Identifying Best-Practices: A Monte Carlo Evaluation of Quantile Regression

Jiannan Wu, Xi'an Jiaotong University

Stuart Bretschneider, Syracuse University

In the field of public management, the concept of best practice implies a sorting out of people, organizations, techniques, etc. in order to find extremely "good" cases. The idea of a best practice has a long established and recognized history in management research dating back to classical works by Frederick Taylor and Elton Mayo to more recent studies by Peters and Waterman and by Osborne and Gaebler (Katorobo, 1998; Overman and Boyd, 1994). Public management interest in best practices has been visible at all levels of government. The State of New York has established a standing legislative committee on Best Practices, while The National Governorís Association, the General Services Administration, and numerous other groups have established awards for best practices in public organizations (Rocheleau, 2000). Effort to measure the performance of public organizations through the regular collection and reporting of information about efficiency, quality, and effectiveness, is arguably the number one research topic in government (Nyhan and Martin, 1999).

Under the combined influence of GPRA,[1] the NPR, state and community benchmarking efforts, and the GASBís SEA reporting, most federal, state, and local government agencies, as well as those private and nonprofit organizations delivering government programs under grants and contracts, are likely to become involved in performance measurement before the end of the decade (Nyhan, Ronald C. and Lawrence L. Martin, 1999). Once governments begin routinely collecting and reporting performance measurement data, policymakers and policy evaluators will be faced with the task of identifying best-practice providers. How, then, can governments go about making comparisons

Abstract: The area of best practice research has only recently begun to embrace statistically based comparisons as a basis for identifying recommended practices. In part motivated by growing interest in performance measurement activities these new approaches hold significant potential for improving out ability to identify and utilize true best practices. Unfortunately, little has been done to study how to apply statistical methods to this task appropriately. In this paper a Monte Carlo evaluation is developed to demonstrate that how Quantile Regression methods can be used to identify best practice. After a brief literature review and a summary of the Quantile Regression technique, the paper develops a specific monte carlo simulation design based on statistical situations with varying numbers of high, medium and low performing organizations. Next, we apply quantile regression to the simulated data and attempts to develop some reasonable guidance about how to apply quantile regression to real world data. The results demonstrate that quantile regression can accurately estimate different models for different types of organizations (e.g. high and low performing) and should be considered as an effective tool for the empirical study of best practices when samples of similar organizations are available. As an order based statistical estimation approach, it also has the virtue of being more robust than typical moment approaches.

Note: This paper is based on a working paper developed at The Maxwell School in June 2001, which was partially funded by a postdoctoral research grant from the Dean of The Maxwell School and the Center for Technology and Information Policy.

across diverse service providers using performance measurement data? Can best-practice providers actually be identified? As noted by Nyhan and Martin (1999), "after encouraging governments to routinely collect and report information on the efficiency, quality, and effectiveness of their programs, the performance measurement literature suddenly becomes silent on how the resulting data might be used in making service provider comparisons."

Most of the standard statistical estimation techniques used to analyze performance data focus attention on typical behavior or average behavior. However, in order to identify a "best" practice case, a different set of techniques is necessary. Several potential approaches have been identified that attempt to sort out cases and identify extreme cases. Data Envelopment Analysis (DEA), for example, converts multiple input and multiple output data on comparable organizations into a relative measure of technical efficiency, which can then identify those units that are best at converting inputs to outputs (Charnes Copper and Rhodes,1978). Quantile regression and substantively weighted least squares (a form of iteratively weighted least squares), have been suggested in situations where a single output is related to multiple inputs, including environmental variables, and the relationship between variables is expected to differ for high performers, typical performers and low performers (Marc-Aurele, DíAmico and Bretschneider, 2000). While both of these approaches directly estimate conditional distribution for the outcomes in the tails of a distribution, the quantile approach allows the user to specify *apriori* a specific conditional quantile (Bassett and Koenker, 1978; Eide and Showalter, 1999; Scharf, Juanes, and Sutherland, 1998). Substantively Weighted Least Squares (SWLS)[2] has also been introduced and applied to exploring how "best practice" organizations differ from typical performers in public administration (Meier and Keiser, 1996; Meier and Gill, 2000). However, SWLS estimation is more heuristic and does not automatically yield a specific conditional quantile, the actual location in the tail of the distribution can only be obtained after estimation and will vary significantly from estimation to estimation. While some analytic results exist to guide the use of quantile regression, recent results from a comparative study on Texas Schools, suggest that application of either quantile regression or SWLS can produce conflicting and un-interpretable results (Marc-Aurele, DíAmico and Bretschneider, 2000)

This paper uses Monte Carlo simulations to develop a deeper understanding of how quantile regression is likely to perform is a wide array of possible realistic settings. It is the view of the authors that such studies are necessary in order to develop appropriate strategies for using any empirical approach to identification and analysis of "best practice cases" from real world data. The next section of the paper describes the quantile regression model, how to estimate the model, and how formal statistical inference is applied. This is followed by a description of the design of the Monte Carlo simulation and the resulting datasets. Section four of the paper discusses the results of applying quantile regression to the data and attempts to develop some reasonable guidance about how to apply quantile regression to real world data. Finally, the paper concludes by summarizing our findings and suggesting directions for future research.

State of the Art of Quantile Regression

In this section of the paper, we review the underlying model associated with Quantile Regression, how to estimate it from data and how to apply statistical inference to the results.

The Model

Quantile Regression is a statistical technique intended to estimate, and conduct inference on, conditional quantile functions (Koenker, 2000a). Its regression model, first introduced by Koener and Bassett (1978b), can be written as

$$y_i = x_i' \beta_\theta + \mu_\theta \ , \ Quant_\theta (y_i \mid x_i) = x_i' \beta_\theta \qquad (1)$$

where (y_i, x_i) $i = 1, \ldots, n$, is a sample from some population, x_i is a $K \leftrightarrow 1$ vector of regressors, and $Quant_\theta (y_i \mid x_i)$ denotes the conditional quantile of y_i, conditional on the regressor vector x_i (Buchinsky, 1998; Eide and Showalter, 1999). Note that as one increases θ continuously from 0 to 1, one traces the entire distribution of conditional on x.

An interesting application of the quantile regression technique is in situations where multiple fixed structures underlay a single dataset. Consider the following situation:

$$Y_i = \begin{cases} a_1 + b_1 X_{1i} + \varepsilon_{1i} \\ a_2 + b_2 X_{2i} + \varepsilon_{2i} \\ \ldots \\ a_m + b_m X_{mi} + \varepsilon_{mi} \end{cases} \quad (2)$$

where an observation Y_i could be generated by any of m alternative equations. Note that in this context we will assume that the set of regressors are the same in each of the equations, only the coefficient values change across equations. In this context, subsets of observations will group together around the different generating models. Note that as m increases to 100, this situation approaches the quantile case where each value of m is a specific quantile. Many real world applications ask the researcher to consider situations were m is some unknown value reflective of different responses or reactions to the same stimuli or inputs. For example, in many fields we are interested in how high medium and low performing organizations or organisms behave and why.

Estimation and Issues

Wagner (1959) provided one of the earliest treatments for estimation of the quantile model by minimizing weighted least absolute deviations. Koenker and Bassett (1978) discuss the theory for estimating quantiles of a variable assumed to be a linear function of other variables. Bloomfield and Steiger (1980) and Koenker and D'Orey (1987) explain the use of linear programming techniques as a means for estimating quantile regression. Specifically, to estimate a Quantile Regression, the problem can be stated as the following linear programming problem:

$$\text{Minimize} = \sum_{t \{t: y_t \geq b\}}^{n} \theta e_t^{-} + \sum_{t \{t: y_t < b\}} (1 - \theta) e_t^{+} \quad (3)$$

Subject to:

$$y_t - x_t\beta - e_t^{+} + e_t^{-} = 0 \quad t = 1, \ldots t$$

Where y_t is the dependent variable, x_t is the explanatory variable, $0 < \theta < 1$ is the specific quantile of interest, and b is a parameter to be estimated, and e_t^{+} and e_t^{-} are positive and negative deviations from the model also estimated by the procedure. When $\theta = .5$, the solution will be identical to the estimates producing the median regression line. Quantile regression had been used in a broad range of applications as a comprehensive approach to the statistical analysis of linear and nonlinear response models (Buchinsky, 1998; Koenker, 2000).

Inference and Issues

Despite early development of the estimation procedure based on minimizing absolute deviations, little could be said about the sampling distributions of the coefficient estimates. This failing is at least partially responsible for the slow diffusion of the method among researchers. In 1978 Basset and Koenker (1978) developed a formal basis for large sample inference for the special case of the median regression and later extended it to include the full set of quantile regression models (Basset and Koenker, 1982). The results demonstrate that in a large sample, estimated parameters are consistent estimators of the true parameters and have an asymptotically normal sampling distribution.

While knowledge of large sample distributions was provided through axiomatic developments, most results surrounding small sample behavior comes from numerous Monte Carlo simulations (Dielman and Pfaffenberger, 1982). Most of the early simulation studies of least absolute deviation estimators (LAV) focused on efficiency of these estimates as compared to least squares estimators for various error distributions. Except for standard normal error distributions, these results found that LAV estimators were consistently more efficient than least squares estimators. The corresponding results from the large sample theory found that for any error distribution where the median is (asymptotically) more efficient than the mean as an estimator of location, LAV estimates would similarly be more efficient than least squares estimation (Dielman and Pfaffenberger, 1982).

Consequently, Quantile Regression provides a powerful alternative to more traditional approaches to estimating relationship for several reasons. While it can be used to estimate standard relationships for typical conditional behavior of phenomena (the median regression), it can also be used to study extreme conditional behavior. Secondly, quantile regression is more robust than traditional methods in the same way a median is more robust than the mean, thus its assumptions are likely to be less restrictive when applied to real world data. Finally, a well-established approach to large sample inference exists permitting formal hypothesis testing. Finally, computational software is now relatively widely diffused making it easy to apply.

While simulations studies have been used to compare efficiency LAV estimates with least square estimates, application of quantile regression to study extreme behavior is relatively new. Several researchers have proposed alternatives to quantile regression based on iterated weighted least squares estimation (Meier and Keiser, 1996; Meier and Gill 2000). To better understand the characteristics of quantile regression as a tool for modeling extreme behaviors, several researchers have compared it with one iterated weighted least squares approach call substantively weighted least squares (SWLS/SWAT) (Marc-Aurele, D' Amico and Bretschneider, 2000; Wu, Bretschneider, Marc-Aurele, 2000). While useful, these prior studies are limited in the set of conditions simulated. In particular, their paper considered relatively symmetrically distributed alternative models. This paper considers a broader set of assumptions in building a set of simulated data, which are more likely to be encountered in the real world.

Monte Carlo Simulation

The purpose behind a Monte Carlo Simulation is to study the behavior of a statistic as it varies across multiple samples. This is done by an artificial "world," or *pseudo-population*, which hopefully resembles the real world in all relevant respects (Mooney, 1997). Our interest is in studying how quantile regression can identify alternative sub-structures within a single sample for high medium and low performing cases.

General Structure (high, middle, low)

Since we are interested in the extreme behavior, we assume there are many organizations that produce some output as a function of a single input where a high performing unit is one that generates a higher level of output for the same level of input than do other units. Similarly, a unit that generates lower output for the same level of input will be considered a lower performer. In order to simplify the situation, we assume that there are only three different types of organizations in our population-high, medium and low performers. Equations 4, 5 and 6, define these specifically in terms of different coefficients to convert inputs to outputs.

$$y_h = 200 + 20\,x_h + e \qquad (4)$$
$$y_m = 120 + 10\,x_m + e \qquad (5)$$
$$y_l = 100 - 5x_l + e \qquad (6)$$

Symmetric vs Non-symmetric sub-samples

The final sample size was set at 600 cases but in order to evaluate a wide range of situations that might be encountered in the real world, several different distributions of cases allocated to each of the three models were generated. First, the values for the three independent variables were generated using a uniform distribution defined in the interval 0 to 10, and three separate sets of random errors were generated based on a normal distribution with zero mean and variance 5. The base case assumed each equation was equally likely to occur so that there were 200 observations for each equation combined to form a single sample of 600 observations. This specific situation was used by Marc-Aurele et. al. (2000) and we refer the reader to that

Table 1: The Asymmetrical Subgroup Structure of Simulated Datasets

Percent of Cases	Corresponding Observations in the simulated Datasets		
High-Medium-Low	High	Medium	Low
10/20/70	60	120	420
20/30/50	120	180	300
50/30/20	300	180	120
60/20/20	480	120	120
70/20/10	420	140	60

study for extensive information on how LAV models faired. Here we move on to consider situations where the relative likelihood of high, medium and low performance is not uniformly distributed and generated several alternative samples as shown in Table 1.

100 iterations for comparisons

For each of the five designs presented in table 1, one hundred separate iterations of a 600 case data set were generated in order to study the sampling distribution of the estimates.

Results and Analysis

Since prior work has focused on what we call symmetric distributions of models within a single sample (Wu and Bretschneider 2000; Marc-Aurele et. al. 2000), here we only present the results from our asymmetric simulations where the ratio of high to medium to low are 10:70:20, 20:30:50, 50:30:20, 60:20:20, and 70:20:10. The results are summarized in a series of

Figure 1: Average Slope Versus Quantile with Ratio 10/70/20

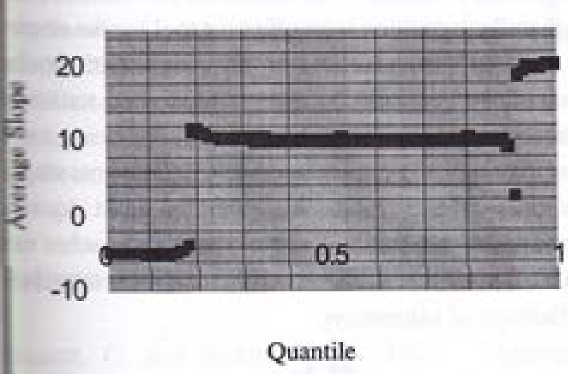


Figure 2: Average Slope Versus Quantile with Ratio 20/30/50

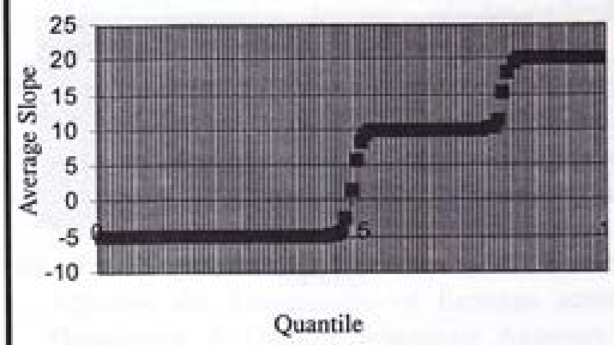


Figure 3: Average Slope Versus Quantile with Ratio 50/30/20

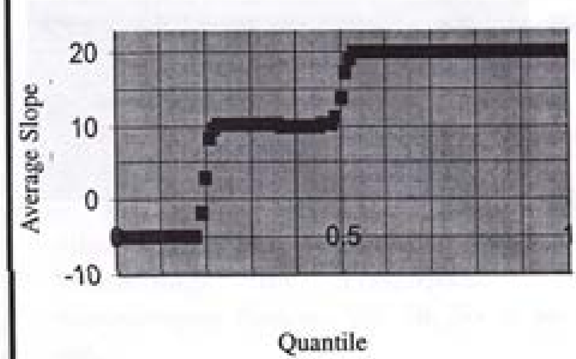


Figure 4: Average Slope Versus Quantile with Ratio 60/20/20

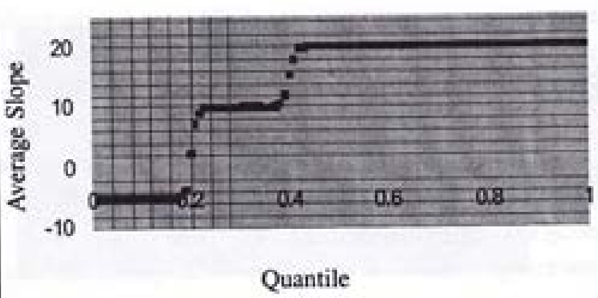
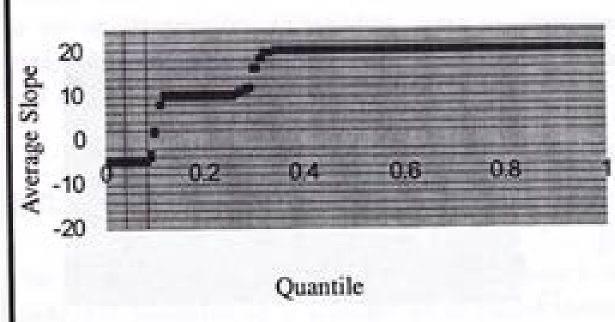


Figure 5: Average Slope Versus Quantile with Ratio 70/20/10



related figures. Each figure plots the average estimate of the slope term for each quantile from the 100 separate samples against the quantile. Each value on the graph then represents the results of 100 estimations and the whole graph covering 100 separate quantiles summarizes 1000 separate estimates. The true values for the high, medium and low slope were 20, 10 and -5 respectively.

In Figure 1 we see a direct correspondence between the percent of cases in each of the 100 samples and the range of quantiles. Quantile one through 18 the average estimate remains relatively constant close to the true value of -5. Then for quantile 19 through 89 the average slope remains close to 10. Finally the estimated average slope for quantiles 91 through 100 remains level at 20. The biggest estimation errors occur at and around the break points. This general pattern hold across all of the various cases presented in figures 1 through 5.

Conclusions

This paper extends the range of situations likely to be encountered in real world data over pervious simulation studies of quantile regression. The results further support the ability of quantile regression to accurately estimate different underlying situations. The particular context for this work is in sorting out how high, medium and low performing organization convert inputs to outputs with particular reference to the best practices area of management research. The results further demonstrate that quantile regression is an effective tool for the empirical study of best practices when samples of similar organizations are available. As an order based statistical estimation approach, it also has the virtue of being more robust than typical moment approaches. The next step is to compare these results with iterated least squares approaches to determine if one of these approaches can be identified as superior at accuracy of estimation (bias) or efficiency of estimation.

Jiannan Wu is an associate Professor and director of Center for Master of Public Administration and chair of Department of Public Management at Xi'an Jiaotong University, China. After working as a post-doctoral research associate at Center for Technology and Information Policy (CTIP) of Maxwell School, he has been appointed as a senior research associate with CTIP since January 2001. His research interest includes performance measurement, innovation policy and administrative discretion.

Stuart Bretschneider is a Professor of Public Administration and Director of the Center for Technology and Information Policy at The Maxwell School of Citizenship and Public Affairs. His primary fields of research have focused on how public organizations make use of information technology and the effects of those technologies on public organizations; how public organizations employ forecasting technology and organize to carry out forecasting activities; and how sector differences affect administrative processes. Dr. Bretschneider is a past Managing Editor of the *Journal of Public Administration Research and Theory*, as well as a past President and Director of the International Institute of Forecasting (IIF).

Endnotes

1 GPRA: the Government Performance and Results Act of 1993 (GPRA, Public law 103-62); NPR: National Performance Review; GASB: the Governmental Accounting Standards Board (GASB).

2 According to its authors, SWLS is just one special form of Substantively Weighted Analysis Techniques.

References

Bassett, G. and Koenker, R., 1978. "Asymptotic Theory of Least Absolute Error Regressions." *Journal of the American Statistical Association,* Vol. 73, pp. 618-622.

Bloomfield, P., and W. Steiger, 1980. "Least Absolute Deviations Curve Fitting." *SIAM Journal on Scientific and Statistical Computing,* Vol. 1, pp. 290-301.

Buchinsky, M., 1998. "Recent Advances in Quantile Regression Models." *The Journal of Human Resources,* Vol. 33, No. 1, pp. 88-126.

Charnes, A., Cooper, W.W. and Rhodes, E. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research,* Vol. 2, 429-44.

Dielman, T and Pfaffenberger, R 1982 "LAV (Least Absolute Value) Estimation In Linear Regression: A Review" Journal *of Studies in the Management Sciences.* Vol. 19, pp. 31-52. (Ask Rick for this paper)

Eide, Eric R. and Mark H. Showalter, 1999. "Factors Affecting the Transmission of Earnings across Generations: A Quantile Regression Approach." *Journal of Human Resources,* Vol. 34, No. 2, p. 253.

Katorobo, James, 1998. "The Study of Best Practices in Civil Service Reforms, Management Development and Governance Division." *http://magnet.undp.org/docs/psm/best7.htm.*

Koenker, R. and Bassett, G. 1978b. "Regression Quantiles." *Econometrica,* Vol. 46, pp. 33-50.

Koenker, R. Galton, Edgeworth, Frisch, 2000. "Prospects for Quantile Regression in econometrics." *Journal of Econometrics,* Vol. 95, pp. 347-374.

Marc-Aurele, Frederick J., Jr., Lynne C. D' Amico, and Stuart Bretschneider, 2000. "Examining the 'Extremes' of Bureaucratic Performance: A Comparative Examination of Quantile Regression and Substantially Weighted Analytic Techniques." Working paper, Center for Technology and Information Policy, Maxwell School of Citizenship and Public Affairs.

Meier, K.J. and Keiser, L.R., 1996. "Public Administration as a Science of the Artificial: A Methodology for Prescription." *Public Administration Review,* Vol. 56, No. 5, pp. 459-466.

Meier, Kenneth J. and Gill, Jeff, 2000. *What Works: A New Approach to Program and Policy Analysis.* Boulder, CO: Westview Press.

Mooney, Christopher Z., 1997. "Monte Carlo Simulation." *Quantitative Applications in the*

Social Sciences. Sage University Papers Series, No. 116.

Nyhan, R. C. and L. L. Martin, 1999. "Assessing the Performance of Municipal Police Departments Using Data Envelopment Analysis: An Exploratory Study." *State and Local Government Review*, vol. 31, pp. 18-30.

Neter, John et al., 1996. *Applied Linear Statistical Models* (Fourth Edition).

Overman, E. Sam and Boyd, Kathy J., 1994. "Best Practice Research and Postbureaucratic Reform." *Journal of Public Administration Research and Theory*, Vol. 4, No. 1, pp. 67-83.

Rocheleau, Bruce, 2000. "Prescriptions for Public-sector Information Management: A Review, Analysis, and Critique, American." *Review of Public Administration*, Vol. 30, No. 4 (December).

Scharf, F.S., Juanes, F., and Sutherland, M., 1998. "Inferring Ecological Relationships From The Edges of Scatter Diagrams: Comparison of Regression Techniques." *Ecology*, Vol. 79, No. 2, pp. 448-460.

Wagner, Harvey, 1959. "Linear Programming Techniques for Regression Analysis." *Journal of the American Statistical Association*, Vol. 54 pp. 206-212.