Robustness, Data Analysis, and Statistical Modeling: The First 50 Years and Beyond

Erniel B. Barrios[1,a]

[a]School of Statistics, University of the Philippines Diliman, Philippines

Abstract

We present a survey of contributions that defined the nature and extent of robust statistics for the last 50 years. From the pioneering work of Tukey, Huber, and Hampel that focused on robust location parameter estimation, we presented various generalizations of these estimation procedures that cover a wide variety of models and data analysis methods. Among these extensions, we present linear models, clustered and dependent observations, times series data, binary and discrete data, models for spatial data, nonparametric methods, and forward search methods for outliers. We also present the current interest in robust statistics and conclude with suggestions on the possible future direction of this area for statistical science.

Keywords: clustered data, forward search algorithm, influence curve, L-Estimator, location parameter, M-estimator, quantile regression, R-estimator, robust statistics, spatial analysis, time series data

1. Introduction

A customer with a basket of commodities queue at the counter to pay. A bar code reader is used to scan the profile of the product (color, length, size, volume, and price), similar process follows for all other commodities in the basket, the customer then decide whether to pay using cash, credit card, debit card, cash coupons, or another instrument, the customer then presents a loyalty card (that contains information of the typical market behavior in the past five years) and then a discount card is presented (also contains information related to the activities leading to the awarding of this discount card). The customer then checks out and the system links information from this transaction to other transactions in the historical profile of this customers. The process is repeated for another customer. Compilation of these information leads to a humongous database size: heterogeneous, complex sources, unknown or purposive samples, self-selection bias, and other features.

Are the observations random, identically distributed, normally distributed? Are there enough observations that will warrant the central limit theorem? Are there enough observations relative to the number of variables to ensure positive error degrees of freedom? These questions are asked on the kind of data extracted from huge compilation databases. Data analysis is dependent on the nature of data available, on how well the classical assumptions can hold, of whether the regularity conditions in asymptotic theory fairly hold. What if the data is from a different family of distributions? If extreme values, outliers, small far away clusters of interesting subjects are present? What if these atypical observations can serve as the anchor of innovation, e.g., customers who would exhibit unique

---

behavior ideal for cross selling of specific package of goods and services? Should one trade statistical optimality over computational efficiency, robustness, and interpretability of the resulting information?

This explosion of information could have been the stimulus of the pioneering work of Tukey, Huber, and Hampel more than five decades ago when robust statistics was introduced. Tukey (1962) noted the unrealistically large data assumption of asymptotic theory, e.g., Mann and Wald (1942), which serves as the foundation of statistical inference. At that time when data are generated mostly within some controlled conditions, Tukey identified the growth areas of data analysis (which he claims to be much broader than statistical inference) to include treatment of "outliers", "wild shots", "blunders", and "large deviations". He further provided some guidelines on how the new data analysis can be initiated: seek out wholly new questions to be answered, tackle old problems in more realistic frameworks (with reference to regulatory conditions in asymptotic theory), find unfamiliar summaries of observational material, establish useful properties, and find rather than evade constraints. He further advices on the use of mathematical argument and mathematical results as basis for judgement rather than as bases for proof or stamps of validity.

There has been several definitions of robustness of estimators. Hampel (1971), defined robustness as follows: $\{T_n\}$ a sequence of estimators in the neighborhood of F, is robust in G if the Prokhorov distance between F and G is small enough. $\{T_n\}$ is further robust in a class of densities if and only if $\{T_n\}$ is robust at all members of that class. Correspondingly, Hampel (1971) characterized robust sequences of estimators to be continuous functionals on the space of probability distributions and defined the breakdown point as a measure of the extent up to where robustness of the sequence of estimators extend.

There were many survey papers on robust statistics so far. Huber (1972), summarized the accomplishments on robust estimation of parameters and has substantially covered various methods of robust estimation of location (M-estimators, L-estimators, and R-estimators). The survey paper also included other issues like order statistic estimations of, scale parameters (by taking logarithm to convert it into a location parameter estimation), multivariate location parameters. He also briefly discussed the implications in robust estimation of regression and analysis of variance, and has pointed out robust estimation in stationary time series data. Hampel (1973) also wrote a survey paper on robust estimation, but focused on stimulating discussions, especially in bridging understanding and cooperation between the pure mathematician and data analyst.

A relatively more recent survey paper of Huber (2002) summarizes the contributions of Tukey to robust statistics. There has less published work on the contribution of Tukey in robust statistics since many of his contributions were included only in lectures, technical reports, and informal discussions with colleagues and students. His 1962 paper 'The Future of Data Analysis' of course, summarizes his thoughts on robust statistics. Huber (2002) presented the contributions of Tukey. This can be gleaned from the philosophical issues dominating his work on data analysis, his views are often expressed in a balanced fashion. Huber (2002) noted that Tukey showed that inference in a sample-to-population sense is only part, not the whole, of statistics and data analysis. Tukey also encouraged simulation rather than on rigorous proof, a similar principle that led towards the growth of computational statistics. Huber (2002) concluded that Tukey wants a data analyst to be more of a scientist than a pure mathematician.

Some books covered a wide range of materials on robust statistics. Rieder, (1996) edited a collection of various papers, while Hampel *et al.* (1986) presented robust statistics focusing on influence functions. More recent materials on robust statistics are included in Huber and Ronchetti (2009).

This paper includes: definitions and basic robust estimates of location as presented in the next section, various issues in robust modeling in Section 3, robustness in spatial analysis as discussed

in Section 4, robustness and nonparametric statistics as presented in Section 5, the Forward Search method as presented in Section 6, applications in Section 7, and; recent development and future directions as presented in Section 8.

2. Definition and Basic Estimates of Location

Robustness can be defined either qualitatively or quantitatively. Let $x_i \sim iid\ F$ and suppose $\{T_n\}$ is a sequence of estimators, $T_n = T_n(x_1, x_2, \ldots, x_n)$, T_n is robust at $F = F_0$ if the sequence of maps of distributions $F \to \mathcal{L}_F(T_n)$, the mapping F to the distribution of T_n, is equicontinuous at F_0. This is the qualitative definition of robustness, (Huber and Ronchetti, 2009, p.11). Huber and Ronchetti (2009, p.12) further defined quantitative robustness by quantifying the effect of change in distribution from F to $\mathcal{L}_F$. Assume $T_n = T(F_n)$ is consistent, i.e., $T_n \to T(F)$ in probability, and asymptotically normal, i.e., $\mathcal{L}_F\{\sqrt{n}[T_n - T(F)]\} \to N(0, A(F, T))$. The quantitative large sample robustness of T is defined in terms of the behavior of asymptotic bias $T(F) - T(F_0)$ and asymptotic variance $A(F, T)$ in some neighborhood $\wp_\varepsilon(F_0)$ of the model F_0.

Huber (1972) discussed three basic estimators of location: M-estimators, L-estimators, and R-estimators.

- M-estimators (maximum likelihood-type)

  Let ρ be a real valued function of a real parameter, $\rho' = \psi$, and $T_n = T_n(x_1, \ldots, x_n)$. The M-estimator T_n is a solution of $\sum_{i=1}^n \psi(x_i - T_n) = 0$. Under general conditions, T_n converges to $T(F)$ defined by $\int \psi(x - T(F))F(dx) = 0$. If $\psi_0(x) = -f_0'(x)/f_0(x)$, then T_n is the maximum likelihood estimator of θ for the true underlying distribution F_0, and under some regularity conditions, is asymptotically efficient for $F = F_0$.

- L-Estimator (linear combination)

  Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$, the ordered sample, the L-estimator is the linear combination $T_n = \sum_{i=1}^n a_i X_{(i)}$, where $a_i = \int_{(i-1)/n}^{i/n} J(t)dt$, for J satisfying $\int_0^1 J(t)dt = 1$. Under some regularity conditions, $T_n \to T(F) = \int J(t)F^{-1}(t)dt$. If $J(t) = (1/I(F_0))\psi_0'(F_0^{-1}(t))$, $I(F_0) = \int \psi_0(x)^2 F_0(dx)$, the Fisher information, T_n is asymptotically efficient for F_0.

- R-Estimators (rank-based)

  Consider a two-sample rank test for shift, $Y_1, \ldots, Y_n \sim F(x)$ and $Z_1, \ldots, Z_n \sim F(x - \Delta)$ are two independent samples. Form the combined samples of size $N = 2n$. In testing for $\Delta = 0$ against $\Delta > 0$, take as test statistic $W(Y_1, \ldots, Y_n; Z_1, \ldots, Z_n) = \sum J(i/(N + 1))V_i$ where $V_i = 1$ if the ith smallest entry in the combined sample is a Y, and $V_i = 0$, otherwise. Determine $T_n(x_1, \ldots, x_n)$ such that $W(X_1 - T_n, \ldots, X_n - T_n; -(X_1 - T_n), \ldots, -(X_n - T_n)) = 0$. The asymptotic behavior of T_n can be determined from the asymptotic power of the rank test. The R-estimator T_n is the solution $T(F)$ of $\int J((F(x) + 1 - F(2T(F) - x))/2)F(dx) = 0$.

The extent of influence of aberrant observations to the flexibility of a statistic is measured through the Influence Curve that account for a scaled differential influence of one additional observation x if $n \to \infty$. Suppose $T \sim F$, Influence Curve of T is defined pointwise by $IC_{T,F}(x) = \lim_{\epsilon \to 0}\{T[(1 - \epsilon)F + \epsilon\delta_x] - T(F)\}/\epsilon$ if this limit is defined for every point x, Hampel (1974).

Example 1.

1. Mean: $T = \int x dF(x)$. Suppose $E(F) = \mu$, then $IC_{T,F}(x) = \lim_{\epsilon \to 0}[(1 - \epsilon)\mu + \epsilon x - \mu]/\epsilon = x - \mu$. An additional observation influences the mean depending on how far this observation is from the true location.

2. Variance: $V = \int (x-\mu)^2 dF(x)$. If $V(F) = \sigma^2$, then $IC_{T,F}(x) = \lim_{\epsilon \to 0}[(1-\epsilon)\sigma^2 + \epsilon(x-\mu)^2 - \sigma^2]/\epsilon = (x-\mu)^2 - \sigma^2$. The influence of an observation to the variance is quantified from the squared deviation on this observation from the true location relative to the true variance of the distribution.

3. M-estimator: $IC_{T,F}(x) = \psi(x; T(F))/(- \int \psi'(x; T(F))F(dx)')$, the influence of an observation is proportional to ψ. Robustness of M-estimator is achieved through the choice of ψ.

There are many themes on various contributions in robust statistics. Some focused on outlier detection (single and multiple outliers). Once the outlier is detected, it is of interest to measure its influence to the statistic or the method in general. Upon confirmation of its influence, the challenge is on how to mitigate the influence of these observations to statistics, models, or methods.

Initially, robust methods focused on the estimation of location parameters with Huber (1964) providing the early illustration of robust estimation of a location parameter. He developed a new approach of estimation from a contaminated normal distribution using the asymptotic theory of estimators. The estimator is a hybrid between the mean and the median and is asymptotically robust among all translation invariant estimators. Sacks and Ylvisaker (1972) showed that Huber estimator works for a more general class of symmetric distributions.

Many robust methods were subsequently introduced on various themes of data analysis that include variance estimation, hypothesis testing, outlier detection vs robust estimation of linear models, generalized linear models, time series models, and more recently, high dimensional data. Another new approach to robustness that is more data dependent called forward search algorithm was introduced in Atkinson (1994).

3. Robust Modeling

Given n observations, we want to fit the model $y_i = x_i'\beta + \epsilon_i$ or equivalently, $Y = X\beta + \epsilon$, $\epsilon \sim N(0, I\sigma^2)$. Ordinary least squares (OLS) estimation consider the solution to the optimization problem: $\min_\beta \|Y - X'\beta\|^2$ which is $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$. OLS generates unbiased estimator, i.e., $E(\hat{\beta}_{OLS}) = \beta$ and $V(\hat{\beta}_{OLS}) = (X'X)^{-1}\sigma^2$. Furthermore, the Gauss-Markov Theorem characterizes $\hat{\beta}_{OLS}$ as the best linear unbiased estimate (BLUE) of β.

The estimate of $E(y|x)$ is given by $\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y$. Define the hat matrix $H = X(X'X)^{-1}X'$ $= [(h_i j)]$, $\hat{y}_i = h_{i1}y_1 + \cdots + h_{in}y_n = h_{ii}y_i + \sum_{j \neq i} h_{ij}y_j \Rightarrow h_{ii}$ is the leverage of the ith observation to its own fit. The studentized residuals are standardization of the residuals based on the hat matrix, defined as $r_i^* = r_i/\sqrt{MSE(1 - h_{ii})}$ where r_i is the raw residual for the ith observation. This measures the influence of outlying observations in modeling. The Cook's Distance $D_i = ((\hat{\beta} - \hat{\beta}_i)'X'X(\hat{\beta} - \hat{\beta}_i))/(pMSE)$, specifically measures the extent of influence of observation i on the estimates of the regression coefficients. Aberrant observations in the design matrix **X** and well as extreme values of the response y easily influence the model estimated through OLS.

Huber (1973) generalized the estimator for location into the parameters of a regression model. The M-estimator of the regression coefficients, $\hat{\beta}_M$ is a solution to the optimization problem of $\min_\beta \sum_{i=1}^{n} \rho((y_i - x_i'\beta)/\sigma)$, for a real function ρ of regression coefficients and a fixed value of the scale parameter σ. In many cases, σ is estimated along with β. Let $F = F_\beta$, $\psi(x, \beta) = \psi_c(y - x'\beta)$,

then $IC_{\hat{\beta}_M, F_\beta}(x) = \psi_c(y - x'\beta)M^{-1}x$, $M = (E\psi'_c)(Exx')$. The influence curve for the M-estimator (also called Huber estimator) of the regression coefficients ($\hat{\beta}_M$) depends on y only through the residuals $r = y - x'\beta$. OLS is a special case where $\rho(r) = r^2/2$. The quantile regression which is a solution to $\min_\beta \|Y - X'\beta\|$ assumes $\rho(r) = |r|$.

3.1. Linear models

The contributions on various aspects of robust estimation of a linear model vary in terms of peculiar behavior of a model, variable selection, hypothesis testing, and hybrid procedures. We now present these contributions on robust linear modeling in this section.

A linear model with variances parameterized with the mean and another parameter θ (ancillary) was analyzed by Carroll and Ruppert (1982) where robust estimate of the regression coefficients was proposed. Assuming that the initial values of θ are available, the robust estimate of β are asymptotically equivalent to the estimate derived with known variances. Parametrization of the variance will not have remarkable impact on robust estimation of the regression coefficients provided that the sample size is fairly large.

The advantages of OLS cannot be discounted, but the impact of aberrant observations also needs to be mitigated. Using a connected double truncated gamma distribution for the error distribution of a linear model, Nassiri and Loris (2013) estimated a generalized quantile regression, the resulting estimates inherits the advantages of both the OLS (differentiable loss function) and quantile regression (robustness to outliers). This illustrates the potential gains in methods resulting from a hybrid of two or more methods which can benefit from the advantages carried by each of the individual methods.

A linear model assumes that the covariates are pre-determined. If instead, measurement errors are present in the covariate data, OLS is known to be inconsistent, even some robust estimates could suffer, e.g., bias. Wei and Carroll (2009) proposed to correct problems with measurement error in the covariate in a quantile regression model though the EM-type estimation algorithm which jointly estimates the equations for all quantile levels.

Variable selection is another big topic in modeling, especially for linear models. Using the least absolute shrinkage and selection operator (LASSO), Li and Zhu (2008) illustrates the advantage of simultaneously controlling the variance of the fitted coefficients and performing automatic variable selection. An efficient algorithm that computes the entire solution path was provided, along with a method of selecting the regularization (penalty) parameter. Alhamzawi (2015) proposed a variable selection method in quantile regression, also based on the LASSO penalty. Simulated and actual data sets provide empirical evidence on the advantage of the method compared with other procedures.

Furno (2004) developed a test for conditional heteroskedasticity in quantile regression, where distributional assumption on the error term nor the pattern that characterizes heteroskedasticity need not be specified. A lack-of-fit test for quantile regression (linear or nonlinear) was developed by He and Zhu (2003). The test does not involve nonparametric smoothing; however, it is consistent for all nonparametric alternatives without any moment conditions on regression error.

There are many other stylized facts about the linear model that are incorporated into the robust estimation procedures. Many other tests are also proposed with themes that revolve around nonparametric approaches and asymptotic methods.

3.2. Clustered and dependent observations

Clustering and dependence (cross-sectional) of observations often occur interchangeably, leading to extreme values among few members of a cluster, members in some clusters, or across all observations

in a few clusters. This typically occurs during epidemics where the response variable is a count of the number of infected individuals.

Given independent but not identically distributed data, e.g., clustered observations, Beran (1982) developed an asymptotically robust estimator within a reasonable choice of contamination neighborhood, defined as the region where the parametric model does not hold true. Moscone and Tosetti (2015) proposed an estimator of the covariance matrix of the fixed effects and mean of the group estimator in a panel data that are cross-sectionally dependent and serially correlated, possibly occurring as a result of clustering.

Heteroskedastic linear models are postulated as a result of clustering at the least, and general dependence structure in the worst case. In testing hypothesis pertaining to heteroskedastic linear models, Zhao and Wang (2009) developed three classes of testing procedures (Wald-type, score-type, and drop-in dispersion tests). For a robustness criterion, the maximum asymptotic bias of the level of the test for distributions in a shrinking contamination neighborhood is used and the most-efficient robust test is derived.

Supposed that data are naturally distributed into groups and that a fixed group effect regression is postulated, Pérez *et al.* (2014) compared three methods in terms of effectiveness in outlier detection and robustness via simulation studies in a wide array of contamination settings. The modified principal sensitivity component (PSC) procedure can detect true outliers and a small number of false outliers. The method is appropriate when contamination is in the error term or in the covariates that also detecte possibly masked high leverage points.

A semiparametric approach using empirical likelihood was used by Kim and Yang (2011) in estimating a random effects quantile regression of clustered data. Assuming independence of the random regression coefficients with a common mean (corresponds to the population-average effects of explanatory variables on the conditional quantile) and the random coefficients that represent cluster-specific deviations in the covariate effects, the estimator of the population-level parameter is asymptotically normal, and that the estimators of the random coefficients are shrunk toward the population-level parameter in the first-order asymptotic sense.

He *et al.* (2005) proposed a robust GEE (bounded scores and leverage-based weights to achieve robustness) in semiparametric generalized partial linear model for longitudinal or clustered data. The model was first approximated with regression splines, robust estimation and inference implemented operationally as in a general linear model.

In lieu of robust methods, Field *et al.* (2010) explored bootstrap methods since data contamination often increases variability of the data, and robust estimates of variance are often smaller than nonrobust counterparts. Simulation studies indicate that transformation and generalized cluster bootstrap procedures perform well and are asymptotically valid under reasonable data contamination conditions.

3.3. Time series data

The nature of autocorrelation structure causes the vulnerability of time series models from outliers. In autoregressive models for instance, Campano and Barrios (2011) noted that parameter estimates using conditional least squares are affected by outliers occurring among more recent observations.

There are some recommendations though on how to robustify estimation in time series models. For example, de Luna and Genton (2001) proposed a robust estimation method of an ARMA model. The proposal is to use the observed time series in estimating the model. Data is then simulated from an ARMA model with assumed parameters chosen so that observation-based and simulation-based auxiliary parameters are comparable. The influence of outlying observations on the parameters are properly mitigated and achieve robustness since the simulation-based estimation procedure inherits

the properties of the auxiliary model estimator.

Chang *et al.* (2013) alternatively proposed a robust nonlinear filtering algorithm to deal with contaminated Gaussian noise based on a robust modification of the derivative-free Kalman filter. The algorithm achieves the robustness of the M-estimation as well as the accuracy and flexibility of the derivative-free Kalman filter for the nonlinear problems. Furthermore, Ursu and Pereau (2014) developed a robust estimation and identification method for periodic vector autoregressive models with linear constraints set on parameters for a given season.

While issues on the influence of outliers in estimation need to be addressed, similar concerns with outliers are as important in hypothesis testing. Xiao (2012) presented a robust inference strategy in unit root and cointegration models, e.g., testing for cointegration which is built on the residuals of models whose parameters are estimated through M-estimators. The test is fairly robust even if the normality assumption is not valid.

3.4. Binary and discrete data

In binary or discrete response models, the likelihood function is often at risk of separation when there is unbalanced distribution of data into the different values of the response or there are extreme values. Methods that would mitigate the impact of the nature of binary/discrete data are available through sample selection (see for example Santos and Barrios (2015)) or through robust estimation, some of which are presented next.

Given continuous and binary regressors, Hubert and Rousseeuw (1997) provide a method of downweighting leverage points by computing robust distances in the space of the continuous regressors. They compute weighted least absolute deviations (LAD) as a function of both the continuous and binary regressors with derived robust estimates of the error scale.

Heteroskedasticity and contaminated data (outliers) can easily influence the resulting estimates since MLE is based on the quadratic norm. Čížek (2008) developed a robust estimation method for binary-choice regression through a maximum symmetrically trimmed likelihood estimator (MSTLE) and design a parameter-free adaptive procedure to choose the amount of trimming. The adaptive MSTLE was noted to preserve the robust properties of the original MSTLE, but significantly improves the finite-sample behavior of MSTLE, and also ensures the asymptotic equivalence of the MSTLE and maximum likelihood estimator if there is no contamination.

Using the quasi-likelihood for a generalized linear model, Cantoni and Ronchetti (2001) defined a robust deviance measure that is used in model selection in the context of classical modeling framework. We investigated asymptotic distributions of robust deviance-based tests were derived and stability of the tests under contamination.

4. Spatial Analysis

There are many themes on robust statistical analysis besides location estimation, model estimation, or hypothesis testing. For example, models that account for spatial interdependencies of adjacent units are estimated in spatial analysis. The variogram has been used to quantify such dependencies and has been a crucial input to estimate spatial models. The estimation of variogram has always been a challenge because there are limited number of data points since measurements are too costly specially in applications in mining, geology, and other geosciences. When outliers are present in already a very small dataset, estimation of spatial models becomes extremely challenging. With a fourth root transformation applied in the data, Cressie and Hawkins (1980) assumed a normal-like central region and possibly, heavy-tails in the periphery, in the construction of a stable and robust estimate of the

variogram.

A spatial autoregressive (SAR) models that is a special parametrization of the spatial model generally yield inconsistent estimates when the error term exhibits unknown form of heteroskedasticity. Doğan and Taşpinar (2014) developed robust estimators based on the generalized method of moments (GMM) and Bayesian Markov Chain Monte Carlo (MCMC) frameworks. Monte Carlo simulation indicates greater bias on the spatial autoregressive parameter when there is a negative spatial dependence in the model; however, finite sample efficiency of MCMC is better than the robust GMM estimator.

5. Robustness and Nonparametric Methods

The natural linkage of robust and nonparametric methods are clearly established in the literature. In fact, Hettmansperger and McKean (1988) summarized all these methods into a book on robust nonparametric statistical methods.

Combining kernel methods for density estimation and robust estimation of location, Hardle (1984) developed estimators that are weak and strongly consistent and asymptotically normal under mild condition on the kernel sequence. The estimator is also minimax robust in the sense of Huber (1964).

Reviewing various robust nonparametric methods, Hettmansperger *et al.* (2000) noted that R-estimates are y-robust but not x-robust (factor space). They further assured that this is not a problem provided that x's do not contain measurement errors, but in case there are measurement errors in the x's, they suggest to use bounded influence procedures which lead to robust but less efficient estimates.

Truncated and censored regression estimated via MLE are naturally sensitive to data contamination like outliers since MLE is based on quadratic norm. Thus, Čížek (2012) proposed a semiparametric general trimmed estimator (GTE) which is imprecise but robust. This is further improved through data-adaptive and one-step trimmed estimators like one-step symmetrically censored least squares (SCLS). This improvement is as robust as GTE and asymptotically equivalent to the original estimator (e.g., SCLS).

Hastie and Tibshirani introduced the generalized additive models; consequently, many variants have been introduced such as the combination of parametric and nonparametric components into a semiparametric model, see for example, Hastie and Tibshirani (1990). Hoshino (2014) introduced a partially linear additive quantile regression model where conditional quantile function comprises a linear parametric component and a nonparametric additive component and proposes a two-step estimation approach. The first step approximates the conditional quantile function using a series estimation method. The second step estimates the nonparametric additive component using either a local polynomial estimator or a weighted Nadaraya-Watson estimator. Consistency and asymptotic normality were established.

Wong *et al* (2014) used M-type estimators to estimate generalized additive models in the presence of aberrant observations. Asymptotic properties served as the motivation in decomposing the overall M-type estimation problem into a sequence of conventional additive model fitting. This yield a computational algorithm that is fast and stable with an automatic smoothing parameter selection, resulting in estimates that are outlier-resistant.

6. Forward Search Methods

From a different framework, forward search methods diagnose the presence of outliers through a highly data-dependent algorithm. The nature of existing data dictates how the search algorithm will progress. The algorithm starts with the selection of a 'clean' set of observations, a subset of the original dataset. A criterion is used to ensure that the initial dataset will not contain observations that

may influence the estimate of the parameters or the method of analysis to be conducted. The initial data and the corresponding estimates or models then serve as a benchmark from where the decision on whether to include the next observation into the dataset will be based on. An observation outside the 'clean' set is included *f* it is very 'close' to the 'clean' set and does not exert influence on the estimates or model that is estimated. Observations are added one-at-a-time until the next observation to be added exerts influence onto the estimates or the model. Then it is said that the algorithm will converge.

There are some issues associated with the implementation of the forward search, e.g., selection of the initial dataset, criteria to be used in deciding whether an observation is allowed to enter the dataset, and the stopping rule, i.e., how to measure the influence of observations that are included into the dataset.

Prior to Atkinson (1994) who formally introduced the forward search, similar data-dependent robust methods were introduced where the 'clean' dataset are identified via random searches. Atkinson (1994) considered a few repeats of a simple forward search starting from a random starting point in order to reveal masked multiple outliers. The method generates observations that sufficiently produced robust parameter estimates. Atkinson (1994) also suggested that parallel computing provides appreciable reduction in computational time for large datasets. Since then, various contributions that modifies the forward search to address specific estimation or modeling problems has been provided in the literature.

Atkinson and Cheng (2000) extended a very robust regression method for missing data to data with several outliers, i.e., outliers are imputed as if they were missing. The method that uses a hybrid of the forward search and the EM algorithm were able to generate robust estimates in the presence of outliers.

The forward search was also used by Bertaccini and Varriale (2007) to detect atypical observations and the analysis of their effect in the ANOVA framework. In clustering multivariate normal data, Atkinson and Riani (2007a) introduced a forward search method based on the Mahalanobis distances of the observations. A new forward plot was introduced, this provided an efficient monitoring device to keep to tract robustness of the resulting clusters.

Atkinson and Riani (2007b) also introduced forward search in variable selection in regression modeling. The forward search is imbedded into the backward elimination procedure (variable selection) while monitoring the influence of observations that might exhibit aberrant behavior that is localized to a particular variable.

Highlighting the importance of forward plots in visualizing the inferential contribution of each observation, Atkinson (2009) introduced the forward search in the robust analysis of econometric data. Mavridis and Moustaki (2009) implement a forward search algorithm for identifying atypical subjects/observations in factor analysis models for binary data.

There were many work that either uses the forward search algorithm singly, or in tandem with another method to estimate various models. Riani (2004) modified the forward search in the analysis of structural time series data. The forward plots provided a good monitoring tool showing how the forward search, can detect the main underlying features of the time series with masked multiple outliers, level shifts, or transitory changes.

Time series and spatio-temporal models are often characterized by temporary structural change that could occur due to some confounding events affecting the random shock to behave wildly. While these are temporary, it can potentially influence the estimates of the models that might fail to keep track of the behavior of the data generating process under 'normal' circumstances. Campano and Barrios (2011) proposed an estimation procedure based on a hybrid of block bootstrap and a modified

forward search algorithm. The procedure yields robust estimates of a time series model with data that exhibits temporary structural change provided that the time series is fairly long. However, Bastero and Barrios (2011) postulated and estimated a spatio-temporal model and estimated it using a procedure that infuses the forward search algorithm and maximum likelihood estimation into the backfitting framework. The forward search algorithm filters the effect of temporary structural change in the estimation of covariate and spatial parameters

7. Some Applications

While many results in robust statistics were generated in the course of trying to solve specific real-life problems, we have discussed some application of robust statistics in various areas in this section.

To measure effective dose (ED), estimators based on parametric models (e.g., probit models) are generally efficient but are not robust to model misspecification. While nonparametric estimators are robust to model misspecification, they are also known to be less efficient. Karunamuni *et al.* (2015) proposed a semiparametric method in estimating ED. The proposed estimators are efficient under the model and have exhibited some robustness properties. Kelly and Lindsey (2002) used M-type estimators in estimating the median lethal dose in biological assays.

Dang *et al.* (2015) estimated a dynamic panel model given financial data. They illustrated robustness of bootstrap to unobserved heterogeneity and serial correlation among the residuals.

Stimulated by irregular return patterns (e.g., volatility) in financial assets, Huang (2012) proposed to use a uniformly spaces series of regression quantiles in volatility forecasting. Six of the seven stock indices with 16 years daily return data yield better volatility forecasts compared to existing methods. However, Gaglianone *et al.* (2011) evaluated Value-at-Risk estimates using quantile regression.

In remote sensing data, Cao and Wang (2015) developed a probabilistic robust learning algorithm for neural networks with random weights (NNRWs) to improve the modeling performance.

8. Recent Developments and Future Directions

Kitromilidou and Fokianos (2015) is interested in robust estimation of a log-linear Poisson model for count time series analysis. MLEs were robustified though the inclusion of intervention terms into the model that was then estimated with bounded-influence estimator.

In image interpolation, Huang and Siu (2015) proposed a robust and precise k-nearest neighbors (k-NN) searching scheme to form an accurate AR model of the local statistic. They make use of both low-resolution (LR) and high-resolution (HR) information obtained from large amounts of training data to form a coherent soft-decision estimation of both AR parameters and high-resolution pixels.

There are several contributions on robustness in the context of high dimensional data (HDD). In high breakdown point estimates of the multivariate location and scatter, computing time increase rapidly with the number of variables, this is impractical for high dimensional data, Maronna and Zamar (2002). They proposed estimator based on modified Gnanadesikan-Kenttenring robust covariance estimate that is much faster especially for large dimension. Wang *et al.* (2010) developed a methodology for high-dimensional pattern regression on medical images via machine learning techniques. Vretos *et al.* (2013) proposed a novel Support Vector Machine (SVM) variant, which makes use of robust statistics. Shahriari and Ahmadi (2015) used S-estimator robust clustering of high dimensional data. Lv *et al.* (2014) simultaneously selected variables and estimated nonlinear models based on modal regression (MR), when the number of coefficients diverges with sample size.

Where is robust statistics going to? As Prof. Huber himself noted during the 58th World Congress of the International Statistical Institute, more interest in robust statistics will be geared towards com-

putational statistics in support of Big Data analysis. As the nature, size, and context of data that is accumulated in various databases becomes more uncertain, unknown to the data scientist, tools in computational statistics becomes handy and in lieu of statistical optimality, model robustness is used as justification for various methods.

There will be more heterogeneity, more complicated data structure, space-time dependency as the size of the data becomes really large, aberrant observations may no longer be rare but non-ignorable in size. The role of additive models and bootstrap methods in better understanding the data generating process is indispensable. Ironically as the data becomes very large, analysis will be gradually less dependent on asymptotic theory and methods will be more dependent on empirical evidence instead.

References

Alhamzawi, R. (2015). Model selection in quantile regression models, *Journal of Applied Statistics*, **42**, 445–458.

Atkinson, A. C. (1994). Fast very robust methods for the detection of multiple outliers, *Journal of the American Statistical Association*, **89**, 1329–1339.

Atkinson, A. C. (2009). Econometric applications of the forward search in regression: Robustness, diagnostics, and graphics, *Econometric Reviews*, **28**, 21–39.

Atkinson, A. C. and Cheng, T. C. (2000). On robust linear regression with incomplete data, *Computational Statistics & Data Analysis*, **33**, 361–380.

Atkinson, A. C. and Riani, M. (2007a). Exploratory tools for clustering multivariate data, *Computational Statistics & Data Analysis*, **52**, 272–285.

Atkinson, A. C. and Riani, M. (2007b). Building regression models with the forward search, *Journal of Computing and Information Technology*, **15**, 287–294.

Bastero, R. F. and Barrios, E. B. (2011). Robust estimation of a spatiotemporal model with structural change, *Communications in Statistics-Simulation and Computation*, **40**, 448–468.

Beran, R. (1982). Robust estimation in models for independent non-identically distributed data, *The Annals of Statistics*, **10**, 415–428.

Bertaccini, B. and Varriale, R. (2007). Robust analysis of variance: An approach based on the forward search, *Computational Statistics & Data Analysis*, **51**, 5172–5183.

Campano, W. Q. and Barrios, E. B. (2011). Robust estimation of a time series model with structural change, *Journal of Statistical Computation and Simulation*, **81**, 909–927.

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models, *Journal of the American Statistical Association*, **96**, 1022–1030.

Cao, F., Ye, H. and Wang, D. (2015). A probabilistic learning algorithm for robust modeling using neural networks with random weights, information sciences, **313**, 62–78.

Carroll, R. J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models, *The Annals of Statistics*, **10**, 429–441.

Chang, L., Hu, B., Chang, G. and Li, A. (2013). Robust derivative-free Kalman filter based on Huber's M-estimation, *Journal of Process Control*, **23**, 1555–1561.

Čížek, P. (2008). Robust and efficient adaptive estimation of binary-choice regression models, *Journal of the American Statistical Association*, **103**, 687–696.

Čížek, P. (2012). Semiparametric robust estimation of truncated and censored regression models, *Journal of Econometrics*, **168**, 347–366.

Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram: I, *Mathematical Geology*, **12**, 115–125.

Dang, V. A., Kim, M. and Shin, Y. (2015). In search of robust methods for dynamic panel data models in empirical corporate finance, *Journal of Banking & Finance*, **53**, 84–98.

de Luna, X. and Genton, M. G. (2001). Robust simulation-based estimation of ARMA models, *Journal of Computational and Graphical Statistics*, **10**, 370–387.

Doğăn, O. and Taşpinar, S. (2014). Spatial autoregressive models with unknown heteroscedasticity: A comparison of Bayesian and robust GMM approach, *Regional Science and Urban Economics*, **45**, 1–21.

Field, C. A., Pang, Z. and Welsh, A. H. (2010). Bootstrapping robust estimates for clustered data, *Journal of the American Statistical Association*, **105**, 1606–1616.

Furno, M. (2004). ARCH tests and quantile regressions, *Journal of Statistical Computation and Simulation*, **74**, 277–292.

Gaglianone, W. P., Lima, L. R., Linton, O. and Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression, *Journal of Business & Economic Statistics*, **29**, 150–160.

Hampel, F. R. (1971). A general qualitative definition of robustness, *The Annals of Mathematical Statistics*, **42**, 1887–1896.

Hampel, F. R. (1973). Robust estimation: A condensed partial survey, *Zeitschrift fur Wahrschein-lichkeitstheorie und Verwandte Gebiete*, **27**, 87–104.

Hampel, F. R. (1974). The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, **69**, 383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, New York.

Härdle, W. (1984). Robust regression function estimation, *Journal of Multivariate Analysis*, **14**, 169–180.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman and Hall, London.

He, X. and Zhu, L. X. (2003). A lack-of-fit test for quantile regression, *Journal of the American Statistical Association*, **98**, 1013–1022.

He, X., Fung, W. Z. and Zhu, Z. (2005). Robust estimation in generalized partial linear models for clustered data, *Journal of the American Statistical Association*, **100**, 1176–1184.

Hettmansperger, T. P. and McKean, J. W. (1988). *Robust Nonparametric Statistical Methods*, Arnold, London.

Hettmansperger, T. P., McKean, J. W., and Sheather, S. J. (2000). Robust nonparametric methods, *Journal of the American Statistical Association*, **95**, 1308–1312.

Hoshino, T. (2014). Quantile regression estimation of partially linear additive models, *Journal of Nonparametric Statistics*, **26**, 509–536.

Huang, A. Y. H. (2012). Volatility forecasting by quantile regression, *Applied Economics*, **44**, 423–433.

Huber, P. J. (1964). Robust estimation of a location parameter, *The Annals of Mathematical Statistics*, **35**, 73–101.

Huber, P. J. (1972). The 1972 wald lecture robust statistics: A review, *The Annals of Mathematical Statistics*, **43**, 1041–1067.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*, **1**, 799–821.

Huber, P. J. (2002). John W. Tukey's contributions to robust statistics, *The Annals of Statistics*, **30**, 1640–1648.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*, 2nd ed., John Wiley and Sons, New York.

Hubert, M. and Rousseeuw, P. J. (1997). Robust regression with both continuous and binary regres-

sors, *Journal of Statistical Planning and Inference*, **57**, 153–163.

Hung, K. W. and Siu, W. C. (2015). Learning-based image interpolation via robust k-NN searching for coherent AR parameters estimation, *Journal of Visual Communication Image Representation*, **31**, 305–311.

Karunamuni, R. J., Tang, Q. and Zhao, B. (2015). Robust and efficient estimation of effective dose, *Computational Statistics & Data Analysis*, **90**, 47–60.

Kelly, G. E. and Lindsey, J. K. (2002). Robust estimation of the median lethal dose, *Journal of Biopharmaceutical Statistics*, **12**, 137–147.

Kitromilidou, S. and Fokianos, K. (2015). Robust estimation methods for a class of log-linear count time series models, *Journal of Statistical Computation and Simulation*, DOI: 10.1080/00949655. 2015.1035271.

Kim, M. O. and Yang, Y. (2011). Semiparametric approach to a random effects quantile regression, *Journal of the American Statistical Association*, **106**, 1405–1417.

Li, Y. and Zhu, J. (2008). L1-norm quantile regression, *Journal of Computational and Graphical Statistics*, **17**, 163–185.

Lv, Z., Zhu, H. and Yu, K. (2014). Robust variable selection for nonlinear models with diverging number of parameters, *Statistics & Probability Letters*, **91**, 90–97.

Mann, H. B. and Wald, A. (1942). On the choice of the number of class intervals in the application of the chi square test, *The Annals of Mathematical Statistics*, **13**, 306–317.

Maronna, R. A. and Zamar, R. H. (2002). Robust estimates of location and dispersion for high dimensional datasets, *Technometrics*, **44**, 307–317.

Mavridis, D. and Moustaki, I. (2009). The forward search algorithm for detecting response patterns in factor analysis for binary data, *Journal of Computational and Graphical Statistics*, **18**, 1016–1034.

Moscone, F. and Tosetti, E. (2015). Robust estimation under error cross section dependence, *Economics Letters*, **133**, 100–104.

Nassiri, V. and Loris, I. (2013). A generalized quantile regression model, *Journal of Applied Statistics*, **40**, 1090–1105.

Pérez, B., Molina, I. and Peña, D. (2014). Outlier detection and robust estimation in linear regression models with fixed group effects, *Journal of Statistical Computation and Simulation*, **84**, 2652–2669.

Riani, M. (2004). Extensions of the forward search to time series, *Studies in Nonlinear Dynamics & Econometrics*, **8**, Article 2.

Rieder, H. (1996). *Robust Statistics, Data Analysis, and Computer Intensive Methods*, Springer-Verlag, New York.

Sacks, J. and Ylvisaker, D. (1972). A note of Huber's robust estimation of a location parameter, *The Annals of Mathematical Statistics*, **43**, 1068–1075.

Santos, K. C. P. and Barrios, E. B. (2015). Improving predictive accuracy of logistic regression model using ranked set samples, *Communications in Statistics-Simulation and Computation*, DOI: 10.1080/03610918.2014.955113.

Shahriari, H. and Ahmadi, O. (2015). Robust estimation of the mean vector for high-dimensional data set using robust clustering, *Journal of Applied Statistics*, **42**, 1183–1205.

Tukey, J. W. (1962). The future of data analysis, *The Annals of Mathematical Statistics*, **33**, 1–67.

Ursu, E. and Pereau, J. C. (2014). Robust modelling of periodic vector autoregressive time series, *Journal of Statistical Planning and Inference*, **155**, 93–106.

Vretos, N., Tefas, A. and Pitas, I. (2013). Using robust dispersion estimation in support vector ma-

chines, *Pattern Recognition*, **46**, 3441–3451.

Wang, Y., Fan, Y., Bhatt, P. and Davatzikos, C. (2010). High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables, *Neuroimage*, **50**, 1519–1535.

Wei, Y. and Carroll, R. J. (2009). Quantile regression with measurement error, *Journal of American Statistical Association*, **104**, 1129–1143.

Wong, R. K. W., Yao, F. and Lee, T. C. M. (2014). Robust estimation for generalized additive models, *Journal of Computational and Graphical Statistics*, **23**, 270–289.

Xiao, Z. (2012). Robust inference in nonstationary time series models, *Journal of Econometrics*, **169**, 211–223.

Zhao, J. and Wang, J. (2009). Robust testing procedures in heteroscedastic linear models, *Communications in Statistics-Simulation and Computation*, **38**, 244–256.