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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 45, No. 3—June 2017

The Annals of Statistics
2017, Vol. 45, No. 3, 923–958
DOI: 10.1214/15-AOS1432

MINIMAX ESTIMATION OF LINEAR AND QUADRATIC FUNCTIONALS ON SPARSITY CLASSES

BY OLIVIER COLLIER*, LAËTITIA COMMINGES† AND ALEXANDRE B. TSYBAKOV‡,[1]

Modal'X, Université Paris-Ouest, Université Paris Dauphine† and CREST-ENSAE‡*

For the Gaussian sequence model, we obtain nonasymptotic minimax rates of estimation of the linear, quadratic and the ℓ_2-norm functionals on classes of sparse vectors and construct optimal estimators that attain these rates. The main object of interest is the class $B_0(s)$ of s-sparse vectors $\theta = (\theta_1, \dots, \theta_d)$, for which we also provide completely adaptive estimators (independent of s and of the noise variance σ) having logarithmically slower rates than the minimax ones. Furthermore, we obtain the minimax rates on the ℓ_q-balls $B_q(r) = \{\theta \in \mathbb{R}^d : \|\theta\|_q \le r\}$ where $0 < q \le 2$, and $\|\theta\|_q = (\sum_{i=1}^d |\theta_i|^q)^{1/q}$. This analysis shows that there are, in general, three zones in the rates of convergence that we call the sparse zone, the dense zone and the degenerate zone, while a fourth zone appears for estimation of the quadratic functional. We show that, as opposed to estimation of θ, the correct logarithmic terms in the optimal rates for the sparse zone scale as $\log(d/s^2)$ and not as $\log(d/s)$. For the class $B_0(s)$, the rates of estimation of the linear functional and of the ℓ_2-norm have a simple elbow at $s = \sqrt{d}$ (boundary between the sparse and the dense zones) and exhibit similar performances, whereas the estimation of the quadratic functional $Q(\theta)$ reveals more complex effects: the minimax risk on $B_0(s)$ is infinite and the sparseness assumption needs to be combined with a bound on the ℓ_2-norm. Finally, we apply our results on estimation of the ℓ_2-norm to the problem of testing against sparse alternatives. In particular, we obtain a nonasymptotic analog of the Ingster–Donoho–Jin theory revealing some effects that were not captured by the previous asymptotic analysis.

REFERENCES

[1] ABRAMOVICH, F. and GRINSHTEIN, V. (2010). MAP model selection in Gaussian regression. *Electron. J. Stat.* **4** 932–949. MR2721039

[2] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII—1983. Lecture Notes in Math.* **1117** 1–198. Springer, Berlin. MR0883646

[3] 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

MSC2010 subject classifications. 62J05, 62G05.

Key words and phrases. Nonasymptotic minimax estimation, linear functional, quadratic functional, sparsity, unknown noise variance, thresholding.

[4] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[5] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946

[6] BIRNBAUM, Z. W. (1942). An inequality for Mill's ratio. *Ann. Math. Stat.* **13** 245–246. MR0006640

[7] BUTUCEA, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.* **35** 1907–1930. MR2363957

[8] BUTUCEA, C. and COMTE, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15** 69–98. MR2546799

[9] CAI, T. T. and LOW, M. G. (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Ann. Statist.* **32** 552–576. MR2060169

[10] CAI, T. T. and LOW, M. G. (2005). On adaptive estimation of linear functionals. *Ann. Statist.* **33** 2311–2343. MR2211088

[11] CAI, T. T. and LOW, M. G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. MR2253108

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

[13] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over ℓ_p-balls for ℓ_q-error. *Probab. Theory Related Fields* **99** 277–303.

[14] DONOHO, D. L. and LIU, R. (1991). Geometrizing rates of convergence. III. *Ann. Statist.* **19** 668–701.

[15] DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323.

[16] EFROMOVICH, S. and LOW, M. L. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **24** 1106–1125.

[17] GOLDENSHLUGER, A. and PEREVERZEV, S. V. (2003). On adaptive inverse estimation of linear functionals. *Bernoulli* **9** 783–807.

[18] GOLUBEV, G. K. (2004). The method of risk envelopes in the estimation of linear functionals. *Problemy Peredachi Informatsii* **40** 58–72.

[19] GOLUBEV, Y. and LEVIT, B. (2004). An oracle approach to adaptive estimation of linear functionals in a Gaussian model. *Math. Methods Statist.* **13** 392–408 (2005). MR2126747

[20] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[21] IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1984). Nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 19–32. MR0739497

[22] INGSTER, YU. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646

[23] INGSTER, Y. I., POUET, C. and TSYBAKOV, A. B. (2009). Classification of sparse high-dimensional vectors. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4427–4448. MR2546395

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

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

[26] JOHNSTONE, I. (2001). Thresholding for weighted χ^2. *Statist. Sinica* **11** 691–704. MR1863157

[27] JOHNSTONE, I. M. (2001). Chi-square oracle inequalities. *IMS Lecture Notes Monogr. Ser.* **36** 399–418. MR1836572

[28] JOHNSTONE, I. M. (2013). *Gaussian Estimation: Sequence and Wavelet Models*. Book draft.

[29] JUDITSKY, A. and NEMIROVSKI, A. (2009). Nonparametric estimation by convex programming. *Ann. Statist.* **37** 2278–2300. MR2543692

[30] KLEMELÄ, J. (2006). Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields* **134** 539–564.

[31] KLEMELÄ, J. and TSYBAKOV, A. B. (2001). Sharp adaptive estimation of linear functionals. *Ann. Statist.* **29** 1567–1600. MR1891739

[32] LAURENT, B., LUDEÑA, C. and PRIEUR, C. (2008). Adaptive estimation of linear functionals by model selection. *Electron. J. Stat.* **2** 993–1020. MR2448602

[33] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785

[34] LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. MR1670867

[35] LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381

[36] NEMIROVSKI, A. (2000). *Topics in Nonparametric Statistics. Ecole d'été de Probabilités de Saint Flour* 1998. *Lecture Notes in Math.* **1738**. Springer, New York.

[37] RIGOLLET, P. and TSYBAKOV, A. B. (2011). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** 731–771. MR2816337

[38] SAMPFORD, M. R. (1953). Some inequalities on Mills ratio and related functions. *Ann. Math. Stat.* **24** 132–134.

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

[40] VERZELEN, N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electron. J. Stat.* **6** 38–90. MR2879672

The Annals of Statistics
2017, Vol. 45, No. 3, 959–990
DOI: 10.1214/16-AOS1470

SUPPORT CONSISTENCY OF DIRECT SPARSE-CHANGE LEARNING IN MARKOV NETWORKS[1]

BY SONG LIU*, TAIJI SUZUKI†,‡, RAISSA RELATOR§, JUN SESE§, MASASHI SUGIYAMA¶ AND KENJI FUKUMIZU*

Institute of Statistical Mathematics, Tokyo Institute of Technology,† PRESTO, Japan Science and Technology Agency,‡ National Institute of Advanced Industrial Science and Technology§ and University of Tokyo¶*

We study the problem of learning sparse structure changes between two Markov networks P and Q. Rather than fitting two Markov networks separately to two sets of data and figuring out their differences, a recent work proposed to learn changes *directly* via estimating the ratio between two Markov network models. In this paper, we give sufficient conditions for *successful change detection* with respect to the sample size n_p, n_q, the dimension of data m and the number of changed edges d. When using an unbounded density ratio model, we prove that the true sparse changes can be consistently identified for $n_p = \Omega(d^2 \log \frac{m^2+m}{2})$ and $n_q = \Omega(n_p^2)$, with an exponentially decaying upper-bound on learning error. Such sample complexity can be improved to $\min(n_p, n_q) = \Omega(d^2 \log \frac{m^2+m}{2})$ when the boundedness of the density ratio model is assumed. Our theoretical guarantee can be applied to a wide range of discrete/continuous Markov networks.

REFERENCES

[1] BANERJEE, O., EL GHAOUI, L. and D'ASPREMONT, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. MR2417243

[2] DANAHER, P., WANG, P. and WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 373–397. MR3164871

[3] DONOHO, D. L. and HUO, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory* **47** 2845–2862. MR1872845

[4] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

[5] HAMMERSLEY, J. M. and CLIFFORD, P. (1971). *Markov Fields on Finite Graphs and Lattices*. Available at http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf.

[6] KOLAR, M. and XING, E. P. (2012). Estimating networks with jumps. *Electron. J. Stat.* **6** 2069–2106. MR3020257

[7] KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, MA. MR2778120

[8] LEE, S.-I., GANAPATHI, V. and KOLLER, D. (2007). Efficient structure learning of Markov networks using l_1-regularization. In *Advances in Neural Information Processing Systems* (B. Schölkopf, J. Platt and T. Hoffman, eds.) **19** 817–824. MIT Press, Cambridge, MA.

MSC2010 subject classifications. Primary 62F12; Secondary 68T99.
Key words and phrases. Markov networks, change detection, density ratio estimation.

[9] LIU, S., QUINN, J. A., GUTMANN, M. U., SUZUKI, T. and SUGIYAMA, M. (2014). Direct learning of sparse changes in Markov networks by density ratio estimation. *Neural Comput*. **26** 1169–1197. MR3234638

[10] LIU, S., SUZUKI, T., RELATOR, R., SESE, J., SUGIYAMA, M. and FUKUMIZU, K. (2016). Supplement to "Support consistency of direct sparse-change learning in Markov networks." DOI:10.1214/16-AOS1470SUPP.

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

[12] NAGASHIMA, T., SHIMODAIRA, H., IDE, K., NAKAKUKI, T., TANI, Y., TAKAHASHI, K., YUMOTO, N. and HATAKEYAMA, M. (2007). Quantitative transcriptional control of erbb receptor signaling undergoes graded to biphasic response for cell differentiation. *Journal of Biological Chemistry* **282** 4045–4056.

[13] NEAL, R. M. (2003). Slice sampling. *Ann. Statist*. **31** 705–767. MR1994729

[14] RASKUTTI, G., YU, B., WAINWRIGHT, M. J. and RAVIKUMAR, P. (2009). Model selection in Gaussian graphical models: High-dimensional consistency of ℓ_1-regularized mle. In *Advances in Neural Information Processing Systems* (D. Koller, D. Schuurmans, Y. Bengio and L. Bottou, eds.) **21** 1329–1336. Curran Associates, Red Hook, NY.

[15] RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist*. **38** 1287–1319. MR2662343

[16] ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York. MR1707311

[17] SUGIYAMA, M., KANAMORI, T., SUZUKI, T., DU PLESSIS, M. C., LIU, S. and TAKEUCHI, I. (2013). Density-difference estimation. *Neural Comput*. **25** 2734–2775. MR3114585

[18] SUGIYAMA, M., NAKAJIMA, S., KASHIMA, H., VON BÜNAU, P. and KAWANABE, M. (2008). Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in Neural Information Processing Systems* 20 (J. C. Platt, D. Koller, Y. Singer and S. T. Roweis, eds.). Curran Associates, Red Hook, NY.

[19] SUGIYAMA, M., SUZUKI, T. and KANAMORI, T. (2012). *Density Ratio Estimation in Machine Learning*. Cambridge Univ. Press, Cambridge. MR2895762

[20] TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **67** 91–108. MR2136641

[21] TOMIOKA, R. and SUZUKI, T. (2014). Spectral norm of random tensors. Preprint. Available at arXiv:1407.1870.

[22] TSUBOI, Y., KASHIMA, H., HIDO, S., BICKEL, S. and SUGIYAMA, M. (2009). Direct density ratio estimation for large-scale covariate shift adaptation. *Journal of Information Processing* **17** 138–155.

[23] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. MR2729873

[24] WASSERMAN, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York. MR2055670

[25] YAMADA, M., SUZUKI, T., KANAMORI, T., HACHIYA, H. and SUGIYAMA, M. (2013). Relative density-ratio estimation for robust distribution comparison. *Neural Comput*. **25** 1324–1370. MR3075782

[26] YANG, E., GENEVERA, A., LIU, Z. and RAVIKUMAR, P. (2012). Graphical models via generalized linear models. In *Advances in Neural Information Processing Systems* (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) **25** 1358–1366. Curran Associates, Red Hook, NY.

[27] YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574
[28] ZHANG, B. and WANG, Y. J. (2010). Learning structural changes of Gaussian graphical models in controlled experiments. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence* (*UAI*2010) 701–708. AUAI Press, Corvallis, OR.
[29] ZHAO, S. D., CAI, T. T. and LI, H. (2014). Direct estimation of differential networks. *Biometrika* **101** 253–268. MR3215346

The Annals of Statistics
2017, Vol. 45, No. 3, 991–1023
DOI: 10.1214/16-AOS1472

RANDOMIZED SKETCHES FOR KERNELS: FAST AND OPTIMAL NONPARAMETRIC REGRESSION[1]

BY YUN YANG*, MERT PILANCI†,2 AND MARTIN J. WAINWRIGHT†

Florida State University and University of California, Berkeley†*

Kernel ridge regression (KRR) is a standard method for performing nonparametric regression over reproducing kernel Hilbert spaces. Given n samples, the time and space complexity of computing the KRR estimate scale as $\mathcal{O}(n^3)$ and $\mathcal{O}(n^2)$, respectively, and so is prohibitive in many cases. We propose approximations of KRR based on m-dimensional randomized sketches of the kernel matrix, and study how small the projection dimension m can be chosen while still preserving minimax optimality of the approximate KRR estimate. For various classes of randomized sketches, including those based on Gaussian and randomized Hadamard matrices, we prove that it suffices to choose the sketch dimension m proportional to the statistical dimension (modulo logarithmic factors). Thus, we obtain fast and minimax optimal approximations to the KRR estimate for nonparametric regression. In doing so, we prove a novel lower bound on the minimax risk of kernel regression in terms of the localized Rademacher complexity.

REFERENCES

[1] AILON, N. and LIBERTY, E. (2009). Fast dimension reduction using Rademacher series on dual BCH codes. *Discrete Comput. Geom.* **42** 615–630. MR2556458
[2] ALAOUI, A. E. and MAHONEY, M. W. (2014). Fast randomized kernel methods with statistical guarantees. Technical Report, UC Berkeley. Available at arXiv:1411.0306.
[3] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453
[4] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404. MR0051437
[5] BACH, F. (2012). Sharp analysis of low-rank kernel matrix approximations. In *International Conference on Learning Theory* (*COLT*). Edinburgh.
[6] BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537. MR2166554
[7] BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic, Boston, MA. MR2239907
[8] BOUTSIDIS, C. and GITTENS, A. (2013). Improved matrix algorithms via the subsampled randomized Hadamard transform. *SIAM J. Matrix Anal. Appl.* **34** 1301–1340. MR3101094
[9] COHEN, M. B., NELSON, J. and WOODRUFF, D. P. (2015). Optimal approximate matrix product in terms of stable rank. Technical report.

MSC2010 subject classifications. Primary 62G08; secondary 68W20.
Key words and phrases. Nonparametric regression, random projection, kernel method, dimensionality reduction, convex optimization.

[10] DRINEAS, P. and MAHONEY, M. W. (2005). On the Nyström method for approximating a Gram matrix for improved kernel-based learning. *J. Mach. Learn. Res.* **6** 2153–2175. MR2249884

[11] GITTENS, A. and MAHONEY, M. W. (2013). Revisiting the nystrom method for improved large-scale machine learning. Preprint. Available at arXiv:1303.1849.

[12] GU, C. (2002). *Smoothing Spline ANOVA Models*. Springer, New York. MR1876599

[13] HALKO, N., MARTINSSON, P. G. and TROPP, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* **53** 217–288. MR2806637

[14] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York. MR1851606

[15] JOHNSTONE, I. M. (2016). *Gaussian Estimation: Sequence and Wavelet Models*. Springer, New York.

[16] KIMELDORF, G. and WAHBA, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95. MR0290013

[17] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. MR2329442

[18] KRAHMER, F. and WARD, R. (2011). New and improved Johnson–Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.* **43** 1269–1281. MR2821584

[19] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[20] MAHONEY, M. W. (2011). Randomized algorithms for matrices and data. *Faund. Trends Mach. Learn.* **3** 123–224.

[21] MATOUŠEK, J. (2002). *Lectures on Discrete Geometry. Graduate Texts in Mathematics* **212**. Springer, New York. MR1899299

[22] MENDELSON, S. (2002). Geometric parameters of kernel machines. In *Computational Learning Theory (Sydney, 2002). Lecture Notes in Computer Science* **2375** 29–43. Springer, Berlin. MR2040403

[23] PILANCI, M. and WAINWRIGHT, M. J. (2015). Randomized sketches of convex programs with sharp guarantees. *IEEE Trans. Inform. Theory* **61** 5096–5115. MR3386504

[24] PILANCI, M. and WAINWRIGHT, M. J. (2016). Iterative Hessian sketch: Fast and accurate solution approximation for constrained least-squares. *J. Mach. Learn. Res.* **17**. Paper No. 53, 38. MR3504613

[25] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.* **13** 389–427. MR2913704

[26] SARLÓS, T. (2006). Improved approximation algorithms for large matrices via random projections. In *Foundations of Computer Science, 2006. FOCS '06. 47th Annual IEEE Symposium*. Berkeley, CA.

[27] SAUNDERS, C., GAMMERMAN, A. and VOVK, V. (1998). Ridge regression learning algorithm in dual variables. In *Proceedings of the Fifteenth International Conference on Machine Learning, ICML '98* 515–521. Morgan Kaufmann, San Francisco, CA.

[28] SHAWE-TAYLOR, J. and CRISTIANINI, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge Univ. Press, Cambridge.

[29] STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. MR0673642

[30] TROPP, J. A. (2011). Improved analysis of the subsampled randomized Hadamard transform. *Adv. Adapt. Data Anal.* **3** 115–126. MR2835584

[31] TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. MR2946459

[32] VAN DE GEER, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.

[33] WAHBA, G. (1990). *Spline Models for Observational Data*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442

[34] WILLIAMS, C. and SEEGER, M. (2001). Using the Nyström method to speed up kernel machines. In *Proceedings of the* 14*th Annual Conference on Neural Information Processing Systems* 682–688. Vancouver, BC, Canada.

[35] ZHANG, Y., DUCHI, J. C. and WAINWRIGHT, M. J. (2013). Divide and conquer kernel ridge regression. In *Computational Learning Theory* (*COLT*) *Conference*. Princeton, NJ.

The Annals of Statistics
2017, Vol. 45, No. 3, 1024–1058
DOI: 10.1214/16-AOS1473

TESTING UNIFORMITY ON HIGH-DIMENSIONAL SPHERES AGAINST MONOTONE ROTATIONALLY SYMMETRIC ALTERNATIVES

BY CHRISTINE CUTTING, DAVY PAINDAVEINE[1] AND THOMAS VERDEBOUT

Université libre de Bruxelles

We consider the problem of testing uniformity on high-dimensional unit spheres. We are primarily interested in nonnull issues. We show that rotationally symmetric alternatives lead to two Local Asymptotic Normality (LAN) structures. The first one is for fixed modal location $\boldsymbol{\theta}$ and allows to derive locally asymptotically most powerful tests under specified $\boldsymbol{\theta}$. The second one, that addresses the Fisher–von Mises–Langevin (FvML) case, relates to the unspecified-$\boldsymbol{\theta}$ problem and shows that the high-dimensional Rayleigh test is locally asymptotically most powerful invariant. Under mild assumptions, we derive the asymptotic nonnull distribution of this test, which allows to extend away from the FvML case the asymptotic powers obtained there from Le Cam's third lemma. Throughout, we allow the dimension p to go to infinity in an arbitrary way as a function of the sample size n. Some of our results also strengthen the local optimality properties of the Rayleigh test in low dimensions. We perform a Monte Carlo study to illustrate our asymptotic results. Finally, we treat an application related to testing for sphericity in high dimensions.

REFERENCES

AZZALINI, A. and CAPITANIO, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 579–602. MR1707862

BANERJEE, A. and GHOSH, J. (2004). Frequency sensitive competitive learning for scalable balanced clustering on high-dimensional hyperspheres. *IEEE T. Neural Networ.* **15** 702–719.

BANERJEE, A., DHILLON, I., GHOSH, J. and SRA, S. (2003). Generative model-based clustering of directional data. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 19–28.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. MR1623559

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. MR1324786

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

CHAUDHURI, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Ann. Statist.* **20** 897–916. MR1165598

CHIKUSE, Y. (1991). High-dimensional limit theorems and matrix decompositions on the Stiefel manifold. *J. Multivariate Anal.* **36** 145–162. MR1096663

MSC2010 subject classifications. Primary 62H11, 62G20; secondary 62H15.

Key words and phrases. Contiguity, directional statistics, high-dimensional statistics, invariance, local asymptotic normality, rotationally symmetric distributions, tests of uniformity.

CHIKUSE, Y. (1993). High-dimensional asymptotic expansions for the matrix Langevin distributions on the Stiefel manifold. *J. Multivariate Anal.* **44** 82–101. MR1208471

CHIKUSE, Y. (2003). *Statistics on Special Manifolds. Lecture Notes in Statistics* **174**. Springer, New York. MR1960435

CUESTA-ALBERTOS, J. A., CUEVAS, A. and FRAIMAN, R. (2009). On projection-based tests for directional and compositional data. *Stat. Comput.* **19** 367–380. MR2565311

CUTTING, C., PAINDAVEINE, D. and VERDEBOUT, T. (2016). Supplement to "Testing uniformity on high-dimensional spheres against monotone rotationally symmetric alternatives." DOI:10.1214/16-AOS1473SUPP.

DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254. MR0501523

DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43. MR0885917

DAVIES, R. B. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika* **89** 484–489. MR1913976

DRYDEN, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Statist.* **33** 1643–1665. MR2166558

EISEN, M. B., SPELLMAN, P. T., BROWN, P. O. and BOTSTEIN, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95** 14863–14868.

FAŸ, G., DELABROUILLE, J., KERKYACHARIAN, G. and PICARD, D. (2013). Testing the isotropy of high energy cosmic rays using spherical needlets. *Ann. Appl. Stat.* **7** 1040–1073. MR3113500

GIRI, N. C. (1996). *Group Invariance in Statistical Inference*. World Scientific, Singapore. MR1626154

HALLIN, M. and PAINDAVEINE, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. MR2329465

HORNIK, K. and GRÜN, B. (2014). movMF: An R package for fitting mixtures of von Mises–Fisher distributions. *J. Statist. Softw.* **58**.

INGSTER, Y. I. (1997). Adaptive chi-square tests. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **244** 150–166, 333. MR1700386

JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. MR0312619

JUAN, J. and PRIETO, F. J. (2001). Using angles to identify concentrated multivariate outliers. *Technometrics* **43** 311–322. MR1943186

KIM, P. T., KOO, J.-Y. and PHAM NGOC, T. M. (2016). Supersmooth testing on the sphere over analytic classes. *J. Nonparametr. Stat.* **28** 84–115. MR3463552

LACOUR, C. and PHAM NGOC, T. M. (2014). Goodness-of-fit test for noisy directional data. *Bernoulli* **20** 2131–2168. MR3263101

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. MR2135927

LIESE, F. and MIESCKE, K.-J. (2008). *Statistical Decision Theory: Estimation, Testing, and Selection*. Springer, New York. MR2421720

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. Wiley, Chichester. MR1828667

MOREIRA, M. J. (2009). A maximum likelihood method for the incidental parameter problem. *Ann. Statist.* **37** 3660–3696. MR2549574

MÖTTÖNEN, J. and OJA, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Statist.* **5** 201–213. MR1346895

OJA, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks. Lecture Notes in Statistics* **199**. Springer, New York. MR2598854

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485

PAINDAVEINE, D. and VERDEBOUT, T. (2016). On high-dimensional sign tests. *Bernoulli* **22** 1745–1769. MR3474832

RAYLEIGH, L. (1919). On the problem of random vibrations and random flights in one, two and three dimensions. *Phil. Mag.* **37** 321–346.

SAW, J. G. (1978). A family of distributions on the m-sphere and some hypothesis tests. *Biometrika* **65** 69–73. MR0497510

SHAO, J. (2003). *Mathematical Statistics*, 2nd ed. Springer, New York. MR2002723

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

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

WATSON, G. S. (1988). The Langevin distribution on high dimensional spheres. *J. Appl. Statist.* **15** 123–130.

ZOU, C., PENG, L., FENG, L. and WANG, Z. (2014). Multivariate sign-based high-dimensional tests for sphericity. *Biometrika* **101** 229–236. MR3180668

The Annals of Statistics
2017, Vol. 45, No. 3, 1059–1095
DOI: 10.1214/16-AOS1475

NONLINEAR SUFFICIENT DIMENSION REDUCTION FOR FUNCTIONAL DATA

BY BING LI[1] AND JUN SONG

Pennsylvania State University

We propose a general theory and the estimation procedures for nonlinear sufficient dimension reduction where both the predictor and the response may be random functions. The relation between the response and predictor can be arbitrary and the sets of observed time points can vary from subject to subject. The functional and nonlinear nature of the problem leads to construction of two functional spaces: the first representing the functional data, assumed to be a Hilbert space, and the second characterizing nonlinearity, assumed to be a reproducing kernel Hilbert space. A particularly attractive feature of our construction is that the two spaces are nested, in the sense that the kernel for the second space is determined by the inner product of the first. We propose two estimators for this general dimension reduction problem, and establish the consistency and convergence rate for one of them. These asymptotic results are flexible enough to accommodate both fully and partially observed functional data. We investigate the performances of our estimators by simulations, and applied them to data sets about speech recognition and handwritten symbols.

REFERENCES

AMATO, U., ANTONIADIS, A. and DE FEIS, I. (2006). Dimension reduction in functional regression with applications. *Comput. Statist. Data Anal.* **50** 2422–2446. MR2225577

AMINI, A. A. and WAINWRIGHT, M. J. (2012). Sampled forms of functional PCA in reproducing kernel Hilbert spaces. *Ann. Statist.* **40** 2483–2510. MR3097610

BACH, F. R. and JORDAN, M. I. (2003). Kernel independent component analysis. *J. Mach. Learn. Res.* **3** 1–48. MR1966051

BAI, J., GOLDSMITH, J., CAFFO, B., GLASS, T. A. and CRAINICEANU, C. M. (2012). Movelets: A dictionary of movement. *Electron. J. Stat.* **6** 559–578. MR2988420

BAKER, C. R. (1973). Joint measures and cross-covariance operators. *Trans. Amer. Math. Soc.* **186** 273–289. MR0336795

BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic, Boston, MA. MR2239907

COOK, R. D. (1998). *Regression Graphics*. Wiley, New York. MR1645673

COOK, R. D. (2007). Fisher lecture: Dimension reduction in regression. *Statist. Sci.* **22** 1–26. MR2408655

COOK, R. D. and LI, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30** 455–474. MR1902895

MSC2010 subject classifications. 62B05, 62G08, 62G20, 62H99.

Key words and phrases. Convergence rate, linear operator, handwriting data, reproducing kernel Hilbert space, speech recognition, Sliced Average Variance Estimator, Sliced Inverse Regression.

COOK, R. D. and SETODJI, C. M. (2003). A model-free test for reduced rank in multivariate regression. *J. Amer. Statist. Assoc.* **98** 340–351. MR1995710

COOK, R. D. and WEISBERG, S. (1991). Sliced inverse regression for dimension reduction: Comment. *J. Amer. Statist. Assoc.* **86** 328–332.

DAUXOIS, J., FERRÉ, L. and YAO, A.-F. (2001). Un modèle semi-paramétrique pour variables aléatoires hilbertiennes. *C. R. Acad. Sci. Paris Sér. I Math.* **333** 947–952. MR1873814

EPIFANIO, I. (2008). Shape descriptors for classification of functional data. *Technometrics* **50** 284–294. MR2528652

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York. MR2229687

FERRÉ, L. and VILLA, N. (2006). Multilayer perceptron with functional inputs: An inverse regression approach. *Scand. J. Stat.* **33** 807–823. MR2300917

FERRÉ, L. and YAO, A. F. (2003). Functional sliced inverse regression analysis. *Statistics* **37** 475–488. MR2022235

FERRÉ, L. and YAO, A.-F. (2005). Smoothed functional inverse regression. *Statist. Sinica* **15** 665–683. MR2233905

FUKUMIZU, K., BACH, F. R. and GRETTON, A. (2007). Statistical consistency of kernel canonical correlation analysis. *J. Mach. Learn. Res.* **8** 361–383. MR2320675

FUKUMIZU, K., BACH, F. R. and JORDAN, M. I. (2009). Kernel dimension reduction in regression. *Ann. Statist.* **37** 1871–1905. MR2533474

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

HORN, R. A. and JOHNSON, C. R. (1985). *Matrix Analysis*. Cambridge Univ. Press, Cambridge. MR0832183

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, New York. MR2920735

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, Chichester. MR3379106

HSING, T. and REN, H. (2009). An RKHS formulation of the inverse regression dimension–reduction problem. *Ann. Statist.* **37** 726–755. MR2502649

KOLTCHINSKII, V. and GINÉ, E. (2000). Random matrix approximation of spectra of integral operators. *Bernoulli* **6** 113–167. MR1781185

LEE, K.-Y., LI, B. and CHIAROMONTE, F. (2013). A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *Ann. Statist.* **41** 221–249. MR3059416

LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. MR1137117

LI, B., ARTEMIOU, A. and LI, L. (2011). Principal support vector machines for linear and nonlinear sufficient dimension reduction. *Ann. Statist.* **39** 3182–3210. MR3012405

LI, B., CHUN, H. and ZHAO, H. (2012). Sparse estimation of conditional graphical models with application to gene networks. *J. Amer. Statist. Assoc.* **107** 152–167. MR2949348

LI, B., CHUN, H. and ZHAO, H. (2014). On an additive semigraphoid model for statistical networks with application to pathway analysis. *J. Amer. Statist. Assoc.* **109** 1188–1204. MR3265690

LI, L., LI, B. and ZHU, L.-X. (2010). Groupwise dimension reduction. *J. Amer. Statist. Assoc.* **105** 1188–1201. MR2752614

LI, B. and SONG, J. (2016). Supplement to "Nonlinear sufficient dimension reduction for functional data." DOI:10.1214/16-AOS1475SUPP.

LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102** 997–1008. MR2354409

LI, B., WEN, S. and ZHU, L. (2008). On a projective resampling method for dimension reduction with multivariate responses. *J. Amer. Statist. Assoc.* **103** 1177–1186. MR2462891

LI, B., ZHA, H. and CHIAROMONTE, F. (2005). Contour regression: A general approach to dimension reduction. *Ann. Statist.* **33** 1580–1616. MR2166556

LLORENS, D., PRAT, F., MARZAL, A., VILAR, J. M., CASTRO, M. J., AMENGUAL, J. C., BARRACHINA, S., CASTELLANOS, A., NA, S. E., GÓMEZ, J. A., GORBE, J., GORDO, A., PALAZÓN, V., PERIS, G., RAMOS-GARIJO, R. and ZAMORA, F. (2008). The ujipenchars database: A pen-based database of isolated handwritten characters. In *Proc. 6th Int. Conf. Language Resources Eval*, Marrakech, Morocco. 2647–2651.

MINH, H. Q. (2010). Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory. *Constr. Approx.* **32** 307–338. MR2677883

MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805. MR2163159

OJA, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Springer, New York. MR2598854

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

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

ROSSI, F. and VILLA, N. (2006). Support vector machine for functional data cliassification. *Neurocomputing* **79** 730–742.

WANG, J.-L., CHIOU, J.-M. and MULLER, H.-G. (2015). Review of functional data analysis. Preprint. Available at arXiv:1507.05135v1.

WANG, G., LIN, N. and ZHANG, B. (2013). Functional contour regression. *J. Multivariate Anal.* **116** 1–13. MR3049886

WANG, Q. and YIN, X. (2008). A nonlinear multi-dimensional variable selection method for high dimensional data: Sparse MAVE. *Comput. Statist. Data Anal.* **52** 4512–4520. MR2432477

WU, H.-M. (2008). Kernel sliced inverse regression with applications to classification. *J. Comput. Graph. Statist.* **17** 590–610. MR2528238

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106

YEH, Y.-R., HUANG, S.-Y. and LEE, Y.-J. (2009). Nonlinear dimension reduction with kernel sliced inverse regression. *IEEE Trans. Knowl. Data Eng.* **11** 1590–1603.

YIN, X. and BURA, E. (2006). Moment-based dimension reduction for multivariate response regression. *J. Statist. Plann. Inference* **136** 3675–3688. MR2256281

ZHU, L., MIAO, B. and PENG, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **101** 630–643. MR2281245

ZWALD, L. and BLANCHARD, G. (2006). On the convergence of eigenspaces in kernel principal component analysis. In *Advances in Neural Information Processing Systems* **18**. MIT Press, Cambridge, MA.

The Annals of Statistics
2017, Vol. 45, No. 3, 1096–1123
DOI: 10.1214/16-AOS1476

NETWORK VECTOR AUTOREGRESSION

BY XUENING ZHU*,[1], RUI PAN†,[2], GUODONG LI‡,[3],
YUEWEN LIU§,[4] AND HANSHENG WANG*,[1]

Peking University, Central University of Finance and Economics,†*
University of Hong Kong‡ and Xi'an Jiaotong University§

We consider here a large-scale social network with a continuous response observed for each node at equally spaced time points. The responses from different nodes constitute an ultra-high dimensional vector, whose time series dynamic is to be investigated. In addition, the network structure is also taken into consideration, for which we propose a network vector autoregressive (NAR) model. The NAR model assumes each node's response at a given time point as a linear combination of (a) its previous value, (b) the average of its connected neighbors, (c) a set of node-specific covariates and (d) an independent noise. The corresponding coefficients are referred to as the momentum effect, the network effect and the nodal effect, respectively. Conditions for strict stationarity of the NAR models are obtained. In order to estimate the NAR model, an ordinary least squares type estimator is developed, and its asymptotic properties are investigated. We further illustrate the usefulness of the NAR model through a number of interesting potential applications. Simulation studies and an empirical example are presented.

REFERENCES

[1] ANSELIN, L. (1999). *Spatial Econometrics*: *Methods and Models*. Springer Science and Business Media, Dordrecht.

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

[3] BEDRICK, E. J. and TSAI, C. L. (1994). Model selection for multivariate regression in small samples. *Biometrics* **50** 226–231.

[4] BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. (1994). *Time Series Analysis*: *Forecasting and Control*, 3rd ed. Prentice Hall, Englewood Cliffs, NJ. MR1312604

[5] BROWN, L. D., HAGERMAN, R. L., GRIFFIN, P. A. and ZMIJEWSKI, M. E. (1987). Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *J. Account*. *Econ*. **9** 61–87.

[6] CHEN, X., CHEN, Y. and XIAO, P. (2013). The impact of sampling and network topology on the estimation of social intercorrelations. *J. Mark*. *Res*. **50** 95–110.

[7] CLAUSET, A., SHALIZI, C. R. and NEWMAN, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* **51** 661–703. MR2563829

[8] CUARESMA, J. C., HLOUSKOVA, J., KOSSMEIER, S. and OBERSTEINER, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models. *Appl*. *Energy* **77** 87–106.

MSC2010 subject classifications. Primary 62M10; secondary 62J05.

Key words and phrases. Multivariate time series, ordinary least squares, social network, vector autoregression.

[9] DANTZIG, G. B. (1998). *Linear Programming and Extensions*, corrected ed. Princeton Univ. Press, Princeton, NJ. MR1658673

[10] DE MOL, C., GIANNONE, D. and REICHLIN, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *J. Econometrics* **146** 318–328. MR2465176

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

[12] FINGLETON, B. (1999). Spurious spatial regression: Some Monte Carlo results with a spatial unit root and spatial cointegration. *J. Reg. Sci.* **39** 1–19.

[13] FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *J. Amer. Statist. Assoc.* **100** 830–840. MR2201012

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

[15] HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press, Princeton, NJ. MR1278033

[16] HAN, F. and LIU, H. (2013). Transition matrix estimation in high dimensional time series. In *Proceedings of the* 30*th International Conference on Machine Learning* (*ICML*-13) 172–180. ACM, Atlanta, GA.

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

[18] HSU, N.-J., HUNG, H.-L. and CHANG, Y.-M. (2008). Subset selection for vector autoregressive processes using Lasso. *Comput. Statist. Data Anal.* **52** 3645–3657. MR2427370

[19] LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663

[20] LEE, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **72** 1899–1925. MR2095537

[21] LEE, L. F. and YU, J. (2009). Spatial nonstationarity and spurious regression: The case with a row-normalized spatial weights matrix. *Spatial Econ. Anal.* **4** 301–327.

[22] LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. MR2172368

[23] MCQUARRIE, A. D. R. and TSAI, C.-L. (1998). *Regression and Time Series Model Selection*. World Scientific, River Edge, NJ. MR1641582

[24] NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348

[25] NEWBOLD, P. and GRANGER, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts. *J. Roy. Statist. Soc. Ser. A* **137** 131–164. MR0451583

[26] NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255

[27] PAN, J. and YAO, Q. (2008). Modelling multiple time series via common factors. *Biometrika* **95** 365–379. MR2521589

[28] PARK, B. U., MAMMEN, E., HÄRDLE, W. and BORAK, S. (2009). Time series modelling with semiparametric factor dynamics. *J. Amer. Statist. Assoc.* **104** 284–298. MR2504378

[29] REINSEL, G. C. and VELU, R. P. (1998). *Multivariate Reduced-Rank Regression*: *Theory and Applications*. *Lecture Notes in Statistics* **136**. Springer, New York. MR1719704

[30] SHUMWAY, R. H. and STOFFER, D. S. (2000). *Time Series Analysis and Its Applications*. Springer, New York. MR1856572

[31] WANG, Y. J. and WONG, G. Y. (1987). Stochastic blockmodels for directed graphs. *J. Amer. Statist. Assoc.* **82** 8–19. MR0883333

[32] WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis*: *Methods and Applications*. Cambridge Univ. Press, London.

[33] ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann*. *Statist*. **40** 2266–2292. MR3059083

[34] ZHOU, J., TU, Y., CHEN, Y. and WANG, H. (2017). Estimating spatial autocorrelation with sampled network data. *J*. *Bus*. *Econ*. *Stat*. **35** 130–138. MR3591541

[35] ZHU, X., PAN, R., LI, G., LIU, Y. and WANG, H. (2016). Supplement to "Network vector autoregression." DOI:10.1214/16-AOS1476SUPP.

The Annals of Statistics
2017, Vol. 45, No. 3, 1124–1151
DOI: 10.1214/16-AOS1477

ON COVERAGE AND LOCAL RADIAL RATES OF CREDIBLE SETS

BY EDUARD BELITSER

VU Amsterdam

In the mildly ill-posed inverse signal-in-white-noise model, we construct confidence sets as credible balls with respect to the empirical Bayes posterior resulting from a certain two-level hierarchical prior. The quality of the posterior is characterized by the contraction rate which we allow to be local, that is, depending on the parameter. The issue of optimality of the constructed confidence sets is addressed via a trade-off between its "size" (the *local radial rate*) and its coverage probability. We introduce *excessive bias restriction* (EBR), more general than *self-similarity* and *polished tail condition* recently studied in the literature. Under EBR, we establish the confidence optimality of our credible set with some local (*oracle*) radial rate. We also derive the oracle estimation inequality and the oracle posterior contraction rate. The obtained local results are more powerful than global: adaptive minimax results for a number of smoothness scales follow as consequence, in particular, the ones considered by Szabó et al. [*Ann. Statist.* **43** (2015) 1391–1428].

REFERENCES

[1] BARAUD, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32** 528–551. MR2060168

[2] BELITSER, E. (2014). On coverage and local radial rates of DDM-credible sets. Available at arXiv:1407.5232.

[3] BELITSER, E. (2016). Supplement to "On coverage and local radial rates of credible sets." DOI:10.1214/16-AOS1477SUPP.

[4] BERAN, R. and DÜMBGEN, L. (1998). Modulation of estimators and confidence sets. *Ann. Statist.* **26** 1826–1856. MR1673280

[5] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946

[6] BULL, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490–1516. MR2988456

[7] BULL, A. D. and NICKL, R. (2013). Adaptive confidence sets in L^2. *Probab. Theory Related Fields* **156** 889–919. MR3078289

[8] CAI, T. T. and LOW, M. G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. MR2275240

[9] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856

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

MSC2010 subject classifications. Primary 62G15, 62C05; secondary 62G99.
Key words and phrases. Credible ball, excessive bias restriction, local radial rate.

[11] CAVALIER, L. and GOLUBEV, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34** 1653–1677. MR2283712

[12] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525

[13] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. MR1740119

[14] GENOVESE, C. and WASSERMAN, L. (2008). Adaptive confidence bands. *Ann. Statist.* **36** 875–905. MR2396818

[15] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007

[16] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707

[17] HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. MR2906872

[18] JOHNSTONE, I. M. (2010). High dimensional Bernstein–von Mises: Simple examples. In *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown. Inst. Math. Stat. Collect.* **6** 87–98. IMS, Beachwood, OH. MR2798513

[19] JUDITSKY, A. and LAMBERT-LACROIX, S. (2003). Nonparametric confidence set estimation. *Math. Methods Statist.* **12** 410–428. MR2054156

[20] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881

[21] LI, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. MR1015135

[22] LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554. MR1604412

[23] NICKL, R. and SZABÓ, B. T. (2014). A sharp adaptive confidence ball for self-similar functions. Available at arXiv:1406.3994.

[24] PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298–335. MR1762913

[25] RAY, K. (2014). Adaptive Bernstein–von Mises theorems in Gaussian white noise. Available at arXiv:1407.3397.

[26] ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241

[27] SERRA, P. and KRIVOBOKOVA, T. (2014). Adaptive empirical Bayesian smoothing splines. Available at arXiv:1411.6860.

[28] SZABÓ, B., VAN DER VAART, A. and VAN ZANTEN, H. (2015). Honest Bayesian confidence sets for the L^2-norm. *J. Statist. Plann. Inference* **166** 36–51. MR3390132

[29] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861

The Annals of Statistics
2017, Vol. 45, No. 3, 1152–1184
DOI: 10.1214/16-AOS1478

TOTAL POSITIVITY IN MARKOV STRUCTURES

BY SHAUN FALLAT[*,1], STEFFEN LAURITZEN[†], KAYVAN SADEGHI[‡,2], CAROLINE UHLER[§,3], NANNY WERMUTH[¶] AND PIOTR ZWIERNIK[‖,4]

University of Regina[*], *University of Copenhagen*[†], *University of Cambridge*[‡], *Massachusetts Institute of Technology*[§], *IST Austria*[§], *Chalmers University of Technology*[¶], *Johannes Gutenberg-University*[¶] *and Universitat Pompeu Fabra*[‖]

We discuss properties of distributions that are multivariate totally positive of order two (MTP_2) related to conditional independence. In particular, we show that any independence model generated by an MTP_2 distribution is a compositional semi-graphoid which is upward-stable and singleton-transitive. In addition, we prove that any MTP_2 distribution satisfying an appropriate support condition is faithful to its concentration graph. Finally, we analyze factorization properties of MTP_2 distributions and discuss ways of constructing MTP_2 distributions; in particular, we give conditions on the log-linear parameters of a discrete distribution which ensure MTP_2 and characterize conditional Gaussian distributions which satisfy MTP_2.

REFERENCES

[1] AHLSWEDE, R. and DAYKIN, D. E. (1978). An inequality for the weights of two families of sets, their unions and intersections. *Z. Wahrsch. Verw. Gebiete* **43** 183–185. MR0491189

[2] ALLMAN, E. S., RHODES, J. A., STURMFELS, B. and ZWIERNIK, P. (2015). Tensors of nonnegative rank two. *Linear Algebra Appl.* **473** 37–53. MR3338324

[3] BARTOLUCCI, F. and BESAG, J. (2002). A recursive algorithm for Markov random fields. *Biometrika* **89** 724–730. MR1929176

[4] BARTOLUCCI, F. and FORCINA, A. (2000). A likelihood ratio test for MTP_2 within binary variables. *Ann. Statist.* **28** 1206–1218. MR1811325

[5] BØLVIKEN, E. (1982). Probability inequalities for the multivariate normal with nonnegative partial correlations. *Scand. J. Stat.* **9** 49–58. MR0651862

[6] CHVÁTAL, V. and WU, B. (2011). On Reichenbach's causal betweenness. *Erkenntnis* **76** 41–48.

[7] COLANGELO, A., SCARSINI, M. and SHAKED, M. (2005). Some notions of multivariate positive dependence. *Insurance Math. Econom.* **37** 13–26. MR2156593

[8] DAWID, A. P. (1979). Conditional independence in statistical theory. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **41** 1–31. MR0535541

[9] DAWID, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8** 598–617. MR0568723

[10] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. MR1241267

MSC2010 subject classifications. Primary 60E15, 62H99; secondary 15B48.
Key words and phrases. Association, concentration graph, conditional Gaussian distribution, faithfulness, graphical models, log-linear interactions, Markov property, positive dependence.

[11] DJOLONGA, J. and KRAUSE, A. (2015). Scalable variational inference in log-supermodular models. In *Proceedings of the International Conference on Machine Learning (ICML)*. Available at ArXiv:1502.06531.

[12] EDWARDS, D. (2000). *Introduction to Graphical Modelling*, 2nd ed. Springer, New York. MR1880319

[13] ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables, with applications. *Ann. Math. Stat.* **38** 1466–1474. MR0217826

[14] FORTUIN, C. M., KASTELEYN, P. W. and GINIBRE, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22** 89–103. MR0309498

[15] HØJSGAARD, S. and LAURITZEN, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 1005–1027. MR2530327

[16] HOLLAND, P. W. and ROSENBAUM, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14** 1523–1543. MR0868316

[17] JOAG-DEV, K. (1983). Independence via uncorrelatedness under certain dependence structures. *Ann. Probab.* **11** 1037–1041. MR0714966

[18] JOE, H. (2006). Generating random correlation matrices based on partial correlations. *J. Multivariate Anal.* **97** 2177–2189. MR2301633

[19] JONES, B. and WEST, M. (2005). Covariance decomposition in undirected Gaussian graphical models. *Biometrika* **92** 779–786. MR2234185

[20] KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498. MR0599685

[21] KARLIN, S. and RINOTT, Y. (1983). M-matrices as covariance matrices of multinormal distributions. *Linear Algebra Appl.* **52/53** 419–438. MR0709364

[22] LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Clarendon, Oxford. MR1419991

[23] LEBOWITZ, J. L. (1972). Bounds on the correlations and analyticity properties of ferromagnetic Ising spin systems. *Comm. Math. Phys.* **28** 313–321. MR0323271

[24] LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Stat.* **37** 1137–1153. MR0202228

[25] MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, London. MR0560319

[26] MÜLLER, A. and SCARSINI, M. (2005). Archimedean copulae and positive dependence. *J. Multivariate Anal.* **93** 434–445. MR2162647

[27] NEWMAN, C. M. (1983). A general central limit theorem for FKG systems. *Comm. Math. Phys.* **91** 75–80. MR0719811

[28] NEWMAN, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In *Inequalities in Statistics and Probability* (*Lincoln, Neb.*, 1982). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **5** 127–140. IMS, Hayward, CA. MR0789244

[29] OSTROWSKI, A. (1937). Über die determinanten mit überwiegender Hauptdiagonale. *Comment. Math. Helv.* **10** 69–96. MR1509568

[30] PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems*: *Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA. MR0965765

[31] PEARL, J. and PAZ, A. (1987). Graphoids: A graph based logic for reasoning about relevancy relations. In *Advances in Artificial Intelligence*, *Vol. II* (B. D. Boulay, D. Hogg and L. Steel, eds.) 357–363. North-Holland, Amsterdam.

[32] PETERS, J. (2015). On the intersection property of conditional independence and its application to causal discovery. *J. Causal Inference* **3** 97–108.

[33] PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252. MR1611693

[34] REICHENBACH, H. (1956). *The Direction of Time*. Univ. California Press, Berkeley, CA.

[35] ROYSTON, P., ALTMAN, D. G. and SAUERBREI, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat. Med.* **25** 127–141. MR2222078

[36] SADEGHI, K. and LAURITZEN, S. (2014). Markov properties for mixed graphs. *Bernoulli* **20** 676–696. MR3178514

[37] SAN MARTÍN, E., MOUCHART, M. and ROLIN, J.-M. (2005). Ignorable common information, null sets and Basu's first theorem. *Sankhyā* **67** 674–698. MR2283009

[38] SARKAR, T. K. (1969). Some lower bounds of reliability. Tech. Report, No. 124, Dept. Operations Research and Dept. Statistics, Stanford Univ., Stanford, CA.

[39] SLAWSKI, M. and HEIN, M. (2015). Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. *Linear Algebra Appl.* **473** 145–179. MR3338330

[40] STEEL, M. and FALLER, B. (2009). Markovian log-supermodularity, and its applications in phylogenetics. *Appl. Math. Lett.* **22** 1141–1144. MR2523016

[41] ŠTEMBERA, Z., ZNAMENÁČEK, K. and POLÁČEK, K. (2012). *High Risk Pregnancy and Child*. Springer, Berlin.

[42] STUDENÝ, M. (2005). *Probabilistic Conditional Independence Structures*. Springer, London. MR3183760

[43] WERMUTH, N. (2012). Traceable regressions. *Int. Stat. Rev.* **80** 415–438. MR3006154

[44] WERMUTH, N. (2015). Graphical Markov models, unifying results and their interpretation. In *Wiley Statsref*: *Statistics Reference Online*. Available at ArXiv:1505.02456.

[45] WERMUTH, N. and COX, D. R. (1998). On association models defined over independence graphs. *Bernoulli* **4** 477–495. MR1679794

[46] WERMUTH, N. and MARCHETTI, G. M. (2014). Star graphs induce tetrad correlations: For Gaussian as well as for binary variables. *Electron. J. Stat.* **8** 253–273. MR3189555

[47] WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester. MR1112133

[48] XIE, X., MA, Z. and GENG, Z. (2008). Some association measures and their collapsibility. *Statist. Sinica* **18** 1165–1183. MR2440406

[49] ZATOŃSKI, W., BECHER, H., LISSOWSKA, J. and WAHRENDORF, J. (1991). Tobacco, alcohol, and diet in the etiology of laryngeal cancer: A population-based case-control study. *Cancer Causes & Control* **2** 3–10.

[50] ZWIERNIK, P. (2016). *Semialgebraic Statistics and Latent Tree Models*. *Monographs on Statistics and Applied Probability* **146**. Chapman & Hall/CRC, Boca Raton, FL. MR3379921

The Annals of Statistics
2017, Vol. 45, No. 3, 1185–1213
DOI: 10.1214/16-AOS1481

TESTS FOR COVARIANCE STRUCTURES WITH HIGH-DIMENSIONAL REPEATED MEASUREMENTS

BY PING-SHOU ZHONG*,1 WEI LAN†,2,
PETER X. K. SONG‡,3 AND CHIH-LING TSAI§

Michigan State University, Southwestern University of Finance and Economics,† University of Michigan‡ and University of California-Davis§*

In regression analysis with repeated measurements, such as longitudinal data and panel data, structured covariance matrices characterized by a small number of parameters have been widely used and play an important role in parameter estimation and statistical inference. To assess the adequacy of a specified covariance structure, one often adopts the classical likelihood-ratio test when the dimension of the repeated measurements (p) is smaller than the sample size (n). However, this assessment becomes quite challenging when p is bigger than n, since the classical likelihood-ratio test is no longer applicable. This paper proposes an adjusted goodness-of-fit test to examine a broad range of covariance structures under the scenario of "large p, small n." Analytical examples are presented to illustrate the effectiveness of the adjustment. In addition, large sample properties of the proposed test are established. Moreover, simulation studies and a real data example are provided to demonstrate the finite sample performance and the practical utility of the test.

REFERENCES

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141. MR0331612

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

BALTAGI, B. H., KAO, C. and PENG, B. (2015). On testing for sphericity with non-normality in a fixed effects panel data model. *Statist. Probab. Lett.* **98** 123–130. MR3306453

BENTLER, P. M. and BONETT, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* **88** 588–606.

BROWNE, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statist. J.* **8** 1–24. MR0370949

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618

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

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

CUI, H., ZHENG, S. and LI, R. (2013). Test of significance on high dimensional covariance matrix structures. Technical report.

MSC2010 subject classifications. Primary 62H15; secondary 62G20, 62G10.
Key words and phrases. Adjusted test, goodness-of-fit test, longitudinal data, panel data.

DAVIS, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer, New York. MR1883764

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

FREES, E. W. (2004). *Longitudinal and Panel Data*: *Analysis and Applications in the Social Sciences*. Cambridge Univ. Press, Cambridge. MR2087947

JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568

JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. MR0312619

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist*. **30** 1081–1102. MR1926169

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

LI, W. and QIN, Y. (2014). Hypothesis testing for high-dimensional covariance matrices. *J. Multivariate Anal*. **128** 108–119. MR3199831

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist*. **41** 1204–1231. MR3113808

ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist*. **42** 225–254. MR3189485

POURAHMADI, M. (2013). *High-Dimensional Covariance Estimation*. Wiley, Hoboken, NJ. MR3235948

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

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

WIESEL, A., BIBI, O. and GLOBERSON, A. (2013). Time varying autoregressive moving average models for covariance estimation. *IEEE Trans. Signal Process*. **61** 2791–2801. MR3064093

ZHAO, W., LI, H., HOU, W. and WU, R. (2007). Wavelet-based parametric functional mapping of developmental trajectories with high-dimensional data. *Genetics* **176** 1879–1892.

ZHONG, P.-S., LAN, W., SONG, P. X. K. and TSAI, C.-L. (2016). Supplement to "Tests for covariance structures with high-dimensional repeated measurements." DOI:10.1214/16-AOS1481SUPP.

ZOU, C., PENG, L., FENG, L. and WANG, Z. (2014). Multivariate sign-based high-dimensional tests for sphericity. *Biometrika* **101** 229–236. MR3180668

The Annals of Statistics
2017, Vol. 45, No. 3, 1214–1253
DOI: 10.1214/16-AOS1482

WEAK SIGNAL IDENTIFICATION AND INFERENCE IN PENALIZED MODEL SELECTION

BY PEIBEI SHI[1,*] AND ANNIE QU[1,†]

University of Michigan and Univeristy of Illinois at Urbana-Champaign†*

Weak signal identification and inference are very important in the area of penalized model selection, yet they are underdeveloped and not well studied. Existing inference procedures for penalized estimators are mainly focused on strong signals. In this paper, we propose an identification procedure for weak signals in finite samples, and provide a transition phase in-between noise and strong signal strengths. We also introduce a new two-step inferential method to construct better confidence intervals for the identified weak signals. Our theory development assumes that variables are orthogonally designed. Both theory and numerical studies indicate that the proposed method leads to better confidence coverage for weak signals, compared with those using asymptotic inference. In addition, the proposed method outperforms the perturbation and bootstrap resampling approaches. We illustrate our method for HIV antiretroviral drug susceptibility data to identify genetic mutations associated with HIV drug resistance.

REFERENCES

[1] ANDREWS, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68** 399–405. MR1748009

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

[3] DONOHO, D. and JIN, J. (2008). Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proc. Natl. Acad. Sci. USA* **105** 14790–14795.

[4] EFRON, B. (2014). Estimation and accuracy after model selection. *J. Amer. Statist. Assoc.* **109** 991–1007. MR3265671

[5] EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability* **57**. Chapman & Hall, New York. MR1270903

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

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

[8] FAN, J. and PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32** 928–961. MR2065194

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

[10] HUANG, J., HOROWITZ, J. L. and MA, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.* **36** 587–613. MR2396808

MSC2010 subject classifications. Primary 62F30, 62J07; secondary 62E15.
Key words and phrases. Model selection, weak signal, finite sample inference, adaptive Lasso.

[11] HUANG, J., MA, S. and ZHANG, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statist*. *Sinica* **18** 1603–1618. MR2469326

[12] HUANG, J. and XIE, H. (2007). Asymptotic oracle properties of SCAD-penalized least squares estimators. In *Asymptotics*: *Particles*, *Processes and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **55** 149–166. IMS, Beachwood, OH. MR2459937

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

[14] JIN, J., ZHANG, C.-H. and ZHANG, Q. (2014). Optimality of graphlet screening in high dimensional variable selection. *J*. *Mach*. *Learn*. *Res*. **15** 2723–2772. MR3270749

[15] JOHNSON, V. A., BRUN-VÉZINET, F., CLOTET, B., KURITZKES, D. R., PILLAY, D., SCHAPIRO, J. M. and RICHMAN, D. D. (2006). Update of the drug resistance mutations in HIV-1: Fall 2006. *Top HIV Med* **14** 125–130.

[16] KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann*. *Statist*. **28** 1356–1378. MR1805787

[17] 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

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

[19] MINNIER, J., TIAN, L. and CAI, T. (2011). A perturbation method for inference on regularized regression estimates. *J*. *Amer. Statist*. *Assoc*. **106** 1371–1382. MR2896842

[20] PÖTSCHER, B. M. and LEEB, H. (2009). On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. *J*. *Multivariate Anal*. **100** 2065–2082. MR2543087

[21] PÖTSCHER, B. M. and SCHNEIDER, U. (2009). On the distribution of the adaptive LASSO estimator. *J*. *Statist*. *Plann*. *Inference* **139** 2775–2790. MR2523666

[22] PÖTSCHER, B. M. and SCHNEIDER, U. (2010). Confidence sets based on penalized maximum likelihood estimators in Gaussian regression. *Electron*. *J*. *Stat*. **4** 334–360. MR2645488

[23] SHEN, X., PAN, W. and ZHU, Y. (2012). Likelihood-based selection and sharp parameter estimation. *J*. *Amer. Statist*. *Assoc*. **107** 223–232. MR2949354

[24] SHI, and QU, (2016). Supplement to "Weak signal identification and inference in penalized model selection." DOI:10.1214/16-AOS1482SUPP.

[25] SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166

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

[27] 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

[28] WANG, H. and LENG, C. (2007). Unified LASSO estimation by least squares approximation. *J*. *Amer. Statist*. *Assoc*. **102** 1039–1048. MR2411663

[29] WANG, H., LI, R. and TSAI, C.-L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* **94** 553–568. MR2410008

[30] WU, M. (2009). A parametric permutation test for regression coefficients in Lasso regularized regression. Ph.D thesis, Dept. Biostatistics, Harvard School of Public Health, Boston, MA.

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

[32] 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

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

[34] ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2007). On the "degrees of freedom" of the lasso. *Ann. Statist.* **35** 2173–2192. MR2363967

[35] 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
2017, Vol. 45, No. 3, 1254–1283
DOI: 10.1214/16-AOS1484

SEMIMARTINGALE DETECTION AND GOODNESS-OF-FIT TESTS

BY ADAM D. BULL[1]

University of Cambridge

In quantitative finance, we often fit a parametric semimartingale model to asset prices. To ensure our model is correct, we must then perform goodness-of-fit tests. In this paper, we give a new goodness-of-fit test for volatility-like processes, which is easily applied to a variety of semimartingale models. In each case, we reduce the problem to the detection of a semimartingale observed under noise. In this setting, we then describe a wavelet-thresholding test, which obtains adaptive and near-optimal detection rates.

REFERENCES

AÏT-SAHALIA, Y. (1996). Testing continuous-time models of the spot interest rate. *Rev. Financ. Stud*. **9** 385–426.

AÏT-SAHALIA, Y. and PARK, J. Y. (2012). Stationarity-based specification tests for diffusions when the process is nonstationary. *J. Econometrics* **169** 279–292. MR2947885

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and EBENS, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* **61** 43–76.

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

BARNDORFF-NIELSEN, O. E. and VERAART, A. (2009). Stochastic volatility of volatility in continuous time. *CREATES Research Paper* **25**.

BIBINGER, M., JIRAK, M. and VETTER, M. (2015). Nonparametric change-point analysis of volatility. Preprint. Available at arXiv:1502.00043.

BULL, A. D. (2016a). Software for "Semimartingale detection and goodness-of-fit tests". Available at http://dx.doi.org/10.17863/CAM.125.

BULL, A. D. (2016b). Supplement to "Semimartingale detection and goodness-of-fit tests." DOI:10.1214/16-AOS1484SUPP.

CHEN, Q., ZHENG, X. and PAN, Z. (2015). Asymptotically distribution-free tests for the volatility function of a diffusion. *J. Econometrics* **184** 124–144. MR3277557

CONT, R. and TANKOV, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Boca Raton, FL. MR2042661

CORRADI, V. and WHITE, H. (1999). Specification tests for the variance of a diffusion. *J. Time Series Anal*. **20** 253–270. MR1693173

DETTE, H. and PODOLSKIJ, M. (2008). Testing the parametric form of the volatility in continuous time diffusion models—A stochastic process approach. *J. Econometrics* **143** 56–73. MR2384433

DETTE, H., PODOLSKIJ, M. and VETTER, M. (2006). Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing. *Scand. J. Stat*. **33** 259–278. MR2279642

MSC2010 subject classifications. Primary 62M02; secondary 60J60, 60J75, 62G10, 65T60.
Key words and phrases. Diffusion, goodness-of-fit, jump process, semimartingale, wavelets.

DETTE, H. and VON LIERES UND WILKAU, C. (2003). On a test for a parametric form of volatility in continuous time financial models. *Finance Stoch*. **7** 363–384. MR1994914

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: Asymptopia? *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **57** 301–369. MR1323344

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

HOFFMANN, M., MUNK, A. and SCHMIDT-HIEBER, J. (2012). Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. *Ann. Inst. Henri Poincaré Probab. Stat*. **48** 1186–1216. MR3052408

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

JACOD, J. and REISS, M. (2014). A remark on the rates of convergence for integrated volatility estimation in the presence of jumps. *Ann. Statist*. **42** 1131–1144. MR3224283

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. and VETTER, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl*. **119** 2249–2276. MR2531091

KLEINOW, T. (2002). Testing the diffusion coefficient. Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.

LEE, S. (2006). The Bickel–Rosenblatt test for diffusion processes. *Statist. Probab. Lett*. **76** 1494–1502. MR2245570

LEE, S. and WEE, I.-S. (2008). Residual empirical process for diffusion processes. *J. Korean Math. Soc*. **45** 683–693. MR2410237

MANCINI, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat*. **36** 270–296. MR2528985

MUNK, A. and SCHMIDT-HIEBER, J. (2010). Lower bounds for volatility estimation in microstructure noise models. In *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown*. *Inst. Math. Stat. Collect*. **6** 43–55. IMS, Beachwood, OH. MR2798510

MYKLAND, P. A. and ZHANG, L. (2006). ANOVA for diffusions and Itô processes. *Ann. Statist*. **34** 1931–1963. MR2283722

NGUYEN, C. M. (2010). Time-changed residual based tests for asset pricing models in continuous time. Unpublished manuscript, Indiana Univ., Bloomington, IN.

PAPANICOLAOU, A. and GIESECKE, K. (2014). Variation-based tests for volatility misspecification. Available at SSRN 2450776.

PAPAPANTOLEON, A. (2008). An introduction to Lévy processes with applications in finance. Preprint. Available at arXiv:0804.0482.

PODOLSKIJ, M. and ZIGGEL, D. (2008). A range-based test for the parametric form of the volatility in diffusion models. CREATES Research Paper 22.

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

REISS, M., TODOROV, V. and TAUCHEN, G. (2014). Nonparametric test for a constant beta over a fixed time interval. Preprint. Available at arXiv:1403.0349.

VETTER, M. (2012). Estimation of integrated volatility of volatility with applications to goodness-of-fit testing. Preprint. Available at arXiv:1206.5761.

VETTER, M. and DETTE, H. (2012). Model checks for the volatility under microstructure noise. *Bernoulli* **18** 1421–1447. MR2995803

The Annals of Statistics
2017, Vol. 45, No. 3, 1284–1311
DOI: 10.1214/16-AOS1485

TESTING FOR TIME-VARYING JUMP ACTIVITY FOR PURE JUMP SEMIMARTINGALES[1]

BY VIKTOR TODOROV

Northwestern University

In this paper, we propose a test for deciding whether the jump activity index of a discretely observed Itô semimartingale of pure-jump type (i.e., one without a diffusion) varies over a fixed interval of time. The asymptotic setting is based on observations within a fixed time interval with mesh of the observation grid shrinking to zero. The test is derived for semimartingales whose "spot" jump compensator around zero is like that of a stable process, but importantly the stability index can vary over the time interval. The test is based on forming a sequence of local estimators of the jump activity over blocks of shrinking time span and contrasting their variability around a global activity estimator based on the whole data set. The local and global jump activity estimates are constructed from the real part of the empirical characteristic function of the increments of the process scaled by local power variations. We derive the asymptotic distribution of the test statistic under the null hypothesis of constant jump activity and show that the test has asymptotic power of one against fixed alternatives of processes with time-varying jump activity.

REFERENCES

AÏT-SAHALIA, Y. and JACOD, J. (2009). Estimating the degree of activity of jumps in high frequency data. *Ann. Statist.* **37** 2202–2244. MR2543690

BULL, A. D. (2016). Near-optimal estimation of jump activity in semimartingales. *Ann. Statist.* **44** 58–86. MR3449762

JACOD, J. and PROTTER, P. (2012). *Discretization of Processes*. Springer, Heidelberg. MR2859096

JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. Springer, Berlin. MR1943877

JING, B.-Y., KONG, X.-B. and LIU, Z. (2011). Estimating the jump activity index under noisy observations using high-frequency data. *J. Amer. Statist. Assoc.* **106** 558–568. MR2847970

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

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

SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Univ. Press, Cambridge, UK.

TODOROV, V. (2013). Power variation from second order differences for pure jump semimartingales. *Stochastic Process. Appl.* **123** 2829–2850. MR3054547

TODOROV, V. (2015). Jump activity estimation for pure-jump semimartingales via self-normalized statistics. *Ann. Statist.* **43** 1831–1864. MR3357880

MSC2010 subject classifications. Primary 62F12, 62M05; secondary 60H10, 60J75.

Key words and phrases. Central limit theorem, high-frequency data, Itô semimartingale, jumps, jump activity index, nonparametric test, stochastic volatility, power variation.

TODOROV, V. (2017). Supplement to "Testing for Time-Varying Jump Activity for Pure Jump Semimartingales." DOI:10.1214/16-AOS1485SUPP.

TODOROV, V. and TAUCHEN, G. (2011). Limit theorems for power variations of pure-jump processes with application to activity estimation. *Ann. Appl. Probab.* **21** 546–588. MR2807966

TODOROV, V. and TAUCHEN, G. (2012). Realized Laplace transforms for pure-jump semimartingales. *Ann. Statist.* **40** 1233–1262. MR2985949

WOERNER, J. H. C. (2003). Variational sums and power variation: A unifying approach to model selection and estimation in semimartingale models. *Statist. Decisions* **21** 47–68. MR1985651

WOERNER, J. H. C. (2007). Inference in Lévy-type stochastic volatility models. *Adv. in Appl. Probab.* **39** 531–549. MR2343676

The Annals of Statistics
2017, Vol. 45, No. 3, 1312–1341
DOI: 10.1214/16-AOS1486

OPERATIONAL TIME AND IN-SAMPLE DENSITY FORECASTING

BY YOUNG K. LEE[1,*], ENNO MAMMEN[2,†],
JENS P. NIELSEN[3,‡] AND BYEONG U. PARK[4,§]

Kangwon National University, Universität Heidelberg and Higher School of Economics*†*, Cass Business School, City University London*‡* and Seoul National University*§*

In this paper, we consider a new structural model for in-sample density forecasting. In-sample density forecasting is to estimate a structured density on a region where data are observed and then reuse the estimated structured density on some region where data are not observed. Our structural assumption is that the density is a product of one-dimensional functions with one function sitting on the scale of a transformed space of observations. The transformation involves another unknown one-dimensional function, so that our model is formulated via a known smooth function of three underlying unknown one-dimensional functions. We present an innovative way of estimating the one-dimensional functions and show that all the estimators of the three components achieve the optimal one-dimensional rate of convergence. We illustrate how one can use our approach by analyzing a real dataset, and also verify the tractable finite sample performance of the method via a simulation study.

REFERENCES

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York. MR1198884

BARAUD, Y. and BIRGÉ, L. (2014). Estimating composite functions by model selection. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 285–314. MR3161532

CHENG, M.-Y. (1997). A bandwidth selector for local linear density estimators. *Ann. Statist.* **25** 1001–1013. MR1447738

FAN, J., HECKMAN, N. E. and WAND, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.* **90** 141–150. MR1325121

HOROWITZ, J. L. and MAMMEN, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35** 2589–2619. MR2382659

JIANG, J., FAN, Y. and FAN, J. (2010). Estimation in additive models with highly or nonhighly correlated covariates. *Ann. Statist.* **38** 1403–1432. MR2662347

JUDITSKY, A. B., LEPSKI, O. V. and TSYBAKOV, A. B. (2009). Nonparametric estimation of composite functions. *Ann. Statist.* **37** 1360–1404. MR2509077

KUANG, D., NIELSEN, B. and NIELSEN, J. P. (2008a). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95** 979–986. MR2461224

KUANG, D., NIELSEN, B. and NIELSEN, J. P. (2008b). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95** 987–991. MR2461225

MSC2010 subject classifications. Primary 62G07; secondary 62G20.
Key words and phrases. Density estimation, kernel smoothing, backfitting, chain Ladder.

KUANG, D., NIELSEN, B. and NIELSEN, J. P. (2009). Chain-ladder as maximum likelihood revisited. *Annals of Actuarial Science* **4** 105–121.
KUANG, D., NIELSEN, B. and NIELSEN, J. P. (2011). Forecasting in an extended chain-ladder-type model. *J. Risk Insur.* **78** 345–359.
LEE, R. D. and CARTER, L. R. (1992). Modeling and forecasting U.S. mortality. *J. Amer. Statist. Assoc.* **87** 659–671.
LEE, Y. K., MAMMEN, E. and PARK, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann. Statist.* **38** 2857–2883. MR2722458
LEE, Y. K., MAMMEN, E. and PARK, B. U. (2012). Flexible generalized varying coefficient regression models. *Ann. Statist.* **40** 1906–1933. MR3015048
LEE, Y. K., MAMMEN, E., NIELSEN, J. P. and PARK, B. U. (2015). Asymptotics for in-sample density forecasting. *Ann. Statist.* **43** 620–651. MR3319138
LEE, Y. K., MAMMEN, E., NIELSEN, J. P. and PARK, B. U. (2016). Supplement to "Operational time and in-sample density forecasting." DOI:10.1214/16-AOS1486SUPP.
MAMMEN, E., MARTÍNEZ MIRANDA, M. D. and NIELSEN, J. P. (2015). In-sample forecasting applied to reserving and mesothelioma mortality. *Insurance Math. Econom.* **61** 76–86. MR3324046
MAMMEN, E. and NIELSEN, J. P. (2003). Generalised structured models. *Biometrika* **90** 551–566. MR2006834
MAMMEN, E., PARK, B. U. and SCHIENLE, M. (2014). Additive models: Extensions and related models. In *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics* (J. S. Racine, L. Su and A. Ullah, eds.) 176–211. Oxford Univ. Press, Oxford. MR3306926
MARTÍNEZ-MIRANDA, M. D., NIELSEN, J. P., SPERLICH, S. and VERRALL, R. J. (2013). Continuous chain ladder: Reformulating and generalising a classical insurance problem. *Expert Syst. Appl.* **40** 5588–5603.
MARTÍNEZ MIRANDA, M. D., NIELSEN, J. P. and VERRALL, R. (2012). Double chain ladder. *Astin Bull.* **42** 59–76. MR2931855
MIKOSCH, T. (2009). *Non-life Insurance Mathematics*, 2nd ed. *Universitext*. Springer, Berlin. MR2503328
PARK, B. U. and MARRON, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.* **85** 66–72.
REID, P. H. (1978). Claims reserves in general insurance. *J. Inst. Actuar.* **105** 211–296.
RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370. MR1311979
TAYLOR, G. C. (1981). Speed finalisation of claims and claims run-off analysis. *Astin Bull.* **12** 81–100.
TAYLOR, G. C. (1982). Zehnwirth's comment on the see-saw method: A reply. *Insurance Math. Econom.* **1** 105–108.
WILKE, R. (2016). Forecasting macroeconomic labour market flows: What can we learn from micro level analysis? (submitted manuscript).
YU, K., PARK, B. U. and MAMMEN, E. (2008). Smooth backfitting in generalized additive models. *Ann. Statist.* **36** 228–260. MR2387970
ZEHNWIRTH, B. (1982). Comments on Taylor's see-saw approach to claims reserving. *Insurance Math. Econom.* **1** 99–103.
ZHANG, X., PARK, B. U. and WANG, J.-L. (2013). Time-varying additive models for longitudinal data. *J. Amer. Statist. Assoc.* **108** 983–998. MR3174678

The Annals of Statistics
2017, Vol. 45, No. 3, 1342–1374
DOI: 10.1214/16-AOS1487

ASYMPTOTICS OF EMPIRICAL EIGENSTRUCTURE FOR HIGH DIMENSIONAL SPIKED COVARIANCE

BY WEICHEN WANG* AND JIANQING FAN[1,*,†]

*Princeton University** *and Fudan University*[†]

We derive the asymptotic distributions of the spiked eigenvalues and eigenvectors under a generalized and unified asymptotic regime, which takes into account the magnitude of spiked eigenvalues, sample size and dimensionality. This regime allows high dimensionality and diverging eigenvalues and provides new insights into the roles that the leading eigenvalues, sample size and dimensionality play in principal component analysis. Our results are a natural extension of those in [*Statist. Sinica* **17** (2007) 1617–1642] to a more general setting and solve the rates of convergence problems in [*Statist. Sinica* **26** (2016) 1747–1770]. They also reveal the biases of estimating leading eigenvalues and eigenvectors by using principal component analysis, and lead to a new covariance estimator for the approximate factor model, called Shrinkage Principal Orthogonal complEment Thresholding (S-POET), that corrects the biases. Our results are successfully applied to outstanding problems in estimation of risks for large portfolios and false discovery proportions for dependent test statistics and are illustrated by simulation studies.

REFERENCES

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *Ann. Statist.* **40** 1171–1197. MR2985947

AMINI, A. A. and WAINWRIGHT, M. J. (2008). High-dimensional analysis of semidefinite relaxations for sparse principal components. In *Information Theory*, 2008. *ISIT* 2008. *IEEE International Symposium on* 2454–2458. IEEE, New York.

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620

ANTONIADIS, A. and FAN, J. (2001). Regularization of wavelet approximations. *J. Amer. Statist. Assoc.* **96** 939–967. MR1946364

BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. MR1711663

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

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, Berlin. MR2567175

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

MSC2010 subject classifications. Primary 62H25; secondary 62H10.

Key words and phrases. Asymptotic distributions, principal component analysis, diverging eigenvalues, approximate factor model, relative risk management, false discovery proportion.

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
BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575
BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201
BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849
BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008
BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** 1055–1084. MR3113803
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., MA, Z. and WU, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281
CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J. ACM* **58** Art. 11, 37. MR2811000
CHAMBERLAIN, G. and ROTHSCHILD, M. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica* **51** 1305–1324.
CHANDRASEKARAN, V., SANGHAVI, S., PARRILO, P. A. and WILLSKY, A. S. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.* **21** 572–596. MR2817479
CHEN, K. H. and SHIMERDA, T. A. (1981). An empirical analysis of useful financial ratios. *Financ. Manag.* 51–60.
DAVIDSON, K. R. and SZAREK, S. J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the Geometry of Banach Spaces* **I** 317–366. North-Holland, Amsterdam. MR1863696
DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450
DE MOL, C., GIANNONE, D. and REICHLIN, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *J. Econometrics* **146** 318–328. MR2465176
DONOHO, D. L., GAVISH, M. and JOHNSTONE, I. M. (2014). Optimal shrinkage of eigenvalues in the spiked covariance model. Preprint. Available at arXiv:1311.0851.
FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991
FAN, J. and HAN, X. (2013). Estimation of false discovery proportion with unknown dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* To appear. Available at arXiv:1305.7007.
FAN, J., HAN, X. and GU, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107** 1019–1035. MR3010887
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
FAN, J., LIAO, Y. and SHI, X. (2015). Risks of large portfolios. *J. Econometrics* **186** 367–387. MR3343792
FAN, J., LIAO, Y. and WANG, W. (2016). Projected principal component analysis in factor models. *Ann. Statist.* **44** 219–254. MR3449767
FAN, J., XUE, L. and YAO, J. (2015). Sufficient forecasting using factor models. Preprint. Available at arXiv:1505.07414.
FAN, J., LIU, H., WANG, W. and ZHU, Z. (2016). Heterogeneity adjustment with applications to graphical model inference. Preprint. Available at arXiv:1602.05455.

HALL, P., MARRON, J. S. and NEEMAN, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 427–444. MR2155347
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
JUNG, S. and MARRON, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37** 4104–4130. MR2572454
KOLTCHINSKII, V. and LOUNICI, K. (2014). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. Preprint. Available at arXiv:1408.4643.
KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768
LANDGREBE, J., WURST, W. and WELZL, G. (2002). Permutation-validated principal components analysis of microarray data. *Genome Biol.* **3** 1–11.
LEE, S., ZOU, F. and WRIGHT, F. A. (2010). Convergence and prediction of principal component scores in high-dimensional settings. *Ann. Statist.* **38** 3605–3629. MR2766862
LEEK, J. T. and STOREY, J. D. (2008). A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci. USA* **105** 18718–18723.
LEEK, J. T., SCHARPF, R. B., BRAVO, H. C., SIMCHA, D., LANGMEAD, B., JOHNSON, W. E., GEMAN, D., BAGGERLY, K. and IRIZARRY, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11** 733–739.
MA, Z. (2013). Sparse principal component analysis and iterative thresholding. *Ann. Statist.* **41** 772–801. MR3099121
ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Econometrics* **168** 244–258. MR2923766
PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865
PESARAN, M. H. and ZAFFARONI, P. (2008). Optimal asset allocation with factor models for large portfolios.
PRICE, A. L., PATTERSON, N. J., PLENGE, R. M., WEINBLATT, M. E., SHADICK, N. A. and REICH, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38** 904–909.
RINGNÉR, M. (2008). What is principal component analysis? *Nat. Biotechnol.* **26** 303–304.
ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* **104** 177–186. MR2504372
SHEN, D., SHEN, H., ZHU, H. and MARRON, J. S. (2016). The statistics and mathematics of high dimension low sample size asymptotics. *Statist. Sinica* **26** 1747–1770. MR3586237
STOCK, J. H. and WATSON, M. W. (2002). Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* **97** 1167–1179. MR1951271
THOMAS, C. G., HARSHMAN, R. A. and MENON, R. S. (2002). Noise reduction in BOLD-based fMRI using component analysis. *Neuroimage* **17** 1521–1537.
VERSHYNIN, R. (2010). Introduction to the non-asymptotic analysis of random matrices. Preprint. Available at arXiv:1011.3027.
VU, V. Q. and LEI, J. (2012). Minimax rates of estimation for sparse PCA in high dimensions. In *AISTATS* **15** 1278–1286.
WANG, W. and FAN, J. (2017). Supplement to "Asymptotics of empirical eigenstructure for high dimensional spiked covariance." DOI:10.1214/16-AOS1487SUPP.
YAMAGUCHI-KABATA, Y., NAKAZONO, K., TAKAHASHI, A., SAITO, S., HOSONO, N., KUBO, M., NAKAMURA, Y. and KAMATANI, N. (2008). Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: Effects on population-based association studies. *Am. J. Hum. Genet.* **83** 445–456.

YATA, K. and AOSHIMA, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* **105** 193–215. MR2877512

YATA, K. and AOSHIMA, M. (2013). PCA consistency for the power spiked model in high-dimensional settings. *J. Multivariate Anal.* **122** 334–354. MR3189327