Part 5b Asymptotic Results for MRR1 Quantal Regression

We are interested in developing final asymptotic results for quantal regression as described in part 5a using the MRR1 estimate with an asymptotically optimal data driven mixing parameter. We have already seen that such a mixing parameter can perform as well as the theoretical asymptotically optimal mixing parameter in the simpler case of chapter 3.

We will set up the MRR1 local quantal regression estimate for the general situation and denote this technique as Generalized Model Robust Quantal Regression (GMRQR) using the notation of part 5a. The parametric and nonparametric estimates are, again, based on the original data, and, after obtaining each separately, combined to form

$$\boldsymbol{l}\hat{f}(\mathbf{x}_i) + (1 - \boldsymbol{l})\hat{g}_{LL}(\mathbf{x}_i) = \hat{\boldsymbol{q}}(\mathbf{x}_i)\text{, for } i = 1 \text{ to } n$$

where $\boldsymbol{l} \in (0,1)$, and $\hat{f}, \hat{g}$ represent the parametric and nonparametric estimates respectively.

Recall that the data driven bandwidth is designated by $\boldsymbol{t}_n^{-1}$ and it is assumed $t_n \to \infty$, and $(\tau_n / n) \to 0$ as $n \to \infty$. The implication here is that the bandwidth goes to zero at a rate slower than n^{-1}. Note that as a result of previous sections, we are concerned with the speed with which the bandwidth descends to zero, but not with the convergence rate to the optimal bandwidth. As seen earlier, this convergence rate plays an important role in the convergence rate of the nonparametric estimate. Also we will again denote by $\boldsymbol{d}_n$, the distance between the unknown regression function, $\boldsymbol{q}$, and the parametric family of continuous (on C) regression functions under consideration.

$$\boldsymbol{d}_n = \inf\{\|\boldsymbol{q} - f(\boldsymbol{b})\| : \boldsymbol{b} \in \Re^{p+1}\}.$$

Similarly, we set $\boldsymbol{g}_n$ as a distance measure for the nonparametric estimate (on C),

$$\boldsymbol{g}_n^2 = \mathrm{E}(\|\hat{g}_{LL}(\mathbf{x}_i) - \boldsymbol{q}\|^2).$$

But now we have for this specific local logistic estimate (from Lemma 5.a.1) that

$$\boldsymbol{g}_n^2 = \boldsymbol{t}_n n^{-1}.$$

Note that our theoretically optimal mixing parameter is now given by

$$\boldsymbol{l}^{*L} = \frac{< \hat{f} - \hat{g}_{LL}, \boldsymbol{q} - \hat{g}_{LL} >}{\left\| \hat{f} - \hat{g}_{LL} \right\|^2} = \frac{< \hat{f} - \hat{g}_{LL}, \boldsymbol{q} > - < \hat{f} - \hat{g}_{LL}, \hat{g}_{LL} >}{\left\| \hat{f} - \hat{g}_{LL} \right\|^2},$$

a quantity which we will attempt to estimate by the following formula.

$$\hat{\boldsymbol{l}}^{*L} = \frac{< \hat{f} - \hat{g}_{LL}, Y_i - \hat{f} >}{\left\| \hat{f} - \hat{g}_{LL} \right\|^2} + 1 = \frac{< \hat{f} - \hat{g}_{LL}, Y_i - \hat{g}_{LL} >}{\left\| \hat{f} - \hat{g}_{LL} \right\|^2}.$$

We will still need one of our previous assumptions.

A1. There exists a function W_1 of two variables which is defined and bounded on C x C , and

$$||\hat{f}(.;\hat{\boldsymbol{b}}) - f(.;\boldsymbol{b}^*) - n^{-1}\sum_{i=1}^{n} W_1(.,x_i)e_i|| = O_P(n^{-1}).$$

Recall that assumption A1 is necessary for the use of Whittle's Inequality (Whittle (1960)) in the proof of Lemma 5.b.3. A2 guarantees that $\|\hat{g}_{LL} - \boldsymbol{q}\| = O_P(\boldsymbol{g}_n) = O_P(\boldsymbol{t}_n^{.5} n^{-.5})$. A2 and A5 are guaranteed by Lemma 5.a.1 and our assumptions on the bandwidth. The existence of W_2 is not necessary for the results of this section, consequently assumptions A3, A4, and A6 are not required either. A1 will hold depending on certain regularity conditions on f. An important concern is the difficulty of obtaining W_1 for assumption A1.

We shall investigate two parametric estimates in this example. Since the biggest transition from chapter 3 is dealing with the nonconstant variance, our first example is a simple estimate that can be set up in that scenario. Suppose a user sets up a model for quantal regression in which the response variance is assumed to be constant for all x_i, and the model is assumed to be linear in x. This is clearly incorrect.

Nevertheless, a weight function, $W_{1\text{I}}$, with property A1 is obtained, where the subscript I denotes “identical variances” for the response variable. Next observe that in quantal regression to properly specify the variance, we need only write

$$W_1(.,x_i) = W_{1\text{I}}(.,x_i)s(f(x_i;\boldsymbol{b}^*)(1-f(x_i;\boldsymbol{b}^*))^{-.5} \quad (5.B.1)$$

(for some real constant s corresponding to the variance of the misspecified model) to obtain the weight function W_1 corresponding to the corrected model having nonconstant variance. The second estimate builds upon the first, and extends to the GLM setting where link functions use information obtained from linear predictors to estimate mean response. Burman and Chaudhuri (1992) state that in the i.i.d. error case, we have

$$\hat{f}(.;\hat{\boldsymbol{b}}) - f(.;\boldsymbol{b}^*) - n^{-1}\sum_{i=1}^{n} W_1(.,x_i)e_i = 0 \quad (5.B.2)$$

provided f is linear in $\boldsymbol{b}$, and continuous in the second argument. If f is given by

$$b_0 + b_1 x$$

then for x_i in C, equation 5.B.2 reduces to

$$x_i(\hat{\boldsymbol{b}}_1 - \boldsymbol{b}_1) = n^{-1}\sum_{j=1}^{n} W_1(.,x_j)\boldsymbol{e}_j \quad (5.B.3)$$

Now if we take F to be a function of this linear predictor that is differentiable in $\boldsymbol{b}$ (note that in this example we are not requiring F to be the standard logistic CDF, although that selection will certainly work) then using equation 5.B.3 and the Taylor series expansion, for x_i in C,

$$\hat{F}(x_i;\hat{\boldsymbol{b}}) - F(x_i;\boldsymbol{b}) = F_{b_1}(x_i;\boldsymbol{b})(\hat{\boldsymbol{b}}_1 - \boldsymbol{b}_1) + o(\hat{\boldsymbol{b}}_1 - \boldsymbol{b}_1)$$

$$= n^{-1}\sum_{j=1}^{n} W_1(x_i,x_j)(F_{b_1}(x_i;\boldsymbol{b}) + c_i)x_i^{-1}\boldsymbol{e}_j \quad (5.B.4)$$

for some real constant c, provided C does not contain 0. Since this was done for arbitrary x_i in C, our new weight function is formed as shown in 5.B.4 with W_1 defined as above. Thus it is quite easy to satisfy A1 for a nonlinear function of a linear predictor in the nonconstant variance case.[].

There is some difficulty involved in verifying that a parametric estimate satisfies A1. It is used here for convenience of proof. However, we feel that the results of this section can be extended to parametric estimates that do not necessarily satisfy A1. We will leave this for future research.

Once more, we give three lemmas dealing with the dichotomy: the parametric estimate is correct ($\boldsymbol{d}_n = 0$) vs. the parametric estimate is incorrect ($\lim_{n\to\infty} \boldsymbol{d}_n \neq 0$). Recall that we are restricting the nonparametric estimate to be that estimate defined in the previous section, but allowing the parametric estimate to take various forms allowed by requirements R1-R4. Recall that this allows for nonlinear functions of x and $\boldsymbol{b}$ to be used for the parametric estimate. We now present the following asymptotic results for the MRR1 quantal regression estimate. Although these results deal with this specific estimate, they may be extended, with a little work on the nonparametric estimate, to the Generalized Model Robust Exponential family Regression (GMRER) case. The proofs for the results in this section may be found in Appendix 5b.

Lemma 5.b.1: Assuming conditions A1, and R1-R4…

$$\left\| \hat{f} - \hat{g}_{LL} \right\| = \begin{cases} O_P(1), & \text{if } \lim_{n\to\infty} \boldsymbol{d}_n \neq 0 \\ O_P(\boldsymbol{t}_n^{.5} n^{-.5}), & \text{if } \boldsymbol{d}_n = 0 \end{cases}.$$

Lemma 5.b.2: Assuming conditions A1, and R1-R4…

$$\boldsymbol{l} *^{L} = \begin{cases} O_P(\boldsymbol{t}_n^{.5} n^{-.5}), & \text{if } \lim_{n\to\infty} \boldsymbol{d}_n \neq 0 \\ 1 + O_P(\boldsymbol{t}_n^{-.5}), & \text{if } \boldsymbol{d}_n = 0 \end{cases}.$$

These lemmas are pertinent to the results that follow. Lemma 5.b.1 is crucial in the proof of Lemma 5.b.3. Lemma 5.b.2 is important in the proofs of the theorems. In Lemma 5.b.1 we deal with the rate of convergence of the vector difference between the parametric and the nonparametric estimates. The rate of $O_P(1)$ is correct in the first instance since we would not expect the parametric estimate to match the nonparametric estimate if the model is misspecified.

If the model is correctly specified the difference will approach zero at the slower nonparametric rate. Lemma 5.b.2 provides the asymptotic properties of the theoretical asymptotically optimal mixing parameter. We would expect $\boldsymbol{l}^{*L}$ to converge to zero if the model is misspecified, and converge to 1 otherwise. In this case, the rate of convergence to 1 is dependent upon the rate at which the bandwidth converges to zero. Notice that in this section we can see the role played by the bandwidth in all convergence rates. Finally, note that if we affiliate $\boldsymbol{t}_n n^{-.5}$ with the former rate of $\boldsymbol{g}_n$, the results are the same as in part 3a. Both proofs are found in the associated appendix.

Lemma 5.b.3 provides the rate of convergence for the difference between the theoretically optimal mixing parameter and the asymptotically optimal data driven mixing parameter. Its proof is found in Appendix 5b.

Lemma 5.b.3: Assuming conditions A1, and R1-R4…

$$\boldsymbol{l}^{*L} - \hat{\boldsymbol{l}}^{*L} = \begin{cases} O_P(\boldsymbol{t}_n^{.5} n^{-1}) + O_P(n^{-.5}), & \text{if } \lim_{n\to\infty} \boldsymbol{d}_n \neq 0 \\ O_P(\boldsymbol{t}_n^{-.5}), & \text{if } \boldsymbol{d}_n = 0 \end{cases}.$$

As we have seen in previous chapters, this lemma drives the process, in that it plays a key role in the proof of the following theorems dealing with estimate convergence. Note that if we make the exchange of rates mentioned previously (for comparison with part 3a results) we have a slightly better result for the logistic regression semiparametric estimate. This is possibly due to not using cross-validation. However, the estimate convergence rates (provided in the theorems) are not affected.

Theorem 5.B.1: Assuming conditions A1 and R1-R4…

$$\left\| \boldsymbol{l}^{*L} \hat{f} + (1 - \boldsymbol{l}^{*L}) \hat{g}_{LL} - \boldsymbol{q} \right\| = \begin{cases} O_P(\boldsymbol{t}_n^{.5} n^{-.5}), & \text{if } \lim_{n\to\infty} \boldsymbol{d}_n \neq 0 \\ O_P(n^{-.5}), & \text{if } \boldsymbol{d}_n = 0 \end{cases}.$$

Theorem 5.B.2: Assuming conditions A1 and R1-R4…

$$\left\| \hat{\lambda}^{*L}\, \hat{f} + (1-\hat{\lambda}^{*L})\hat{g}_{LL} - \theta \right\| = \begin{cases} O_P(\tau_n^{.5} n^{-.5}), & \text{if } \lim_{n\to\infty} \delta_n \neq 0 \\ O_P(n^{-.5}), & \text{if } \delta_n = 0 \end{cases}.$$

Theorem 5.B.1 gives the convergence rate for the artificial case in which θ (and consequently, λ^{*L}) is known. Theorem 5.B.2 gives the convergence rate for a data driven MRR1 procedure. Once again, we have another form of "The Golden Result of Model Robust Regression". This is clear by applying the exchange of rates mentioned earlier. In these theorems, however, the result has been extended to include a combination of parametric and nonparametric estimates not previously covered in the framework of chapter 3. The estimate here is much more general and complex than those presented earlier. Here we have a local linear logistic estimate for the nonparametric part, and have allowed for a generalized nonlinear estimate in the parametric part. We will say more about Theorem 5.B.2 in the last part of the section. The proofs of both theorems are found in Appendix 5b.

We will demonstrate the usefulness of the quantal MRR1 estimate with two examples.

First suppose a user is estimating the mean response function θ, for a quantal response variable Y, using MRR1 and attempting to model the function parametrically with the standard logistic regression procedure, and nonparametrically by the form of local linear logistic regression specified in this chapter using the constant bandwidth, h_{AMISE}, as outlined in Fan, Heckman, and Wand (1995). The estimate will also utilize the Epanechnikov Kernel and $\hat{\lambda}^{*L}$ for the mixing parameter. From our earlier example, condition A1 is met.

Fan, Heckman, and Wand (1995) also state that

$$h_{AMISE} = \boldsymbol{t}_n^{-1} = o_P(n^{-\frac{1}{5}}) .$$

Then using Theorem 5.B.2, we have that asymptotically the user has an estimate such that

$$\left\| \hat{\boldsymbol{l}} *^L \ \hat{f} + (1 - \hat{\boldsymbol{l}} *^L)\hat{g}_{LL} - \boldsymbol{q} \right\| = \begin{cases} O_P(n^{-.4}), & if \lim_{n \to \infty} \boldsymbol{d}_n \neq 0 \\ O_P(n^{-.5}), & if \ \boldsymbol{d}_n = 0 \end{cases} ,$$

and will thus converge to the true mean function at a rate no slower than $O_P(n^{-.4})$ if the model is misspecified, and as fast as $O_P(n^{-.5})$ if $\boldsymbol{q}(\mathbf{x})$ truly follows the canonical parametric form on C.[].

The second example is similar to the first except that we consider the parametric estimate that Nottingham, Birch, and Bodt (1999) used to model armor penetration probabilities corresponding to various muzzle velocities measured in feet/second (fps). In this paper, C = (1432, 2773)fps. The parametric estimate is given by

$$\hat{f}(x; \hat{\boldsymbol{b}}) = (1 - F_T(x; \hat{\boldsymbol{b}}_T))F_I(x; \hat{\boldsymbol{b}}_I) + F_T(x; \hat{\boldsymbol{b}}_T)F_S(x; \hat{\boldsymbol{b}}_S)$$

where

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_I \\ \boldsymbol{b}_S \\ \boldsymbol{b}_T \end{pmatrix} \in \Re^{+6} ,$$

x represents the shell muzzle velocity, and

$$F_T(x; \boldsymbol{b}_T) = \left(1 + e^{\frac{-(x - \boldsymbol{b}_{T_1})}{\boldsymbol{b}_{T_2}}} \right)^{-1} .$$

It is not too hard to show that this estimate satisfies requirements R1-R4, and A1 since each of F_T, F_S, and F_I are bounded, continuous in all arguments, can be expressed as a function of a linear predictor, and f is a sum of products of F_T, F_S, and F_I.

Then using Theorem 5.B.2, we have that asymptotically

$$\left\| \hat{\lambda}^{*L} \hat{f} + (1 - \hat{\lambda}^{*L}) \hat{g}_{LL} - \theta \right\| = \begin{cases} O_P(n^{-.4}), & \text{if } \lim_{n\to\infty} \delta_n \neq 0 \\ O_P(n^{-.5}), & \text{if } \delta_n = 0 \end{cases},$$

and will thus converge to the true mean function at a rate no slower than $O_P(n^{-.4})$ (clearly the local linear logistic estimate using h_{AMISE} performs admirably asymptotically in light of the previous nonparametric estimates) if the model is misspecified, and as fast as $O_P(n^{-.5})$ if $\theta(\mathbf{x})$, the penetration probability, truly follows the parametric form implied above on C.[].

Concluding Remarks

These results demonstrate again the main advantage of MRR1 outlined earlier. If the user's model is incorrect, MRR1 can achieve a consistent estimate at the asymptotic convergence rate of the nonparametric estimate. On the other hand, if the model is correct, MRR1 achieves consistency at the parametric rate, faster than a purely nonparametric estimate. Once again we see that the MRR1 estimate with the asymptotically optimal data driven mixing parameter $\hat{\lambda}^{*L}$ performs just as well as the MRR1 estimate with the asymptotically optimal theoretical mixing parameter λ^{*L}. In addition, the theorems demonstrate the flexibility of the estimate as well. It is not unrealistic to believe that the convex combination of estimates given by MRR1 can produce final estimates that are asymptotically superior to either of the component estimates in a wide variety of settings.