Chapter 6 Summary and Future Work

Summary

Model Robust Regression is clearly a technique of the future. The results presented in this dissertation provide evidence of its asymptotic superiority over individual parametric and nonparametric estimates. The "Golden Result of Model Robust Regression" confirms that the MRR user obtains the best component estimate rate of convergence asymptotically, regardless of which type (parametric or nonparametric) of component is appropriate for the true mean response function q. We have seen that this result occurs in a variety of settings to include (but not necessarily limited to) the MRR1 and MRR2 procedures in which OLS and some GLS parametric estimates are paired with local polynomial estimates obtained through LS or GLM. There are various ways that this research can be extended. We will look at a few of these areas below.

Comparison of Mixing Parameter Estimates at Small Samples

Mays (1995) states in his 9th Chapter (Future Research), "The main area of future research in terms of combining the separate parametric and nonparametric fits is in choosing the mixing parameter l". Initial studies done by Mays (in conjunction with this research) indicate that the asymptotically optimal data driven mixing parameters $\hat{l}^{*c}$ and $\hat{l}^{*}$ perform rather well in small sample settings (n = 10, 19, 50) for MRR1 and MRR2 respectively. These data sets correspond in size to those used in Mays (1995). In each case the predictor values are spaced uniformly over the support interval for the function in question. We would do well to learn from these results and possibly consider other models that are solid asymptotically in the small sample setting. Similarly, one might also investigate small sample properties of $\hat{l}^{*P}$ in the MRR1 Logistic estimate. If asymptotically sound models perform well for small samples, then one should select the estimate that has optimal performance in both directions.

Asymptotic Results for Dual Model Robust Regression

Robinson (1997) researched the application of MRR to the non-constant variance case and developed what is now called Dual Model Robust Regression (DMRR). We feel that a study in MRR asymptotics is not complete without considering the asymptotic properties of these important models as well.

Specifying the Results

According to a conversation with Steve Marron, we have considered only the first level of asymptotic rates of convergence. It would be interesting to see how the MRR estimates perform at higher levels of asymptotic investigation where more specific rates of convergence based on values of constants and bandwidth convergence rates are obtained.

Generalizing the Results

We have restricted ourselves in chapter 5 to the local linear nonparametric estimate for h. With some work we could possibly generalize this to include other types of local polynomial estimates; particularly the local cubic estimate. To begin with, the model set forth in this chapter is restricted to the univariate case. The extension to multiple regression would be worth investigating. It would also be interesting to formalize the extension to kernel functions other than the Epanechnikov (mentioned by Fan). In various lemmas we have considered convergence rates of mixing parameter estimates to the theoretically optimal mixing parameter that is based on the Least Squares criterion. It may be worthwhile to consider a theoretically optimal mixing parameter based on another criterion such as Maximum Likelihood. Throughout the dissertation, the assumption of independent (although not necessarily identically distributed) responses has been used. One possible (but difficult) research problem would be to find the rates of convergence of the MRR estimates when the responses are correlated. Not quite so hard, is the problem of extending the results to include all types of bandwidth selectors, no matter how slowly they diminish. Finally, as mentioned in Chapter 5, it would be nice to generalize the results of section 5b in two directions.

First, the generalization of all results so that assumption A1 is not necessary would be more convenient for users of MRR. Second, the GMRQR procedures are fairly broad, however the generalization of results to include the GMRER procedures would not be too difficult to attain, and would cover a much larger group of estimates.

Closing Thoughts

Although theoretically based, rates of convergence will continue to play a greater role in the study of Statistics. One reason for this is that statisticians are encountering larger data sets than ever before. Sample sizes in excess of 1000 warrant measures that are not only asymptotically sound, but optimal. It is my hope that the work contained in this dissertation will serve to advance the use of the semiparametric process known as Model Robust Regression, as well as provide a platform for further asymptotic work in regression and related fields.