Part 2c A Review of Model Robust Regression

In this section we shall look briefly at the techniques called model robust regression (MRR). MRR1 was developed by Einsporn (1987) and Einsporn and Birch (1993) to address the shortcomings of parametric and nonparametric regression. The MRR1 technique allows the user to vary the weights on the parametric and nonparametric portions via a mixing parameter. The MRR1 estimate is given by

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

where $l \in [0,1]$, and $\hat{f}, \hat{g}$ represent the parametric and nonparametric estimates respectively. Originally the estimate had the mixing parameter as a coefficient of the nonparametric fit, but we have reversed this notation for now for purposes of comparison (with the Burman and Chaudhuri (1992) results). The key idea is the introduction of l, which ranges from 0 to 1 according to the amount of misspecification in the user's parametric estimate. MRR1 was originally entitled "Hatlink" because it was constructed in terms of hat (or projection) matrices. Olkin and Spiegelman (1987) introduced the same technique for density estimation, and provided asymptotic results that are germane to this dissertation.

The approach in MRR1 is to construct a parametric and nonparametric estimate separately and then combine them using the mixing parameter. Unfortunately, this produces one drawback. If there are locations in the data where both estimates give fitted values too high or too low, then the MRR1 estimate cannot compensate for the error, since any combination using one or both of the parametric/nonparametric estimates will also be too high or too low.

MRR2 was developed by Mays (1995) to address shortcomings in the MRR1 and partial linear regression (PLR) (Speckman (1988)) estimates. In particular, the PLR estimate always passes through the origin on a 2-dimensional graph, and always includes a nonparametric portion even when the parametric model is correct.

The MRR2 estimate is given by

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

where $l \in [0,1]$, and $\hat{f}, \hat{g}$ represent the parametric and nonparametric estimates respectively. The estimates are similar in their simplicity and use of l. However, the approach of the MRR2 estimate differs somewhat from that of MRR1. In this case, the parametric estimate is formed first from the original data. Obtaining the residuals from the parametric fit, the user then forms a new data set consisting of the original $\mathbf{x}$ values and a new response variable, realized by the residuals. So the new ordered pairs appear as $(\mathbf{x}_i, r_i)$ for $i = 1$ to n. A nonparametric estimate is then developed from this data and added to the parametric estimate according to the value of l. Additionally, notice that the MRR2 estimate allows for a maximum contribution of 50% from the nonparametric estimate. Although the case $l = 1$ represents 100% use of the residual fit, in reality the contribution is only 50% since the contribution from the parametric fit always has a coefficient of one. In this way, the parametric estimate always plays an important role in the MRR2 estimate so that if, in fact, the true function q has a parametric form in the family of forms under consideration, it can be captured by MRR2. If q has no functional form (or has a parametric form outside the family under consideration) the nonparametric estimate can still capture this; it is just constructed from a different baseline (namely the parametric estimate). Mays (1995) went on to demonstrate that the MRR2 estimate was superior theoretically to both of the previous estimates in that the user was free to use any parametric and nonparametric estimate, could eliminate the nonparametric contribution if necessary, and could successfully circumvent the MRR1 fitting problem (mentioned above) since the nonparametric fit would be based on the residuals. These improvements have merit, as indicated by the success of the MRR2 estimate in a number of small sample simulations (Mays (1995)).

We have not discussed to this point (for either estimate) the selection process for λ .

As with the bandwidth problem in nonparametric regression, the selection of λ has been pondered extensively by the developers of Model Robust Regression. Essentially, they felt that certain criteria might result in λ being chosen too large or too small. Specifically, Einsporn (1987) developed several criteria used to select λ , namely PRESS* and four methods based on Mallows's C_p statistic. PRESS* and the third C_p criterion proved to be the best of these.

PRESS* was developed as an alternative to the PRESS statistic, to protect against the tendency of PRESS to select λ favoring the nonparametric estimate (overfitting). The form for PRESS* is given by

$$\text{PRESS*}(\lambda) = \frac{\sum_{i=1}^{n}(Y_i - Y_{i,-i}^{MRR})^2}{n - tr(H^{MRR})}$$

where $Y_{i,-i}^{MRR}$ is the MRR estimate (either MRR1 or MRR2) of the mean response at $\mathbf{x}_i$ obtained using the data set with the pair ($\mathbf{x}_i$, Y_i) removed, and H^{MRR} is the MRR hat (or projection) matrix. Mays (1995), in order to protect against the PRESS* criterion selecting λ favoring the parametric estimate (underfitting), developed PRESS**. Its form is given by

$$\text{PRESS**}(\lambda) = \frac{\sum_{i=1}^{n}(Y_i - Y_{i,-i}^{MRR})^2}{n - tr(H^{MRR}) + (n-1)\dfrac{SSE_{MAX} - SSE_{\lambda}}{SSE_{MAX}}}$$

where SSE_{MAX} is the maximum SSE obtained for a value of λ (λ = 1 for MRR1 and λ = 0 for MRR2), and SSE_{λ} is the specific SSE for a given λ . Among these selectors Mays (1995) determined that PRESS** generally provided the best results for each MRR estimate at small samples.

In subsequent chapters we will develop asymptotic results for the MRR1 estimate using PRESS, and PRESS* as selectors for l, and for MRR2 using PRESS as a selector for l. In addition we will investigate the asymptotic properties of MRR estimates using other data driven l selectors that are asymptotically optimal in the non-cross validated estimate. The results will be discussed in those chapters.