The Annals of Statistics
2016, Vol. 44, No. 5, 1837–1853
DOI: 10.1214/16-AOS1490

PETER HALL'S CONTRIBUTIONS TO NONPARAMETRIC FUNCTION ESTIMATION AND MODELING

BY MING-YEN CHENG[1] AND JIANQING FAN[2]

National Taiwan University and Princeton University

Peter Hall made wide-ranging and far-reaching contributions to nonparametric modeling. He was one of the leading figures in the developments of nonparametric techniques with over 300 published papers in the field alone. This article gives a selective overview on the contributions of Peter Hall to nonparametric function estimation and modeling. The focuses are on density estimation, nonparametric regression, bandwidth selection, boundary corrections, inference under shape constraints, estimation of residual variances, analysis of wavelet estimators, multivariate regression and applications of nonparametric methods.

1. Introduction. Peter Hall made wide ranging and far-reaching contributions to nonparametric function estimation and modeling. His work not only broke new ground in methodology, but also had a profound influence on statistical theory. He significantly altered the toolbox available for studying nonparametric function estimation, and as a result of his contributions, new areas of nonparametric modeling became tractable for rigorous theoretical analysis. He was one of most influential figures in leading the developments of nonparametric techniques, as exemplified by over 300 papers in the field. In addition to his phenomenal contributions to the core of nonparametric modeling, he also played a leading role in applying nonparametric techniques to various areas of statistics such as deconvolution and measurement error models, classification, functional data analysis, high-dimensional statistical learning, which accounts for over 50 papers. As his contributions to these areas will be highlighted in other papers, they will not be covered here.

It is Hall's contributions on various vital issues in nonparametric function estimation, both of practical and intellectual significance, that the era of neoclassic nonparametric modeling was born, resulting in an active and broad field of statistical modeling techniques. Hall was always in the forefront of these developments.

Received June 2016.
[1]Supported by Ministry of Science and Technology Grant 104-2118-M-002-005-MY3.
[2]Supported by NSF Grants DMS-12-06464 and DMS-14-06266.
MSC2010 subject classifications. 62G07, 62G08, 62G10, 62G20.
Key words and phrases. Nonparametric, density, regression, projection pursuit, trigonometric, boundary, endpoints, edge, cross-validation, mode estimator, exponent, bandwidth, smoothing parameter, variance function, fractal, tilting, local linear.

These developments are the dawn of sparse inferences and high-dimensional statistics, which are the foundation of statistical analysis of big data nowadays.

Hall had a very distinguished career. In addition to garnering various awards and honors outlined in the preface by Runze Li and contributing seminally to nonparametric smoothing (this article), Hall had also made fundamental contributions to various areas such as bootstrap [Chen (2016)], deconvolution [Delaigle (2016)], functional data analysis and random objects [Mueller (2016)], high-dimensional data and classification [Samworth (2016)], as well as probability and stochastic processes.

In addition to his extraordinary scholarly achievements and his dedicated professional services such as President of the Institute of Mathematical Statistics (2011) and Co-Editor of the *Annals of Statistics* (2013–2015), Peter Hall was also a great citizen and mentor. He fostered many generations of statisticians through collaborations, guidance, mentoring and encouragements. In 1992, Peter Hall approached Jianqing Fan, the co-author of this article, on the collaboration of a minimax estimation under the L_1-loss [Fan and Hall (1994)]. What an honor and shock to a then postdoctoral fellow at the institute. This shows his efforts of promoting and encouraging a young mind and did his homework. Figure 1 is a photo of Peter Hall, Jianqing Fan and Irene Gijbels, taken at Berkeley in 1992. Hall's scholastic thinking and attitude, his humbleness and politeness, and his persistency, intense scientific interest and hard-work have everlasting impact on this then-young man's career. Ming-Yen Cheng, another co-author of this article, benefited equally from Peter Hall's mentoring. When she worked on her first joint paper with Peter Hall [Cheng, Hall and Titterington (1997)] in 1995, she was a fresh PhD on a visit to the Australian National University, where Peter Hall spent most of his career. She always remembers clearly that she was very much surprised when Peter Hall happily took into account her suggestions to modify the methodology. Later, during her postdoctoral study under his supervision in 1996–1997 and years afterward, she gradually realized that the reason Peter Hall reacted to young people's suggestions in such an encouraging way was because he was truly an open-minded scholar, a great mentor and a very kind person. Figure 2 is a photo of Peter Hall, Jeannie Hall, Shu-Hui Chang, David Siegmund and Ming-Yen Cheng, taken in Taipei in 2011.

2. Nonparametric density estimation. Hall ventured into the realm of nonparametric curve estimation as early as in 1980. His very first two papers in this field are "Estimating a density on a positive half-line by the method of orthogonal series" and "On trigonometrics series estimates of a density" published respectively in *Annals of the Institute of Statistical Mathematics* [Hall (1980)] and the *Annals of Statistics* [Hall (1981a)]. The former concerns nonparametric density estimation on $(0, \infty)$ using Hermite and Laguerre polynomials with focus on the mean integrated square errors and the latter studies a similar problem using the Fourier methods, correcting several mistakes and results in Walter and Blum

FIG. 1. *Peter Hall, Jianqing Fan, and Irene Gijbels were at Mathematical Sciences Research Institute at Berkeley* 1992.

(1979). In both papers, Hall was among the first to note the importance of the boundary effect. These are also the first two papers that Peter Hall published in statistical journals. Starting from then, his interest in statistics, with emphasis on

FIG. 2. *Peter Hall, Jeannie Hall, Shu-Hui Chang, David Siegmund and Ming-Yen Cheng in Taipei in* 2011.

nonparametric curve estimation, surged. His work touched virtually all theoretical and practical aspects of nonparametric density estimation.

In Hall (1981b), he unveiled beautifully the laws of the iterated logarithm for kernel density estimators, trigonometric series estimators and orthogonal polynomial estimators. He established convincingly the central limit theorem for the mean integrated error [Hall (1984a)], pioneered the work on the estimation of integrated squared functions derivatives [Hall and Marron (1987a)] and discovered some interesting phenomenon on parametric and nonparametric behavior of such functionals, and introduced and studied the two classes of kernel density estimators for spherical data [Hall, Watson and Cabrera (1987)]. Furthermore, he studied the choice of the order of kernel from both mean integrated squared error and asymptotic optimality with cross-validation choice of bandwidths [Hall and Marron (1988)], addressed the challenging issue on constructing confidence intervals for probability densities [Hall (1992a)] and investigated the binning effect on the kernel density estimation [Hall and Wand (1996)]. In addition, he pioneered a number of important and practical issues of nonparametric density such as bandwidth selection, boundary bias correlation, discontinuities, shape constraints, among others. These will be addressed in separate sessions.

3. Nonparametric regression. Hall's earliest work on nonparametric regression was analysis of integrated square error of kernel regression and the use of cross-validation as an estimate [Hall (1984b, 1984c)]. Although it was noted that nonparametric regression and bandwidth selection for kernel methods can be affected by dependent errors, the effect of long-range dependence on the convergence rate was not clear until Hall and Hart (1990). The paper established minimax convergence rates in the presence of dependent errors, and gave necessary and sufficient condition for the minimax convergence rates when the errors are independent to be maintained in the dependent case. In particular, the convergence rate is slower in the case of long-range dependence. This work led to subsequent developments of bandwidth selection rules with dependent errors. For example, Hall, Lahiri and Polzehl (1995) analyzed the effect of dependence on optimal bandwidth, and suggested ways to adjust the block length in block bootstrap and the leave-out number in cross-validation in order to obtain a first-order optimal data-driven bandwidth. He further exploited various ideas for construction of confidence intervals and simultaneous confidence bands based on kernel regression, including interpolation [Hall and Titterington (1988)], undersmoothing [Hall (1992a)], bias correction [Hall (1992b)], pivoting [Hall (1993)] and averaging [Hall and Horowitz (2013)]. He pioneered this topic and established many of the key techniques. These are just a few examples to remind us how remarkable and unusual Peter Hall was in that he mastered various areas in probability and statistics and that he was able to make connections between seemingly unrelated problems in different areas and come up with innovative theory and methodology. Besides, although he did not produce many papers on nonparametric change-point detection and estimation, his works in this direction are highly pioneering and influential. Realizing the estimated curve may not be smooth, Hall and Titterington (1992) introduced the use of left-, central- and right-smooths in kernel methods for diagnosis and estimation of jumps and peaks, and the use of kernel derivative estimates to achieve optimal n^{-1} convergence in estimation of jump points [Gijbels, Hall and Kneip (1999)].

The popular local linear regression enjoys many theoretical and numerical advantages [Fan (1992, 1993)], and it has been extensively used in many semiparametric regression models such as varying coefficient models [Fan and Zhang (1999, 2008), Hastie and Tibshirani (1993)]. Seifert and Gasser (1996) pointed out that its finite sample variance is often infinite, due to design sparseness, and suggested a shrinkage approach to remedy the problem. Hall and Marron (1997) provided necessary and sufficient conditions on the shrinkage parameter to guarantee the traditional mean squared error formula. At the same time, Cheng, Hall and Titterington (1997) suggested shrinking toward another local linear estimate based on an infinitely supported kernel with sufficiently heavy tails, and Hall and Turlach (1997a) suggested to create pseudo data points by interpolation and apply local linear regression to the union of the observed and pseudo data. These simple methods are less sensitive to the choices of the tuning parameters, and are effective in ensuring superior performance of local linear regression in finite sample situations.

In nonparametric regression, the optimal convergence rate depends on both the degree of smoothness of the unknown curve and that of the local regression model. Determination of the order of the local model is thus somehow subjective, and the convention in practice is to employ lower order local models to avoid erratic numerical behaviors due to over-fitting. In the case of local polynomial regression, usually local linear regression is used. It has several theoretical and numerical advantages over local constant regression, the Nadaraya–Watson estimator. On the other hand, when the fourth-order derivative exists, it has slower convergence rate than local quadratic and local cubic estimators. In that case, Choi and Hall (1998) proposed a novel skewing approach to reduce the bias by two orders of magnitude at the expense of a slight increase in variance. Hall and Turlach (1999) suggested to use the biased bootstrap [Hall and Presnell (1999a)] to achieve bias reduction. Note that these two methods assume the existence of higher order derivatives. To exploit the unknown degree of smoothness of the curve, Hall and Racine (2015) gave deep insights into the theoretical performance with infinite order local polynomial, and showed that leave-one-out cross-validation can be used to determine simultaneously the order of the local polynomial and the bandwidth.

Penalized spline regression is a popular alternative to regression spline models, and smoothing spline regression may be viewed as a special case. However, compared to regression spline and smoothing spline regression, there existed very limited theory for penalized spline regression. Hall and Opsomer (2005) was the first to derive explicit expressions for asymptotic bias and variance of penalized spline regression. Moreover, the paper showed that it achieves the optimal nonparametric convergence rate, and provided useful insights into the role of the penalty.

Traditionally, it is assumed that the error term is uncorrelated with the regressor. The case where this assumption is violated, that is, the regressor is endogenous, had been much less studied despite the fact that the situation occurs often in social science such as economics and sociology. Hall and Horowitz (2005) introduced kernel and orthogonal series methods in a general nonparametric setup with available instrumental variables, based on the inversion of a linear operator on the space of square-integrable functions. The case where an additional exogenous variable is present was also studied, and thorough theoretical justifications were given for both cases.

Group testing for collecting data with a binary response is a common technique in large screening studies. Work on nonparametric regression modeling of group testing data is limited, however. Delaigle and Hall (2012) and Delaigle, Hall and Wishart (2014) developed kernel regression models when the pooling is homogeneous, and investigated the effect of over-pooling on the convergence rate.

4. Bandwidth selection. Hall made fundamental contributions to smoothing parameter estimation in nonparametric curve estimation. His methods basically can be classified as "cross-validation" based approaches and the plug-in based approaches. He led the developments in both areas.

Hall pioneered theoretical analysis on the kernel density estimation with bandwidth selected by cross-validation. His very first paper on this topic discovered the surprising suboptimality of a version of cross-validation in kernel density estimation [Hall (1982a)]. This led him to analyze the Bowman and Dudemo version of cross-validation. There, he demonstrated for the first time that a cross-validatory procedure for density estimation is asymptotically optimal in terms of mean integrated error [Hall (1983)]. Realizing that the Kullback–Leibler loss is a more appropriate measure, he made determined efforts [Hall (1987)] to show how kernel function should be appropriately chosen so that the likelihood cross-validation does result in asymptotic minimization of the Kullback–Leibler loss. Instead of studying the Integrated Square Errors (ISE) or Mean ISE (MISE), Hall and Marron (1987b) investigated the bandwidth that minimizes ISE or MISE and argued that the former bandwidth should be the benchmark. It is convincingly and surprisingly demonstrated that in comparison to the benchmark bandwidth, the bandwidth selected by the least-squares criterion performs as well as the "oracle" selector, both in the first- and the second-order. The results and technical proofs are extremely remarkable. In the very article by Härdle, Hall and Marron (1988), Hall derived further the asymptotic normality for a family of generalized cross-validation type of bandwidth selectors to their benchmarks, not only in rates but also in the asymptotic distributions, for both bandwidth selectors and MISE. The results are truly remarkable and useful. To further improve the performance and stability of cross-validatory bandwidths, Hall, Marron and Park (1992) introduced a smoothed cross-validation bandwidth and demonstrated that it achieves the optimal rates of convergence in terms of the relative errors. Hall and Marron (1991) investigated in depth why the cross-validation function has multiple local minima through modeling the cross-validation function as a Gaussian stochastic process. The results are very insightful.

Hall's contributions to the theory and methods of cross-validation do not limit to the kernel density estimation. He also analyzed the methods in the image analysis [Hall and Koch (1992)], kernel regression [Härdle, Hall and Marron (1988)], short-range and long-range dependent data [Hall, Lahiri and Polzehl (1995)], estimation of conditional density [Hall, Racine and Li (2004)], among others.

Realizing the high variance of cross-validation methods in bandwidth selection, Hall was also instrumental in developing plug-in bandwidth selection. Hall et al. (1991) developed a plug-in estimator of bandwidth and demonstrated that it has $n^{-1/2}$ rate of convergency, a surprising feast by the nonparametric standard.

5. Nonparametric boundary issues. Boundary effects in nonparametric density and regression estimators are serious problems in both theory and applications. Boundary kernels [Gasser and Müller (1979)] and the generalized jackknife method [Rice (1984)] apply to the Nadaraya–Watson regression estimator and kernel density estimators to correct the boundary biases, and Schuster (1985) suggested a reflection principle for kernel density estimation. The former two methods can cope with the boundary effects and result in the same order of bias as in the interior, and the straight line reflection corrects only for jumps at the boundary points and the bias is of the same order as the bandwidth. Boundary kernels tend to inflate the variance, and it is necessary to decide the ratio of the two bandwidths used in Rice's generalized jackknife method.

Hall created novel and simple alternatives that do not have the above side issues when coping with boundary effects. Hall and Wehrly (1991) suggested a reflection method for Nadaraya–Watson regression: pseudodata are produced by reflecting the observed data asymmetrically with respect to the Nadaraya–Watson estimator evaluated at the endpoints. The final estimator based on the combined data has the same order of bias across the entire support, because the pseudodata are created in a way that they centered around a curve that differs from the true curve by an amount of the same order. For kernel density estimation, noting that mean of an order statistic is asymptotically a respective quantile of the true density, Cowling and Hall (1996) suggested to estimate an extension of the quantile function beyond the design interval in a way that the estimate is sufficiently smooth. The extended quantile function, the pseudodata are based on, is a polynomial fit to data close to the endpoint. This method is very simple to implement, and to adopt it to higher order kernels we need only to change the order of the fitted polynomial. Cross-validation can be used when choosing bandwidth for the new estimator and there is no need to downweight at the ends of the support. Interestingly, there are numerical evidences that these simple pseudodata methods do not have the variance inflation problem the boundary kernel approach suffers from. The intuition is clear: there are the same number of observations near the two sides of the boundary. Hall and Park (2002) observed the connection between density support boundary estimation and density estimation near the boundary, and based a new boundary bias corrected density estimator on the translation idea in boundary estimation.

6. Nonparametric estimation with shape constraints. Often the function of interest is known to follow certain shape constraints such as unimodality, monotonicity, convexity, etc. In the earlier days, Hall was interested in estimation of the mode of a unimodal density function, and he developed rigorous theory for the convergence of Grenander's and kernel estimators [Grund and Hall (1995), Hall (1982b)]. Later, he became interested in the problems of testing unimodality and monotonicity, and nonparametric estimation under shape constraints. For the problem of testing unimodality of a density function, Cheng and Hall (1998) and Cheng and Hall (1999) established asymptotic properties of the excess mass and Silverman's bandwidth tests and proposed methods to calibrate the tests, based on asymptotic distributions of the test statistics or bootstrap. Cheng and Hall (1999) noted that when the unimodal density has a shoulder, the asymptotic distributions of the test statistics are very different, and to make unimodality tests adaptive they need to be calibrated for the most difficult null hypothesis where the density has one mode and one shoulder. Hall and York (2001) extended this approach to the problem of testing multimodality using Silverman's bandwidth test. The convention in Silverman's bandwidth test for multimodality uses the Gaussian kernel because the number of modes is monotonely nonincreasing in this case. Hall, Minnotte and Zhang (2004) showed theoretical and numerical justifications for the use of non-Gaussian kernels, including the popular Epanechnikov, biweight and triweight kernels. The main reason is that the nonmonotonicity has negligible effects. Compared to testing multimodality, testing monotonicity was a much less addressed issue although it is an important problem. Hall was one of the first to study this topic. Hall and Heckman (2000) introduced a test for monotonicity of a regression function based on running gradients, and calibrated it so that it overcomes problems caused by flat parts of the curve. Gijbels et al. (2000) developed another test for monotonicity using signs of differences of observations on the response variable. It is robust against heavy-tailed errors.

The main reason nonparametric kernel estimation is popular is its simplicity. On the other hand, it is difficult to force kernel estimators to satisfy qualitative constraints such as unimodality and monotonicity. Cheng, Gasser and Hall (1999) successfully constructed unimodal and monotone kernel density estimators by an iterative transformation algorithm. In the same year, Hall and Presnell (1999b) discussed an alternative tilting approach originated from the biased bootstrap techniques suggested by Hall and Presnell (1999a) but no numerical or theoretical results were provided. Hall and Huang (2001, 2002) employed quadratic programming in the implementation of the tilting approach in kernel estimation of a monotone regression and a monotone density function, respectively, and provided theoretical justifications. Braun and Hall (2001) suggested a general data sharpening technique to improve the performance of a statistical method, which involves perturbing the data, and applied it to the estimation of monotone and unimodal density functions and other problems including bias reduction. Hall and Kang (2005) showed that, when the true density is unimodal, the unimodal kernel density estimator obtained by data sharpening generally improves on the usual kernel density estimator in terms of mean integrated squared error. The numerical comparisons also suggested that the data sharpening approach is advantageous to the titling approach as the latter requires to remove spurious wiggles in the tails of the traditional kernel density estimator.

7. Estimation of residual variances. Residual variance is fundamentally important in statistical inference and bandwidth selection and provides a benchmark for forecasting. How can this parameter be estimated in the nonparametric regression model? In the nonparametric environment, there are always biases in the estimation of the mean function. What is then the impact of the estimation of the mean function on the residual variance or variance function in general? Hall and Carroll (1989) pioneered the study on this issue. They provided elegantly the results on the extent to which the smoothness of the mean regression function impacts on the best rate of convergence for estimating variance function. In particular, they showed that as long as the mean function satisfies a Lipschitz condition of order $1/3$ or more, the nonparametric variance function can be estimated with the rate of convergence $O(n^{-2/5})$. If the residual variance is a parametric function, it can be estimated with root-n consistency as long as f is Lipchitz of order $1/2$ or more.

Realizing bias is important to residual variance estimation, Hall, Kay and Titterington (1990) proposed and computed the optimal difference sequences for estimating error variance in homoscedastic nonparametric regression. This corresponds to undersmoothing estimates for nonparametric mean function. It is shown that the optimal difference sequences do not depend on unknown parameters. This results in a very easily implementable procedure. The asymptotic efficiency of such a simple method was also established, which is $2m/(2m+1)$ for the mth-order difference.

To gain more insightful results, Hall and Marron (1990) considered again the homoscedastic nonparametric regression model with the aim of estimating the homogeneous variance σ^2. Based on the kernel density regression estimator, Hall and Marron (1990) defined the estimator $\widehat{\sigma}^2$ as the residual sum of squares with degree of freedom properly corrected so that it will be an unbiased estimator when the mean regression is zero. It was nicely demonstrated that

$$E(\widehat{\sigma}^2 - \sigma^2)^2 = n^{-1}\{\operatorname{var}(\varepsilon^2) + O(n^{-(4r-1)/(4r+1)})\},$$

where r is the degree of smoothness of the regression function and ε is the noise random variable. Furthermore, it is shown that such an estimator is asymptotically optimal to the first- and the second-order.

8. Analysis of wavelet estimators. Hall made seminal contributions to nonparameteric function estimation. Soon after wavelets techniques were introduced to statistics for denoising via thresholding [Donoho and Johnstone (1994), Donoho et al. (1995)], Hall took on a different path on analyzing the properties of wavelets from a lens of traditional smoothing point of view. Hall and Patil (1995a) unveiled the MISE of nonlinear wavelets estimator for estimating nonparametric regression function and its derivatives for both cases where the nonlinear part is negligible. "This MISE formula is relatively unaffected by assumptions of continuity," pointed out correctly by Hall. After this derivation, Hall continued to quest the effect and performance of wavelet thresholding for smoothing and spatial adaptation in a series of work.

Hall and Patil (1995b) pointed out that the linear part of wavelet estimator is very analogous to the kernel smoothing and illustrated the adaptive qualities of the nonlinear component of a wavelet estimator by describing its performance when the target function is smooth but has high-frequency oscillations. He set to understand the property of spatial adaptation of nonlinear wavelet estimator via analyzing its local property. In 1993, he derived the pointwise property of the nonlinear wavelet thresholding estimator to demonstrate the spatial adaptation through explicit local modeling and also showed that such spatial adaptations can also be achieved via adaptive local smoothing. The papers were published respectively in Fan et al. (1996, 1999). Hall, McKay and Turlach (1996) investigated the limits to which wavelet methods can be pushed for adaptation to discontinuity by allowing the number of discontinuities to increase and their sizes to decrease with the sample size.

After deeply understanding wavelet estimators, Hall started to address a number of practical challenges when applying wavelets to statistical smoothing problems. Hall and Patil (1996) developed theory and methods for nonlinear wavelet estimators of regression means, in the context of general error distributions and general designs, and addressed the choice of threshold and truncation parameters. Hall and Turlach (1997b) introduced two interpolation methods for wavelet estimators to be applicable to nonparametric regression with stochastic design, or nondyadic regular design. Hall and Nason (1997) addressed the issue of choosing a noninteger resolution level for wavelet methods.

Recognizing the term-by-term thresholding is not optimal, Hall, Kerkyacharian and Picard (1998) proposed and analyzed the block-thresholding rules using kernel and wavelet methods and demonstrated the advantages of these rules. Hall, Kerkyacharian and Picard (1999) showed further that such procedures are indeed minimax optimal for a broad class of functions.

9. Multivariate nonparametric regression. It is well known that nonparametric regression suffers from serious variance inflation in the high-dimensional case, due to data sparseness, and it is common practice to perform dimension reduction or assume some model structure. Hall (1989) introduced the projection pursuit regression model constructed via kernel estimation of the first projective approximation to the regression function. He showed that the projection direction estimator can be estimated at the $\sqrt{n}$ rate using a two-stage estimation scheme which employs two different bandwidths in the two stages. A single-index model is similar to projection pursuit regression but differs from the latter in the sense that the model is exact. Before, it was not clear whether or not it is necessary to use two different bandwidths in the estimation. Härdle, Hall and Ichimura (1993) suggested a simultaneous estimation method for the index parameter and the nonparametric function, and showed that the same bandwidth can be used to achieve $\sqrt{n}$-consistent estimation of the index parameter. The objective function involves both the index vector and the bandwidth. Härdle, Hall and Ichimura (1993) pointed out that the $\sqrt{n}$-consistency is attained because the objective function admits an asymptotic expansion that is decomposed into two terms with one depending only on the index parameter and the other depending only on the bandwidth. Hall et al. (1997) analyzed the design sparseness problem with local linear regression when a high-dimensional design is projected into a lower dimensional space, which is the case in projection pursuit regression and estimation in single index models. The theoretical study led to an adaptive local bandwidth method to deal with the problem.

Miller and Hall (2010) suggested an appealing approach to multivariate nonparametric regression. Instead of making model structure or sparsity assumptions, it performs variable selection locally on the usual multivariate local linear regression. Then, locally, those variables judged to be irrelevant are downweighted by extending the bandwidths in the corresponding directions. This approach allows for locally redundant variables and permits relevant variables to have zero gradient, and it attains a nonparametric oracle property on the entire domain.

10. Applications of nonparametric techniques. Peter Hall was interested not only on the foundation of nonparametric function estimation, but also on its various applications. An example of this is his important contributions to the estimation of fractal dimensionality with applications to understanding how the surface roughness relates to the amount of water that is available for plant growth [Constantine and Hall (1994), Davies and Hall (1999)]. In the paper read before the Royal Statistical Society [Davies and Hall (1999)], Hall demonstrated that the fractal index of line transects of a random field can vary with orientation is very limited: for any three orientations, the two lowest fractal indices must be the same. He then proceeded to provide novel estimate of the fractal index for isotropic case and the antitropic case and obtained their asymptotic distributions. In a similar vein, Hall contributed to understanding of the degree of long range dependence via estimating the Hurst index [Hall, Koul and Turlach (1997), Hall et al. (2000)].

Conditional distribution functions are very important for constructing predictive intervals. Yet, a simple application of the local linear smoother to the indicator function will result in a possibly negative and nonmonotonic estimate of the distribution function. To remedy this problem, in Hall, Wolff and Yao (1999), two new methods for estimating conditional distributions were proposed. The first one is based on locally fitting a logistic model, which always takes on values in the interval [0, 1]. The second method is based on an adjusted form of the Nadaraya–Watson estimator, which results in a distribution function itself and preserves the traditional bias and variance properties. Hall, Racine and Li (2004) addressed the problem of variable selection and bandwidth selection for estimating the conditional density. It is demonstrated surprisingly that cross-validation automatically determines which components are relevant and which are not, through assigning large smoothing parameters to the latter and removing irrelevant components from contention and that the cross-validation smoothes the relevant components by assigning their smoothing parameters of optimal size.

The problems of estimating the endpoint of a univariate distribution and the boundary of a bivariate distribution have many applications. In the former problem, Hall (1982c) suggested to use an increasing number of extreme order statistics to improve on the existing methods that are based on finite number of extremes. Fisher et al. (1997) addressed the problem of estimating the support function of a convex set on the plane by employing local linear smoothing with an extended version of the von Mises density on the circle as the kernel. A closely related problem is frontier estimation in productivity analysis. Production frontier models are important and widely used in econometrics. There are two branches in modeling productions: one is deterministic frontier models and the other is stochastic frontier ones. The former is simpler and easier to work with in the sense that the latter requires some parametric assumptions on the inefficiency and the errors which are confounded with each other. Popular nonparametric approaches to estimation of deterministic frontier curveness include data envelope analysis. Hall, Park and Stern (1998) viewed the data as coming from a Poisson process on the plane, and suggested a local maximum likelihood method to estimate the nonparametric frontier function. The asymptotic analysis by Hall and Van Keilegom (2009) led to a new approach to frontier estimation with multiple covariates by treating the errors as being positioned at the frontier and by modeling their conditional distributions given the covariates from an extreme value viewpoint.

In applications, the data we analyze often have periodic patterns such as seasonal effects and observations on variable stars. Hall, Reimann and Rice (2000) suggested a nonparametric method to estimate a periodic regression function with unknown period. The period length is estimated by minimizing an least squares objective function obtained by kernel regression given the period parameter. Sometimes, the period length is not a fixed constant. Genton and Hall (2007) suggested to approximate it by a linear model or an exponential model and approximate the amplitude function by local polynomials. Then the changing period length and amplitude functions are estimated simultaneously by a kernel method.

REFERENCES

BRAUN, W. J. and HALL, P. (2001). Data sharpening for nonparametric inference subject to constraints. *J. Comput. Graph. Statist.* **10** 786–806. MR1938980

CHEN, X. (2016). Peter Hall's contribution to the bootstrap. *Ann. Statist.* **44** 1821–1836.

CHENG, M.-Y., GASSER, T. and HALL, P. (1999). Nonparametric density estimation under unimodality and monotonicity constraints. *J. Comput. Graph. Statist.* **8** 1–21. MR1705906

CHENG, M.-Y. and HALL, P. (1998). Calibrating the excess mass and dip tests of modality. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 579–589. MR1625938

CHENG, M.-Y. and HALL, P. (1999). Mode testing in difficult cases. *Ann. Statist.* **27** 1294–1315. MR1740110

CHENG, M. Y., HALL, P. and TITTERINGTON, D. M. (1997). On the shrinkage of local linear curve estimators. *Stat. Comput.* **7** 11–17.

CHOI, E. and HALL, P. (1998). On bias reduction in local linear smoothing. *Biometrika* **85** 333–345. MR1649117

CONSTANTINE, A. G. and HALL, P. (1994). Characterizing surface smoothness via estimation of effective fractal dimension. *J. Roy. Statist. Soc. Ser. B* **56** 97–113. MR1257799

COWLING, A. and HALL, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **58** 551–563. MR1394366

DAVIES, S. and HALL, P. (1999). Fractal analysis of surface roughness by using spatial data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 3–37. MR1664088

DELAIGLE, A. (2016). Peter Hall's main contributions to deconvolution. *Ann. Statist.* **44** 1854–1866.

DELAIGLE, A. and HALL, P. (2012). Nonparametric regression with homogeneous group testing data. *Ann. Statist.* **40** 131–158. MR3013182

DELAIGLE, A., HALL, P. and WISHART, J. R. (2014). New approaches to nonparametric and semiparametric regression for univariate and multivariate group testing data. *Biometrika* **101** 567–585. MR3254901

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089

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

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004. MR1209561

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216. MR1212173

FAN, J. and HALL, P. (1994). On curve estimation by minimizing mean absolute deviation and its implications. *Ann. Statist.* **22** 867–885. MR1292544

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

FAN, J. and ZHANG, W. (2008). Statistical methods with varying coefficient models. *Stat. Interface* **1** 179–195. MR2425354

FAN, J., HALL, P., MARTIN, M. A. and PATIL, P. (1996). On local smoothing of nonparametric curve estimators. *J. Amer. Statist. Assoc.* **91** 258–266. MR1394080

FAN, J., HALL, P., MARTIN, M. and PATIL, P. (1999). Adaptation to high spatial inhomogeneity using wavelet methods. *Statist. Sinica* **9** 85–102. MR1678882

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

GASSER, T. and MÜLLER, H. G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.). *Lecture Notes in Math.* **757** 23–68. Springer, Berlin. MR0564251

GENTON, M. G. and HALL, P. (2007). Statistical inference for evolving periodic functions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 643–657. MR2370073

GIJBELS, I., HALL, P. and KNEIP, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51** 231–251. MR1707773

GIJBELS, I., HALL, P., JONES, M. C. and KOCH, I. (2000). Tests for monotonicity of a regression mean with guaranteed level. *Biometrika* **87** 663–673. MR1789816

GRUND, B. and HALL, P. (1995). On the minimisation of L^p error in mode estimation. *Ann. Statist.* **23** 2264–2284. MR1389874

HALL, P. (1980). Estimating a density on the positive half line by the method of orthogonal series. *Ann. Inst. Statist. Math.* **32** 351–362. MR0609028

HALL, P. (1981a). On trigonometric series estimates of densities. *Ann. Statist.* **9** 683–685. MR0615446

HALL, P. (1981b). Laws of the iterated logarithm for nonparametric density estimators. *Z. Wahrsch. Verw. Gebiete* **56** 47–61. MR0612159

HALL, P. (1982a). Cross-validation in density estimation. *Biometrika* **69** 383–390. MR0671976

HALL, P. (1982b). Asymptotic theory of Grenander's mode estimator. *Z. Wahrsch. Verw. Gebiete* **60** 315–334. MR0664420

HALL, P. (1982c). On estimating the endpoint of a distribution. *Ann. Statist.* **10** 556–568. MR0653530

HALL, P. (1983). Large sample optimality of least squares cross-validation in density estimation. *Ann. Statist.* **11** 1156–1174. MR0720261

HALL, P. (1984a). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. MR0734096

HALL, P. (1984b). Integrated square error properties of kernel estimators of regression functions. *Ann. Statist.* **12** 241–260. MR0733511

HALL, P. (1984c). Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function. *Z. Wahrsch. Verw. Gebiete* **67** 175–196. MR0758072

HALL, P. (1987). On Kullback–Leibler loss and density estimation. *Ann. Statist.* **15** 1491–1519. MR0913570

HALL, P. (1989). On projection pursuit regression. *Ann. Statist.* **17** 573–588. MR0994251

HALL, P. (1992a). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694. MR1165587

HALL, P. (1992b). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20** 695–711. MR1165588

HALL, P. (1993). On Edgeworth expansions and bootstrap confidence bands in nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **55** 291–304.

HALL, P. and CARROLL, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *J. Roy. Statist. Soc. Ser. B* **51** 3–14. MR0984989

HALL, P. and HART, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36** 339–351. MR1084984

HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39. MR1762902

HALL, P. and HOROWITZ, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Ann. Statist.* **33** 2904–2929. MR2253107

HALL, P. and HOROWITZ, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Ann. Statist.* **41** 1892–1921. MR3127852

HALL, P. and HUANG, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* **29** 624–647. MR1865334

HALL, P. and HUANG, L.-S. (2002). Unimodal density estimation using kernel methods. *Statist. Sinica* **12** 965–990. MR1947056

HALL, P. and KANG, K.-H. (2005). Unimodal kernel density estimation by data sharpening. *Statist. Sinica* **15** 73–98. MR2125721

HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. MR1087842

HALL, P., KERKYACHARIAN, G. and PICARD, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.* **26** 922–942. MR1635418

HALL, P., KERKYACHARIAN, G. and PICARD, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Statist. Sinica* **9** 33–49. MR1678880

HALL, P. and KOCH, I. (1992). On the feasibility of cross-validation in image analysis. *SIAM J. Appl. Math.* **52** 292–313. MR1148329

HALL, P., KOUL, H. L. and TURLACH, B. A. (1997). Note on convergence rates of semiparametric estimators of dependence index. *Ann. Statist.* **25** 1725–1739. MR1463572

HALL, P., LAHIRI, S. N. and POLZEHL, J. (1995). On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *Ann. Statist.* **23** 1921–1936. MR1389858

HALL, P. and MARRON, J. S. (1987a). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6** 109–115. MR0907270

HALL, P. and MARRON, J. S. (1987b). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probab. Theory Related Fields* **74** 567–581. MR0876256

HALL, P. and MARRON, J. S. (1988). Choice of kernel order in density estimation. *Ann. Statist.* **16** 161–173. MR0924863

HALL, P. and MARRON, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77** 415–419. MR1064818

HALL, P. and MARRON, J. S. (1991). Local minima in cross-validation functions. *J. Roy. Statist. Soc. Ser. B* **53** 245–252. MR1094284

HALL, P. and MARRON, J. S. (1997). On the role of the shrinkage parameter in local linear smoothing. *Probab. Theory Related Fields* **108** 495–516. MR1465639

HALL, P., MARRON, J. S. and PARK, B. U. (1992). Smoothed cross-validation. *Probab. Theory Related Fields* **92** 1–20. MR1156447

HALL, P., MCKAY, I. and TURLACH, B. A. (1996). Performance of wavelet methods for functions with many discontinuities. *Ann. Statist.* **24** 2462–2476. MR1425961

HALL, P., MINNOTTE, M. C. and ZHANG, C. (2004). Bump hunting with non-Gaussian kernels. *Ann. Statist.* **32** 2124–2141. MR2102505

HALL, P. and NASON, G. P. (1997). On choosing a non-integer resolution level when using wavelet methods. *Statist. Probab. Lett.* **34** 5–11. MR1457490

HALL, P. and OPSOMER, J. D. (2005). Theory for penalised spline regression. *Biometrika* **92** 105–118. MR2158613

HALL, P. and PARK, B. U. (2002). New methods for bias correction at endpoints and boundaries. *Ann. Statist.* **30** 1460–1479. MR1936326

HALL, P., PARK, B. U. and STERN, S. E. (1998). On polynomial estimators of frontiers and boundaries. *J. Multivariate Anal.* **66** 71–98. MR1648521

HALL, P. and PATIL, P. (1995a). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.* **23** 905–928. MR1345206

HALL, P. and PATIL, P. (1995b). On wavelet methods for estimating smooth functions. *Bernoulli* **1** 41–58. MR1354455

HALL, P. and PATIL, P. (1996). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by non-linear wavelet methods. *J. Roy. Statist. Soc. Ser. B* **58** 361–377. MR1377838

HALL, P. and PRESNELL, B. (1999a). Intentionally biased bootstrap methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 143–158. MR1664116

HALL, P. and PRESNELL, B. (1999b). Density estimation under constraints. *J. Comput. Graph. Statist.* **8** 259–277. MR1706365

HALL, P. G. and RACINE, J. S. (2015). Infinite order cross-validated local polynomial regression. *J. Econometrics* **185** 510–525. MR3311836

HALL, P., RACINE, J. and LI, Q. (2004). Cross-validation and the estimation of conditional probability densities. *J. Amer. Statist. Assoc.* **99** 1015–1026. MR2109491

HALL, P., REIMANN, J. and RICE, J. (2000). Nonparametric estimation of a periodic function. *Biometrika* **87** 545–557. MR1789808

HALL, P. and TITTERINGTON, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27** 228–254. MR0971184

HALL, P. and TITTERINGTON, D. M. (1992). Edge-preserving and peak-preserving smoothing. *Technometrics* **34** 429–440. MR1190262

HALL, P. and TURLACH, B. A. (1997a). Interpolation methods for adapting to sparse design in nonparametric regression. *J. Amer. Statist. Assoc.* **92** 466–477. MR1467841

HALL, P. and TURLACH, B. A. (1997b). Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Ann. Statist.* **25** 1912–1925. MR1474074

HALL, P. and TURLACH, B. A. (1999). Reducing bias in curve estimation by use of weights. *Comput. Statist. Data Anal.* **30** 67–86. MR1681455

HALL, P. and VAN KEILEGOM, I. (2009). Nonparametric "regression" when errors are positioned at end-points. *Bernoulli* **15** 614–633. MR2555192

HALL, P. and WAND, M. P. (1996). On the accuracy of binned kernel density estimators. *J. Multivariate Anal.* **56** 165–184. MR1379525

HALL, P., WATSON, G. S. and CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–762. MR0919843

HALL, P. and WEHRLY, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *J. Amer. Statist. Assoc.* **86** 665–672. MR1147090

HALL, P., WOLFF, R. C. L. and YAO, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.* **94** 154–163. MR1689221

HALL, P. and YORK, M. (2001). On the calibration of Silverman's test for multimodality. *Statist. Sinica* **11** 515–536. MR1844538

HALL, P., SHEATHER, S. J., JONES, M. C. and MARRON, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **78** 263–269. MR1131158

HALL, P., MARRON, J. S., NEUMANN, M. H. and TITTERINGTON, D. M. (1997). Curve estimation when the design density is low. *Ann. Statist.* **25** 756–770. MR1439322

HALL, P., HÄRDLE, W., KLEINOW, T. and SCHMIDT, P. (2000). Semiparametric bootstrap approach to hypothesis tests and confidence intervals for the Hurst coefficient. *Stat. Inference Stoch. Process.* **3** 263–276 (2001). MR1819399

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

HÄRDLE, W., HALL, P. and MARRON, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.* **83** 86–101. MR0941001

HASTIE, T. and TIBSHIRANI, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55** 757–796. MR1229881

MILLER, H. and HALL, P. (2010). Local polynomial regression and variable selection. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown. Inst. Math. Stat. Collect.* **6** 216–233. IMS, Beachwood, OH. MR2798521

MUELLER, H. G. (2016). Peter Hall, functional data analysis and random objects. *Ann. Statist.* **44** 1867–1887.

RICE, J. (1984). Boundary modification for kernel regression. *Comm. Statist. Theory Methods* **13** 893–900. MR0745507

SAMWORTH, R. J. (2016). Peter Hall's work on high-dimensional data and classification. *Ann. Statist.* **44** 1888–1895.

SCHUSTER, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. Theory Methods* **14** 1123–1136. MR0797636

SEIFERT, B. and GASSER, T. (1996). Finite-sample variance of local polynomials: Analysis and solutions. *J. Amer. Statist. Assoc.* **91** 267–275. MR1394081

WALTER, G. and BLUM, J. (1979). Probability density estimation using delta sequences. *Ann. Statist.* **7** 328–340. MR0520243

DEPARTMENT OF MATHEMATICS
NATIONAL TAIWAN UNIVERSITY
TAIPEI 106
TAIWAN
E-MAIL: cheng@math.ntu.edu.tw

DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: jqfan@princeton.edu