ISSN 0090-5364 (print)
ISSN 2168-8966 (online)

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Vol. 46, No. 3—June 2018

THE ANNALS OF STATISTICS Vol. 46, No. 3, pp. 925–1382 June 2018

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 46, Number 3, June 2018. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

The Annals of Statistics
2018, Vol. 46, No. 3, 925–957
https://doi.org/10.1214/17-AOS1570

HIGH-DIMENSIONAL *A*-LEARNING FOR OPTIMAL DYNAMIC TREATMENT REGIMES

BY CHENGCHUN SHI[1], AILIN FAN, RUI SONG[1] AND WENBIN LU[2]

North Carolina State University

Precision medicine is a medical paradigm that focuses on finding the most effective treatment decision based on individual patient information. For many complex diseases, such as cancer, treatment decisions need to be tailored over time according to patients' responses to previous treatments. Such an adaptive strategy is referred as a dynamic treatment regime. A major challenge in deriving an optimal dynamic treatment regime arises when an extraordinary large number of prognostic factors, such as patient's genetic information, demographic characteristics, medical history and clinical measurements over time are available, but not all of them are necessary for making treatment decision. This makes variable selection an emerging need in precision medicine.

In this paper, we propose a penalized multi-stage *A*-learning for deriving the optimal dynamic treatment regime when the number of covariates is of the nonpolynomial (NP) order of the sample size. To preserve the double robustness property of the *A*-learning method, we adopt the Dantzig selector, which directly penalizes the A-leaning estimating equations. Oracle inequalities of the proposed estimators for the parameters in the optimal dynamic treatment regime and error bounds on the difference between the value functions of the estimated optimal dynamic treatment regime and the true optimal dynamic treatment regime are established. Empirical performance of the proposed approach is evaluated by simulations and illustrated with an application to data from the STAR*D study.

REFERENCES

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

CANDÈS, E. and TAO, T. (2007). Rejoinder: "The Dantzig selector: Statistical estimation when p is much larger than n" [Ann. Statist. **35** (2007), 2313–2351; MR2382644]. *Ann. Statist.* **35** 2392–2404. MR2382651

CHAKRABORTY, B., MURPHY, S. and STRECHER, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat. Methods Med. Res.* **19** 317–343. MR2757118

CHEN, J. and CHEN, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95** 759–771. MR2443189

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

MSC2010 subject classifications. Primary 62C99; secondary 62J07.

Key words and phrases. A-learning, Dantzig selector, NP-dimensionality, model misspecification, optimal dynamic treatment regime, oracle inequality.

FAN, J. and LV, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inform. Theory* **57** 5467–5484. MR2849368

FAN, Y. and TANG, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 531–552. MR3065478

FAVA, M., RUSH, A. J., TRIVEDI, M. H., NIERENBERG, A. A., THASE, M. E., SACKEIM, H. A., QUITKIN, F. M., WISNIEWSKI, S., LAVORI, P. W., ROSENBAUM, J. F. et al. (2003). Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr. Clin. North Am.* **26** 457–494.

LU, W., ZHANG, H. H. and ZENG, D. (2013). Variable selection for optimal treatment decision. *Stat. Methods Med. Res.* **22** 493–504. MR3190671

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.* **44** 713–742. MR3476615

LV, J. and FAN, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.* **37** 3498–3528. MR2549567

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*, 2nd ed. Chapman & Hall, London. MR3223057

MENDELSON, S., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2007). Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.* **17** 1248–1282. MR2373017

MENDELSON, S., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2008). Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx.* **28** 277–289. MR2453368

MILMAN, V. D. and PAJOR, A. (1989). Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. In *Geometric Aspects of Functional Analysis* (1987–1988). *Lecture Notes in Math.* **1376** 64–104. Springer, Berlin. MR1008717

MILMAN, V. D. and PAJOR, A. (2003). Regularization of star bodies by random hyperplane cut off. *Studia Math.* **159** 247–261. MR2052221

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752

QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351

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

RUSH, A. J., FAVA, M., WISNIEWSKI, S. R., LAVORI, P. W., TRIVEDI, M. H., SACKEIM, H. A., THASE, M. E., NIERENBERG, A. A., QUITKIN, F. M., KASHNER, T. M. et al. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Control. Clin. Trials* **25** 119–142.

SHI, C., SONG, R. and LU, W. (2016). Robust learning for optimal treatment decision with NP-dimensionality. *Electron. J. Stat.* **10** 2894–2921. MR3557316

SHI, C., FAN, A., SONG, R. and LU, W. (2018). Supplement to "High-dimensional A-learning for optimal dynamic treatment regimes." DOI:10.1214/17-AOS1570SUPP.

TIBSHIRANI, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 273–282. MR2815776

WATKINS, C. J. C. H. and DAYAN, P. (1992). Q-Learning. *Mach. Learn.* **8** 279–292.

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018. MR3040007

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. MR3094445

ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898

ZHAO, Y.-Q., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **110** 583–598. MR3367249

ZHOU, S. (2009). Restricted eigenvalue conditions on subgaussian random matrices. Available at arxiv:0912.4045.

ZHOU, X., MAYER-HAMBLETT, N., KHAN, U. and KOSOROK, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *J. Amer. Statist. Assoc.* **112** 169–187. MR3646564

The Annals of Statistics
2018, Vol. 46, No. 3, 958–988
https://doi.org/10.1214/17-AOS1573

TEST FOR HIGH-DIMENSIONAL REGRESSION COEFFICIENTS USING REFITTED CROSS-VALIDATION VARIANCE ESTIMATION

BY HENGJIAN CUI[*,1], WENWEN GUO[*] AND WEI ZHONG[†,2]

Capital Normal University[*] and Xiamen University[†]

Testing a hypothesis for high-dimensional regression coefficients is of fundamental importance in the statistical theory and applications. In this paper, we develop a new test for the overall significance of coefficients in high-dimensional linear regression models based on an estimated U-statistics of order two. With the aid of the martingale central limit theorem, we prove that the asymptotic distributions of the proposed test are normal under two different distribution assumptions. Refitted cross-validation (RCV) variance estimation is utilized to avoid the overestimation of the variance and enhance the empirical power. We examine the finite-sample performances of the proposed test via Monte Carlo simulations, which show that the new test based on the RCV estimator achieves higher powers, especially for the sparse cases. We also demonstrate an application by an empirical analysis of a microarray data set on Yorkshire gilts.

REFERENCES

[1] BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

[2] CAI, T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870

[3] CHEN, S. X. and QIN, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

[4] CHEN, S. X., ZHANG, L. X. and ZHONG, P. S. (2010). Tests for high dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

[5] CUI, H., GUO, W. and ZHONG, W. (2018). Supplement to "Test for high-dimensional regression coefficients using refitted cross-validation variance estimation." DOI:10.1214/17-AOS1573SUPP.

[6] FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 37–65. MR2885839

[7] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and it oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

[8] FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322

[9] FANG, K. T., KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London. MR1071174

MSC2010 subject classifications. 62F03, 62H15.

Key words and phrases. High-dimensional regression, hypothesis testing, martingale central limit theorem, refitted cross-validation variance estimation, U-statistics.

[10] GOEMAN, J. J., FINOS, L. and VAN HOUWELINGEN, J. C. (2011). Testing against a high dimensional alternative in the generalized linear model: Asymptotic alpha-level control. *Biometrika* **98** 381–390. MR2806435

[11] GOEMAN, J. J., VAN DE GEER, S. and VAN HOUWELINGEN, J. C. (2006). Testing against a high dimensional alternative. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 477–493. MR2278336

[12] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

[13] LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. MR3010900

[14] LKHAGVADORJ, S., QU, L., CAI, W., COUTURE, O. P., BARB, C. R., HAUSMAN, G. J., NETTLETON, D., ANDERSON, L. L., DEKKERS, J. C. M. and TUGGLE, C. K. (2009). Microarray gene expression profiles of fasting induced changes in liver and adipose tissues of pigs expressing the melanocortin-4 receptor D298N variant. *Physiol. Genomics* **38** 98–111.

[15] RAO, C. R., TOUTEBURG, H., SHALABH and HEUMANN, C. (2008). *Linear Models and Generalizations*. Springer, New York. MR2370506

[16] SCHMIDT, R. (2001). Tail dependence for elliptically contoured distributions. *Math. Methods Oper. Res.* **55** 301–327. MR1919580

[17] SRIVASTAVA, M. S. and DU, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99** 386–402. MR2396970

[18] TIBSHIRANI, R. (1996). Regression shrinkage and selection via LASSO. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. MR1379242

[19] WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.* **110** 1658–1669. MR3449062

[20] WANG, S. and CUI, H. (2013). Generalized F test for high dimensional linear regression coefficients. *J. Multivariate Anal.* **117** 134–149. MR3053539

[21] WANG, S. and CUI, H. (2015). A new test for part of high dimensional regression coefficients. *J. Multivariate Anal.* **137** 187–203. MR3332807

[22] YATA, K. and AOSHIMA, M. (2013). Correlation tests for high-dimensional data using extended cross-data-matrix methodology. *J. Multivariate Anal.* **117** 313–331. MR3053550

[23] ZHANG, C. H. and HUANG, J. (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. MR2435448

[24] ZHONG, P. S. and CHEN, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.* **106** 260–274. MR2816719

[25] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

The Annals of Statistics
2018, Vol. 46, No. 3, 989–1017
https://doi.org/10.1214/17-AOS1575

ARE DISCOVERIES SPURIOUS? DISTRIBUTIONS OF MAXIMUM SPURIOUS CORRELATIONS AND THEIR APPLICATIONS

BY JIANQING FAN[*,†,1], QI-MAN SHAO[‡,2] AND WEN-XIN ZHOU[†,§,3]

Fudan University[*], *Princeton University*[†], *Chinese University of Hong Kong*[‡] *and University of California, San Diego*[§]

Over the last two decades, many exciting variable selection methods have been developed for finding a small group of covariates that are associated with the response from a large pool. Can the discoveries from these data mining approaches be spurious due to high dimensionality and limited sample size? Can our fundamental assumptions about the exogeneity of the covariates needed for such variable selection be validated with the data? To answer these questions, we need to derive the distributions of the maximum spurious correlations given a certain number of predictors, namely, the distribution of the correlation of a response variable Y with the best s linear combinations of p covariates $\mathbf{X}$, even when $\mathbf{X}$ and Y are independent. When the covariance matrix of $\mathbf{X}$ possesses the restricted eigenvalue property, we derive such distributions for both a finite s and a diverging s, using Gaussian approximation and empirical process techniques. However, such a distribution depends on the unknown covariance matrix of $\mathbf{X}$. Hence, we use the multiplier bootstrap procedure to approximate the unknown distributions and establish the consistency of such a simple bootstrap approach. The results are further extended to the situation where the residuals are from regularized fits. Our approach is then used to construct the upper confidence limit for the maximum spurious correlation and to test the exogeneity of the covariates. The former provides a baseline for guarding against false discoveries and the latter tests whether our fundamental assumptions for high-dimensional model selection are statistically valid. Our techniques and results are illustrated with both numerical examples and real data analysis.

REFERENCES

ARLOT, S., BLANCHARD, G. and ROQUAIN, E. (2010). Some nonasymptotic results on resampling in high dimension. I. Confidence regions. *Ann. Statist.* **38** 51–82. MR2589316

BARRETT, G. F. and DONALD, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica* **71** 71–104. MR1956856

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

BRUSCO, M. J. and STAHL, S. (2005). *Branch-and-Bound Applications in Combinatorial Data Analysis*. Springer, New York. MR2172059

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. MR2807761

MSC2010 subject classifications. Primary 62H10, 62H20; secondary 62E17, 62F03.
Key words and phrases. High dimension, spurious correlation, bootstrap, false discovery.

CAI, T., FAN, J. and JIANG, T. (2013). Distributions of angles in random packing on spheres. *J. Mach. Learn. Res.* **14** 1837–1864. MR3104497

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210

CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870

CHANG, J., ZHENG, C., ZHOU, W.-X. and ZHOU, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300–1310. MR3744543

CHATTERJEE, S. and BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33** 414–436. MR2157808

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461

DAVYDOV, YU. A., LIFSHITS, M. A. and SMORODINA, N. V. (1998). *Local Properties of Distributions of Stochastic Functionals. Translations of Mathematical Monographs* **173**. Amer. Math. Soc., Providence, RI. Translated from the 1995 Russian original by V. E. Nazaĭkinskiĭ and M. A. Shishkova. MR1604537

DUDOIT, S. and VAN DER LAAN, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics*. Springer, New York. MR2373771

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

FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 37–65. MR2885839

FAN, J., HAN, F. and LIU, H. (2014). Challenges of big data analysis. *Natl. Sci. Rev.* **1** 293–314.

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

FAN, J. and LIAO, Y. (2014). Endogeneity in high dimensions. *Ann. Statist.* **42** 872–917. MR3210990

FAN, J. and LV, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist. Sinica* **20** 101–148. MR2640659

FAN, J., SHAO, Q.-M. and ZHOU, W.-X. (2018). Supplement to "Are discoveries spurious? Distributions of maximum spurious correlations and their applications." DOI:10.1214/17-AOS1575SUPP.

FAN, J., XUE, L. and ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *Ann. Statist.* **42** 819–849. MR3210988

GOEMAN, J. J., VAN DE GEER, S. A. and VAN HOUWELINGEN, H. C. (2006). Testing against a high dimensional alternative. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 477–493. MR2278336

HANSEN, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64** 413–430. MR1375740

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York. MR2722294

SHAO, Q.-M. and ZHOU, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab.* **42** 623–648. MR3178469

STRANGER, B. E., NICA, A. C., FORREST, M. S., DIMAS, A., BIRD, C. P., BEAZLEY, C., INGLE, C. E., DUNNING, M., FLICEK, P., KOLLER, D., MONTGOMERY, S., TAVARÉ, S., DELOUKAS, P. and DERMITZAKIS, E. T. (2007). Population genomics of human gene expression. *Nat*. *Genet*. **39** 1217–1224.

THORGEIRSSON, T. E. et al. (2010). Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat*. *Genet*. **42** 448–453.

THORISSON, G. A., SMITH, A. V., KRISHNAN, L. and STEIN, L. D. (2005). The international HapMap project web site. *Genome Res*. **15** 1592–1593.

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

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. Springer, New York. MR1385671

VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* (Y. Eldar and G. Kutyniok, eds.) 210–268. Cambridge Univ. Press, Cambridge. MR2963170

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann*. *Statist*. **38** 894–942. MR2604701

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

The Annals of Statistics
2018, Vol. 46, No. 3, 1018–1049
https://doi.org/10.1214/17-AOS1576

ADAPTIVE ESTIMATION OF PLANAR CONVEX SETS

BY T. TONY CAI*,[1], ADITYANAND GUNTUBOYINA†,[2] AND YUTING WEI†

University of Pennsylvania and University of California, Berkeley†*

In this paper, we consider adaptive estimation of an unknown planar compact, convex set from noisy measurements of its support function. Both the problem of estimating the support function at a point and that of estimating the whole convex set are studied. For pointwise estimation, we consider the problem in a general nonasymptotic framework, which evaluates the performance of a procedure at each individual set, instead of the worst case performance over a large parameter space as in conventional minimax theory. A data-driven adaptive estimator is proposed and is shown to be optimally adaptive to every compact, convex set. For estimating the whole convex set, we propose estimators that are shown to adaptively achieve the optimal rate of convergence. In both of these problems, our analysis makes no smoothness assumptions on the boundary of the unknown convex set.

REFERENCES

[1] ALEXANDROFF, A. D. (1939). Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. *Leningrad State Univ. Annals* [*Uchenye Zapiski*] *Math. Ser.* **6** 3–35. MR0003051

[2] BARAUD, Y. and BIRGÉ, L. (2015). Rates of convergence of rho-estimators for sets of densities satisfying shape constraints. Preprint. Available at arXiv:1503.04427.

[3] BRUNEL, V.-E. (2014). Non-parametric estimation of convex bodies and convex polytopes. Ph.D. thesis, Univ. Pierre et Marie Curie–Paris VI; Univ. of Haifa.

[4] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc. Sympos.*, *Indiana Univ.*, *Bloomington*, *Ind.*, 1969) 177–197. Cambridge Univ. Press, London. MR0277070

[5] CAI, T. T. and LOW, M. G. (2015). A framework for estimation of convex functions. *Statist. Sinica* **25** 423–456. MR3379081

[6] CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119

[7] CAROLAN, C. and DYKSTRA, R. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *Canad. J. Statist.* **27** 557–566. MR1745821

[8] CATOR, E. (2011). Adaptivity and optimality of the monotone least-squares estimator. *Bernoulli* **17** 714–735. MR2787612

[9] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. MR3357878

[10] DEZA, M. M. and DEZA, E. (2009). *Encyclopedia of Distances*. Springer, Berlin. MR2538177

[11] FISHER, N. I., HALL, P., TURLACH, B. A. and WATSON, G. S. (1997). On the estimation of a convex set from noisy data on its support function. *J. Amer. Statist. Assoc.* **92** 84–91. MR1436100

MSC2010 subject classifications. Primary 62G07; secondary 52A20.

Key words and phrases. Adaptive estimation, circle convexity, convex set, Hausdorff distance, minimax rate of convergence, support function.

[12] GARDNER, R. J. (2006). *Geometric Tomography*, 2nd ed. *Encyclopedia of Mathematics and Its Applications* **58**. Cambridge Univ. Press, New York. MR2251886

[13] GARDNER, R. J. and KIDERLEN, M. (2009). A new algorithm for 3D reconstruction from support functions. *IEEE Trans. Pattern Anal. Mach. Intell.* **31** 556–562. MR2527431

[14] GARDNER, R. J., KIDERLEN, M. and MILANFAR, P. (2006). Convergence of algorithms for reconstructing convex bodies and directional measures. *Ann. Statist.* **34** 1331–1374. MR2278360

[15] GREGOR, J. and RANNOU, F. R. (2002). Three-dimensional support function estimation and application for projection magnetic resonance imaging. *Int. J. Imaging Syst. Technol.* **12** 43–50.

[16] GROENEBOOM, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* **11** 1016–1027. MR0714964

[17] GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* (*Berkeley, Calif.*, 1983). 539–555. Wadsworth, Belmont, CA. MR0822052

[18] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293

[19] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29** 1620–1652. MR1891741

[20] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742

[21] GUNTUBOYINA, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Ann. Statist.* **40** 385–411. MR3014311

[22] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621

[23] HANSON, D. L. and PLEDGER, G. (1976). Consistency in concave regression. *Ann. Statist.* **4** 1038–1050. MR0426273

[24] JANKOWSKI, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.* **42** 625–653. MR3210981

[25] LELE, A. S., KULKARNI, S. R. and WILLSKY, A. S. (1992). Convex-polygon estimation from support-line measurements and applications to target reconstruction from laser-radar data. *Journal of the Optical Society of America, Series A* **9** 1693–1714.

[26] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. MR0856411

[27] MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19** 741–759. MR1105842

[28] MCCLURE, D. E. and VITALE, R. A. (1975). Polygonal approximation of plane convex bodies. *J. Math. Anal. Appl.* **51** 326–358. MR0385714

[29] PRINCE, J. L. and WILLSKY, A. S. (1990). Reconstructing convex sets from support line measurements. *IEEE Trans. Pattern Anal. Mach. Intell.* **12** 377–389.

[30] SCHNEIDER, R. (1993). *Convex Bodies: The Brunn–Minkowski Theory*. *Encyclopedia of Mathematics and Its Applications* **44**. Cambridge Univ. Press, Cambridge. MR1216521

[31] STARK, H. and YANG, Y. (1998). *Vector Space Projections*. Wiley, New York.

[32] CAI, T. T., GUNTUBOYINA, A. and WEI, Y. (2018). Supplement to "Adaptive estimation of planar convex sets." DOI:10.1214/17-AOS1576SUPP.

[33] VITALE, R. A. (1979). Support functions of plane convex sets. Technical report, Claremont Graduate School, Claremont, CA.

[34] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann*. *Statist*. **9** 443–448. MR0606630

[35] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann*. *Statist*. **30** 528–555. MR1902898

The Annals of Statistics
2018, Vol. 46, No. 3, 1050–1076
https://doi.org/10.1214/17-AOS1577

CONSISTENCY OF AIC AND BIC IN ESTIMATING THE NUMBER OF SIGNIFICANT COMPONENTS IN HIGH-DIMENSIONAL PRINCIPAL COMPONENT ANALYSIS

BY ZHIDONG BAI[1], KWOK PUI CHOI[2] AND YASUNORI FUJIKOSHI[3]

Northeast Normal University, National University of Singapore and Hiroshima University

In this paper, we study the problem of estimating the number of significant components in principal component analysis (PCA), which corresponds to the number of dominant eigenvalues of the covariance matrix of p variables. Our purpose is to examine the consistency of the estimation criteria AIC and BIC based on the model selection criteria by Akaike [In *2nd International Symposium on Information Theory* (1973) 267–281, Akadémia Kiado] and Schwarz [*Estimating the dimension of a model* **6** (1978) 461–464] under a high-dimensional asymptotic framework. Using random matrix theory techniques, we derive sufficient conditions for the criterion to be strongly consistent for the case when the dominant population eigenvalues are bounded, and when the dominant eigenvalues tend to infinity. Moreover, the asymptotic results are obtained without normality assumption on the population distribution. Simulation studies are also conducted, and results show that the sufficient conditions in our theorems are essential.

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (B. N. Petrov and F. Csáki, eds.) 267–281, Budapest: Akadémia Kiado. MR0483125

BAI, Z. D., MIAO, B. Q. and RAO, C. R. (1990). Estimation of direction of arrival of signals: Asymptotic results. In *Advances in Spectrum Analysis and Array Processing* (S. Haykins, ed.) **2** 327–347. Prentice Hall, Englewood, Cliffs, NJ.

BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051

BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686

BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

FERRÉ, L. (1995). Selection of components in principal component analysis: A comparison of methods. *Comput. Statist. Data Anal.* **19** 669–682. MR1342614

FUJIKOSHI, Y. and SAKURAI, T. (2016a). Some properties of estimation criteria for dimensionality in principal component anlysis. *Amer. J. Math. Management Sci.* **35** 133–142.

MSC2010 subject classifications. Primary 62H12; secondary 62H30.

Key words and phrases. AIC, BIC, consistency, dimensionality, high-dimensional framework, number of significant components, principal component analysis, random matrix theory, signal processing, spiked model.

FUJIKOSHI, Y. and SAKURAI, T. (2016b). High-dimensional consistency of rank estimation criteria in multivariate linear model. *J. Multivariate Anal.* **149** 199–212. MR3507324

FUJIKOSHI, Y., SAKURAI, T. and YANAGIHARA, H. (2014). Consistency of high-dimensional AIC-type and C_p-type criteria in multivariate linear regression. *J. Multivariate Anal.* **123** 184–200. MR3130429

FUJIKOSHI, Y., ULYANOV, V. V. and SHIMIZU, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, NJ. MR2640807

FUJIKOSHI, Y., YAMADA, T., WATANABE, D. and SUGIYAMA, T. (2007). Asymptotic distribution of the LR statistic for equality of the smallest eigenvalues in high-dimensional principal component analysis. *J. Multivariate Anal.* **98** 2002–2008. MR2396951

GUNDERSON, B. K. and MUIRHEAD, R. J. (1997). On estimating the dimensionality in canonical correlation analysis. *J. Multivariate Anal.* **62** 121–136. MR1467877

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448

JOLLIFFE, I. T. (2002). *Principal Component Analysis*, 2nd ed. Springer, New York. MR2036084

JOLLIFFE, I. T., TRENDAFILOV, N. T. and UDDIN, M. (2003). A modified principal component technique based on the LASSO. *J. Comput. Graph. Statist.* **12** 531–547. MR2002634

KIM, Y., KWON, S. and CHOI, H. (2012). Consistent model selection criteria on high dimensions. *J. Mach. Learn. Res.* **13** 1037–1057. MR2930632

NISHII, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.* **12** 758–765. MR0740928

NISHII, R., BAI, Z. D. and KRISHNAIAH, P. R. (1988). Strong consistency of the information criterion for model selection in multivariate analysis. *Hiroshima Math. J.* **18** 451–462. MR0991240

PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

RAO, C. R. and RAO, M. B. (1998). *Matrix Algebra and Its Applications to Statistics and Econometrics*. World Scientific, River Edge, NJ. MR1660868

SCHOTT, J. R. (2006). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *J. Multivariate Anal.* **97** 827–843. MR2256563

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014

SHAO, J. (1997). An asymptotic theory for linear model selection. *Statist. Sinica* **7** 221–264. MR1466682

SHIBATA, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117–126. MR0403130

SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408

YANAGIHARA, H., WAKAKI, H. and FUJIKOSHI, Y. (2015). A consistency property of the AIC for multivariate linear models when the dimension and the sample size are large. *Electron. J. Stat.* **9** 869–897. MR3338666

YANG, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika* **92** 937–950. MR2234196

ZHAO, L. C., KRISHNAIAH, P. R. and BAI, Z. D. (1986). On detection of the number of signals in presence of white noise. *J. Multivariate Anal.* **20** 1–25. MR0862239

ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2006). Sparse principal component analysis. *J. Comput. Graph. Statist.* **15** 265–286. MR2252527

The Annals of Statistics
2018, Vol. 46, No. 3, 1077–1108
https://doi.org/10.1214/17-AOS1578

ON THE SYSTEMATIC AND IDIOSYNCRATIC VOLATILITY WITH LARGE PANEL HIGH-FREQUENCY DATA[1]

BY XIN-BING KONG

Nanjing Audit University

In this paper, we separate the integrated (spot) volatility of an individual Itô process into integrated (spot) systematic and idiosyncratic volatilities, and estimate them by aggregation of local factor analysis (localization) with large-dimensional high-frequency data. We show that, when both the sampling frequency n and the dimensionality p go to infinity and $p \geq C\sqrt{n}$ for some constant C, our estimators of High dimensional Itô process; common driving process; specific driving process, integrated High dimensional Itô process, common driving process, specific driving process, systematic and idiosyncratic volatilities are $\sqrt{n}$ ($n^{1/4}$ for spot estimates) consistent, the best rate achieved in estimating the integrated (spot) volatility which is readily identified even with univariate high-frequency data. However, when $Cn^{1/4} \leq p < C\sqrt{n}$, aggregation of $n^{1/4}$-consistent local estimates of systematic and idiosyncratic volatilities results in p-consistent (not $\sqrt{n}$-consistent) estimates of integrated systematic and idiosyncratic volatilities. Even more interesting, when $p < Cn^{1/4}$, the integrated estimate has the same convergence rate as the spot estimate, both being p-consistent. This reveals a distinctive feature from aggregating local estimates in the low-dimensional high-frequency data setting. We also present estimators of the integrated (spot) idiosyncratic volatility matrices as well as their inverse matrices under some sparsity assumption. We finally present a factor-based estimator of the inverse of the spot volatility matrix. Numerical studies including the Monte Carlo experiments and real data analysis justify the performance of our estimators.

REFERENCES

AÏT-SAHALIA, Y., MYKLAND, P. A. and ZHANG, L. (2016). How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Financ. Stud.* **18** 3519–416. DOI:10.1007/0-387-28359-5_1.

AÏT-SAHALIA, Y. and XIU, D. (2016). Principal component analysis of high frequency data. NBER Working Paper No. 21584. Available at http://www.nber.org/papers/w21584.

AÏT-SAHALIA, Y. and XIU, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *J. Econometrics*. To appear. DOI:10.1016/j.jeconom.2017.08.015.

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2003). Modeling and forecasting realized volatility. *Econometrica* **71** 579–625. MR1958138

BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171. MR1956857

MSC2010 subject classifications. Primary 62M05, 62G20; secondary 60J75, 60G20.
Key words and phrases. High dimensional Itô process, common driving process, specific driving process.

BAI, J. and NG, S. (2003). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76** 1481–1536. MR2468558

BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 253–280. MR1904704

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008

CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. MR2847949

CHAMBERLAIN, G. and ROTHSCHILD, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51** 1281–1304. MR0736050

CHEN, S. X. and XU, Z. (2014). On implied volatility for options—Some reasons to smile and more to correct. *J. Econometrics* **179** 1–15. MR3153645

CHRISTENSEN, K., KINNEBROCK, S. and PODOLSKIJ, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics* **159** 116–133. MR2720847

DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450

FAN, J., FURGER, A. and XIU, D. (2016). Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *J. Bus. Econom. Statist.* **34** 489–503. MR3547991

FAN, J., LI, Y. and YU, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Amer. Statist. Assoc.* **107** 412–428. MR2949370

FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M., and VETTER, M. (2013). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* **119** 2249–2276. MR2531091

JACOD, J. and ROSENBAUM, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Ann. Statist.* **41** 1462–1484. MR3113818

JACOD, J. and TODOROV, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann. Statist.* **42** 1029–1069. MR3210995

JING, B.-Y., KONG, X.-B. and LIU, Z. (2012). Modeling high-frequency financial data by pure jump processes. *Ann. Statist.* **40** 759–784. MR2933665

JING, B.-Y., KONG, X.-B., LIU, Z. and MYKLAND, P. (2012). On the jump activity index for semimartingales. *J. Econometrics* **166** 213–223. MR2862961

KIM, D., KONG, X. B., LI, C. and WANG, Y. (2015). Adaptive thresholding for large volatility matrix estimation based on high-frequency financial data. Unpublished manuscript. Available at https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxkb25nZ3l1a2ltMDMyOXxneDozZjNmMTQxZDNiNDc4OGNl.

KONG, X.-B. (2018). Supplement to "On the systematic and idiosyncratic volatility with large panel high-frequency data." DOI:10.1214/17-AOS1578SUPP.

KONG, X.-B., LIU, Z. and JING, B.-Y. (2015). Testing for pure-jump processes for high-frequency data. *Ann. Statist.* **43** 847–877. MR3325712

KONG, X. B. (2017). On the number of common factors with high-frequency data. *Biometrika* **104** 397–410.

LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938

MYKLAND, P. A., SHEPHARD, N. and SHEPHARD, K. (2012). Efficient and feasible inference for the components of financial variation using blocked multipower variation. Technical report. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2008690.

MYKLAND, P. A. and ZHANG, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* **77** 1403–1445. MR2561071

PELGER, M. (2016). Large-dimensional factor modeling based on high-frequency observations. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2584172.

QIU, Y. and CHEN, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.* **40** 1285–1314. MR3015026

ROSS, S. A. (1976). The arbitrage theory of capital asset pricing. *J. Econom. Theory* **13** 341–360. MR0429063

TAO, M., WANG, Y. and CHEN, X. (2013). Fast convergence rates in estimating large volatility matrices using high-frequency financial data. *Econometric Theory* **29** 838–856. MR3092465

TAO, M., WANG, Y. and ZHOU, H. H. (2013). Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *Ann. Statist.* **41** 1816–1864. MR3127850

TODOROV, V. and TAUCHEN, G. (2012a). The realized Laplace transform of volatility. *Econometrica* **80** 1105–1127. MR2963883

TODOROV, V. and TAUCHEN, G. (2012b). Inverse realized Laplace transforms for nonparametric volatility density estimation in jump-diffusions. *J. Amer. Statist. Assoc.* **107** 622–635. MR2980072

WANG, Y. and ZOU, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.* **38** 943–978. MR2604708

ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411. MR2236450

The Annals of Statistics
2018, Vol. 46, No. 3, 1109–1137
https://doi.org/10.1214/17-AOS1579

BALL DIVERGENCE: NONPARAMETRIC TWO SAMPLE TEST

BY WENLIANG PAN*,[1], YUAN TIAN*, XUEQIN WANG*,[2] AND HEPING ZHANG*,†,[3]

Sun Yat-sen University and Yale University*†

In this paper, we first introduce Ball Divergence, a novel measure of the difference between two probability measures in separable Banach spaces, and show that the Ball Divergence of two probability measures is zero if and only if these two probability measures are identical without any moment assumption. Using Ball Divergence, we present a metric rank test procedure to detect the equality of distribution measures underlying independent samples. It is therefore robust to outliers or heavy-tail data. We show that this multivariate two sample test statistic is consistent with the Ball Divergence, and it converges to a mixture of χ^2 distributions under the null hypothesis and a normal distribution under the alternative hypothesis. Importantly, we prove its consistency against a general alternative hypothesis. Moreover, this result does not depend on the ratio of the two imbalanced sample sizes, ensuring that can be applied to imbalanced data. Numerical studies confirm that our test is superior to several existing tests in terms of Type I error and power. We conclude our paper with two applications of our method: one is for virtual screening in drug development process and the other is for genome wide expression analysis in hormone replacement therapy.

REFERENCES

[1] ANDERSEN, L., FRIIS, S., HALLAS, J., RAVN, P., SCHRØDER, H. D. and GAIST, D. (2014). Hormone replacement therapy increases the risk of cranial meningioma. *Neurology* **82** P3.325.

[2] BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

[3] BOGACHEV, V. I. (2007). *Measure Theory, Vol. I*. Springer, Berlin. MR2267655

[4] CHEN, L., DOU, W. W. and QIAO, Z. (2013). Ensemble subsampling for imbalanced multivariate two-sample tests. *J. Amer. Statist. Assoc.* **108** 1308–1323. MR3174710

[5] CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

[6] CHIU, S. N. and LIU, K. I. (2009). Generalized Cramér–von Mises goodness-of-fit tests for multivariate distributions. *Comput. Statist. Data Anal.* **53** 3817–3834. MR2749926

[7] DENTI, L. (2009). The hormone replacement therapy (HRT) of menopause: Focus on cardiovascular implications. *Acta Biomed. Atenei Parmensis* **81** 73–76.

[8] DUMEAUX, V., JOHANSEN, J., BORRESENDALE, A. L. and LUND, E. (2006). Gene expression profiling of whole-blood samples from women exposed to hormone replacement therapy. *Mol. Cancer Ther.* **5** 868–876.

MSC2010 subject classifications. Primary 62H15; secondary 62G10.
Key words and phrases. Ball Divergence, Banach space, metric rank, permutation procedure.

[9] GEHAN, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52** 650–653. MR0207131

[10] GRETTON, A., BORGWARDT, K. M., RASCH, M., SCHÖLKOPF, B. and SMOLA, A. J. (2006). A kernel method for the two-sample-problem. In *Advances in Neural Information Processing Systems* 513–520.

[11] HOU, N., HONG, S., WANG, W., OLOPADE, O. I., DIGNAM, J. J. and HUO, D. (2013). Hormone replacement therapy and breast cancer: Heterogeneous risks by race, weight, and breast density. *J. Natl. Cancer Inst.* **105** 1365–1372.

[12] JACKSON, S. and MAULDIN, R. D. (1999). On the σ-class generated by open balls. *Math. Proc. Cambridge Philos. Soc.* **127** 99–108. MR1692499

[13] JUSTEL, A., PEÑA, D. and ZAMAR, R. (1997). A multivariate Kolmogorov–Smirnov test of goodness of fit. *Statist. Probab. Lett.* **35** 251–259. MR1484961

[14] KOSOROK, M. R. and MA, S. (2007). Marginal asymptotics for the "large p, small n" paradigm: With applications to microarray data. *Ann. Statist.* **35** 1456–1486. MR2351093

[15] LEE, A. J. (1990). *U-Statistics: Theory and Practice. Statistics: Textbooks and Monographs* **110**. Dekker, Inc., New York. MR1075417

[16] NEUHAUS, G. (1977). Functional limit theorems for U-statistics in the degenerate case. *J. Multivariate Anal.* **7** 424–439. MR0455084

[17] PREISS, D. and TIŠER, J. (1991). Measures in Banach spaces are determined by their values on balls. *Mathematika* **38** 391–397. MR1147839

[18] RESHEF, D. N., RESHEF, Y. A., FINUCANE, H. K., GROSSMAN, S. R., MCVEAN, G., TURNBAUGH, P. J., LANDER, E. S., MITZENMACHER, M. and SABETI, P. C. (2011). Detecting novel associations in large data sets. *Science* **334** 1518–1524.

[19] SCHIERZ, A. C. (2009). Virtual screening of bioassay data. *J. Cheminform.* **1** 21.

[20] SCHOENBERG, I. J. (1937). On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Ann. of Math.* (2) **38** 787–793. MR1503370

[21] SCHOENBERG, I. J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.* **44** 522–536. MR1501980

[22] SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866

[23] SZÉKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *InterStat* **5**.

[24] VAN DER LAAN, M. J. and BRYAN, J. (2001). Gene expression analysis with the parametric bootstrap. *Biostatistics* **2** 445–461.

[25] ZHANG, Q., PAN, W. and WANG, X. (2017). Distribution free multiple change point detection in multivariate time series. Preprint.

The Annals of Statistics
2018, Vol. 46, No. 3, 1138–1166
https://doi.org/10.1214/17-AOS1580

A SMOOTH BLOCK BOOTSTRAP FOR QUANTILE REGRESSION WITH TIME SERIES

BY KARL B. GREGORY[1], SOUMENDRA N. LAHIRI[2] AND DANIEL J. NORDMAN[3]

University of South Carolina, North Carolina State University and Iowa State University

Quantile regression allows for broad (conditional) characterizations of a response distribution beyond conditional means and is of increasing interest in economic and financial applications. Because quantile regression estimators have complex limiting distributions, several bootstrap methods for the independent data setting have been proposed, many of which involve smoothing steps to improve bootstrap approximations. Currently, no similar advances in smoothed bootstraps exist for quantile regression with dependent data. To this end, we establish a smooth tapered block bootstrap procedure for approximating the distribution of quantile regression estimators for time series. This bootstrap involves two rounds of smoothing in resampling: individual observations are resampled via kernel smoothing techniques and resampled data blocks are smoothed by tapering. The smooth bootstrap results in performance improvements over previous unsmoothed versions of the block bootstrap as well as normal approximations based on Powell's kernel variance estimator, which are common in application. Our theoretical results correct errors in proofs for earlier and simpler versions of the (unsmoothed) moving blocks bootstrap for quantile regression and broaden the validity of block bootstraps for this problem under weak conditions. We illustrate the smooth bootstrap through numerical studies and examples.

REFERENCES

[1] ARCONES, M. A. and GINÉ, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990) (R. LePage and L. Billard, eds.) 13–47. Wiley, New York. MR1197777

[2] BUCHINSKY, M. (1994). Changes in the U.S. wage structure 1963–1987: Application of qunatile regression. *Econometrica* **62** 405–458.

[3] DE ANGELIS, D., HALL, P. and YOUNG, G. A. (1993). A note on coverage error of bootstrap confidence intervals for quantiles. *Math*. *Proc*. *Cambridge Philos*. *Soc*. **114** 517–531. MR1235999

[4] DE ANGELIS, D., HALL, P. and YOUNG, G. A. (1993). Analytical and bootstrap approximations to estimator distributions in L^1 regression. *J. Amer. Statist*. *Assoc*. **88** 1310–1316. MR1245364

[5] DOUKHAN, P. (1994). *Mixing*: *Properties and Examples*. *Lecture Notes in Statistics* **85**. Springer, New York. MR1312160

MSC2010 subject classifications. Primary 62G09; secondary 62G20, 62J05, 62M10.

Key words and phrases. Kernel smoothing, jackknife after bootstrap, moving blocks, tapering, value at risk.

[6] ENGLE, R. F. and MANGANELLI, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econom. Statist.* **22** 367–381. MR2091566

[7] FENG, X., HE, X. and HU, J. (2011). Wild bootstrap for quantile regression. *Biometrika* **98** 995–999. MR2860339

[8] FITZENBERGER, B. (1998). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *J. Econometrics* **82** 235–287. MR1613422

[9] GAGLIANONE, W. P., LIMA, L. R., LINTON, O. and SMITH, D. R. (2011). Evaluating value-at-risk models via quantile regression. *J. Bus. Econom. Statist.* **29** 150–160. MR2789438

[10] GREGORY, K. B., LAHIRI, S. N. and NORDMAN, D. J. (2015). A smooth block bootstrap for statistical functionals and time series. *J. Time Series Anal.* **36** 442–461. MR3343010

[11] GREGORY, K. B., LAHIRI, S. N. and NORDMAN, D. J. (2018). Supplement to "A smooth block bootstrap for quantile regression with time series." DOI:10.1214/17-AOS1580SUPP.

[12] GUTENBRUNNER, C., JUREČKOVÁ, J., KOENKER, R. and PORTNOY, S. (1993). Tests of linear hypotheses based on regression rank scores. *J. Nonparametr. Stat.* **2** 307–331. MR1256383

[13] HAHN, J. (1995). Bootstrapping quantile regression estimators. *Econometric Theory* **11** 105–121. MR1325103

[14] HASAN, M. N. and KOENKER, R. W. (1997). Robust rank tests of the unit root hypothesis. *Econometrica* **65** 133–161. MR1433687

[15] HE, X., ZHU, Z.-Y. and FUNG, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89** 579–590. MR1929164

[16] HOROWITZ, J. L. (1998). Bootstrap methods for median regression models. *Econometrica* **66** 1327–1351. MR1654307

[17] 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

[18] KATO, K. (2012). Asymptotic normality of Powell's kernel estimator. *Ann. Inst. Statist. Math.* **64** 255–273. MR2878905

[19] KOENKER, R. (1994). Confidence intervals for regression quantiles. In *Asymptotic Statistics* (*Prague*, 1993) (P. Mandl and M. Hušková, eds.) 349–359. Physica, Heidelberg. MR1311953

[20] KOENKER, R. (2013). quantreg: Quantile regression. R package version 5.05.

[21] KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[22] KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147

[23] LAHIRI, S. N. (2002). On the jackknife-after-bootstrap method for dependent data and its consistency properties. *Econometric Theory* **18** 79–98. MR1885351

[24] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York. MR2001447

[25] LAHIRI, S. N., FURUKAWA, K. and LEE, Y.-D. (2007). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Stat. Methodol.* **4** 292–321. MR2380557

[26] LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing, MI*, 1990) 225–248. Wiley, New York. MR1197787

[27] PAPARODITIS, E. and POLITIS, D. N. (2001). Tapered block bootstrap. *Biometrika* **88** 1105–1119. MR1872222

[28] PARZEN, M. I., WEI, L. J. and YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81** 341–350. MR1294895

[29] POLLARD, D. (1985). New ways to prove central limit theorems. *Econometric Theory* **1** 295–313.
[30] POWELL, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (*Durham*, *NC*, 1988) 357–384. Cambridge Univ. Press, Cambridge. MR1174980
[31] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165
[32] SHAO, X. (2010). Extended tapered block bootstrap. *Statist*. *Sinica* **20** 807–821. MR2682643
[33] SHEATHER, S. J. and JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **53** 683–690. MR1125725
[34] STOFFER, D. (2014). astsa: Applied statistical time series analysis. R package version 1.3.
[35] UMANTSEV, L. and CHERNOZHUKOV, V. (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empir*. *Econom*. **26** 271–292.
[36] VAN DEN GOORBERGH, R. W. J. and VLAAR, P. J. G. (1999). Value-at-risk analysis of stock returns historical simulation, variance techniques or tail index estimation? DNB Staff Reports (discontinued) No. 40, Netherlands Central Bank. Available at http://ideas.repec.org/p/dnb/staffs/40.html.
[37] WEISS, A. A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory* **7** 46–68. MR1101211
[38] ZHOU, Z. and SHAO, X. (2013). Inference for linear models with dependent errors. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **75** 323–343. MR3021390

The Annals of Statistics
2018, Vol. 46, No. 3, 1167–1196
https://doi.org/10.1214/17-AOS1581

ASYMPTOTIC DISTRIBUTION-FREE TESTS FOR SEMIPARAMETRIC REGRESSIONS WITH DEPENDENT DATA

BY JUAN CARLOS ESCANCIANO[1], JUAN CARLOS PARDO-FERNÁNDEZ[2] AND INGRID VAN KEILEGOM[3]

Indiana University, Universidade de Vigo and KU Leuven

This article proposes a new general methodology for constructing nonparametric and semiparametric Asymptotically Distribution-Free (ADF) tests for semiparametric hypotheses in regression models for possibly dependent data coming from a strictly stationary process. Classical tests based on the difference between the estimated distributions of the restricted and unrestricted regression errors are not ADF. In this article, we introduce a novel transformation of this difference that leads to ADF tests with well-known critical values. The general methodology is illustrated with applications to testing for parametric models against nonparametric or semiparametric alternatives, and semiparametric constrained mean–variance models. Several Monte Carlo studies and an empirical application show that the finite sample performance of the proposed tests is satisfactory in moderate sample sizes.

REFERENCES

[1] BEKAERT, G. and HARVEY, C. R. (1995). Time-varying world market integration. *J. Finance* **50** 403–444.

[2] CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics* (J. J. Heckman and E. E. Leamer, eds.) **6** 5549–5632.

[3] DEDECKER, J. and LOUHICHI, S. (2002). Maximal inequalities and empirical central limit theorems. In *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch and M. Sørensen, eds.) 137–159. Birkhäuser, Boston, MA. MR1958779

[4] DELGADO, M. A. and GONZÁLEZ MANTEIGA, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.* **29** 1469–1507. MR1873339

[5] DETTE, H., MARCHLEWSKI, M. and WAGENER, J. (2012). Testing for a constant coefficient of variation in nonparametric regression by empirical processes. *Ann. Inst. Statist. Math.* **64** 1045–1070. MR2960957

[6] DETTE, H., NEUMEYER, N. and VAN KEILEGOM, I. (2007). A new test for the parametric form of the variance function in non-parametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 903–917. MR2368576

[7] DETTE, H., PARDO-FERNÁNDEZ, J. C. and VAN KEILEGOM, I. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scand. J. Stat.* **36** 782–799. MR2573308

[8] DE SANTIS, G. and GERARD, B. (1997). International asset pricing and portfolio diversification with time-varying risk. *J. Finance* **52** 1881–1912.

[9] DOUKHAN, P. (1994). *Mixing. Lecture Notes in Statistics: Properties and Examples* **85**. Springer, New York. MR1312160

MSC2010 subject classifications. 62E20, 62G08, 62G20, 62H15.

Key words and phrases. Beta-mixing, error distribution, goodness-of-fit tests, local polynomial estimation, nonparametric regression.

[10] EINMAHL, U. and MASON, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33** 1380–1403. MR2195639

[11] ESCANCIANO, J. C. (2009). On the lack of power of omnibus specification tests. *Econometric Theory* **25** 162–194. MR2472049

[12] ESCANCIANO, J. C., JACHO-CHÁVEZ, D. T. and LEWBEL, A. (2014). Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *J. Econometrics* **178** 426–443. MR3132442

[13] ESCANCIANO, J. C., PARDO-FERNÁNDEZ, J. C. and VAN KEILEGOM, I. (2017). Semiparametric estimation of risk-return relationships. *J. Bus. Econom. Statist.* **35** 40–52. MR3591536

[14] ESCANCIANO, J. C., PARDO-FERNÁNDEZ, J. C. and VAN KEILEGOM, I. (2018). Supplement to "Asymptotic distribution-free tests for semiparametric regressions with dependent data." DOI:10.1214/17-AOS1581SUPP.

[15] FAN, J. and YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660. MR1665822

[16] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York. MR1964455

[17] FERSON, W. (1989). Changes in expected security returns, risk and level of interest rates. *J. Finance* **44** 1191–1217.

[18] FERSON, W., FOERSTER, S. R. and KEIM, D. B. (1993). General tests of latent variable models and mean–variance spanning. *J. Finance* **48** 131–156.

[19] GONZÁLEZ-MANTEIGA, W. and CRUJEIRAS, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST* **22** 361–411. MR3093195

[20] HANSEN, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24** 726–748. MR2409261

[21] HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171

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

[23] HARVEY, C. R. (1989). Time-varying conditional covariances in tests of asset pricing models. *J. Financ. Econom.* **24** 289–317.

[24] JANSSEN, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.* **28** 239–253. MR1762910

[25] KHMALADZE, E. V. and KOUL, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32** 995–1034. MR2065196

[26] KOUL, H. L. and STUTE, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27** 204–236. MR1701108

[27] LEE, T.-H., TU, Y. and ULLAH, A. (2015). Forecasting equity premium: Global historical average versus local historical average and constraints. *J. Bus. Econom. Statist.* **33** 393–402. MR3372666

[28] MASRY, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Series Anal.* **17** 571–599. MR1424907

[29] MERTON, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica* **41** 867–887. MR0441271

[30] MERTON, R. C. (1980). On estimating the expected return on the market. An explanatory investigation. *J. Financ. Econom.* **8** 323–361.

[31] NEUMEYER, N. and VAN KEILEGOM, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.* **101** 1067–1078. MR2595293

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

[33] ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762

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

[35] STUTE, W., XU, W. L. and ZHU, L. X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika* **95** 451–467. MR2521592

[36] TANG, Y. and WHITELAW, R. F. (2011). Time-varying sharpe ratios and market timing. *Quarterly J. Finance* **1** 465–493.

[37] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

[38] VAN KEILEGOM, I., GONZÁLEZ MANTEIGA, W. and SÁNCHEZ SELLERO, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST* **17** 401–415. MR2434335

The Annals of Statistics
2018, Vol. 46, No. 3, 1197–1224
https://doi.org/10.1214/17-AOS1582

GRADIENT-BASED STRUCTURAL CHANGE DETECTION FOR NONSTATIONARY TIME SERIES M-ESTIMATION

BY WEICHI WU AND ZHOU ZHOU[1]

University College London and University of Toronto

We consider structural change testing for a wide class of time series M-estimation with nonstationary predictors and errors. Flexible predictor-error relationships, including exogenous, state-heteroscedastic and autoregressive regressions and their mixtures, are allowed. New uniform Bahadur representations are established with nearly optimal approximation rates. A CUSUM-type test statistic based on the gradient vectors of the regression is considered. In this paper, a simple bootstrap method is proposed and is proved to be consistent for M-estimation structural change detection under both abrupt and smooth nonstationarity and temporal dependence. Our bootstrap procedure is shown to have certain asymptotically optimal properties in terms of accuracy and power. A public health time series dataset is used to illustrate our methodology, and asymmetry of structural changes in high and low quantiles is found.

REFERENCES

[1] ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

[2] ARCONES, M. A. (1996). The Bahadur–Kiefer representation of L_p regression estimators. *Econometric Theory* **12** 257–283. MR1395032

[3] AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012

[4] BABU, G. J. (1989). Strong representations for LAD estimators in linear models. *Probab. Theory Related Fields* **83** 547–558. MR1022629

[5] BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Stat.* **37** 577–580. MR0189095

[6] BAI, J. (1996). Testing for parameter constancy in linear regressions: An empirical distribution function approach. *Econometrica* **64** 597–622. MR1385559

[7] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

[8] BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR0853051

[9] BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192. MR0378310

[10] CAI, Z., FAN, J. and LI, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* **95** 888–902. MR1804446

[11] DETTE, H., WU, W. and ZHOU, Z. (2015). Change point analysis of second order characteristics in non-stationary time series. Available at arXiv:1503.08610.

MSC2010 subject classifications. 62J20, 62M10, 62G09, 62G10.
Key words and phrases. M-estimation, piecewise local stationarity, bootstrap, structural change.

[12] ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121

[13] FAN, J. and ZHANG, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497

[14] FAN, J. and ZHANG, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Stat.* **27** 715–731. MR1804172

[15] HANSEN, B. E. (2000). Testing for structural change in conditional models. *J. Econometrics* **97** 93–115. MR1788819

[16] HE, X. and ZHU, L.-X. (2003). A lack-of-fit test for quantile regression. *J. Amer. Statist. Assoc.* **98** 1013–1022. MR2041489

[17] JUHL, T. and XIAO, Z. (2009). Tests for changing mean with monotonic power. *J. Econometrics* **148** 14–24. MR2494814

[18] KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Stat.* **38** 1323–1342. MR0217844

[19] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York. MR2001447

[20] MCCABE, B. P. M. and HARRISON, M. J. (1980). Testing the constancy of regression relationships over time using least squares residuals. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **29** 142–148.

[21] PLOBERGER, W. and KRÄMER, W. (1992). The CUSUM test with OLS residuals. *Econometrica* **60** 271–285. MR1162619

[22] POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York. MR1707286

[23] PORTNOY, S. (1991). Asymptotic behavior of regression quantiles in nonstationary, dependent cases. *J. Multivariate Anal.* **38** 100–113. MR1128939

[24] POWELL, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (*Durham, NC*, 1988). *Internat. Sympos. Econom. Theory Econometrics* 357–384. Cambridge Univ. Press, Cambridge. MR1174980

[25] PRÁŠKOVÁ, Z. and CHOCHOLA, O. (2014). M-procedures for detection of a change under weak dependence. *J. Statist. Plann. Inference* **149** 60–76. MR3199894

[26] QU, Z. (2008). Testing for structural change in regression quantiles. *J. Econometrics* **146** 170–184. MR2459652

[27] SU, L. and WHITE, H. (2010). Testing structural change in partially linear models. *Econometric Theory* **26** 1761–1806. MR2738016

[28] TONG, H. (1990). *Nonlinear Time Series*: *A Dynamical System Approach*. *Oxford Statistical Science Series* **6**. Oxford Univ. Press, New York. MR1079320

[29] WU, W. and ZHOU, Z. (2018). Supplement to "Gradient-based structural change detection for nonstationary time series M-estimation." DOI:10.1214/17-AOS1582SUPP.

[30] WU, W. B. (2007). M-estimation of linear models with dependent errors. *Ann. Statist.* **35** 495–521. MR2336857

[31] ZHANG, T. and WU, W. B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40** 1376–1402. MR3015029

[32] ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655

[33] ZHOU, Z. and WU, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 513–531. MR2758526

The Annals of Statistics
2018, Vol. 46, No. 3, 1225–1254
https://doi.org/10.1214/17-AOS1583

MODERATE DEVIATIONS AND NONPARAMETRIC INFERENCE FOR MONOTONE FUNCTIONS

BY FUQING GAO[1], JIE XIONG[2] AND XINGQIU ZHAO[3]

Wuhan University, University of Macau and The Hong Kong Polytechnic University

This paper considers self-normalized limits and moderate deviations of nonparametric maximum likelihood estimators for monotone functions. We obtain their self-normalized Cramér-type moderate deviations and limit distribution theorems for the nonparametric maximum likelihood estimator in the current status model and the Grenander-type estimator. As applications of the results, we present a new procedure to construct asymptotical confidence intervals and asymptotical rejection regions of hypothesis testing for monotone functions. The theoretical results can guarantee that the new test has the probability of type II error tending to 0 exponentially. Simulation studies also show that the new nonparametric test works well for the most commonly used parametric survival functions such as exponential and Weibull survival distributions.

REFERENCES

ABREVAYA, J. and HUANG, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245

BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693

BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743

BANERJEE, M. and WELLNER, J. A. (2005a). Confidence intervals for current status data. *Scand. J. Stat.* **32** 405–424. MR2204627

BANERJEE, M. and WELLNER, J. A. (2005b). Score statistics for current status data: Comparisons with likelihood ratio and Wald statistics. *Int. J. Biostat.* **1** Article ID 3. MR2232228

CHANG, J., SHAO, Q.-M. and ZHOU, W.-X. (2016). Cramér-type moderate deviations for Studentized two-sample U-statistics with applications. *Ann. Statist.* **44** 1931–1956. MR3546439

DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036

DE LA PEÑA, V. H., LAI, T. L. and SHAO, Q.-M. (2009). *Self-Normalized Processes*: *Limit Theory and Statistical Applications*. Springer, Berlin. MR2488094

DÜMBGEN, L., WELLNER, J. A. and WOLFF, M. (2016). A law of the iterated logarithm for Grenander's estimator. *Stochastic Process. Appl.* **126** 3854–3864. MR3565482

DUROT, C. (2002). Sharp asymptotics for isotonic regression. *Probab. Theory Related Fields* **122** 222–240. MR1894068

MSC2010 subject classifications. Primary 60F10, 62G20; secondary 62G07.

Key words and phrases. Grenander estimator, interval censored data, large deviations, moderate deviations, nonparametric MLE, self-normalized limit, strong approximation, Talagrand inequality.

DUROT, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. MR2341699

DUROT, C., KULIKOV, V. N. and LOPUHAÄ, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *Ann. Statist.* **40** 1578–1608. MR3015036

GAO, F. (2003). Moderate deviations and large deviations for kernel density estimators. *J. Theoret. Probab.* **16** 401–418. MR1982035

GAO, F., XIONG, J. and ZHAO, X. (2018). Supplement to "Moderate deviations and nonparametric inference for monotone functions." DOI:10.1214/17-AOS1583SUPP.

GAO, F. and ZHAO, X. (2011). Delta method in large deviations and moderate deviations for estimators. *Ann. Statist.* **39** 1211–1240. MR2816352

GINÉ, E. and GUILLOU, A. (2001). On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 503–522. MR1876841

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* (*Berkeley, Calif.*, 1983) 539–555. Wadsworth, Belmont, CA. MR0822052

GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR0981568

GROENEBOOM, P. (1996). Lectures on inverse problems. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1994). *Lecture Notes in Math.* **1648** 67–164. Springer, Berlin. MR1600884

GROENEBOOM, P. (2014). Maximum smoothed likelihood estimators for the interval censoring model. *Ann. Statist.* **42** 2092–2137. MR3262478

GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.* **27** 1316–1347. MR1740109

GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators*, *Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293

GROENEBOOM, P. and JONGBLOED, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43** 2019–2054. MR3375875

GROENEBOOM, P., JONGBLOED, G. and WITTE, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.* **38** 352–387. MR2589325

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321

GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400. MR1939706

HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1998). An extremal limit theorem for the argmax process of Brownian motion minus a parabolic drift. *Extremes* **1** 215–240. MR1814624

HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. MR1334065

LIU, W. and SHAO, Q.-M. (2010). Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series. *Ann. Statist.* **38** 1913–1935. MR2662363

LIU, W. and SHAO, Q.-M. (2013). A Carmér moderate deviation theorem for Hotelling's T^2-statistic with applications to global tests. *Ann. Statist.* **41** 296–322. MR3059419

MAUMY, M. (2004). Strong approximations for the compound empirical process. *Ann. I.S.U.P.* **48** 69–83. MR2083781

PETROV, V. V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und ihrer Grenzgebiete* **82**. Springer, New York. Translated from the Russian by A. A. Brown. MR0388499

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

SUN, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer, New York. MR2287318

TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

The Annals of Statistics
2018, Vol. 46, No. 3, 1255–1287
https://doi.org/10.1214/17-AOS1584

UNIFORM ASYMPTOTIC INFERENCE AND THE BOOTSTRAP AFTER MODEL SELECTION

BY RYAN J. TIBSHIRANI[*,1], ALESSANDRO RINALDO[*], ROB TIBSHIRANI[†,2] AND LARRY WASSERMAN[*]

Carnegie Mellon University[*] *and Stanford University*[†]

Recently, Tibshirani et al. [*J. Amer. Statist. Assoc.* **111** (2016) 600–620] proposed a method for making inferences about parameters defined by model selection, in a typical regression setting with normally distributed errors. Here, we study the large sample properties of this method, without assuming normality. We prove that the test statistic of Tibshirani et al. (2016) is asymptotically valid, as the number of samples n grows and the dimension d of the regression problem stays fixed. Our asymptotic result holds uniformly over a wide class of nonnormal error distributions. We also propose an efficient bootstrap version of this test that is provably (asymptotically) conservative, and in practice, often delivers shorter intervals than those from the original normality-based approach. Finally, we prove that the test statistic of Tibshirani et al. (2016) does not enjoy uniform validity in a high-dimensional setting, when the dimension d is allowed grow.

REFERENCES

BACHOC, F., LEEB, H. and POTSCHER, B. (2014). Valid confidence intervals for post-model-selection predictors. Available at arXiv:1412.4605.

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

CHOI, Y., TAYLOR, J. and TIBSHIRANI, R. (2014). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. Available at arXiv:1410.8260.

DONOHO, D. L. (1988). One-sided inference about functionals of a density. *Ann. Statist.* **16** 1390–1420. MR0964930

FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. Available at arXiv:1410.2597.

HYUN, S., G'SELL, M. and TIBSHIRANI, R. J. (2016). Exact post-selection inference for change-point detection and other generalized lasso problems. Available at arXiv:1606.03552.

KASY, M. (2015). Uniformity and the delta method. Unpublished manuscript.

LEE, J. and TAYLOR, J. (2014). Exact post model selection inference for marginal screening. *Adv. Neural Inf. Process. Syst.* **27** 136–144.

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

LEEB, H. and PÖTSCHER, B. M. (2003). The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory* **19** 100–142. MR1965844

MSC2010 subject classifications. 62F05, 62F35, 62J05, 62J07.

Key words and phrases. Post-selection inference, selective inference, asymptotics, bootstrap, forward stepwise regression, lasso.

LEEB, H. and PÖTSCHER, B. M. (2006). Can one estimate the conditional distribution of post-model-selection estimators? *Ann. Statist.* **34** 2554–2591. MR2291510

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

LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. MR3210970

LOFTUS, J. and TAYLOR, J. (2014). A significance test for forward stepwise model selection. Available at arXiv:1405.3920.

O'HAGAN, A. and LEONARD, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika* **63** 201–203. MR0428571

REID, S., TAYLOR, J. and TIBSHIRANI, R. (2017). Post-selection point and interval estimation of signal sizes in Gaussian samples. *Canad. J. Statist.* **45** 128–148. MR3646193

TAYLOR, J. E., LOFTUS, J. R. and TIBSHIRANI, R. J. (2016). Inference in adaptive regression via the Kac–Rice formula. *Ann. Statist.* **44** 743–770. MR3476616

TIAN, X. and TAYLOR, J. (2017). Asymptotics of selective inference. *Scand. J. Stat.* **44** 480–499.

TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. MR3538689

TIBSHIRANI, R. J., RINALDO, A., TIBSHIRANI, R. and WASSERMAN, L. (2018). Supplement to "Uniform asymptotic inference and the bootstrap after model selection." DOI:10.1214/17-AOS1584SUPP.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

WASSERMAN, L. (2014). Discussion: "A significance test for the lasso" [MR3210970]. *Ann. Statist.* **42** 501–508. MR3210975

The Annals of Statistics
2018, Vol. 46, No. 3, 1288–1317
https://doi.org/10.1214/17-AOS1585

DETECTION THRESHOLDS FOR THE β-MODEL ON SPARSE GRAPHS

BY RAJARSHI MUKHERJEE, SUMIT MUKHERJEE[1] AND SUBHABRATA SEN[2]

University of California, Berkeley, Columbia University and Microsoft Research

In this paper, we study sharp thresholds for detecting sparse signals in β-models for potentially sparse random graphs. The results demonstrate interesting interplay between graph sparsity, signal sparsity and signal strength. In regimes of moderately dense signals, irrespective of graph sparsity, the detection thresholds mirror corresponding results in independent Gaussian sequence problems. For sparser signals, extreme graph sparsity implies that all tests are asymptotically powerless, irrespective of the signal strength. On the other hand, sharp detection thresholds are obtained, up to matching constants, on denser graphs. The phase transitions mentioned above are sharp. As a crucial ingredient, we study a version of the higher criticism test which is provably sharp up to optimal constants in the regime of sparse signals. The theoretical results are further verified by numerical simulations.

REFERENCES

ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. MR2722464

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877

ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. MR2246369

ARIAS-CASTRO, E. and VERZELEN, N. (2013). Community detection in random networks. Available at arXiv:1302.7099.

ARIAS-CASTRO, E. and WANG, M. (2015). The sparse Poisson means model. *Electron. J. Stat.* **9** 2170–2201. MR3406276

ARIAS-CASTRO, E., CANDÈS, E. J., HELGASON, H. and ZEITOUNI, O. (2008). Searching for a trail of evidence in a maze. *Ann. Statist.* **36** 1726–1757. MR2435454

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

BARNETT, I., MUKHERJEE, R. and LIN, X. (2017). The generalized higher criticism for testing SNP-set effects in genetic association studies. *J. Amer. Statist. Assoc.* **112** 64–76. MR3646553

BARVINOK, A. and HARTIGAN, J. A. (2013). The number of graphs and a random graph with a given degree sequence. *Random Structures Algorithms* **42** 301–348. MR3039682

BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868

BLITZSTEIN, J. and DIACONIS, P. (2010). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6** 489–522. MR2809836

MSC2010 subject classifications. 62G10, 62G20, 62C20.

Key words and phrases. Detection boundary, sparse random graphs, beta model, higher criticism, sparse signals.

BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

CAI, T. T. and YUAN, M. (2014). Rate-optimal detection of very short signal segments. Available at arXiv:1407.2812.

CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

FIENBERG, S. E. and WASSERMAN, S. (1981). Categorical data analysis of single sociometric relations. *Sociol. Method.* **12** 156–192.

GOODREAU, S. M. (2007). Advances in exponential random graph (p*) models applied to a large social network. *Soc. Netw.* **29** 231–248.

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

HARA, H. and TAKEMURA, A. (2010). Connecting tables with zero-one entries by a subset of a Markov basis. In *Algebraic Methods in Statistics and Probability II*. *Contemp. Math.* **516** 199–213. Amer. Math. Soc., Providence, RI. MR2730750

HILLAR, C. and WIBISONO, A. (2013). Maximum entropy distributions on graphs. Available at arXiv:1301.3321.

HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. MR1991446

INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131

KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107. MR2788206

KARWA, V. and SLAVKOVIĆ, A. (2016). Inference using noisy degrees: Differentially private β-model and synthetic graphs. *Ann. Statist.* **44** 87–112. MR3449763

LAURITZEN, S. L. (2002). Rasch models with exchangeable rows and columns. Research Report Series, No. R-02-2005, Dept. Mathematical Sciences, Aalborg Univ.

LAURITZEN, S. L. (2008). Exchangeable Rasch matrices. *Rend. Mat. Appl.* (7) **28** 83–95. MR2463441

MUKHERJEE, R., MUKHERJEE, S. and SEN, S. (2018). Supplement to "Detection thresholds for the β-model on sparse graphs." DOI:10.1214/17-AOS1585SUPP.

MUKHERJEE, R., PILLAI, N. S. and LIN, X. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43** 352–381. MR3311863

OGAWA, M., HARA, H. and TAKEMURA, A. (2013). Graver basis for an undirected graph and its application to testing the beta model of random graphs. *Ann. Inst. Statist. Math.* **65** 191–212. MR3011620

PERRY, P. O. and WOLFE, P. J. (2012). Null models for network data. Available at arXiv:1201.5871.

PETROVIĆ, S., RINALDO, A. and FIENBERG, S. E. (2010). Algebraic statistics for a directed random graph model with reciprocation. In *Algebraic Methods in Statistics and Probability II*. *Contemp. Math.* **516** 261–283. Amer. Math. Soc., Providence, RI. MR2730754

RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804

ROBINS, G., PATTISON, P., KALISH, Y. and LUSHER, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Soc. Netw.* **29** 173–191.

VERZELEN, N. and ARIAS-CASTRO, E. (2015). Community detection in sparse random networks. *Ann. Appl. Probab.* **25** 3465–3510. MR3404642

WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393** 440–442.

YAN, T., QIN, H. and WANG, H. (2016). Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statist. Sinica* **26** 273–293. MR3468353

YAN, T. and XU, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika* **100** 519–524. MR3068452

YAN, T., ZHAO, Y. and QIN, H. (2015). Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters. *J. Multivariate Anal.* **133** 61–76. MR3282018

YAN, X., SHALIZI, C., JENSEN, J. E., KRZAKALA, F., MOORE, C., ZDEBOROVÁ, L., ZHANG, P. and ZHU, Y. (2014). Model selection for degree-corrected block models. *J. Stat. Mech. Theory Exp.* **2014** P05007.

The Annals of Statistics
2018, Vol. 46, No. 3, 1318–1351
https://doi.org/10.1214/17-AOS1586

ADAPTIVE SUP-NORM ESTIMATION OF THE WIGNER FUNCTION IN NOISY QUANTUM HOMODYNE TOMOGRAPHY

BY KARIM LOUNICI*,†,[1], KATIA MEZIANI‡,§,[2] AND GABRIEL PEYRÉ¶,[3]

Georgia Institute of Technology and LJAD, Université Côte d'Azur, UMR CNRS 7351 †, CEREMADE, UMR CNRS 7534, Université Paris Dauphine, PSL Research University‡ and CREST-ENSAE§ and CNRS and DMA, École Normale Supérieure ¶*

In quantum optics, the quantum state of a light beam is represented through the Wigner function, a density on $\mathbb{R}^2$, which may take negative values but must respect intrinsic positivity constraints imposed by quantum physics. In the framework of noisy quantum homodyne tomography with efficiency parameter $1/2 < \eta \leq 1$, we study the theoretical performance of a kernel estimator of the Wigner function. We prove that it is minimax efficient, up to a logarithmic factor in the sample size, for the $\mathbb{L}_\infty$-risk over a class of infinitely differentiable functions. We also compute the lower bound for the $\mathbb{L}_2$-risk. We construct an adaptive estimator, that is, which does not depend on the smoothness parameters, and prove that it attains the minimax rates for the corresponding smoothness of the class of functions up to a logarithmic factor in the sample size. Finite sample behaviour of our adaptive procedure is explored through numerical experiments.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A., eds. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York. MR1225604

ALQUIER, P., MEZIANI, K. and PEYRÉ, G. (2013). Adaptive estimation of the density matrix in quantum homodyne tomography with noisy data. *Inverse Probl.* **29** 075017, 20. MR3080478

ARTILES, L. M., GILL, R. D. and GUŢĂ, M. I. (2005). An invitation to quantum tomography. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 109–134. MR2136642

AUBRY, J.-M., BUTUCEA, C. and MEZIANI, K. (2009). State estimation in quantum homodyne tomography with noisy data. *Inverse Probl.* **25** 015003, 22. MR2465335

AVERBUCH, A., COIFMAN, R. R., DONOHO, D. L., ISRAELI, M., SHKOLNISKY, Y. and SEDELNIKOV, I. (2008). A framework for discrete integral transformations. II. The 2D discrete Radon transform. *SIAM J. Sci. Comput.* **30** 785–803. MR2385885

BARNDORFF-NIELSEN, O. E., GILL, R. D. and JUPP, P. E. (2003). On quantum statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 775–816. MR2017871

BERGH, J. and LÖFSTRÖM, J. (1976). *Interpolation Spaces. An Introduction*. Springer, Berlin. MR0482275

BISSANTZ, N. and HOLZMANN, H. (2008). Statistical inference for inverse problems. *Inverse Probl.* **24** 034009, 17. MR2421946

MSC2010 subject classifications. 62G05, 81V80.

Key words and phrases. Nonparametric minimax estimation, adaptive estimation, inverse problem, $\mathbb{L}_2$ and $\mathbb{L}_\infty$ risks, quantum homodyne tomography, Wigner function, Radon transform, quantum state.

BISSANTZ, N., DÜMBGEN, L., HOLZMANN, H. and MUNK, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 483–506. MR2323764

BOURDAUD, G., LANZA DE CRISTOFORIS, M. and SICKEL, W. (2006). Superposition operators and functions of bounded p-variation. *Rev. Mat. Iberoam.* **22** 455–487. MR2294787

BOUSQUET, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** 495–500. MR1890640

BUTUCEA, C., GUŢĂ, M. and ARTILES, L. (2007). Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data. *Ann. Statist.* **35** 465–494. MR2336856

BUTUCEA, C. and TSYBAKOV, A. B. (2008a). Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.* **52** 24–39.

BUTUCEA, C. and TSYBAKOV, A. B. (2008b). Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52** 237–249. MR2742504

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186. MR0997599

CAVALIER, L. (2008). Nonparametric statistical inverse problems. *Inverse Probl.* **24** 034004, 19. MR2421941

D'ARIANO, G. M., MACCHIAVELLO, C. and PARIS, M. G. (1994). Detection of the density matrix through optical homodyne tomography without filtered back projection. *Phys. Rev. A* (3) **50** 4298–4302.

DELAIGLE, A. and GIJBELS, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* **45** 249–267. MR2045631

DIGGLE, P. J. and HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55** 523–531. MR1224414

DONOHO, D. L. and LOW, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970. MR1165601

ERDÉLYI, A., MAGNUS, W., OBERHETTINGER, F. and TRICOMI, F. G. (1953). *Higher Transcendental Functions. Vols. I, II.* McGraw-Hill, New York.

FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324

FAN, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.* **21** 600–610. MR1232507

FEINGOLD, D. G. and VARGA, R. S. (1962). Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem. *Pacific J. Math.* **12** 1241–1250. MR0151473

GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921. MR1955344

GINÉ, E. and NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605–1646. MR2546757

GOLDENSHLUGER, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5** 907–925. MR1715444

GUŢĂ, M. and ARTILES, L. (2007). Minimax estimation of the Wigner function in quantum homodyne tomography with ideal detectors. *Math. Methods Statist.* **16** 1–15. MR2319467

HELSTROM, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press, New York.

HESSE, C. H. and MEISTER, A. (2004). Optimal iterative density deconvolution. *J. Nonparametr. Stat.* **16** 879–900. MR2094745

HOLEVO, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory. North-Holland Series in Statistics and Probability* **1**. North-Holland, Amsterdam. MR0681693

JOHNSTONE, I. M. and RAIMONDO, M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Ann. Statist.* **32** 1781–1804. MR2102493

JOHNSTONE, I. M. and SILVERMAN, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann*. *Statist*. **18** 251–280. MR1041393

JOHNSTONE, I. M., KERKYACHARIAN, G., PICARD, D. and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **66** 547–573. MR2088290

KLEMELÄ, J. and MAMMEN, E. (2010). Empirical risk minimization in inverse problems. *Ann*. *Statist*. **38** 482–511. MR2589328

KOROSTELËV, A. P. and TSYBAKOV, A. B. (1991). Optimal rates of convergence of estimates in a probabilistic formulation of the tomography problem. *Problemy Peredachi Informatsii* **27** 92–103. MR1294566

KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statistics* **82**. Springer, New York. MR1226450

KRASIKOV, I. (2007). Inequalities for orthonormal Laguerre polynomials. *J*. *Approx*. *Theory* **144** 1–26. MR2287374

LEONHARDT, U. (1997). *Measuring the Quantum State of Light*. Cambridge Univ. Press, Cambridge.

LEONHARDT, U., PAUL, H. and D'ARIANO, G. M. (1995). Tomographic reconstruction of the density matrix via pattern functions. *Phys*. *Rev*. *A* (3) **52** 4899–4907.

LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor*. *Veroyatn*. *Primen*. **36** 645–659. MR1147167

LEPSKIĬ, O. V. (1992). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teor*. *Veroyatn*. *Primen*. **37** 468–481. MR1214353

LOUNICI, K., MEZIANI, K. and PEYRÉ, G. (2018). Supplement to "Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography." DOI:10.1214/17-AOS1586SUPP.

LOUNICI, K. and NICKL, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann*. *Statist*. **39** 201–231. MR2797844

MEISTER, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Probl*. **24** 015003, 14. MR2384762

MÉZIANI, K. (2007). Nonparametric estimation of the purity of a quantum state in quantum homodyne tomography with noisy data. *Math*. *Methods Statist*. **16** 354–368. MR2378280

MEZIANI, K. (2008). Nonparametric goodness-of fit testing in quantum homodyne tomography with noisy data. *Electron*. *J*. *Stat*. **2** 1195–1223. MR2461899

MUCKENHOUPT, B. (1970). Asymptotic forms for Laguerre polynomials. *Proc*. *Amer*. *Math*. *Soc*. **24** 288–292. MR0251272

PENSKY, M. and SAPATINAS, T. (2009). Functional deconvolution in a periodic setting: Uniform case. *Ann*. *Statist*. **37** 73–104. MR2488345

PENSKY, M. and VIDAKOVIC, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann*. *Statist*. **27** 2033–2053. MR1765627

RICHTER, T. (2000). Realistic pattern functions for optical homodyne tomography and determination of specific expectation values. *Phys*. *Rev*. *A* **61**.

STEFANSKI, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist*. *Probab*. *Lett*. **9** 229–235. MR1045189

STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184. MR1054861

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359

VOGEL, K. and RISKEN, H. (1989). Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys*. *Rev*. *A* **40** 2847–2849.

WATSON, G. N. (1995). *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press, Cambridge. MR1349110

WIGNER, E. (1932). On the quantum correction for thermodynamic equations. *Phys. Rev.* **40** 749–759.

ZORICH, V. A. (2016). *Mathematical Analysis. II*, 2nd ed. Springer, Heidelberg. MR3445604

The Annals of Statistics
2018, Vol. 46, No. 3, 1352–1382
https://doi.org/10.1214/17-AOS1587

DISTRIBUTED TESTING AND ESTIMATION UNDER SPARSE HIGH DIMENSIONAL MODELS

BY HEATHER BATTEY[*,†,1], JIANQING FAN[†,‡,2], HAN LIU[†], JUNWEI LU[†] AND ZIWEI ZHU[†]

Imperial College London[*], *Princeton University*[†] *and Fudan University*[‡]

This paper studies hypothesis testing and parameter estimation in the context of the divide-and-conquer algorithm. In a unified likelihood-based framework, we propose new test statistics and point estimators obtained by aggregating various statistics from k subsamples of size n/k, where n is the sample size. In both low dimensional and sparse high dimensional settings, we address the important question of how large k can be, as n grows large, such that the loss of efficiency due to the divide-and-conquer algorithm is negligible. In other words, the resulting estimators have the same inferential efficiencies and estimation rates as an oracle with access to the full sample. Thorough numerical results are provided to back up the theory.

REFERENCES

BATTEY, H., FAN, J., LIU, H., LU, J. and ZHU, Z. (2018). Supplement to "Distributed testing and estimation under sparse high dimensional models." DOI:10.1214/17-AOS1587SUPP.

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434. MR0386168

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. MR2807761

CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

CHEN, X. and XIE, M. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statist. Sinica* **24** 1655–1684. MR3308656

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London. MR0370837

FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 37–65. MR2885839

FAN, J., HAN, F. and LIU, H. (2014). Challenges of big data analysis. *Nat. Sci. Rev.* **1** 293–314.

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

FAN, J. and LV, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inform. Theory* **57** 5467–5484. MR2849368

FAN, J. and SONG, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist.* **38** 3567–3604. MR2766861

MSC2010 subject classifications. Primary 62F05, 62F10; secondary 62F12.
Key words and phrases. Divide and conquer, debiasing, massive data, thresholding.

JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

KALLENBERG, O. (1997). *Foundations of Modern Probability*. Springer, New York. MR1464694

KLEINER, A., TALWALKAR, A., SARKAR, P. and JORDAN, M. I. (2014). A scalable bootstrap for massive data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 795–816. MR3248677

LEE, J. D., LIU, Q., SUN, Y. and TAYLOR, J. E. (2017). Communication-efficient sparse regression. *J. Mach. Learn. Res.* **18** Paper No. 5. MR3625709

LIU, Q. and IHLER, A. T. (2014). Distributed estimation, information loss and exponential families. In *Advances in Neural Information Processing Systems* 27 (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, eds.) 1098–1106. MIT Press, Cambridge, MA.

LOH, P.-L. and WAINWRIGHT, M. J. (2013). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. In *Advances in Neural Information Processing Systems* 26 (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) 476–484. Curran Associates Inc., Red Hook, NY.

LOH, P.-L. and WAINWRIGHT, M. J. (2015). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* **16** 559–616. MR3335800

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

NEGAHBAN, S. N., YU, B., WAINWRIGHT, M. J. and RAVIKUMAR, P. (2009). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. In *Advances in Neural Information Processing Systems* 22 (Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams and A. Culotta, eds.) 1348–1356. Curran Associates Inc., Red Hook, NY.

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489

ROSENBLATT, J. D. and NADLER, B. (2016). On the optimality of averaging in distributed statistical learning. *Inf. Inference* **5** 379–404. MR3609865

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

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285

WANG, Z., LIU, H. and ZHANG, T. (2014). Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. *Ann. Statist.* **42** 2164–2201. MR3269977

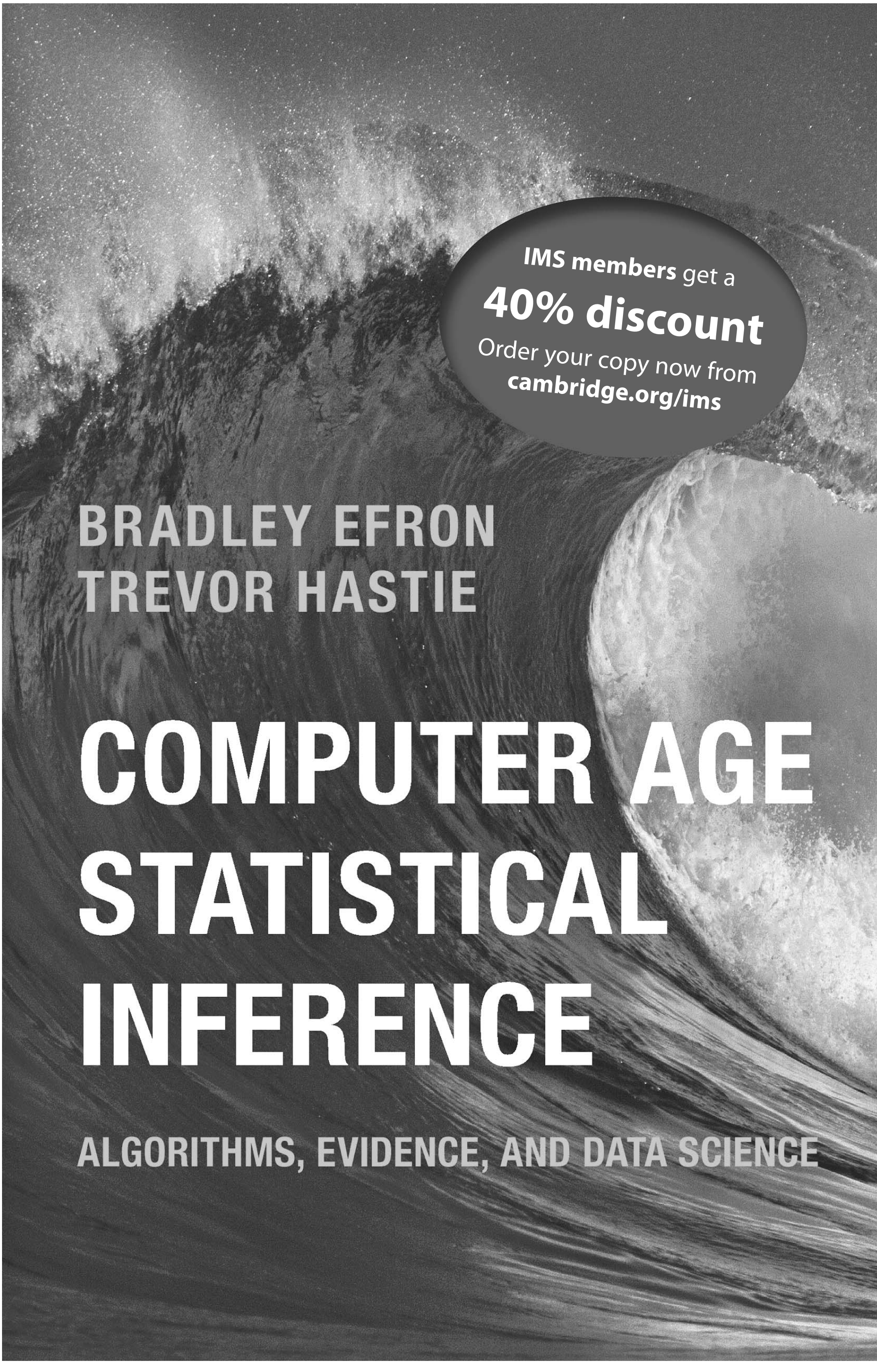
ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701

ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *J. Mach. Learn. Res.* **16** 3299–3340. MR3450540

ZHANG, C.-H. and ZHANG, T. (2012). A general theory of concave regularization for high-dimensional sparse estimation problems. *Statist. Sci.* **27** 576–593. MR3025135

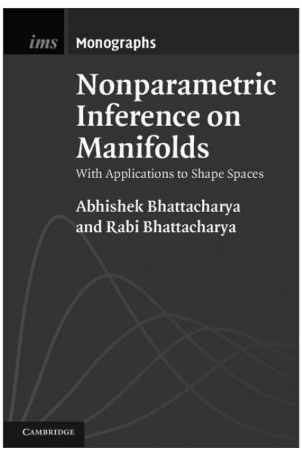
ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940

ZHAO, T., CHENG, G. and LIU, H. (2016). A partially linear framework for massive heterogeneous data. *Ann. Statist.* **44** 1400–1437. MR3519928


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

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



Nonparametric Inference on Manifolds

With Applications to Shape Spaces

Abhishek Bhattacharya, Rabi Bhattacharya

This book introduces in a systematic manner a general nonparametric theory of statistics on manifolds, with emphasis on manifolds of shapes. The theory has important and varied applications in medical diagnostics, image analysis, and machine vision. An early chapter of examples establishes the effectiveness of the new methods and demonstrates how they outperform their parametric counterparts. Inference is developed for both intrinsic and extrinsic Fréchet means of probability distributions on manifolds, then applied to shape spaces defined as orbits of landmarks under a Lie group of transformations—in particular, similarity, reflection similarity, affine and projective transformations. In addition, nonparametric Bayesian theory is adapted and extended to manifolds for the purposes of density estimation, regression and classification. Ideal for statisticians who analyze manifold data and wish to develop their own methodology, this book is also of interest to probabilists, mathematicians, computer scientists and morphometricians with mathematical training.

IMS member? Claim your 40% discount:
www.cambridge.org/ims

Hardback price US$51.00
(non-member price $85.00)

Cambridge University Press, in conjunction with the Institute of Mathematical Statistics, established the IMS Monographs and IMS Textbooks series of high-quality books. The Series Editors are Xiao-Li Meng, Susan Holmes, Ben Hambly, D. R. Cox and Alan Agresti.