepwise Bayesian procedure. *Ann. Statist.* **7** 860–868. MR0532249

HWANG, J. T. and BROWN, L. D. (1991). Estimated confidence under the validity constraint. *Ann. Statist.* **19** 1964–1977. MR1135159

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, *Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

JOHNSON, B. MCK. (1971). On the admissible estimators for certain fixed sample binomial problems. *Ann. Inst. Statist. Math.* **42** 1579–1587. MR0418300

JOHNSON, O. (2004). *Information Theory and the Central Limit Theorem*. Imperial College Press, London. MR2109042

JOHNSTONE, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Ann. Statist.* **12** 1173–1198. MR0760682

JOHNSTONE, I. and LALLEY, S. (1984). On independent statistical decision problems and products of diffusions. *Z. Wahrsch. Verw. Gebiete* **68** 29–47. MR0767442

KAGAN, A. M., LINNIK, Y. V. and RAO, C. R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York. MR0346969

KARLIN, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Inst. Statist. Math.* **29** 406–436. MR0124101

KIEFER, J. (1977). Conditional confidence statements and confidence estimators. *J. Amer. Statist. Assoc.* **72** 789–827. MR0518611

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883

KOROSTELEV, A. and KOROSTELEVA, O. (2011). *Mathematical Statistics*. *Graduate Studies in Mathematics*, *Asymptotic Minimax Theory* **119**. Amer. Math. Soc., Providence, RI. MR2767163

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411

LEEB, H. and PÖTSCHER, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory* **24** 338–376. MR2422862

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York. MR0107933

LEHMANN, E. L. and ROMANO, J. P. (2008). *Testing Statistical Hypotheses*. Springer, New York.

LEHMANN, E. L. and STEIN, C. M. (1953). The admissibility of certain invariant statistical tests involving a translation parameter. *Ann. Math. Stat.* **24** 473–479. MR0056249

LEVIT, B. Y. (1987). On bounds for the minimax risk. In *Probability Theory and Mathematical Statistics*, *Vol. II* (*Vilnius*, 1985) 203–216. VNU Sci. Press, Utrecht. MR0901534

LIU, R. C. and BROWN, L. D. (1993). Nonexistence of informative unbiased estimators in singular problems. *Ann. Statist.* **21** 1–13. MR1212163

MEEDEN, G., GHOSH, M., SRINIVASAN, C. and VARDEMAN, S. (1989). The admissibility of the Kaplan–Meier and other maximum likelihood estimators in the presence of censoring. *Ann. Statist.* **17** 1509–1531. MR1026297

MORRIS, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10** 65–80. MR0642719

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959

OLKIN, I. and SOBEL, M. (1979). Admissible and minimax estimation for the multinomial distribution and for k independent binomial distributions. *Ann. Statist.* **7** 284–290. MR0520240

PÖTSCHER, B. M. (1991). Effects of model selection on inference. *Econometric Theory* **7** 163–185. MR1128410

RUKHIN, A. L. (1995). Admissibility: Survey of a concept in progress. *Int. Stat. Rev.* **63** 95–115.

SINHA, B. K. and GUPTA, A. D. (1984). Admissibility of generalized Bayes and Pitman estimates in the nonregular family. *Comm. Statist. Theory Methods* **13** 1709–1721. MR0742524

SKIBINSKY, M. and RUKHIN, A. L. (1989). Admissible estimators of binomial probability and the inverse Bayes rule map. *Ann. Inst. Statist. Math.* **41** 699–716. MR1039400

SOBEL, M. (1953). An essentially complete class of decision functions for certain standard sequential problems. *Ann. Math. Stat.* **24** 319–337. MR0056899

STEIN, C. (1959). The admissibility of Pitman's estimator of a single location parameter. *Ann. Inst. Statist. Math.* **30** 970–979. MR0109392

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

TSYBAKOV, A. B. (2004). *Introduction à L'estimation Nonparamétrique*. *Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **41**. Springer, Berlin. MR2013911

WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. MR0013275

Statistical Science
2019, Vol. 34, No. 4, 635–656
https://doi.org/10.1214/19-STS718

Gaussianization Machines for Non-Gaussian Function Estimation Models

T. Tony Cai

Abstract. A wide range of nonparametric function estimation models have been studied individually in the literature. Among them the homoscedastic nonparametric Gaussian regression is arguably the best known and understood. Inspired by the asymptotic equivalence theory, Brown, Cai and Zhou (*Ann. Statist.* **36** (2008) 2055–2084; *Ann. Statist.* **38** (2010) 2005–2046) and Brown et al. (*Probab. Theory Related Fields* **146** (2010) 401–433) developed a unified approach to turn a collection of non-Gaussian function estimation models into a standard Gaussian regression and any good Gaussian nonparametric regression method can then be used.

These Gaussianization Machines have two key components, binning and transformation. When combined with BlockJS, a wavelet thresholding procedure for Gaussian regression, the procedures are computationally efficient with strong theoretical guarantees. Technical analysis given in Brown, Cai and Zhou (*Ann. Statist.* **36** (2008) 2055–2084; *Ann. Statist.* **38** (2010) 2005–2046) and Brown et al. (*Probab. Theory Related Fields* **146** (2010) 401–433) shows that the estimators attain the optimal rate of convergence adaptively over a large set of Besov spaces and across a collection of non-Gaussian function estimation models, including robust nonparametric regression, density estimation, and nonparametric regression in exponential families. The estimators are also spatially adaptive.

The Gaussianization Machines significantly extend the flexibility and scope of the theories and methodologies originally developed for the conventional nonparametric Gaussian regression. This article aims to provide a concise account of the Gaussianization Machines developed in Brown, Cai and Zhou (*Ann. Statist.* **36** (2008) 2055–2084; *Ann. Statist.* **38** (2010) 2005–2046), Brown et al. (*Probab. Theory Related Fields* **146** (2010) 401–433).

Key words and phrases: Adaptivity, asymptotic equivalence, block thresholding, density estimation, exponential family, mean matching, nonparametric function estimation, quadratic variance function, quantile coupling, robust regression, variance stabilizing transformation, wavelets.

REFERENCES

ANSCOMBE, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika* **35** 246–254. MR0028556

BARTLETT, M. S. (1936). The square root transformation in analysis of variance. *Suppl. J. R. Stat. Soc.* **3** 68–78.

BERK, R. and MACDONALD, J. M. (2008). Overdispersion and Poisson regression. *J. Quant. Criminol.* **24** 269–284.

BESBEAS, P., DE FEIS, I. and SAPATINAS, T. (2004). A comparative simulation study of wavelet shrinkage estimators for Poisson counts. *Int. Stat. Rev.* **72** 209–237.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **9**. IMS, Hayward, CA. MR0882001

BROWN, L. D., CAI, T. T. and ZHOU, H. H. (2008). Robust nonparametric estimation via wavelet median regression. *Ann.*

Tony Cai is Daniel H. Silberberg Professor of Statistics, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: tcai@wharton.upenn.edu).

Statist. **36** 2055–2084. MR2458179

BROWN, L. D., CAI, T. T. and ZHOU, H. H. (2010). Nonparametric regression in exponential families. *Ann*. *Statist*. **38** 2005–2046. MR2676882

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann*. *Statist*. **24** 2384–2398. MR1425958

BROWN, L. D., WANG, Y. and ZHAO, L. H. (2003). On the statistical equivalence at suitable frequencies of GARCH and stochastic volatility models with the corresponding diffusion model. *Statist*. *Sinica* **13** 993–1013.

BROWN, L. D. and ZHANG, C.-H. (1998). Asymptotic nonequivalence of nonparametric experiments when the smoothness index is $1/2$. *Ann*. *Statist*. **26** 279–287. MR1611772

BROWN, L. D., CAI, T. T., LOW, M. G. and ZHANG, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann*. *Statist*. **30** 688–707.

BROWN, L. D., CARTER, A. V., LOW, M. G. and ZHANG, C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *Ann*. *Statist*. **32** 2074–2097. MR2102503

BROWN, L., CAI, T. T., ZHANG, R., ZHAO, L. and ZHOU, H. (2010). The root-unroot algorithm for density estimation as implemented via wavelet block thresholding. *Probab*. *Theory Related Fields* **146** 401–433. MR2574733

CAI, T. T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann*. *Statist*. **27** 898–924. MR1724035

CAI, T. T. and ZHOU, H. H. (2009). Asymptotic equivalence and adaptive estimation for robust nonparametric regression. *Ann*. *Statist*. **37** 3204–3235. MR2549558

CAI, T. T. and ZHOU, H. H. (2010). Nonparametric regression in natural exponential families. In *Borrowing Strength*: *Theory Powering Applications—a Festschrift for Lawrence D*. *Brown*. *Inst*. *Math*. *Stat*. (*IMS*) *Collect*. **6** 199–215. IMS, Beachwood, OH. MR2798520

DALALYAN, A. and REISS, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab*. *Theory Related Fields* **134** 248–282. MR2222384

DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab*. *Theory Related Fields* **137** 25–47. MR2278451

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. SIAM, Philadelphia, PA. MR1162107

DELATTRE, S. and HOFFMANN, M. (2002). Asymptotic equivalence for a null recurrent diffusion. *Bernoulli* **8** 139–174. MR1895888

DONOHO, D. L. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Different Perspectives on Wavelets* (*San Antonio*, *TX*, 1993). *Proc*. *Sympos*. *Appl*. *Math*. **47** 173–205. Amer. Math. Soc., Providence, RI. MR1268002

EFROMOVICH, S. and SAMAROV, A. (1996). Asymptotic equivalence of nonparametric regression and white noise model has its limits. *Statist*. *Probab*. *Lett*. **28** 143–145. MR1394666

FRYZLEWICZ, P. and NASON, G. P. (2004). A Haar–Fisz algorithm for Poisson intensity estimation. *J*. *Comput*. *Graph*. *Statist*. **13** 621–638. MR2087718

GENON-CATALOT, V., LAREDO, C. and NUSSBAUM, M. (2002). Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift. *Ann*. *Statist*. **30** 731–753.

GOLUBEV, G. K., NUSSBAUM, M. and ZHOU, H. H. (2010). Asymptotic equivalence of spectral density estimation and Gaussian white noise. *Ann*. *Statist*. **38** 181–214. MR2589320

GRAMA, I. G. and NEUMANN, M. H. (2006). Asymptotic equivalence of nonparametric autoregression and nonparametric regression. *Ann*. *Statist*. **34** 1701–1732. MR2283714

GRAMA, I. and NUSSBAUM, M. (2002). Asymptotic equivalence for nonparametric regression. *Math*. *Methods Statist*. **11** 1–36. MR1900972

HOYLE, M. H. (1973). Transformations–an introduction and a bibliography. *Int*. *Stat*. *Rev*. **41** 203–223. MR0423611

JOHNSTONE, I. M. (2011). *Gaussian Estimation*: *Sequence and Wavelet Models*. Unpublished manuscript.

KOLACZYK, E. D. (1999a). Bayesian multiscale models for Poisson processes. *J*. *Amer*. *Statist*. *Assoc*. **94** 920–933. MR1723303

KOLACZYK, E. D. (1999b). Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statist*. *Sinica* **9** 119–135. MR1678884

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z*. *Wahrsch*. *Verw*. *Gebiete* **32** 111–131. MR0375412

KROLL, M. (2019). Non-parametric Poisson regression from independent and weakly dependent observations by model selection. *J*. *Statist*. *Plann*. *Inference* **199** 249–270. MR3857826

LE CAM, L. (1974). On the information contained in additional observations. *Ann*. *Statist*. **2** 630–649. MR0436400

LOW, M. G. and ZHOU, H. H. (2007). A complement to Le Cam's theorem. *Ann*. *Statist*. **35** 1146–1165. MR2341701

MASON, D. M. and ZHOU, H. H. (2012). Quantile coupling inequalities and their applications. *Probab*. *Surv*. **9** 439–479. MR3007210

MEISTER, A. (2011). Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann*. *Statist*. **39** 1471–1495. MR2850209

MILSTEIN, G. and NUSSBAUM, M. (1998). Diffusion approximation for nonparametric autoregression. *Probab*. *Theory Related Fields* **112** 535–543. MR1664703

MORRIS, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann*. *Statist*. **10** 65–80. MR0642719

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann*. *Statist*. **24** 2399–2430. MR1425959

REISS, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann*. *Statist*. **36** 1957–1982. MR2435461

REISS, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *Ann*. *Statist*. **39** 772–802. MR2816338

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR0848134

TRIEBEL, H. (1983). *Theory of Function Spaces*. *Monographs in Mathematics* **78**. Birkhäuser, Basel. MR0781540

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. MR2724359

VER HOEF, J. M. and BOVENG, P. L. (2007). Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology* **88** 2766–2772.

WANG, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *Ann. Statist.* **30** 754–783.

WINKELMANN, R. (2003). *Econometric Analysis of Count Data*, 4th ed. Springer, Berlin. MR2148271

Statistical Science
2019, Vol. 34, No. 4, 657–668
https://doi.org/10.1214/19-STS744

Larry Brown's Work on Admissibility

Iain M. Johnstone

Abstract. Many papers in the early part of Brown's career focused on the admissibility or otherwise of estimators of a vector parameter. He established that inadmissibility of invariant estimators in three and higher dimensions is a general phenomenon, and found deep and beautiful connections between admissibility and other areas of mathematics. This review touches on several of his major contributions, with a focus on his celebrated 1971 paper connecting admissibility, recurrence and elliptic partial differential equations.

Key words and phrases: Complete class theorems, inadmissibility, Blyth's method, elliptic partial differential equation, recurrence, Brownian diffusion, loss function, best invariant estimator, James–Stein estimator, variational problem, differential inequality.

REFERENCES

BERGER, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.* **8** 545–571. MR0568720

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR0804611

BERGER, J. and DASGUPTA, A. (2019). Larry Brown's contributions to parametric inference, decision theory and foundations: A survey. *Statist. Sci.* **34** 621–634.

BERGER, J. O. and SRINIVASAN, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.* **6** 783–801. MR0478426

BERGER, J. O. and STRAWDERMAN, W. E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *Ann. Statist.* **24** 931–951. MR1401831

BERGER, J. O., STRAWDERMAN, W. and TANG, D. (2005). Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Ann. Statist.* **33** 606–646. MR2163154

BERK, R. H., BROWN, L. D. and COHEN, A. (1981). Properties of Bayes sequential tests. *Ann. Statist.* **9** 678–682. MR0615445

BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309. MR0630112

BLACKWELL, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Stat.* **22** 393–399. MR0043418

BLYTH, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Stat.* **22** 22–42. MR0039966

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Stat.* **37** 1087–1136. MR0216647

BROWN, L. (1968). Inadmissiblity of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Stat.* **39** 29–48. MR0222992

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **42** 855–903. MR0286209

BROWN, L. D. (1973). Correction to: "Admissible estimators, recurrent diffusions, and insoluble boundary value problems" (*Ann. Math. Statist.* **42** (1971), 855–903). *Ann. Statist.* **1** 594.

BROWN, L. D. (1975). Estimation with incompletely specified loss functions (the case of several location parameters). *J. Amer. Statist. Assoc.* **70** 417–427. MR0373082

BROWN, L. D. (1977). Closure theorems for sequential-design processes. In *Statistical Decision Theory and Related Topics, II* (*Proc. Sympos., Purdue Univ., Lafayette, Ind.*, 1976) 57–91. MR0461738

BROWN, L. D. (1979a). Counterexample—an inadmissible estimator which is generalized Bayes for a prior with "light" tails. *J. Multivariate Anal.* **9** 332–336. MR0538413

BROWN, L. D. (1979b). A heuristic method for determining admissibility of estimators—with applications. *Ann. Statist.* **7** 960–994. MR0536501

BROWN, L. (1980a). Examples of Berger's phenomenon in the estimation of independent normal means. *Ann. Statist.* **8** 572–585. MR0568721

BROWN, L. D. (1980b). A necessary condition for admissibility. *Ann. Statist.* **8** 540–544. MR0568719

BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9** 1289–1300. MR0630111

BROWN, L. D. (1988a). The differential inequality of a statistical estimation problem. In *Statistical Decision Theory and Related Topics, IV, Vol.* 1 (*West Lafayette, Ind.*, 1986) 299–324. Springer, New York. MR0927109

Iain M. Johnstone is Professor of Statistics and of Biomedical Data Science, Department of Statistics, Stanford University, Sequoia Hall, 390 Jane Stanford Way, Stanford, California 94305-4020, USA (e-mail: imj@stanford.edu).

BROWN, L. D. (1988b). Admissibility in discrete and continuous invariant nonparametric estimation problems and in their multinomial analogs. *Ann. Statist.* **16** 1567–1593. MR0964939

BROWN, L. D. (1990). The 1985 Wald Memorial Lectures. An ancillarity paradox which appears in multiple linear regression. *Ann. Statist.* **18** 471–538. MR1056325

BROWN, L. D., CHOW, M. S. and FONG, D. K. H. (1992). On the admissibility of the maximum-likelihood estimator of the binomial variance. *Canad. J. Statist.* **20** 353–358. MR1208348

BROWN, L. D. and COHEN, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.* **2** 963–976. MR0356334

BROWN, L. D. and COHEN, A. (1981). Inadmissibility of large classes of sequential tests. *Ann. Statist.* **9** 1239–1247. MR0630106

BROWN, L. D. and COHEN, A. (1995). Complete classes for confidence set estimation. *Statist. Sinica* **5** 291–301. MR1329299

BROWN, L. D., COHEN, A. and SAMUEL-CAHN, E. (1983). A sharp necessary condition for admissibility of sequential tests—necessary and sufficient conditions for admissibility of SPRTs. *Ann. Statist.* **11** 640–653. MR0696075

BROWN, L. D., COHEN, A. and STRAWDERMAN, W. E. (1976). A complete class theorem for strict monotone likelihood ratio with applications. *Ann. Statist.* **4** 712–722. MR0415821

BROWN, L. D., COHEN, A. and STRAWDERMAN, W. E. (1979a). On the admissibility or inadmissibility of fixed sample size tests in a sequential setting. *Ann. Statist.* **7** 569–578. MR0527492

BROWN, L. D., COHEN, A. and STRAWDERMAN, W. E. (1979b). Monotonicity of Bayes sequential tests. *Ann. Statist.* **7** 1222–1230. MR0550145

BROWN, L. D., COHEN, A. and STRAWDERMAN, W. E. (1980). Complete classes for sequential tests of hypotheses. *Ann. Statist.* **8** 377–398. Erratum **17** (1989), 1414–1416. MR0560735

BROWN, L. D. and FARRELL, R. H. (1985a). Complete class theorems for estimation of multivariate Poisson means and related problems. *Ann. Statist.* **13** 706–726. MR0790567

BROWN, L. D. and FARRELL, R. H. (1985b). All admissible linear estimators of a multivariate Poisson mean. *Ann. Statist.* **13** 282–294. MR0773168

BROWN, L. D. and FARRELL, R. H. (1988). Proof of a necessary and sufficient condition for admissibility in discrete multivariate problems. *J. Multivariate Anal.* **24** 46–52. MR0925128

BROWN, L. D. and FOX, M. (1974a). Admissibility of procedures in two-dimensional location parameter problems. *Ann. Statist.* **2** 248–266. MR0420929

BROWN, L. D. and FOX, M. (1974b). Admissibility in statistical problems involving a location or scale parameter. *Ann. Statist.* **2** 807–814. MR0370850

BROWN, L. D., GEORGE, E. I. and XU, X. (2008). Admissible predictive density estimation. *Ann. Statist.* **36** 1156–1170. MR2418653

BROWN, L. D. and GREENSHTEIN, E. (1992). Two-sided sequential tests. *Ann. Statist.* **20** 555–561. MR1150361

BROWN, L. D. and HWANG, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics, III, Vol.* 1 (*West Lafayette, Ind.*, 1981) 205–230. Academic Press, New York. MR0705290

BROWN, L. D. and HWANG, J. T. (1986). Some limitation on Stein's phenomenon. *Comm. Statist. Theory Methods* **15** 2025–2042. MR0851855

BROWN, L. D. and HWANG, J. T. (1989). Universal domination and stochastic domination: U-admissibility and U-inadmissibility of the least squares estimator. *Ann. Statist.* **17** 252–267. MR0981448

BROWN, L. D. and MARDEN, J. I. (1989). Complete class results for hypothesis testing problems with simple null hypotheses. *Ann. Statist.* **17** 209–235. MR0981446

BROWN, L. D. and MARDEN, J. I. (1992). Local admissibility and local unbiasedness in hypothesis testing problems. *Ann. Statist.* **20** 832–852. MR1165595

BROWN, L. D. and ZHAO, L. H. (2012). A geometrical explanation of Stein shrinkage. *Statist. Sci.* **27** 24–30. MR2953493

BROWN, L., DASGUPTA, A., HAFF, L. R. and STRAWDERMAN, W. E. (2006). The heat equation and Stein's identity: Connections, applications. *J. Statist. Plann. Inference* **136** 2254–2278. MR2235057

DASGUPTA, A. (2005). A conversation with Larry Brown. *Statist. Sci.* **20** 193–203. MR2183449

EATON, M. L. (1992). A statistical diptych: Admissible inferences—recurrence of symmetric Markov chains. *Ann. Statist.* **20** 1147–1179. MR1186245

EFRON, B. (2011). Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.* **106** 1602–1614. MR2896860

FARRELL, R. H. (1968a). On a necessary and sufficient condition for admissibility of estimators when strictly convex loss is used. *Ann. Math. Stat.* **39** 23–28. MR0231484

FARRELL, R. H. (1968b). Towards a theory of generalized Bayes tests. *Ann. Math. Stat.* **39** 1–22. MR0224211

GEORGE, E. I., LIANG, F. and XU, X. (2006). Improved minimax predictive densities under Kullback–Leibler loss. *Ann. Statist.* **34** 78–91. MR2275235

GUTMANN, S. (1984). Decisions immune to Stein's effect. *Sankhyā Ser. A* **46** 186–194. MR0778869

HARTIGAN, J. A. (2012). Asymptotic admissibility of priors and elliptic differential equations. In *Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman. Inst. Math. Stat.* (*IMS*) *Collect.* **8** 117–130. IMS, Beachwood, OH. MR3202507

HODGES, J. L. JR. and LEHMANN, E. L. (1951). Some applications of the Cramér–Rao inequality. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1950 13–22. Univ. California Press, Berkeley and Los Angeles. MR0044795

IGHODARO, A., SANTNER, T. and BROWN, L. (1982). Admissibility and complete class results for the multinomial estimation problem with entropy and squared error loss. *J. Multivariate Anal.* **12** 469–479. MR0680522

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

JOHNSTONE, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Ann. Statist.* **12** 1173–1198. MR0760682

JOHNSTONE, I. (1986). Admissible estimation, Dirichlet principles and recurrence of birth–death chains on $\mathbf{Z}_+^p$. *Probab. Theory Related Fields* **71** 231–269. MR0816705

JOHNSTONE, I. and LALLEY, S. (1984). On independent statistical decision problems and products of diffusions. *Z. Wahrsch. Verw. Gebiete* **68** 29–47. MR0767442

KAKUTANI, S. (1944). On Brownian motions in n-space. *Proc. Imp. Acad.* (*Tokyo*) **20** 648–652. MR0014646

LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875

LÉVY, P. (1940). Le mouvement brownien plan. *Amer. J. Math*. **62** 487–550. MR0002734

ROBBINS, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol*. *I* 157–163. Univ. California Press, Berkeley and Los Angeles. MR0084919

RUKHIN, A. L. (1995). Admissibility: Survey of a concept in progress. *Int*. *Stat*. *Rev*. 95–115.

SACKS, J. (1963). Generalized Bayes solutions in estimation problems. *Ann*. *Math*. *Stat*. **34** 751–768. MR0150908

SRINIVASAN, C. (1981). Admissible generalized Bayes estimators and exterior boundary value problems. *Sankhya*, *Ser. A* **43** 1–25. MR0656265

SRINIVASAN, C. (1982). A relation between normal and exponential families through admissibility. *Sankhya*, *Ser. A* **44** 423–435. MR0705465

STEIN, C. (1955). A necessary and sufficient condition for admissibility. *Ann*. *Math*. *Stat*. **26** 518–522. MR0070929

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol*. *I* 197–206. Univ. California Press, Berkeley and Los Angeles. MR0084922

STEIN, C. (1959). The admissibility of Pitman's estimator of a single location parameter. *Ann*. *Math*. *Stat*. **30** 970–979. MR0109392

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann*. *Inst*. *Statist*. *Math*. **16** 155–160. MR0171344

STEIN, C. (1974). Estimation of the mean of a multivariate normal distribution. In *Proceedings of the Prague Symposium on Asymptotic Statistics* (*Charles Univ*., *Prague*, 1973), *Vol*. *II* 345–381. MR0381062

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann*. *Statist*. **9** 1135–1151. MR0630098

TWEEDIE, M. C. K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proc*. *Camb. Philos*. *Soc*. **43** 41–49. MR0018381

Statistical Science
2019, Vol. 34, No. 4, 669–691
https://doi.org/10.1214/19-STS754

Statistical Theory Powering Data Science

Junhui Cai, Avishai Mandelbaum, Chaitra H. Nagaraja, Haipeng Shen and Linda Zhao

Dedicated to Lawrence D. Brown

Abstract. Statisticians are finding their place in the emerging field of data science. However, many issues considered "new" in data science have long histories in statistics. Examples of using statistical thinking are illustrated, which range from exploratory data analysis to measuring uncertainty to accommodating nonrandom samples. These examples are then applied to service networks, baseball predictions and official statistics.

Key words and phrases: Service networks, queueing theory, empirical Bayes, nonparametric estimation, sports statistics, decennial census, house price index.

REFERENCES

ADLER, P. S., MANDELBAUM, A., NGUYEN, V. and SCHWERER, E. (1995). From project to process management: An empirically-based framework for analyzing product development time. *Manage. Sci.* **41** 458–484.

ALDOR-NOIMAN, S., FEIGIN, P. D. and MANDELBAUM, A. (2009). Workload forecasting for a call center: Methodology and a case study. *Ann. Appl. Stat.* **3** 1403–1447. MR2752140

ANDERSON, M. (2015). *The American Census: A Social History*, 2nd ed. Yale University Press, New Haven.

ARMONY, M., ISRAELIT, S., MANDELBAUM, A., MARMOR, Y. N., TSEYTLIN, Y. and YOM-TOV, G. B. (2015). On patient flow in hospitals: A data-based queueing-science perspective. *Stoch. Syst.* **5** 146–194. MR3442392

AZRIEL, D., FEIGIN, P. and MANDELBAUM, A. (2014). Erlang-S: A data-based model of servers in queueing networks. *Manage. Sci.* **65** 4607–4635.

BACCELLI, F., KAUFFMANN, B. and VEITCH, D. (2009). Inverse problems in queueing theory and Internet probing. *Queueing Syst.* **63** 59–107. MR2576007

BAILEY, M. J., MUTH, R. F. and NOURSE, H. O. (1963). A regression method for real estate price index construction. *J. Amer. Statist. Assoc.* **58** 933–942.

BENDER-DEMOLL, S. and MCFARLAND, D. A. (2006). The art and science of dynamic network visualization. *J. Soc. Struct.* **7** 1–38.

BERK, R., BROWN, L. D., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122

BORST, S., MANDELBAUM, A. and REIMAN, M. I. (2004). Dimensioning large call centers. *Oper. Res.* **52** 17–34. MR2066238

BRAMSON, M. (1998). State space collapse with application to heavy traffic limits for multiclass queueing networks. *Queueing Syst.* **30** 89–148. MR1663763

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **42** 855–903. MR0286209

BROWN, L. D. (2008). In-season prediction of batting averages: A field test of empirical Bayes and Bayes methodologies. *Ann. Appl. Stat.* **2** 113–152. MR2415597

BROWN, L. D. (2015). Comments on "Methodological issues and challenges in the production of official statistics." *J. Surv. Statist. Methodol.* **3** 478–481.

BROWN, L. D. and GREENSHTEIN, E. (2009). Nonparametric empirical Bayes and compound decision approaches to estimation

Junhui Cai is Ph.D. candidate, Department of Statistics, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104, USA (e-mail: junhui@wharton.upenn.edu). Avishai Mandelbaum is Professor Emeritus, Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Room 518, Bloomfield Building, Technion City, Haifa 3200003, Israel. Chaitra H. Nagaraja is Associate Professor of Statistics, Strategy and Statistics Area, Gabelli School of Business, Fordham University, Martino Hall, 45 Columbus Avenue, New York, New York 10023, USA. Haipeng Shen is Patrick S C Poon Professor in Analytics and Innovation, Faculty of Business and Economics, The University of Hong Kong, Room 815, K. K. Leung Building, Pok Fu Lam Road, Hong Kong. Linda Zhao is Professor of Statistics, Department of Statistics, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104, USA.

of a high-dimensional vector of normal means. *Ann. Statist.* **37** 1685–1704. MR2533468

BROWN, L., GANS, N., MANDELBAUM, A., SAKOV, A., SHEN, H., ZELTYN, S. and ZHAO, L. (2005). Statistical analysis of a telephone call center: A queueing-science perspective. *J. Amer. Statist. Assoc.* **100** 36–50. MR2166068

CAI, J. and ZHAO, L. (2019). Nonparametric empirical Bayes method for sparse noisy signals. Preprint.

CALHOUN, C. (1996). OFHEO House Price Indices: HPI Technical Description. Available at https://www.fhfa.gov/PolicyProgramsResearch/Research/Pages/HPI-Technical-Description.aspx.

CASE, K. E. and SHILLER, R. J. (1987). Prices of single-family homes since 1970: New indexes for four cities. *N. Engl. Econ. Rev.* **Sept/Oct** 45–56.

CASE, K. E. and SHILLER, R. J. (1989). The efficiency of the market for single family homes. *Am. Econ. Rev.* **79** 125–137.

CHAN, W. and L'ECUYER, P. CCOptim: Call Center Optimization Java Library. Available at http://simul.iro.umontreal.ca/contactcenters/index.html.

CHEN, N., LEE, D. and SHEN, H. (2018). Can Customer Arrival Rates Be Modelled by Sine Waves? Submitted. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3125120

CHEN, H. and YAO, D. D. (2001). *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization, Stochastic Modelling and Applied Probability. Applications of Mathematics (New York)* **46**. Springer, New York. MR1835969

CHEN, H., HARRISON, J. M., MANDELBAUM, A., VAN ACKERE, A. and WEIN, L. (1988). Empirical evaluation of a queueing network model for semiconductor wafer fabrication. *Oper. Res.* **36** 202–215.

CITRO, C. F. (2016). The US federal statistical system's past, present, and future. *Annu. Rev. Stat. Appl.* **3** 347–373.

COWLING, A., HALL, P. and PHILLIPS, M. J. (1996). Bootstrap confidence regions for the intensity of a Poisson point process. *J. Amer. Statist. Assoc.* **91** 1516–1524. MR1439091

DAI, J. G. and HE, S. (2010). Customer abandonment in many-server queues. *Math. Oper. Res.* **35** 347–362. MR2674724

DAI, J. G., YEH, D. H. and ZHOU, C. (1997). The QNet method for re-entrant queueing networks with priority disciplines. *Oper. Res.* **45** 610–623.

DAVENPORT, T. H. and PATIL, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*.

DEO, S. and LIN, W. (2013). The impact of size and occupancy of hospital on the extent of ambulance diversion: Theory and evidence. *Oper. Res.* **61** 544–562.

DICKER, L. H. and ZHAO, S.D. (2016). High-dimensional classification via nonparametric empirical Bayes and maximum likelihood inference. *Biometrika* **103** 21–34.

DONG, J., YOM-TOV, E. and YOM-TOV, G. B. (2018). The impact of delay announcements on hospital network coordination and waiting times. *Manage. Sci.* **65** 1969–1994.

EBERSTADT, N., NUNN, R., SCHANZENBACK, D. W. and STRAIN, M. R. (2017). "In order that they might rest their arguments on facts": The vital role of government-collected data. The Hamilton Project at Brookings and the American Enterprise Institute.

EFRON, B. and MORRIS, C. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *J. Amer. Statist. Assoc.* **70** 311–319.

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597

EFRON, B. and MORRIS, C. (1977). Stein's paradox in statistics. *Sci. Am.* **236** 119–127.

ERLANG, A. K. (1948). On the rational determination of the number of circuits. In *The Life and Works of A. K. Erlang* (E. Brockmeyer, H. L. Halstrom and A. Jensen, eds.) 216–221. The Copenhagen Telephone Company, Copenhagen.

FEDERAL HOUSING AND FINANCE AGENCY. House Price Index, Quarterly Purchase-Only Indexes (Estimated Using Sales Price Data), 100 Largest Metropolitan Statistical Areas (Seasonally Adjusted and Unadjusted). Available at https://www.fhfa.gov/DataTools/Downloads/pages/house-price-index.aspx; accessed 30 January 2019.

FELDMAN, Z. and MANDELBAUM, A. (2010). Using simulation-based stochastic approximation to optimize staffing of systems with skills-based-routing. In *Proceedings—Winter Simulation Conference*. 3307–3317.

FELDMAN, Z., MANDELBAUM, A., MASSEY, W. A. and WHITT, W. (2008). Staffing of time-varying queues to achieve time-stable performance. *Manage. Sci.* **54** 324–338.

GANS, N., KOOLE, G. and MANDELBAUM, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manuf. Serv. Oper. Manag.* **5** 79–141.

GANS, N., LIU, N., MANDELBAUM, A., SHEN, H. and YE, H. (2010). Service times in call centers: Agent heterogeneity and learning with some operational consequences. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown. Inst. Math. Stat. (IMS) Collect.* **6** 99–123. IMS, Beachwood, OH. MR2798514

GANS, N., SHEN, H., ZHOU, Y. P., KOROLEV, N., MCCORD, A. and RISTOCK, H. (2015). Parametric forecasting and stochastic programming models for call-center workforce scheduling. *Manuf. Serv. Oper. Manag.* **17** 571–588.

GARNETT, O., MANDELBAUM, A. and REIMAN, M. (2002). Designing a call center with impatient customers. *Manuf. Serv. Oper. Manag.* **4** 208–227.

GERSHWIN, G. and GERSHWIN, I. (1937). Let's Call the Whole Thing Off. Shall We Dance?

GLASSERMAN, P. (2004). *Monte Carlo Methods in Financial Engineering: Stochastic Modelling and Applied Probability. Applications of Mathematics (New York)* **53**. Springer, New York. MR1999614

GLYNN, P. W. and IGLEHART, D. L. (1989). Importance sampling for stochastic simulations. *Manage. Sci.* **35** 1367–1392. MR1024494

GROVES, R. M. (2011). Three eras of survey reseach. *Public Opin. Q.* **75** 861–871.

GU, J. and KOENKER, R. (2017). Empirical Bayesball remixed: Empirical Bayes methods for longitudinal data. *J. Appl. Econometrics* **32** 575–599. MR3631958

GURVICH, I. and WHITT, W. (2009). Queue-and-idleness-ratio controls in many-server service systems. *Math. Oper. Res.* **34** 363–396. MR2554064

IBRAHIM, R. (2018). Sharing delay information in service systems: A literature survey. *Queueing Syst.* **89** 49–79. MR3803957

IBRAHIM, R. and L'ECUYER, P. (2013). Forecasting call center arrivals: Fixed-effects, mixed-effects, and bivariate models. *Manuf. Serv. Oper. Manag.* **15** 72–85.

IBRAHIM, R. and WHITT, W. (2011). Wait-time predictors for customer service systems with time-varying demand and capacity. *Oper. Res.* **59** 1106–1118. MR2864327

IBRAHIM, R., YE, H., L'ECUYER, P. and SHEN, H. (2016a). Modeling and forecasting call center arrivals: A literature survey and a case study. *Int. J. Forecast.* **32** 865–874.

IBRAHIM, R., L'ECUYER, P., SHEN, H. and THIONGANE, M. (2016b). Inter-dependent, heterogeneous, and time-varying service-time distributions in call centers. *European J. Oper. Res.* **250** 480–492. MR3435009

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

JIANG, W. and ZHANG, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37** 1647–1684. MR2533467

JIANG, W. and ZHANG, C.-H. (2010). Empirical Bayes in-season prediction of baseball batting averages. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown. Inst. Math. Stat. (IMS) Collect.* **6** 263–273. IMS, Beachwood, OH. MR2798524

KANG, W., PANG, G. (2013). Fluid limit of a many-server queueing network with abandonment. Preprint. http://scripts.cac.psu.edu/users/g/u/gup3/fluidnetwork2013r.pdf.

KASPI, H. and RAMANAN, K. (2011). Law of large numbers limits for many-server queues. *Ann. Appl. Probab.* **21** 33–114. MR2759196

KIM, S. H. and WHITT, W. (2014). Are call center and hospital arrivals well modeled by nonhomogeneous Poisson processes? *Manuf. Serv. Oper. Manag.* **16** 464–480.

KOENKER, R. and MIZERA, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *J. Amer. Statist. Assoc.* **109** 674–685. MR3223742

KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models. Springer Series in Statistics*. Springer, New York. MR2724362

LAUGER, A., WISNIEWSKI, B. and MCKENNA, L. (2014). Disclosure Avoidance Techniques at the U.S. Census Bureau: Current Practices and Research. Research Report Series (Disclosure Avoidance #2014-02).

LI, G., HUANG, J. Z. and SHEN, H. (2018). To wait or not to wait: Two-way functional hazards model for understanding waiting in call centers. *J. Amer. Statist. Assoc.* **113** 1503–1514. MR3902225

LINDLEY, D. V. (1962). Discussion on professor Stein's paper. *J. R. Stat. Soc.* **24** 285–287.

MADISON, J. (1790). Census of the Union. In *Annals of Congress, House of Representatives, 1st Congress, 2nd Session*.

MAMAN, S. (2009). Uncertainty in the demand for service: The case of call centers and emergency departments Ph.D. thesis Technion-Israel Institute of Technology, Faculty of Industrial. http://ie.technion.ac.il/serveng/References/Thesis_Shimrit.pdf.

MANDELBAUM, A. and MOMČILOVIĆ, P. (2012). Queues with many servers and impatient customers. *Math. Oper. Res.* **37** 41–65. MR2891146

MANDELBAUM, A. and ZELTYN, S. (2009). Staffing many-server queues with impatient customers: Constraint satisfaction in call centers. *Oper. Res.* **57** 1189–1205. MR2583510

MANDELBAUM, A. and ZELTYN, S. (2013). Data-stories about (im)patient customers in tele-queues. *Queueing Syst.* **75** 115–146. MR3110635

MANDELBAUM, A., MOMČILOVIĆ, P., TRICHAKIS, N., KADISH, S., LEIB, R. and BUNNELL, C. (2017). Data-driven appointment-scheduling under uncertainty: The case of an infusion unit in a cancer center. Under Revision to Management Science.

MATTESON, D. S., MCLEAN, M. W., WOODARD, D. B. and HENDERSON, S. G. (2011). Forecasting emergency medical service call arrival rates. *Ann. Appl. Stat.* **5** 1379–1406. MR2849778

MENG, X.-L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *Ann. Appl. Stat.* **12** 685–726. MR3834282

MURALIDHARAN, O. (2010). An empirical Bayes mixture method for effect size and false discovery rate estimation. *Ann. Appl. Stat.* **4** 422–438. MR2758178

MUTHURAMAN, K. and ZHA, H. (2008). Simulation-based portfolio optimization for large portfolios with transaction costs. *Math. Finance* **18** 115–134. MR2380942

NAGARAJA, C. H. (2019). *Measuring Society*. CRC Press, Boca Raton, FL.

NAGARAJA, C. H. and BROWN, L. D. (2013). Constructing and evaluating an autoregressive house price index. In *Topics in Applied Statistics* (M. Hu, Y. Liu and J. Lin, eds.). *Springer Proceedings in Mathematics & Statistics* **55** 3–12. Springer, Berlin.

NAGARAJA, C. H., BROWN, L. D. and WACHTER, S. (2014). Repeat sales house price index methodology. *J. Real Estate Lit.* **22** 23–46.

NAGARAJA, C. H., BROWN, L. D. and ZHAO, L. H. (2011). An autoregressive approach to house price modeling. *Ann. Appl. Stat.* **5** 124–149. MR2810392

NEWMAN, M. E. J. (2008). *The Mathematics of Networks. The New Palgrave Encyclopedia of Economics*. Palgrave Macmillan, Basingstoke, UK.

NEWMAN, M. (2018). *Networks*. Oxford Univ. Press, Oxford. MR3838417

PFEFFERMANN, D. (2015). Methdological issues and challenges in the production of official statistics: 24th Annual Morris Hansen Lecture. *J. Surv. Statist. Methodol.* **3** 425–477.

PIDGIN, C. F. (1888). *Practical Statistics: A Handbook for the Use of the Statistician at Work, Students in Colleges and Academies, Agents, Census Enumerators, Etc*. The W.E. Smythe Company.

PUHALSKII, A. A. and REIMAN, M. I. (2000). The multiclass $GI/PH/N$ queue in the Halfin–Whitt regime. *Adv. in Appl. Probab.* **32** 564–595. MR1778580

RAYKAR, V. and ZHAO, L. (2010). Nonparametric prior for adaptive sparsity. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* 629–636.

REED, J. and TEZCAN, T. (2012). Hazard rate scaling of the abandonment distribution for the $GI/M/n + GI$ queue in heavy traffic. *Oper. Res.* **60** 981–995. MR2979435

REICH, M. (2011). The workload process: Modelling, inference and applications. M. Sc. research proposal.

ROBERT, P. (2003). *Stochastic Networks and Queues: Stochastic Modelling and Applied Probability*, French ed. *Applications of Mathematics (New York)* **52**. Springer, Berlin. MR1996883

SANGALLI, L. M. (2018). The role of statistics in the era of big data. *Statist. Probab. Lett.* **136** 1–3. MR3806826

SEELAB (Service Enterprise Engineering Laboratory). Available at https://web.iem.technion.ac.il/en/service-enterprise-engineering-see-lab/general-information.html.

SENDEROVICH, A. (2016). Queue Mining: Service Perspectives in Process Mining Ph.D. thesis Technion-Israel Institute of Technology, Faculty of Industrial. http://ie.technion.ac.il/serveng/References/Thesis_Submission_Arik_Senderovich.pdf.

SENDEROVICH, A., WEIDLICH, M., GAL, A. and MANDELBAUM, A. (2015). Queue mining for delay prediction in multi-class service processes. *Inf. Syst.* **53** 278–295.

SHEN, H. and BROWN, L. D. (2006). Non-parametric modelling for time-varying customer service time at a bank call centre. *Appl. Stoch. Models Bus. Ind.* **22** 297–311. MR2275576

SHEN, H. and HUANG, J. Z. (2008a). Interday forecasting and intraday updating of call center arrivals. *Manuf. Serv. Oper. Manag.* **10** 391–410.

SHEN, H. and HUANG, J. Z. (2008b). Forecasting time series of inhomogeneous Poisson processes with application to call center workforce management. *Ann. Appl. Stat.* **2** 601–623. MR2524348

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 197–206. Univ. California Press, Berkeley and Los Angeles. MR0084922

STEIN, C. M. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. Ser. B* **24** 265–296. MR0148184

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Stat.* **42** 385–388. MR0397939

STRAWDERMAN, W. E. (1973). Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. *Ann. Statist.* **1** 1189–1194. MR0365806

TAYLOR, J. W. (2012). Density forecasting of intraday call center arrivals using models based on exponential smoothing. *Manage. Sci.* **58** 534–549.

TORRIERI, N., ACSO, DSSD and SEHSD PROGRAM STAFF (2014). American Community Survey Design and Methodology. U.S. Census Bureau.

TUKEY, J. W. (1977). *Exploratory Data Analysis*. *Addison-Wesley Series in Behavioral Science*. Addison-Wesley Pub. Co., Reading, MA.

VAN DYK, D., FUENTES, M., JORDAN, M. I., NEWTON, M., RAY, B. R., TEMPLE LANG, D. and WICKHAM, H. (2015). ASA Statement on the Role of Statistics in Data Science. *Amstat news*.

VARDI, Y. (1996). Network tomography: Estimating source-destination traffic intensities from link data. *J. Amer. Statist. Assoc.* **91** 365–377. MR1394093

VARIAN, H. (2009). Hal Varian on how the Web challenges managers. McKinsey & Company.

WEINBERG, J., BROWN, L. D. and STROUD, J. R. (2007). Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. *J. Amer. Statist. Assoc.* **102** 1185–1198. MR2412542

WEINSTEIN, A., MA, Z., BROWN, L. D. and ZHANG, C.-H. (2018). Group-linear empirical Bayes estimates for a heteroscedastic normal mean. *J. Amer. Statist. Assoc.* **113** 698–710. MR3832220

WHITT, W. (1983). The queueing network analyzer. *Bell Syst. Tech. J.* **62** 2779–2815.

WHITT, W. (1992). Understanding the efficiency of multi-server service systems. *Manage. Sci.* **38** 708–723.

WHITT, W. (2002a). *Stochastic-Process Limits*: *An Introduction to Stochastic-Process Limits and Their Application to Queues*. *Springer Series in Operations Research*. Springer, New York. MR1876437

WHITT, W. (2002b). Stochastic models for the design and management of customer contact centers: Some research directions. Department of Industrial Engineering and Operations Research, Columbia Univ., New York.

WHITT, W. (2012). Fitting birth-and-death queueing models to data. *Statist. Probab. Lett.* **82** 998–1004. MR2910048

WRIGHT, C. D. and HUNT, W. O. (1900). The history and growth of the United States census: Prepared for the Senate Committee on the Census. In 56*th Congreess*, 1*st Session*; *Document No*. 194.

XIE, X., KOU, S. C. and BROWN, L. D. (2012). SURE estimates for a heteroscedastic hierarchical model. *J. Amer. Statist. Assoc.* **107** 1465–1479. MR3036408

YE, H., LUEDTKE, J. and SHEN, H. (2019). Call center arrivals: When to jointly forecast multiple streams? *Prod. Oper. Manag.* **28** 27–42.

YOM-TOV, G. and MANDELBAUM, A. (2014). Erlang-R: A time-varying queue with reentrant customers, in support of healthcare staffing. *Manuf. Serv. Oper. Manag.* **16** 283–299.

ZELTYN, S. and MANDELBAUM, A. (2005). Call centers with impatient customers: Many-server asymptotics of the $M/M/n+G$ queue. *Queueing Syst.* **51** 361–402. MR2189598

ZELTYN, S., MARMOR, Y. N., MANDELBAUM, A., CARMELI, B., GREENSHPAN, O., MESIKA, Y., WASSERKRUG, S., VORTMAN, P., SCHWARTZ, D. et al. (2011). Simulation-based models of emergency departments: Real-time control, operations planning and scenario analysis. *ACM Trans. Model. Comput. Simul.* **21** 3.

ZHANG, P. and SERBAN, N. (2007). Discovery, visualization and performance analysis of enterprise workflow. *Comput. Statist. Data Anal.* **51** 2670–2687. MR2338995

U.S. CENSUS BUREAU (1907). Heads of Families at the First Census of the United States Taken in the Year 1790. Government Printing Office, Washington, DC.

U.S. CENSUS BUREAU (2009). TIGER/Line Shapefiles. Available at https://www.census.gov/geo.

U.S. CENSUS BUREAU. Decennial census of population and housing. Available at https://www.census.gov/programs-surveys/decennial-census.html.

U.S. CENSUS BUREAU (2010). Census Bureau Launches 2010 Census Advertising Campaign: Communication Effort Seeks to Boost Nation's Mail-Back Participation Rates. Available at https://www.census.gov/newsroom/releases/archives/2010_census/cb10-cn08.html, January 2010.

U.S. CENSUS BUREAU (2017a). "Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2017—Table PEPANNRES." Population Estimates Program.

U.S. CENSUS BUREAU (2017b). "Geographic Mobility by Selected Characteristics in the United States—Table S0701." American Community Survey 1-Year Estimates. Available at https://factfinder.census.gov.

U.S. CENSUS BUREAU (2017c). "Citizen, Voting-age Population by Age—Table B29001." American Community Survey 1-Year Estimates. Available at https://factfinder.census.gov.

U.S. Census Bureau (2017d). "Field of Bachelor's Degree for First Major—Table S1502." American Community Survey 1-Year Estimates. Available at https://factfinder.census.gov.

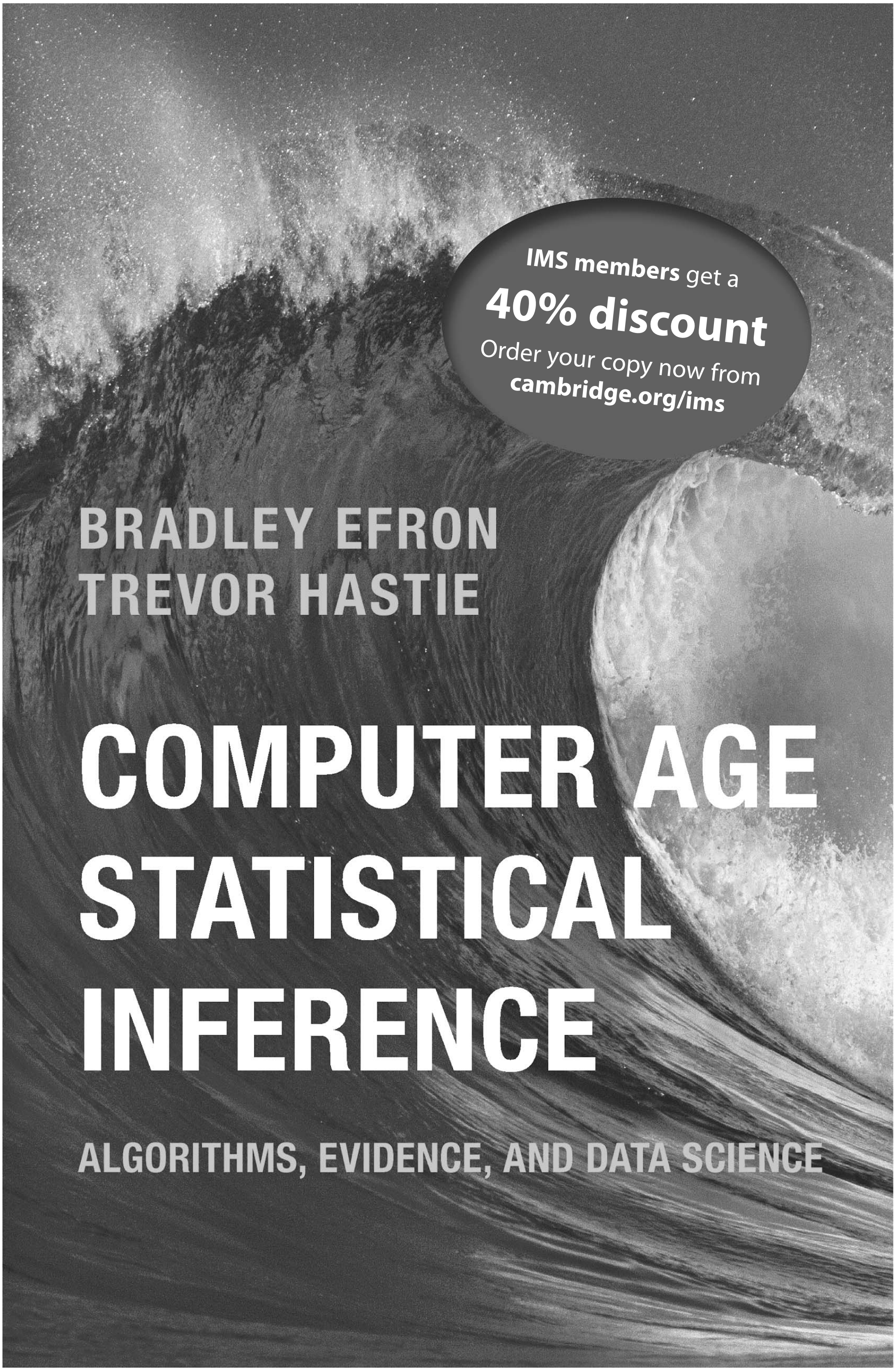
U.S. Census Bureau (2017e). "Commuting Characteristics by Sex—Table S0801." American Community Survey 1-Year Estimates. Available at https://factfinder.census.gov.

U.S. Census Bureau (2017f). "Veteran Status—Table S2101." American Community Survey 1-Year Estimates. Available at https://factfinder.census.gov.

U.S. Census Bureau—Census History Staff (2017g). Title 13, U.S. Code. Available at https://www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html. Last revised: July 18, 2017.

U.S. Census Bureau—Census History Staff (2017h). Title 26, U.S. Code. Available at https://www.census.gov/history/www/reference/privacy_confidentiality/title_26_us_code_1.html. Last revised July 18, 2017.

U.S. Office of the Secretary of State (1793). Return of the Whole Number of Persons Within the Several Districts of the United States According to, "An Act Providing for the Enumeration of the Inhabitants of the United States."


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