Chapter 5 Asymptotic Results for Semiparametric Quantal Regression

We wish to extend our asymptotic work in Model Robust Regression to include parametric and nonparametric estimates that coincide with more recent and pertinent applications. There are many different areas that warrant investigation, but one that stands out particularly is the study of quantal regression. In parametric regression, quantal regression is used most often when the response variable is binary in nature. Either the response is positive or negative, successful or unsuccessful or (stochastically) 1 or 0. The quantal regression estimate is designed to ascertain the probability that the response is 1. Military target analysis, dose-response problems in the health sciences, load-failure problems in engineering, and coupon redemption problems in marketing are all areas of application for quantal regression.

Again, our model is written as

$$Y_i = \boldsymbol{q}(\mathbf{x}_i) + \boldsymbol{e}_i \text{ for } i = 1 \text{ to } n.$$

The $\boldsymbol{e}_i$'s, however, are now independent with $E(\boldsymbol{e}_i) = 0$ and $V(\boldsymbol{e}_i) = \boldsymbol{s}_i^2 = \boldsymbol{q}(\mathbf{x}_i)(1-\boldsymbol{q}(\mathbf{x}_i))$. Note that $\boldsymbol{q}$ represents a probability and is therefore restricted to the interval (0,1). We will assume in this case that the x_i's are fixed uniformly on a compact set C in $\Re$. We will also assume that $\boldsymbol{q}(\mathbf{x}_i)$ is continuous and bounded by constants B_1 and B_2 $(B_1, B_2 \in (0,1))$ on C. Although we express $\mathbf{x}$ as a vector, we do so to take into account any powers of the independent variable the user may wish to employ. It is assumed again that the regression in this chapter is univariate, and $\mathbf{x}$ represents $(1, x, x^2, \ldots, x^p)$.

Quantal regression is a particular form of generalized linear models (GLM)(using the likelihood function) put forth by Nelder and Wedderburn (1972), and McCullagh and Nelder (1988). Within this framework, the quantal linear model is one in which

$$\boldsymbol{q}(x) = F(\mathbf{x}_i^T \boldsymbol{b}).$$

At the same time logistic regression results if Y_i = 0 or 1 and

$$F(\mathbf{x}_i^T \boldsymbol{b}) = \frac{e^{\mathbf{x}_i^T \boldsymbol{b}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{b}}}.$$

Other forms of F are possible in quantal regression, such as the standard Gaussian cumulative distribution function (CDF) and the standard Weibull CDF. In this chapter we investigate both the parametric and nonparametric forms of logistic regression. We then obtain asymptotic results for a particular nonparametric estimate and a general parametric estimate. Finally, these results are combined (using results from chapter 3) to obtain asymptotic properties of an MRR1 logistic estimate.

Part 5a Quantal Regression Asymptotic Preliminaries

We begin by describing the local logistic regression procedure. Local logistic regression is a term employed by Nottingham, Birch, and Bodt (1999) to describe a specific form of nonparametric regression set forth by Fan, Heckman, and Wand (1995). We will use this specific form as our nonparametric estimate in our semiparametric mixture and address this estimate asymptotically.

The local logistic regression estimate is formed the same way most nonparametric estimates are; locally forming a low degree polynomial using weighting via a kernel function. Fan, Heckman, and Wand (1995) provide general asymptotic results from which we will draw. Since this procedure relies heavily on GLM theory we give a brief overview below as it pertains to logistic regression.

In the GLM we are trying to model the mean response corresponding to a specific independent vector, given by $q(\mathbf{x})$, the regression function, where the distribution of Y belongs to the Exponential family. If the error distribution is non-Gaussian, the user is often faced with estimating a nonlinear function q.

In GLM this is accomplished by modeling a transformation of a linear predictor of order p,

$$\boldsymbol{h}(\mathbf{x}) = \boldsymbol{b}_0 + \boldsymbol{b}_1 x + \ldots + \boldsymbol{b}_p x^p ,$$

and maximizing a likelihood function L with respect to $\boldsymbol{b}$ over the entire data set. The transformation from $\boldsymbol{h}$ to $\boldsymbol{q}$ is given by F. Thus,

$$\hat{\boldsymbol{q}}(\mathbf{x}) = F(\hat{\boldsymbol{h}}(\mathbf{x}; \hat{\boldsymbol{b}})) ,$$

at any specific $\mathbf{x}$, where $\hat{\boldsymbol{b}} = (\hat{\boldsymbol{b}}_0, \hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_p)^T$ maximizes

$$\prod_{i=1}^{n} (L(F(\mathbf{x}; \boldsymbol{b}), Y_i)) ,$$

where L is the likelihood function of Y_i. Fan, Heckman, and Wand (1995) suggest using the same procedure locally to produce a GLM type nonparametric estimate.

In local logistic regression we wish to estimate $\boldsymbol{h}(\mathbf{x})$ locally. The mean response is estimated through the response function F by

$$\hat{g}_{LL}(x) = \hat{\boldsymbol{q}}(x) = F(\hat{\boldsymbol{h}}(x; \hat{\boldsymbol{b}}_x)) ,$$

at any specific x. Here the subscript "LL" represents the local logistic nonparametric estimate. Note that the subscript x is used in the argument of F. This is because the estimate of the mean response in the nonparametric estimate, is germane to a particular x. Here, at each x, $\hat{\boldsymbol{b}}_x = (\hat{\boldsymbol{b}}_{0x}, \hat{\boldsymbol{b}}_{1x}, \ldots, \hat{\boldsymbol{b}}_{px})^T$ maximizes

$$\sum_{i=1}^{n} (\log L(F(\boldsymbol{b}_{0x} + \boldsymbol{b}_{1x}(x_i) + \ldots + \boldsymbol{b}_{px}(x_i)^p), Y_i) K_h(x_i - x))$$

where K_h is the usual kernel function defined in chapter 2, the additional subscripts on the $\boldsymbol{b}$'s indicate values of x for which they are germane, and $\log L$ is the log-likelihood function using the natural logarithm. $h = \boldsymbol{t}_n^{-1}$ represents the corresponding bandwidth.

In order to invoke the results of Fan, Heckman and Wand (1995) we need to meet certain conditions given in the Appendix of that article.

For the purpose of this chapter we will use the local logistic estimate having $\hat{h}(x)$ obtained by the method described above for the case p = 1 ("local linear logistic regression"). In addition, the estimate will have the following attributes. First, we will set the design density as uniform on the domain in question. That is

$$d(x) = \frac{1}{b-a} I_x(a,b)$$

for C = (a,b) with 0 $< a < b$, and C having finite length. In spite of the fact that the x_i's are fixed, we know that, at least asymptotically, $d(x)$ is apropos. This issue is discussed in more detail at the end of this section. Second, we will be using the Epanechnikov (1969) kernel (see Chapter 2). Third, we are not particularly interested in the derivatives of $h(\mathbf{x})$, so the derivative order is r = 0. Fourth, as we have already noted for this particular application we will have p = 1.

Fifth, we must assume that the conditional variance at a specific $\mathbf{x}$ is correctly specified (meaning the mean response is correctly specified for local logistic regression). Sixth, we will use the quantal regression canonical link:

$$h(x) = F^{-1}(q(x)) = \ln\left(\frac{q(x)}{1-q(x)}\right),\ q(x) \neq 1,$$

the logit function.

This estimate to a large extent coincides with the Nottingham, Birch, and Bodt(1999) local logistic estimate. We will confirm below that this estimate meets the conditions given in the appendix of Fan, Heckman and Wand (1995). The conditions are given first.

1. The canonical link is used, and the variance is correctly specified.

2. The functions $d'(x),\ h^{(p+2)},\ V(Y|X=x),\ V^{(2)},\ and\ g^{(3)}$, are all continuous on their respective domains.

3. For each x in the support of d, the following functions are nonzero:

$$\boldsymbol{r}(x) = \{g'(\boldsymbol{q}(x))^2 V(\boldsymbol{q}(x))\}^{-1},\ V(Y|X = x)\ \ and\ \ g'(\boldsymbol{q}(x)),$$

4. The kernel function $K(u)$ is a symmetric probability density function with support $(-1,1)$.

5. For each boundary value x_d on the support of d, there exists an interval C containing x_d, having nonnull interior, such that $\inf_{x \in C} d(x) > 0$.

Next observe that the following hold for $\hat{g}_{LL}(\mathbf{x})$,

LL1. The conditional variance will be correctly specified provided both the link function and the distribution are correct. The Logit link is used, and the distribution is clearly Bernoulli, so that the variance is correctly specified.

LL2. The functions

$$d'(x) = 0,$$

$$g'''(u) = F^{-1(3)}(u) = [(6u^2 - 6u + 2)/(u - u^2)^3],\ \ u \neq 0,1,$$

$$V(Y|X = x) = \boldsymbol{q}(x)(1 - \boldsymbol{q}(x)),\ \ \boldsymbol{q}(x) \neq 0,1$$

$$V''(u) = -2$$

$$\boldsymbol{h}'''(x) = 0$$

are all continuous on their respective domains.

LL3. For each $x \in (a,b)$, the functions

$$V(Y|X = x) = \boldsymbol{q}(x)(1 - \boldsymbol{q}(x)),\ \ \boldsymbol{q}(x) \neq 0,1,$$

$$\boldsymbol{r}(x) = g'(\boldsymbol{q}(x)) = F^{-1(1)}(\boldsymbol{q}(x)) = \frac{\P F(x)}{\P \boldsymbol{q}} = (\boldsymbol{q}(x)(1 - \boldsymbol{q}(x)))^{-1},\ \ \boldsymbol{q}(x) \neq 1,$$

are nonzero.

LL4. The Epanechnikov kernel function

$$K(u) = .75(1 - u^2) I_u(-1,1)$$

is a symmetric probability density function with support $(-1,1)$.

LL5. Since the support of $d(x)$ is (a,b), we can say that for any boundary value x_d, there exists an open interval C containing x_d such that $\inf_{x \in C} d(x) > 0$.

Having satisfied these conditions we are now able to present the following asymptotic result. This will be the only result presented for the nonparametric estimate at this time, since this is the estimate we are interested in for the MRR1 procedure in section 5b. Similar results may be obtained for other variations of the local logistic estimate using the same machinery.
The following result is obtained utilizing Theorem 2 (and Theorem 1a) of Fan, Heckman and Wand (1995). The proofs for this section are found in Appendix 5a.

Lemma 5.a.1 Assuming that $\boldsymbol{t}_n > \frac{2}{b-a}$, we have that for the local linear logistic regression estimate, $\hat{g}_{LL}$ (where as before this symbol represents the vector estimate)

$$\left\| \hat{g}_{LL} - \mathbf{q} \right\|^2 = O_P(\boldsymbol{t}_n n^{-1}) .$$

This result is similar (although at first glance it does not appear to be so) to the asymptotic result established in Fan and Gijbels (1996) (and first by Ruppert and Wand (1994)) for the local linear estimate) and to the tacit assumption of Burman and Chaudhuri (1992) for the nonparametric rate of convergence $O_P(g_n^2)$ in the proof of their Lemma 5.3. Indeed, throughout the remainder of this chapter the reader is encouraged to substitute the two representations to verify this fact and to compare asymptotic results across the chapters. Once again, we note that this rate of convergence is slower than the usual parametric rate of $O_P(n^{-.5})$. Having established the nonparametric rate of convergence we now proceed with finding a similar result for the parametric estimate.

Our focus will now be on the parametric part of the semiparametric estimate. We will begin by giving a general result and then illustrate that result using the usual GLM logistic regression and another example. In our general parametric estimate we will not require any specific form for the mean response as a function of the independent variable (as in GLM), except that f be a continuous function of both x and $\boldsymbol{\beta}$. Thus, we write

$$\theta(\mathbf{x}_i) = f(\mathbf{x}_i;\boldsymbol{\beta})$$

Notice that we do not require f to be linear in $\boldsymbol{\beta}$, so that we are now operating with the generalized non-linear model, which contains the generalized linear model. Additionally, we will make the following four requirements for the parametric estimate.

R1. The observed responses are independent (but not necessarily identically distributed), and are functions of x_i values that are fixed uniformly on C = (a,b) with $0 < a < b$, and C having finite length.

R2. $V(x_i)$ is bounded both from above and below.

R3. For every $n > p+1$, The matrix D is of full rank $(p+1)$ where D is given by

$$\begin{bmatrix} \frac{\partial f(x_1;\boldsymbol{\beta})}{\partial \beta_0} & \frac{\partial f(x_1;\boldsymbol{\beta})}{\partial \beta_1} & \cdots & \frac{\partial f(x_1;\boldsymbol{\beta})}{\partial \beta_p} \\ \frac{\partial f(x_2;\boldsymbol{\beta})}{\partial \beta_0} & \frac{\partial f(x_2;\boldsymbol{\beta})}{\partial \beta_1} & \cdots & \frac{\partial f(x_2;\boldsymbol{\beta})}{\partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(x_n;\boldsymbol{\beta})}{\partial \beta_0} & \frac{\partial f(x_n;\boldsymbol{\beta})}{\partial \beta_1} & \cdots & \frac{\partial f(x_n;\boldsymbol{\beta})}{\partial \beta_p} \end{bmatrix}$$

And by this, we mean that

$$\frac{\partial f(x;\boldsymbol{\beta})}{\partial \beta_i} = f_{\beta_i}(x;\boldsymbol{\beta}) = \sum_{j\neq i} c_j f_{\beta_j}(x;\boldsymbol{\beta}),\ for\ a \le x \le b$$

(for some $i \in \{0,1,\dots p\}$) implies that $c_j = 0,\ j = 0,1,\dots p,\ j \neq i$.

Note, that this also implies that there is no β_i (for some $i \in \{0,1,\dots p\}$) such that

$$\frac{\partial f(x;\boldsymbol{\beta})}{\partial \beta_i} = f_{\beta_i}(x;\boldsymbol{\beta}) \equiv 0,\ for\ a \le x \le b\,.$$

R4. For $i = 0,1,\ldots p, \dfrac{\partial f(x;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_i} = f_{\beta_i}(x;\boldsymbol{\beta})$ is continuous and bounded for both variables on C = (a,b) and $\Re^{p+1}$ respectively.

Next, we are ready to make some pertinent comments about a matrix which constitutes an important term that must be dealt with in the second lemma which gives us our parametric rate of convergence. Define

$$S_{Gn} = n^{-1}\sum_{i=1}^{n}\left(w_i \frac{\partial f(x_i;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \left(\frac{\partial f(x_i;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right)^T \right)$$

and

$$S_G = \lim_{n\to\infty} S_{Gn},$$

where w_i is a weighting constant related to $V(x_i)$. We are interested in whether or not this last matrix exists and is defined and finite (i.e. both $\det(S_G^{-1})$ and $\det(S_G)$ are nonzero and finite). The next lemma gives an important result. Once again, the proofs for this section can be found in Appendix 5a.

Lemma 5.a.2 Assuming requirements R1-R4 above S_{Gn} has a finite limit asymptotically; that is, S_G^{-1} is defined and finite, where $S_G^{-1} = (\lim_{n\to\infty} S_{Gn})^{-1}$.

The preceding lemma allows us to make a very strong asymptotic statement in the following result. Carroll and Ruppert (1988) and Bishop, Feinberg and Holland (1975) provide the machinery for Lemma 5.a.3, the capstone result for this section. The following lemma gives the convergence rate for almost any type of non-linear parametric regression estimate formed by using the IRLS iteration scheme (see Chapter 2). The proof is found in Appendix 5a.

Lemma 5.a.3 Assuming requirements R1-R4 above, then for the parametric regression estimate, $\hat{f}$, obtained through IRLS (with a $n^{.5}$ consistent starting estimate for $\boldsymbol{b}$),

$$\left\|\hat{f} - f\right\|^2 = O_P(n^{-1}).$$

Note that this is the same parametric rate of convergence obtained previously (since we are viewing $\left\|\hat{f} - f\right\|^2$). Now, however, it applies to a much larger spectrum of parametric regression estimates. As long as requirements R1-R4 are met, the user is free to concoct nearly any type of mean response estimate with parameters. The lemma also serves notice that the rate $O_P(n^{-.5})$ is more than likely, the best rate possible for any regression estimate (which would coincide with the result of the Central Limit Theorem).

We now present an example to show that parametric logistic regression in one regressor using only the first order term has this rate of convergence. Recall that we have already noted a number of the basics of logistic regression. Particularly,

$$\hat{\boldsymbol{q}}(\mathbf{x}) = F(\hat{\boldsymbol{b}}_0 + \hat{\boldsymbol{b}}_1 x)$$

where F is the standard logistic CDF. And for consistency with earlier work we write

$$\hat{\boldsymbol{q}}(\mathbf{x}_i) = \hat{f}(\mathbf{x}_i^T \hat{\boldsymbol{b}}) = \hat{f}(\hat{\boldsymbol{b}}_0 + \hat{\boldsymbol{b}}_1 x) = F(\hat{\boldsymbol{b}}_0 + \hat{\boldsymbol{b}}_1 x).$$

Using the standard logistic CDF we obtain

$$\hat{f}(\mathbf{x}_i^T \hat{\boldsymbol{b}}) = \frac{e^{\mathbf{x}_i^T \hat{\boldsymbol{b}}}}{(1 + e^{\mathbf{x}_i^T \hat{\boldsymbol{b}}})}.$$

Computationally, we cannot obtain the estimated mean response in the usual least squares fashion. We have noted that this particular estimate falls under the category of GLM.
In addition to having a function of the linear predictor, and having a non-Gaussian error distribution (shifted Bernoulli), we also have a non-constant variance due to the variation in the mean response. Fortunately, getting the GLM estimate (as mentioned in Chapter 2) can be done using Generalized Least Squares (GLS), and in particular IRLS. This is also the case in getting

the nonparametric estimate mentioned earlier. Note also the following qualities of this estimate.

1. The x_i values are fixed uniformly on C = (a,b) with 0 $< a < b$, and C having finite length. Also for Lemma 5.a.2 we will require that $x_i \neq x_j$ for some i, j.

2. In dealing with the weights (w_i) for the nonconstant variance in the IRLS procedure, we will assume that $\mathbf{x}_i^T \boldsymbol{b}$ is finite which implies that $\hat{w}_i > 0$, where $\hat{w}_i = (\hat{F}_i(1-\hat{F}_i))$ is an estimate of the variance of response at x_i. We will also assume (in conjunction with quality 1) that on C,

$$0 < f_L < f < f_U < 1,$$

where f_L, and f_U represent the lower and upper bounds of f, respectively.

3. The estimate $\hat{f}$ is obtained through IRLS using a LS starting estimate for $\boldsymbol{b}$.

4. We will also assume that the Y_i's are independently distributed.

With these notes, we continue the present example. The Logistic GLM estimate meets requirements R1-R4 in the following way. R1 and R2 are satisfied by qualities 4 and 1 respectively. R3 is met by observing that

$$f_{b_0}(x_i;\boldsymbol{b}) = \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{b}}}{(1+e^{\mathbf{x}_i^T \boldsymbol{b}})^2}\right),\ f_{b_1}(x_i;\boldsymbol{b}) = \left(\frac{x_i e^{\mathbf{x}_i^T \boldsymbol{b}}}{(1+e^{\mathbf{x}_i^T \boldsymbol{b}})^2}\right)$$

(where the subscripts indicate the variable for which the partial derivative is taken), so that

$$x_i f_{b_0}(x_i;\boldsymbol{b}) = f_{b_1}(x_i;\boldsymbol{b}),\ \forall\ a < x_i < b\,.$$

Clearly, one function is not a linear combination of the other. R4 is satisfied by observing that $f_{b_0}(x;\boldsymbol{b})$ and $f_{b_1}(x;\boldsymbol{b})$ are both continuous and bounded for both variables on C = (a,b) and $\Re^2$ respectively (for $\Re^2$ note that 0 < $f_{b_0}(x;\boldsymbol{b})$ < 1, and that 0 < $f_{b_1}(x;\boldsymbol{b})$ < b).

So by Lemma 5.a.3

$$\left\|\hat{f} - f\right\|^2 = O_P(n^{-1}),$$

as desired.[].

Comments

Some comments are in order. We will begin by addressing a couple of technical issues. Observe that we took a deterministic approach to the values of x in earlier chapters, but now we are using a uniform distribution. This is due, in large part, to condition requirements placed by Fan, Heckman and Wand (1995) on the LLR estimate. But there are other reasons to consider. Observe that asymptotically, a uniform distribution for x, and deterministically spacing the values uniformly over C is the same. So that for the asymptotic results presented in this context, it doesn't matter which way the x values are dispersed. There is a very fine line between the terms deterministic and stochastic. In fact, many would claim that there is no difference philosophically, and that setting out x values uniformly is essentially the same as having them set out by "Exterior Influences". Myers (1990) demonstrates that at least for simple linear regression, the same estimate is obtained in both cases. Finally, however, Burman and Chaudhuri (1992), state that all results can be generalized from deterministic to stochastic predictor values, but at the cost of more mathematical technicalities. So the results in this chapter are actually more generalized (in terms of the dispersion of the independent variable) than those of previous chapters. And, as we shall see, this enables us to more freely apply the techniques of those chapters to obtain asymptotic results for the MRR1 quantal regression procedure.

Observe also that we require use of the Epanechnikov kernel, primarily resulting from condition LL4 in Fan, Heckman, and Wand (1995), where they require a kernel that is a symmetric probability density with support (-1,1).

We have confirmed with Fan (personal communication), however, that this condition can be weakened (with some additional effort in the proof) to include symmetric kernels that are supported on $\Re$. This would allow for the Gaussian kernel to be substituted for the Epanechnikov kernel in Lemma 5.a.1.

With these results we can obtain asymptotic results for the MRR1 quantal estimate.