Chapter 4 : Dual Model Robust Regression (DMRR)

The goal of this research is to provide a procedure for estimating a dual model which incorporates some (parametric) knowledge of the underlying models, and is flexible enough to capture trends in the data which deviate from the specified parametric forms. Such a procedure should be capable of handling cases ranging from one or both of the specified parametric models being the true underlying models, to cases in which one or both of the specified models is / are inappropriate. The approach developed here provides a robust generalized least squares (GLS) algorithm which combines a robust means fit with a robust variance fit. Variance estimation will be residual-based where the residuals from the robust means fit serve as building blocks for the variance model. The proposed methodology will be detailed in this chapter and comparisons, based on MSE criteria, of the proposed methodology with other dual modeling techniques will be presented in Chapter 5.

4.A The Model

As mentioned, the current research seeks to obtain a robust dual modeling procedure that has two main capabilities: first, the method should be able to incorporate the user's parametric knowledge of the process and second, the method should remain flexible enough to capture trends in the data where parametric model(s) are inadequate. Thus, it is useful to think of the process mean and variance as functions which can be expressed as two components: a user supplied parametric component and a "lack of fit" component which represents the portion of the process mean and variance which cannot be captured parametrically. We write the underlying dual model as follows:

Means Model :
$$\begin{aligned} y_i &= h(\mathbf{x}_i) + g^{1/2}(\mathbf{z}_i)\varepsilon_i \qquad (4.A.1.1)\\ &= m(\mathbf{x}_i;\boldsymbol{\beta}) + f(\mathbf{x}_i) + g^{1/2}(\mathbf{z}_i)\varepsilon_i \end{aligned}$$

Variance Model :
$$\begin{aligned} \sigma_i^2 &= g(\mathbf{z}_i) \qquad (4.A.1.2)\\ &= v(\mathbf{z}_i;\boldsymbol{\theta}) + l(\mathbf{z}_i) \end{aligned}$$

Recall that, for purposes of this research, we assume a one-regressor dual model, implying that $\mathbf{x}_i = (1\ x_i\ x_i^2\ \cdots)' = \mathbf{z}_i = (1\ z_i\ z_i^2\ \cdots)'$. Regarding notation, the user-specified parametric components are $m(\mathbf{x}_i;\boldsymbol{\beta})$ and $v(\mathbf{z}_i;\boldsymbol{\theta})$ for the mean and variance models respectively and the lack of fit components are given by $f(\mathbf{x}_i)$ in the means model, and $l(\mathbf{z}_i)$ in the variance model. In this research, we limit ourselves to discussion of the family of dual models in which

$$m(\mathbf{x}_i;\boldsymbol{\beta}) = \mathbf{x}_i^{'}\boldsymbol{\beta} \qquad (4.A.1.3)$$

and

$$v(\mathbf{z}_i;\boldsymbol{\theta}) = \exp\left\{\mathbf{z}_i^{'}\boldsymbol{\theta}\right\}. \qquad (4.A.1.4)$$

However, the results of this research may be easily extended to the case in which $m(\mathbf{x}_i;\boldsymbol{\beta})$ and $v(\mathbf{z}_i;\boldsymbol{\theta})$ are arbitrary functions.

Since the mean and variance models can be thought of as being composed of both a parametric portion and a nonparametric, lack of fit portion, the approach here is to estimate each individual model (mean and variance) by combining parametric and nonparametric estimation procedures. Estimation of the models will occur within the context of an iterative, generalized least squares algorithm. In the next section, model robust means estimation is discussed and this will be followed by a section on model robust variance estimation.

4.B Means Model Robust Regression (MMRR)

The *means model robust regression* (MMRR) procedure proposed here is a simple extension of the MRR procedure developed by Mays and Birch (1995). Recall from Section 2.D that their approach was to first obtain an ordinary least squares (parametric) fit to the data, and then to augment this with a portion of the local linear (nonparametric) fit to the residuals. The MRR estimate was given as:

$$\hat{\mathbf{y}}^{(\mathrm{mrr})} = \hat{\mathbf{y}}^{(\mathrm{ols})} + \lambda\mathbf{H}^{(\mathrm{llr})}\left(\mathbf{y} - \hat{\mathbf{y}}^{(\mathrm{ols})}\right) \qquad (4.B.1)$$

where $\lambda \in [0,1]$, $\mathbf{H}^{(\mathrm{llr})}$ is the local linear hat matrix used to fit the means model residuals, and $\hat{\mathbf{y}}^{(\mathrm{ols})}$ represents the *n*x1 vector of ordinary least squares estimates of the mean. This estimate of the mean proposed by Mays and Birch (1995) can be extended to the dual modeling setting simply by accounting for heterogeneity of variance in the raw data. This can be done by replacing the ordinary least squares estimate with an estimated weighted least squares estimate. The means model robust estimate is then given as:

$$\hat{\mathbf{y}}^{(\mathrm{mmrr})} = \hat{\mathbf{y}}^{(\mathrm{ewls})} + \lambda_\mu\mathbf{H}_{b_\mu}^{(\mathrm{llr})}\left(\mathbf{y} - \hat{\mathbf{y}}^{(\mathrm{ewls})}\right) \qquad (4.B.2)$$
$$\text{where } \lambda_\mu \in [0,1]$$

$$= \left[\mathbf{H}^{(\text{ewls})}\mathbf{y} + \lambda_\mu \mathbf{H}_{b_\mu}^{(\text{llr})}\left(\mathbf{I} - \mathbf{H}^{(\text{ewls})}\right)\right]\mathbf{y}$$

$$\text{where } \mathbf{H}^{(\text{ewls})} = \mathbf{X}\left(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}$$

$$\text{and } \hat{\mathbf{V}} = \text{diag}\left\langle \hat{\sigma}_1^2, \cdots, \hat{\sigma}_n^2 \right\rangle$$

$$= \mathbf{H}^{(\text{mmrr})}\mathbf{y} \tag{4.B.3}$$

where

$$\mathbf{H}^{(\text{mmrr})} = \mathbf{H}^{(\text{ewls})} + \lambda_\mu \mathbf{H}_{b_\mu}^{(\text{llr})}\left(\mathbf{I} - \mathbf{H}^{(\text{ewls})}\right) \tag{4.B.4}$$

is the MMRR "hat matrix". The individual fitted observations are obtained as

$$\hat{y}_i^{(\text{mmrr})} = \sum_{j=1}^{n} h_{ij}^{(\text{mmrr})} y_j . \tag{4.B.5}$$

Before moving on to discuss robust variance estimation it is important to make a few remarks regarding the notation above. First, the estimated variances that compose the diagonal of the matrix $\hat{\mathbf{V}}$ will be obtained from the robust variance estimation which is discussed in the next section. Second, the mixing parameter λ_μ serves the same purpose as that of λ from the MRR procedure in that it increases from 0 to 1 as the amount of model misspecification in the parametric portion increases. The reason for the subscript "μ" is to distinguish this mixing parameter from the one that will be used in the robust variance estimate. The subscript "b_μ" used with the local linear "hat matrix" represents the bandwidth that is used for smoothing the residuals from the parametric means fit.

4.C Variance Model Robust Regression (VMRR)

Recall from the introduction to this chapter that the dual model robust procedure proposed in this research is one that involves not only robust means fit but also one that provides a robust, residual-based, variance estimate. In chapter 3 it was noted that the success of residual-based variance estimation hinges on how well one is able to estimate the mean. For instance, in parametric dual modeling, if the specified means model is insufficient, then the residuals not only contain information regarding process variance but they are also "contaminated" with lack of fit. The result is an estimate of variance that could contain substantial bias. At the other extreme is a result that is often a consequence of nonparametric estimation of the mean. In many situations, the nonparametric estimate can fit the data too closely, thus leaving meager residuals with which to model the variance. This scenario also results in bias problems for the variance model as the model is "underfit". Compounding these two potential problems is the fact that the user still must decide whether to specify a parametric form for the variance and estimate the variance parametrically or to adopt a purely nonparametric form of estimation. The former choice runs the risk of model misspecification and the latter choice runs the risk of obtaining an estimate of variance that is too variable. Thus, in offering a robust variance estimation procedure it is important to first provide data for variance estimation which is robust to means misspecification and second, to offer a method of estimating the variance function which is robust to functional misspecification.

The problem of obtaining "robust data" for variance estimation is a simple problem to solve. Since MMRR is intended to offer an estimate of the mean which is robust to model misspecification, it seems natural to use the squared MMRR residuals as variance model data. The residuals from the MMRR estimate are given as

$$\mathbf{e}^{(\text{mmrr})} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(\text{ewls})} - \lambda_{\mu}\, \mathbf{H}_{b_{\mu}}^{(\text{llr})}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(\text{ewls})}\right) \qquad (4.\text{C}.1)$$

Notice that if there is no misspecification of the means model then λ_{μ} should be close to zero and the MMRR residuals will simply be the parametric residuals. However, if there is model misspecification, λ_{μ} should be closer to 1 to allow for a proportion of the extra means structure to be captured. Now that a source of data has been proposed for the variance model, it is necessary to provide a procedure of variance estimation which is robust to variance function misspecification.

Like the model robust estimate used in the means model, the variance model robust regression (VMRR) procedure is an extension of the MRR procedure of Mays and Birch (1996). Recall that MRR involves combining parametric and nonparametric estimates of the raw data to form the model robust estimate. The proposed VMRR procedure works similarly but the raw data, as mentioned above, is taken to be the residuals from the MMRR estimate.

The parametric portion of VMRR is obtained as a parametric fit to the squared, robust residuals from the means fit. This parametric fit mimics the parametric method discussed in Section 3.B.1 in which Chi-Square regression was used to fit the model

$$\mathbf{e}^{2(\text{ewls})} = v\{\mathbf{Z};\boldsymbol{\theta}\} + \boldsymbol{\eta}. \qquad (4.\text{C}.2)$$

Since we take $v(\mathbf{Z};\boldsymbol{\theta})$ to be given as

$$v(\mathbf{Z};\boldsymbol{\theta}) = \exp\{\mathbf{Z}\boldsymbol{\theta}\}\ , \qquad (4.\text{C}.3)$$

the parametric estimate in VMRR involves a Chi-Square analysis of the model

$$\mathbf{e}^{2(\text{mmrr})} = \exp\left\{ \mathbf{Z\theta} \right\} + \boldsymbol{\eta}. \qquad (4.C.4)$$

It is useful to note that the model written in 4.C.4 falls under the umbrella of generalized linear models discussed by McCullagh and Nelder (1994). The criterion for classification as a generalized linear model is that the data's error distribution must be one that is a member of the exponential family and there must exist some function of the data which is linear in the model parameters (called the link function). In (4.C.4) we assume that the error distribution of the squared robust residuals is Chi-Square and the link function is the logarithmic function. Parameter estimation of the model in (4.C.4) is accomplished by weighted least squares, the details of which are provided in Appendix A.2. From Appendix A.2, the parametric variance estimate is written as

$$\hat{\boldsymbol{\sigma}}^{2\,(\text{glm})} = \exp\left\{\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{glm})}\right\} \qquad (4.C.5)$$

where

$$\hat{\boldsymbol{\theta}}^{(\text{glm})} \approx \boldsymbol{\theta}^{(\text{glm})} + \left(\mathbf{Z}'\boldsymbol{\Delta}\mathbf{V}_{\text{e}}^{-1}\boldsymbol{\Delta}\mathbf{Z}\right)^{-1}\mathbf{Z}'\boldsymbol{\Delta}\mathbf{V}_{\text{e}}^{-1}\left(\mathbf{e}^{2(\text{mmrr})} - \exp\left\{\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\right\}\right). \qquad (4.C.6)$$

Regarding notation, $\hat{\boldsymbol{\theta}}^{(\text{glm})}$ is the estimated vector of parameters from the Chi-Square analysis of the model given in (4.C.4), $\boldsymbol{\theta}^{(\text{glm})}$ denotes the vector of converged values for $\hat{\boldsymbol{\theta}}^{(\text{glm})}$ in the nonlinear analysis of (4.C.4), $\boldsymbol{\Delta} = \text{diag}\left\langle \left[\left.\frac{\partial \exp\left(\mathbf{z}_i'\hat{\boldsymbol{\theta}}^{(\text{glm})}\right)}{\partial\left(\mathbf{z}_i'\hat{\boldsymbol{\theta}}^{(\text{glm})}\right)}\right|_{\hat{\boldsymbol{\theta}}^{(\text{glm})} = \boldsymbol{\theta}^{(\text{glm})}} \right] \right\rangle$, $\mathbf{V}_{\text{e}}$ is a diagonal matrix comprised of the variances of the squared MMRR residuals, and $\mathbf{e}^{2\,(\text{mmrr})}$ is the nx1 vector of squared MMRR residuals.

The nonparametric portion of the VMRR estimate involves a local linear fit to the residuals from the parametric variance fit. Denoting these residuals as $\mathbf{r}_{\boldsymbol{\sigma}} = \mathbf{e}^{2(\text{mmrr})} - \exp\left\{\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{glm})}\right\}$, the local linear fit may be expressed as $\hat{\mathbf{r}}_{\sigma} = \mathbf{H}_{b_\sigma}^{(\text{llr})}\mathbf{r}_{\sigma}$. The subscript "$b_\sigma$" used with $\mathbf{H}_{b_\sigma}^{(\text{llr})}$ represents the bandwidth that is used for smoothing the residuals from the parametric variance fit.

The VMRR variance estimate is now given as

$$\hat{\boldsymbol{\sigma}}^{2\,(\text{vmrr})} = \exp\left\{\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{glm})}\right\} + \lambda_\sigma \mathbf{H}_{b_\sigma}^{(\text{llr})}\left(\mathbf{e}^{2(\text{mmrr})} - \exp\left\{\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{glm})}\right\}\right) \qquad (4.C.7)$$

where $\lambda_\sigma \in [0,1]$ and . The individual fitted observations are obtained as

$$\hat{\sigma}_i^{2\,(vmrr)} = \exp\left\{\mathbf{z}_i'\hat{\boldsymbol{\theta}}^{(glm)}\right\} + \lambda_\sigma \mathbf{h}_{i_{b_\sigma}}^{'\,(llr)}\left(\mathbf{e}^{2(mmrr)} - \exp\left\{\mathbf{Z}\hat{\boldsymbol{\theta}}^{(glm)}\right\}\right) \qquad (4.C.8)$$

where $\mathbf{h}_{i_{b_\sigma}}^{'\,(llr)}$ is the i^{th} row of the matrix $\mathbf{H}_{b_\sigma}^{(llr)}$.

4.D Dual Model Robust Regression (DMRR)

Recall from the development of MMRR that the parametric portion of MMRR required an estimate of the variances at the *n* data points. In the development of VMRR, the data was taken to be the squared residuals from the robust means fit. Since the mean and variance estimates are interdependent, it seems logical that estimation of the dual model should take place within a single, iterative algorithm. The proposed methodology, Dual Model Robust Regression (DMRR), combines the robust means fit (MMRR) with the robust variance fit (VMRR) in a single, generalized least squares algorithm. The model robust algorithm follows the same type of outline as the algorithm used in parametric dual modeling (discussed in Chapter 3) and proceeds as follows:

1. Let $\hat{V} = I$ where $\hat{V} = \operatorname{diag}\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \cdots, \hat{\sigma}_n^2\right)$.
2. Using weighted least squares, obtain the parametric estimate of the means model :
   $\hat{y}_i^{(ewls)} = \mathbf{x}_i'\hat{\boldsymbol{\beta}}^{(ewls)} = \mathbf{x}_i'\left(X'\hat{V}^{-1}X\right)^{-1}X'\hat{V}^{-1}\mathbf{y}$
3. Form the residuals from the fit found in Step 2 , $e_i^{(ewls)} = (y_i - \hat{y}_i^{(ewls)})$, and perform local linear regression on this set of residuals, obtaining $\hat{r}_{i_\mu} = \mathbf{h}_{i_{b_\mu}}^{'\,(llr)}\mathbf{e}^{(ewls)}$, where $\mathbf{h}_{i_{b_\mu}}^{'\,(llr)}$ is the i^{th} row of the local linear hat matrix and $\mathbf{e}^{(ewls)}$ is the *n* x 1 vector of EWLS residuals.
4. Obtain the MMRR fit to the means model, written as : $\hat{y}_i^{(mmrr)} = \hat{y}_i^{(ewls)} + \lambda_\mu \hat{r}_{i_\mu}$.
5. Form the squared residuals from the MMRR fit to the mean, obtaining : $e_i^{2\,(mmrr)} = \left(y_i - \hat{y}_i^{(mmrr)}\right)^2$.
6. Obtain the parametric estimate of variance by estimating the parameters in the regression model : $\hat{\sigma}_i^{2\,(glm)} = \exp\left\{\mathbf{z}_i'\hat{\boldsymbol{\theta}}^{(glm)}\right\}$.

7. Form the residuals from the parametric fit to the variance model, $r_{i_\sigma} = \left(e_i^{(mmrr)} - \exp\left\{ \mathbf{z}_i' \hat{\boldsymbol{\theta}}^{(glm)} \right\} \right)$, and perform local linear regression on this set of variance residuals, obtaining $\hat{r}_{i_\sigma} = \mathbf{h}_{i_{b_\sigma}}^{'(llr)} \mathbf{r}_\sigma$, where $\mathbf{h}_{i_{b_\sigma}}^{'(llr)}$ is the i^{th} row of the local linear hat matrix used to fit the *n* x 1 vector of parametric variance residuals, $\mathbf{r}_\sigma$.
8. Obtain the VMRR estimates of variance which are written as: $\hat{\sigma}_i^{2\,(VMRR)} = \hat{\sigma}_\mathrm{i}^{2\,(glm)} + \lambda_\sigma \hat{r}_{i_\sigma}$ where $\lambda_\sigma \in [0,1]$ is the variance model mixing parameter.
9. Return to step 2 with $\hat{\sigma}_i^{2\,(vmrr)}$.
10. Cycle through steps 2 - 9 until convergence of the means model parameters.

4.E Advantages

The purpose of this section is to summarize the different methods of dual modeling which have been discussed thus far and to express the need for a model robust dual modeling approach such as DMRR. Throughout this chapter it has been pointed out that in dual modeling, the mean and variance estimates are interdependent. This is especially true in parametric dual modeling where the means estimate depends on weights provided by the variance estimate and the data for the variance model is the set of residuals from the means fit. As long as the underlying mean and variance functions are adequately described by the user's prescribed models, parametric dual modeling is very effective. However, any amount of trend in the data that is not captured by the specified functions can have serious implications on the analysis. For instance, if the specified parametric model for the mean is inadequate for even part of the data, the variation in the means model residuals cannot be assumed to be due to process variance alone. Instead, part of the variation is due to "lack of fit" from the means model. As a result of this "lack of fit" contamination, the variance estimate becomes distorted.

Another type of misspecification that can adversely affect dual model analysis is misspecification of the functional form of the underlying process variance. Misspecification of the variance function does not result in bias problems for the means estimate but inference regarding the estimated mean will be affected.

When the researcher assumes no parametric knowledge regarding the process mean and/or variance functions, nonparametric dual modeling procedures are used. Strictly nonparametric methods include the nonparametric residual-based method of (Hall and Carroll (1988)) and the difference-based method of (Müller and Stadtmüller (1987)). These methods have been shown to be effective in capturing the trend of the data but they often result in estimates which capture the trend too closely and as a result, are more variable than what is necessary.

In many data sets, the researcher is confident in the structure of the means function but the form of the variance function cannot be specified. Solutions to this type of scenario have been proposed by Ruppert and Carroll (1988) and Müller and Stadtmüller (1987 and 1993). Both propose estimated weighted least squares analysis of the mean but a nonparametric variance estimate. Ruppert and Carroll suggest estimating the variance function via a smoothing technique on the squared, means model residuals. Müller and Stadtmüller estimate the variance via a smoothing technique on the squared, pseudo-residuals.

It is clear from the discussion thus far that current dual modeling techniques have been developed under the following assumption: the researcher either has *complete* confidence in the forms of one or both of the specified models or the researcher has *no* confidence in a specification of one or both of the models. This research contends that there are many cases in which the researcher's state of knowledge is not binary. Rather, the researcher is confident that a specified parametric function may be appropriate across part of the data but realizes that there may be certain trends in the data which cannot be captured parametrically. Instead of forcing the researcher to make an extreme declaration such as *no* parametric knowledge or *full* parametric knowledge, DMRR seeks to utilize as much of the researcher's parametric knowledge as possible while still allowing for specific deviations in the data to be captured. The flexibility that DMRR affords to the researcher is due to the mixing parameters λ_μ and λ_σ. By varying the values of λ_μ and λ_σ, DMRR can successfully accommodate states of nature ranging from both models (mean and variance) being correctly specified, to processes in which only one model is correctly specified, to situations where neither model has been correctly specified. The next chapter develops MSE criteria which will be used as a basis for theoretical comparisons among the various dual modeling procedures. In Chapters 6 and 7, several examples will be discussed which illustrate the performance of the dual modeling techniques under different degrees of mean and/or variance model misspecification.