

Chapter 5: Theoretical Comparisons

The previous two chapters have described various dual model regression techniques for fitting a set of data. The purpose of this chapter is to provide the MSE criteria which will be used to compare the various dual modeling procedures. This chapter will be organized into two sections: the first section will motivate and provide the general set-up of the underlying dual model and the section to follow provides the MSE criteria which are based on the general framework.

5.A Underlying Dual Model

Before providing the MSE criteria, it is necessary to state the general, underlying dual model from which the criteria are derived. The general form of the underlying dual model must be such that any data set can be generated from it. Since the methods being compared involve procedures ranging from strictly parametric estimation, to mixing parametric and nonparametric estimates, to purely nonparametric estimation, it seems logical to think of the underlying dual model as a combination of parametric and nonparametric functions in both the mean and variance models. Such a model was given in Section 4.A and is written now in matrix notation as follows:

$$\begin{aligned} \text{Means Model :} \quad \mathbf{y} &= \mathbf{h}(\mathbf{x}) + \mathbf{g}^{1/2}(\mathbf{z})\boldsymbol{\varepsilon} & (5.A.1) \\ &= \mathbf{m}(\mathbf{X}; \boldsymbol{\beta}) + \mathbf{f} + \mathbf{g}^{1/2}(\mathbf{Z})\boldsymbol{\varepsilon} \end{aligned}$$

$$\begin{aligned} \text{Variance Model :} \quad \boldsymbol{\sigma}^2 &= \mathbf{g}(\mathbf{z}) & (5.A.2) \\ &= \mathbf{v}(\mathbf{Z}; \boldsymbol{\theta}) + \mathbf{1} \end{aligned}$$

where $\mathbf{h}(\mathbf{x}) = [h(x_1), \dots, h(x_n)]'$ and $\mathbf{v}(\mathbf{z}) = [v(z_1), \dots, v(z_n)]'$, with h and v as *general* regressions functions. For purposes of this research we consider $\mathbf{m}(\mathbf{X}; \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{v}(\mathbf{Z}; \boldsymbol{\theta}) = \exp\{\mathbf{Z}\boldsymbol{\theta}\}$. The matrix \mathbf{X} is the $n \times (k+1)$ matrix of mean regressors (augmented with a column of ones), \mathbf{Z} is taken to be the $n \times (q+1)$ matrix of variance regressors (augmented with a column of ones), $\mathbf{f} = [f(x_1), \dots, f(x_n)]'$ and $\mathbf{1} = [l(z_1), \dots, l(z_n)]'$. The two functions, f and l , are considered to be unknown (smooth) regression functions. The means model parameters are given by the $(k+1) \times 1$ vector $\boldsymbol{\beta}$ and the variance model parameters are given by the $(q+1) \times 1$ vector $\boldsymbol{\theta}$. Notice the flexibility that the dual model given in (5.A.1) and (5.A.2) affords the researcher. By assuming that $\mathbf{m}(\mathbf{X}; \boldsymbol{\beta}) = \mathbf{v}(\mathbf{Z}; \boldsymbol{\theta}) = \mathbf{0}$, the researcher assumes no parametric knowledge of the underlying process and the analysis is conducted from a nonparametric standpoint. If the researcher assumes $\mathbf{f} = \mathbf{1} = \mathbf{0}$, then the researcher proceeds from a purely parametric standpoint. In DMRR, the approach is to assume that any underlying

function can be decomposed into a parametric portion and a “remainder” portion. Applying this point of view to the models given above, the $n \times 1$ vectors \mathbf{f} and \mathbf{l} can be thought of as “remainder” portions from specified parametric models, since $\mathbf{f} = \mathbf{h}(\mathbf{X}) - \mathbf{m}(\mathbf{X}; \boldsymbol{\beta})$ and $\mathbf{l} = \mathbf{g}(\mathbf{Z}) - \mathbf{v}(\mathbf{Z}; \boldsymbol{\theta})$. Given the general form of the underlying dual model, the next section provides the bias and variance expressions for the mean and variance estimates. From these expressions, one can obtain the MSE criteria which will be used as the basis for procedural comparisons.

5.B MSE Criteria

The bias, variance and MSE expressions, for each of the dual modeling techniques being compared have been derived in Appendix B. It is important to note that these expressions are asymptotic and thus, for small samples, they are approximations to the true expressions. In comparing dual modeling procedures, there are several issues that the researcher may be interested in. First, since there are two phenomena being estimated (the mean and variance), there are two separate criteria of interest: the *means* MSE and the *variance* MSE. Depending on the practical situation, one of these criteria may be of more interest to the researcher than the other. In this research, we will make comparisons on the premise that the mean and variance estimates are of equal importance to the researcher.

The researcher must also decide if comparisons among the procedures should be based on a summary of individual data points or in comparing entire fits. If concerned only with individual points, the expressions provided can be used to calculate both the *means* MSE and the *variance* MSE at each of the n data points. Comparisons among the procedures could then be done by summarizing the n MSE's in the means model and n MSE's in the variance model in terms of the average means MSE (AVEMMSE) and average variance MSE (AVEVMSE). In this research, these two numbers will not be utilized for procedural comparisons but for determining the “optimal” bandwidth and mixing parameters (when applicable) for the nonparametric and model robust dual modeling procedures. This process will be discussed in detail later in the chapter. To compare entire fits among the various techniques, it is more appropriate to compute the approximate integrated MSE for both the mean and variance fits. These quantities are based on the entire range of the data and involve averaging the MSE calculations over a fine grid (1000 points) of the data. The two approximate integrated MSE values (called MIMSE for the means model integrated MSE and VIMSE for the variance model integrated MSE), will serve as final criterion for comparing the various dual modeling techniques. The MSE formulas (for parametric dual modeling, residual-based dual modeling, difference-based dual modeling, and DMRR) are all derived from the general model written in (5.A.1) and (5.A.2). The formulas provided here are for the bias and variance of the mean and variance fits. The MSE expressions can be obtained as a straightforward extension by squaring the bias and adding the variance.

Parametric Dual Modeling

In parametric dual modeling the estimated mean response across the n data points is given by $\hat{\mathbf{y}}^{(ewls)} = \mathbf{H}^{(ewls)} \mathbf{y}$ and the bias and variance of $\hat{\mathbf{y}}^{(ewls)}$ can be expressed by

$$\text{Bias}\left(\hat{\mathbf{y}}^{(ewls)}\right) = -\left(\mathbf{I} - \mathbf{H}_0^{(ewls)}\right)\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{f}\right), \quad (5.B.1)$$

$$\text{Var}\left(\hat{\mathbf{y}}^{(ewls)}\right) = \mathbf{H}_0^{(ewls)} \mathbf{V} \mathbf{H}_0'^{(ewls)}. \quad (5.B.2)$$

Notice that as the true means model deviates further from $\mathbf{X}\boldsymbol{\beta}$, the value of \mathbf{f} becomes large and the bias increases. The bias and variance of $\hat{y}_o^{(ewls)}$ at any location $\mathbf{x}_o' = (1 \ x_o \ x_o^2 \ \dots)$ can be expressed as

$$\text{Bias}\left(\hat{y}_o^{(ewls)}\right) = \mathbf{h}_o'^{(ewls)} \mathbf{f} - f(x_o), \quad (5.B.3)$$

$$\text{Var}\left(\hat{y}_o^{(ewls)}\right) = \mathbf{h}_o'^{(ewls)} \mathbf{V} \mathbf{h}_o^{(ewls)}, \quad (5.B.4)$$

where $\mathbf{h}_o'^{(ewls)} = \mathbf{x}_o' \left(\mathbf{X}' \mathbf{V}_0^{-1}(\text{glm}) \mathbf{X}\right)^{-1} \mathbf{X}' \mathbf{V}_0^{-1}(\text{glm})$.

The parametric variance estimate is written as $\hat{\boldsymbol{\sigma}}^{2(\text{glm})} = \exp\left(\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{glm})}\right)$ and the bias of $\hat{\boldsymbol{\sigma}}^{2(\text{glm})}$ can be expressed as

$$\begin{aligned} \text{Bias}\left(\hat{\boldsymbol{\sigma}}^{2(\text{glm})}\right) &\approx \left\{ \exp\left\{\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\right\} - \left[\exp\left\{\mathbf{Z}\boldsymbol{\theta}\right\} - \mathbf{I} \right] \right\} \\ &\quad + \mathbf{H}^{(\text{glm})} \left(\mathbf{E}\left(\mathbf{e}^{2(\text{ewls})}\right) - \exp\left\{\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\right\} \right) \end{aligned} \quad (5.B.5)$$

where $\mathbf{E}\left(\mathbf{e}^{2(\text{ewls})}\right) = \left[\mathbf{E}\left(e_1^{2(\text{ewls})}\right) \dots \mathbf{E}\left(e_n^{2(\text{ewls})}\right) \right]'$. From Appendix C we have that

$\mathbf{E}\left(e_i^{2(\text{ewls})}\right)$ is given by :

$$\begin{aligned} \mathbf{E}\left(e_i^{2(\text{ewls})}\right) &= \left[\left(1 - h_{ii}^{(\text{ewls})}\right)^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{ewls})} \sigma_j^2 \right] \\ &\quad + \left[\left(1 - h_{ii}^{(\text{ewls})}\right) f(x_i) - \sum_{j \neq i}^n h_{ji}^{(\text{mmrr})} f(x_j) \right]^2 \end{aligned} \quad (5.B.6)$$

where $\sigma_i^2 = \exp\{\mathbf{z}_i' \boldsymbol{\theta}\} + l(z_i)$. Notice that the bias of the estimated variance is effected by two phenomena: the bias in the means fit (quantified by \mathbf{f}) and the amount of misspecification in the variance function (quantified by \mathbf{I}). If there is no means or variance model misspecification, ($\mathbf{f} = \mathbf{I} = \mathbf{0}$), then apart from leverage, $E\left(e_i^{2(\text{ewls})}\right) \approx \sigma_i^2 = \exp\{\mathbf{z}_i' \boldsymbol{\theta}\}$. The amount that $E\left(\exp\left(\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{glm})}\right)\right)$ differs from $\exp\left(\mathbf{Z}\boldsymbol{\theta}\right)$ depends on two things: the amount of misspecification in the mean and the amount to misspecification in the variance model. Thus, if $\mathbf{f} = \mathbf{I} = \mathbf{0}$, we have that (asymptotically) $E\left(\exp\left(\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{glm})}\right)\right) = \exp\left(\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\right) = \exp\left(\mathbf{Z}\boldsymbol{\theta}\right)$ and, consequently, the bias of $\hat{\boldsymbol{\sigma}}^{2(\text{glm})}$ is zero.

The variance of $\hat{\boldsymbol{\sigma}}^{2(\text{glm})}$ can be expressed as

$$\text{Var}\left(\hat{\boldsymbol{\sigma}}^{2(\text{glm})}\right) \approx \mathbf{H}^{(\text{glm})} \mathbf{V}_{e^{2(\text{ewls})}} \mathbf{H}^{(\text{glm})'} \quad (5.B.7)$$

where $\mathbf{V}_{e^{2(\text{ewls})}}$ is assumed to be a diagonal matrix with the $\text{Var}\left(e_i^{2(\text{ewls})}\right)$ as the i^{th} diagonal element. From Appendix C.1 that:

$$\begin{aligned} \text{Var}\left(e_i^{2(\text{ewls})}\right) &= 2 \left[\left(1 - h_{ii}^{(\text{ewls})}\right)^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{ewls})} \sigma_j^2 \right]^2 + \quad (5.B.8) \\ &4 \left\{ \left[\left(1 - h_{ii}^{(\text{ewls})}\right)^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{ewls})} \sigma_j^2 \right] \left[\left(1 - h_{ii}^{(\text{ewls})}\right) f(x_i) - \sum_{j \neq i}^n h_{ji}^{(\text{ewls})} \right]^2 f(x_j) \right\} \end{aligned}$$

Notice that the only type of model misspecification which influences the $\text{Var}\left(\hat{\boldsymbol{\sigma}}^{2(\text{glm})}\right)$ is misspecification of the means model. In conclusion, the bias and variance of $\hat{\boldsymbol{\sigma}}_0^{2(\text{glm})}$ at any location $\mathbf{z}_0' = (1 \ z_0 \ z_0^2 \ \dots)$ can be expressed as

$$\begin{aligned} \text{Bias}\left(\hat{\boldsymbol{\sigma}}_0^{2(\text{glm})}\right) &\approx \left\{ \exp\left\{\mathbf{z}_0' \boldsymbol{\theta}^{(\text{glm})}\right\} - \left[\exp\left\{\mathbf{z}_0' \boldsymbol{\theta}\right\} - l(z_0) \right] \right\} \quad (5.B.9) \\ &+ \mathbf{h}_0^{(\text{glm})} \left(E\left(e^{2(\text{ewls})}\right) - \exp\left\{\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\right\} \right), \end{aligned}$$

$$\text{Var}\left(\hat{\sigma}_o^2(\text{glm})\right) \approx \mathbf{h}_o'(\text{glm}) \mathbf{V}_{e^2(\text{ewls})} \mathbf{h}_o(\text{glm}), \quad (5.B.10)$$

where $\mathbf{h}_o'(\text{glm}) = \delta_o \mathbf{z}_o' (\mathbf{Z}' \Delta \mathbf{V}_e^{-1} \Delta \mathbf{Z})^{-1} \mathbf{Z}' \Delta \mathbf{V}_e^{-1}$ and δ_o is as defined in Appendix A.1.

Difference-Based Dual Modeling

In difference-based dual modeling the mean is estimated by a nonparametric smoothing technique such as local linear regression and the estimate can be written as $\hat{\mathbf{y}}^{(\text{llr})} = \mathbf{H}_b^{(\text{llr})} \mathbf{y}$. The bias and variance of $\hat{\mathbf{y}}^{(\text{llr})}$ are given by

$$\text{Bias}\left(\hat{\mathbf{y}}^{(\text{llr})}\right) = -\left(\mathbf{I} - \mathbf{H}_b^{(\text{llr})}\right) (\mathbf{X}\boldsymbol{\beta} + \mathbf{f}), \quad (5.B.11)$$

$$\text{Var}\left(\hat{\mathbf{y}}^{(\text{llr})}\right) = \mathbf{H}_b^{(\text{llr})} \mathbf{V} \mathbf{H}_b'(\text{llr}). \quad (5.B.12)$$

Notice that the bias and variance of $\hat{\mathbf{y}}^{(\text{llr})}$ are both written in terms of the smoothing parameter b . This brings up an important issue that must be addressed for the nonparametric and model robust dual modeling techniques. Namely, how to treat the selection of the bandwidth b , and smoothing parameter, λ (when applicable). Here we use the general notation " b " and " λ " to refer to any bandwidth or mixing parameter. Since the bandwidth and mixing parameter are sample specific, they are random quantities. The approach taken in this research follows that of Speckman (1988). Specifically, b and λ are both considered as fixed quantities when deriving the bias and variance expressions. We take b and λ to be the "optimal" bandwidth and mixing parameter for each procedure, where "optimal" implies the values of b and λ which minimize the average mean square error of the estimate being considered (AVEMMSE when estimating the mean and AVEVMSE when estimating the variance). In general, " b_o " and " λ_o " will be used to denote the optimal bandwidth and mixing parameter. For the local linear estimate of the mean used by the difference-based procedure, b_o denotes the bandwidth which minimizes the AVEMSE calculated from the bias and variance expressions in (5.B.11) and (5.B.12).

The bias and variance of $\hat{y}_o^{(\text{llr})}$ at any given point \mathbf{x}_o' can be obtained as

$$\text{Bias}\left(\hat{y}_o^{(\text{llr})}\right) = \left[\mathbf{h}_o'(\text{llr}) \mathbf{X} - \mathbf{x}_o'\right] \boldsymbol{\beta} + \mathbf{h}_o'(\text{llr}) \mathbf{f} - f(x_o), \quad (5.B.13)$$

$$\text{Var}\left(\hat{y}_o^{(\text{llr})}\right) = \mathbf{h}_o'(\text{llr}) \mathbf{V} \mathbf{h}_o(\text{llr}), \quad (5.B.14)$$

where $\mathbf{h}_o^{(llr)}$ is the row of a local linear hat matrix in which the kernel weights are determined by the distances that the data points (the x values) are to x_o .

The difference-based variance estimate is given by $\hat{\sigma}^{2(\text{diff})} = \mathbf{H}_{b_{\tilde{e}}}^{(llr)} \tilde{\mathbf{e}}^{2(\text{pseud})}$ where $\tilde{\mathbf{e}}^{2(\text{pseud})}$ is the $n \times 1$ vector of squared pseudo residuals. The bias and variance of $\hat{\sigma}^{2(\text{diff})}$ are expressed as follows

$$\text{Bias} \left(\hat{\sigma}^{2(\text{diff})} \right) \approx - \left(\mathbf{I} - \mathbf{H}_{b_{\tilde{e}}}^{(llr)} \right) \left(\exp \left\{ \mathbf{Z}\boldsymbol{\theta} \right\} + \mathbf{1} \right), \quad (5.B.15)$$

$$\text{Var} \left(\hat{\sigma}^{2(\text{diff})} \right) = \mathbf{H}_{b_{\tilde{e}}}^{(llr)} \mathbf{V}_{\tilde{\mathbf{e}}} \mathbf{H}_{b_{\tilde{e}}}^{(llr)'} , \quad (5.B.16)$$

where $\mathbf{V}_{\tilde{\mathbf{e}}}$ denotes the dispersion matrix of the $n \times 1$ vector of squared pseudo residuals. The bias and variance of $\hat{\sigma}_o^{2(\text{ewls})}$ at any location z_o' can be expressed as

$$\text{Bias} \left(\hat{\sigma}_o^{2(\text{diff})} \right) = \mathbf{h}_{b_{\tilde{e},o}}^{(llr)} \left[\exp \left\{ \mathbf{Z}\boldsymbol{\theta} \right\} + \mathbf{1} \right] - \exp \left\{ \mathbf{z}_o' \boldsymbol{\theta} \right\} - l(z_o), \quad (5.B.17)$$

$$\text{Var} \left(\hat{\sigma}_o^{2(\text{diff})} \right) = \mathbf{h}_{b_{\tilde{e},o}}^{(llr)} \mathbf{V}_{\tilde{\mathbf{e}}} \mathbf{h}_{b_{\tilde{e},o}}^{(llr)'} , \quad (5.B.18)$$

where $\mathbf{h}_{b_{\tilde{e},o}}^{(llr)}$ is the row of a local linear hat matrix in which the kernel weights are determined by the distances that the data points (the z values) are to z_o . Keep in mind that the bandwidth is taken to be fixed at $b_{\tilde{e}} = b_{\tilde{e}_o}$ where $b_{\tilde{e}_o}$ is the bandwidth which minimizes the AVEMSE calculated from the bias and variance expressions given in (5.B.15) and (5.B.16).

Residual-Based Dual Modeling

Recall that the estimate of the mean in residual-based dual modeling is the same as that used in difference-based dual modeling, namely the local linear fit to the raw data. Thus, to avoid being redundant, the bias and variance expressions for $\hat{\mathbf{y}}^{(llr)}$ will not be repeated. The difference in the two procedures lies in the estimate of variance. The residual-based variance estimate is written as $\hat{\sigma}^{2(\text{res})} = \mathbf{H}_{b_e}^{(llr)} \mathbf{e}^{2(\text{llr})}$ where $\mathbf{e}^{2(\text{res})}$ is the $n \times 1$ vector of local linear squared residuals (residuals resulting from a local linear fit to the raw data). The bias of $\hat{\sigma}^{2(\text{res})}$ is expressed as

$$\text{Bias} \left(\hat{\sigma}^{2(\text{res})} \right) = \mathbf{H}_{b_e}^{(llr)} \mathbf{E} \left(\mathbf{e}^{2(\text{llr})} \right) - \exp \left\{ \mathbf{Z}\boldsymbol{\theta} \right\} - \mathbf{1} \quad (5.B.19)$$

where $E(\mathbf{e}^{2(\text{llr})}) = \left[E(e_1^{2(\text{llr})}) \dots E(e_n^{2(\text{llr})}) \right]'$. From Appendix C.2 we have that $E(e_i^{2(\text{llr})})$ is given by

$$E(e_i^{2(\text{llr})}) = \left[(1 - h_{ii}^{(\text{llr})})^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{llr})} \sigma_j^2 \right] + \left[(1 - h_{ii}^{(\text{llr})})(\mathbf{x}_i' \boldsymbol{\beta} + f(x_i)) - \sum_{j \neq i}^n h_{ji}^{(\text{llr})}(\mathbf{x}_j' \boldsymbol{\beta} + f(x_j)) \right]^2 \quad (5.B.20)$$

The variance of $\hat{\sigma}^{2(\text{res})}$ can be expressed as

$$\text{Var}(\hat{\sigma}^{2(\text{res})}) \approx \mathbf{H}_{b_e}^{(\text{llr})} \mathbf{V}_{e^{2(\text{llr})}} \mathbf{H}_{b_e}^{(\text{llr})'} \quad (5.B.21)$$

where $\mathbf{V}_{e^{2(\text{llr})}}$ is assumed to be a diagonal matrix with the $\text{Var}(e_i^{2(\text{llr})})$ as the i^{th} diagonal element. From Appendix C.2 we have that:

$$\text{Var}(e_i^{2(\text{llr})}) = 2 \left[(1 - h_{ii}^{(\text{llr})})^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{llr})} \sigma_j^2 \right]^2 + 4 \left\{ \left[(1 - h_{ii}^{(\text{mmrr})})^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{mmrr})} \sigma_j^2 \right] \times \left[(1 - h_{ii}^{(\text{llr})})(\mathbf{x}_i' \boldsymbol{\beta} + f(x_i)) - \sum_{j \neq i}^n h_{ji}^{(\text{llr})}(\mathbf{x}_j' \boldsymbol{\beta} + f(x_j)) \right]^2 \right\}. \quad (5.B.22)$$

Here, b_e is fixed at b_{e_0} , where b_{e_0} is the bandwidth that minimizes the AVEMSE based on the bias and variance expressions in (5.B.19) and (5.B.20). In conclusion, the bias and variance of $\hat{\sigma}_0^{2(\text{res})}$ at any location \mathbf{z}_0' can be expressed as

$$\text{Bias}(\hat{\sigma}_0^{2(\text{res})}) = \mathbf{h}_0'^{(\text{llr})} E(\mathbf{e}^{2(\text{llr})}) - \exp\left\{ \mathbf{z}_0' \boldsymbol{\theta} \right\} - l(z_0), \quad (5.B.23)$$

$$\text{Var}(\hat{\sigma}_0^{2(\text{res})}) \approx \mathbf{h}_0'^{(\text{llr})} \mathbf{V}_{e^{2(\text{llr})}} \mathbf{h}_0^{(\text{llr})}, \quad (5.B.24)$$

where $\mathbf{h}'_o^{(llr)}$ is the row of a local linear hat matrix determined by z'_o when fitting the residuals from the squared, local linear residuals from the means model fit.

Dual Model Robust Regression

The model robust means estimate, known as MMRR, is given by $\hat{\mathbf{y}}^{(mmrr)} = \mathbf{H}^{(mmrr)} \mathbf{y}$ and the bias and variance of $\hat{\mathbf{y}}^{(mmrr)}$ (with b_μ and λ_μ held constant) can be expressed as follows:

$$\text{Bias}(\hat{\mathbf{y}}^{(mmrr)}) = -(\mathbf{I} - \mathbf{H}^{(mmrr)})\mathbf{f} \quad (5.B.25)$$

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

$$\text{Var}(\hat{\mathbf{y}}^{(mmrr)}) = \mathbf{H}^{(mmrr)} \mathbf{V} \mathbf{H}'^{(mmrr)}. \quad (5.B.26)$$

Note from (5.B.25) that the bias is independent of the linear term $\mathbf{X}\boldsymbol{\beta}$ and is only affected by the form of \mathbf{f} . For the MMRR results above, b_μ is fixed at b_{μ_o} where b_{μ_o} is the bandwidth which minimizes the AVEMSE of the local linear fit to the EWLS residuals. The local linear fit to the EWLS residuals can be expressed as $\hat{\mathbf{r}}_\mu = \mathbf{H}_{b_\mu}^{(llr)} \mathbf{r}_\mu$ where \mathbf{r}_μ denotes the $n \times 1$ vector of EWLS residuals. The bias and variance of $\hat{\mathbf{r}}_\mu$ are expressed as:

$$\text{Bias}(\hat{\mathbf{r}}_\mu) = -(\mathbf{I} - \mathbf{H}_{b_\mu}^{(llr)}) \left(\mathbf{I} - \mathbf{H}^{(ewls)} \right) \begin{pmatrix} \mathbf{f} \\ \end{pmatrix}, \quad (5.B.27)$$

$$\text{Var}(\hat{\mathbf{r}}_\mu) = \mathbf{H}_{b_\mu}^{(llr)} \left(\mathbf{I} - \mathbf{H}^{(ewls)} \right) \mathbf{V} \left(\mathbf{I} - \mathbf{H}^{(ewls)} \right)' \mathbf{H}_{b_\mu}^{(llr)'}. \quad (5.B.28)$$

The means model mixing parameter λ_μ is fixed at λ_{μ_o} , which is the value of λ_μ which minimizes the AVEMSE based on the bias and variance expressions given in (5.B.25) and (5.B.26) with b_μ is fixed at b_{μ_o} .

The MMRR estimate at an individual point \mathbf{x}_o is written as $\hat{y}_o^{(mmrr)} = \mathbf{h}'_o^{(mmrr)} \mathbf{y}$ where $\mathbf{h}'_o^{(mmrr)} = \mathbf{h}'_o^{(ewls)} + \lambda_\mu \mathbf{h}'_{b_\mu, o}^{(llr)} (\mathbf{I} - \mathbf{H}^{(ewls)})$. The vector $\mathbf{h}'_o^{(mmrr)}$ can be thought of as the row of the MMRR hat matrix that would be determined by \mathbf{x}'_o . The term $\mathbf{h}'_{b_\mu, o}^{(llr)}$ refers to the

row of the local linear hat matrix determined by \mathbf{x}'_o when fitting the residuals from the EWLS means fit. The bias and variance of $\hat{y}_o^{(\text{mmrr})}$ are expressed as follows:

$$\text{Bias} \left(\hat{y}_o^{(\text{mmrr})} \right) = \mathbf{h}'_o^{(\text{mmrr})} \mathbf{f} - f(x_o), \quad (5.B.29)$$

$$\text{Var} \left(\hat{y}_o^{(\text{mmrr})} \right) = \mathbf{h}'_o^{(\text{mmrr})} \mathbf{V} \mathbf{h}_o^{(\text{mmrr})}. \quad (5.B.30)$$

The model robust variance estimate, VMRR, is given as $\hat{\sigma}^{2(\text{vmrr})} = \exp\{\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{glm})}\} + \lambda_\sigma \mathbf{H}_{b_\sigma}^{(\text{llr})} \left(\mathbf{e}^{2(\text{mmrr})} - \exp\{\mathbf{Z}\hat{\boldsymbol{\theta}}^{(\text{mmrr})}\} \right)$ and the bias of $\hat{\sigma}^{2(\text{vmrr})}$ (with b_σ and λ_σ held constant) can be expressed as:

$$\begin{aligned} \text{Bias} \left(\hat{\sigma}^{2(\text{vmrr})} \right) \approx & \exp\{\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\} - \left[\exp\{\mathbf{Z}\boldsymbol{\theta}\} + \mathbf{1} \right] + \\ & \mathbf{H}^{(\text{vmrr})} \left(\mathbf{E} \left(\mathbf{e}^{2(\text{mmrr})} \right) - \exp\{\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\} \right) \end{aligned} \quad (5.B.31)$$

where $\mathbf{E} \left(\mathbf{e}^{2(\text{mmrr})} \right) = \left[\mathbf{E} \left(e_1^{2(\text{mmrr})} \right) \dots \mathbf{E} \left(e_n^{2(\text{mmrr})} \right) \right]'$. From Appendix C.1 we have that $\mathbf{E} \left(e_i^{2(\text{mmrr})} \right)$ is given by

$$\begin{aligned} \mathbf{E} \left(e_i^{2(\text{mmrr})} \right) = & \left[\left(1 - h_{ii}^{(\text{mmrr})} \right)^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{mmrr})} \sigma_j^2 \right] \\ & + \left[\left(1 - h_{ii}^{(\text{mmrr})} \right) f(x_i) - \sum_{j \neq i}^n h_{ji}^{(\text{mmrr})} f(x_j) \right]^2 \end{aligned} \quad (5.B.32)$$

where $\sigma_i^2 = \exp\{\mathbf{z}'_i \boldsymbol{\theta}\} + l(z_i)$. Notice that the bias of $\hat{\sigma}^{2(\text{vmrr})}$ is affected not only by the functional misspecification of the variance model (quantified by $l(z_i)$) but also by the misspecification of the means model.

The variance of $\hat{\sigma}^{2(\text{vmrr})}$ (with b_σ and λ_σ held constant) can be expressed as:

$$\text{Var} \left(\hat{\sigma}^{2(\text{vmrr})} \right) \approx \mathbf{H}^{(\text{vmrr})} \mathbf{V}_{e^{2(\text{mmrr})}} \mathbf{H}^{(\text{vmrr})'} \quad (5.B.33)$$

where $\mathbf{H}^{(\text{vmrr})} = \mathbf{H}^{(\text{glm})} + \lambda_{\sigma} \mathbf{H}_{b_{\sigma}}^{(\text{llr})} (\mathbf{I} - \mathbf{H}^{(\text{glm})})$, $\mathbf{V}_{\mathbf{e}^2(\text{mmrr})}$ is assumed to be a diagonal matrix with the $\text{Var}(\mathbf{e}_i^{2(\text{mmrr})})$ as the i^{th} diagonal element. From Appendix we have C.1 that:

$$\text{Var}(\mathbf{e}_i^{2(\text{mmrr})}) = 2 \left[\left(1 - h_{ii}^{(\text{mmrr})}\right)^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{mmrr})} \sigma_j^2 \right]^2 + \quad (5.B.34)$$

$$4 \left\{ \left[\left(1 - h_{ii}^{(\text{mmrr})}\right)^2 \sigma_i^2 + \sum_{j \neq i}^n h_{ji}^{2(\text{mmrr})} \sigma_j^2 \right] \left[\left(1 - h_{ii}^{(\text{mmrr})}\right) f(x_i) - \sum_{j \neq i}^n h_{ji}^{(\text{mmrr})} f(x_j) \right]^2 \right\}.$$

For the VMRR results above, b_{σ} is fixed at b_{σ_0} where b_{σ_0} is the bandwidth which minimizes the AVMSE of the local linear fit to the GLM residuals. The local linear fit to the GLM residuals can be expressed as $\hat{\mathbf{r}}_{\sigma} = \mathbf{H}_{b_{\sigma}}^{(\text{llr})} \mathbf{r}_{\sigma}$ where \mathbf{r}_{σ} denotes the $n \times 1$ vector of GLM residuals. The bias and variance of $\hat{\mathbf{r}}_{\sigma}$ are as follows:

$$\text{Bias}(\hat{\mathbf{r}}_{\sigma}) \approx - \left(\mathbf{I} - \mathbf{H}_{b_{\sigma}}^{(\text{llr})} \right) \left(\mathbf{I} - \mathbf{H}^{(\text{glm})} \right) \left(\mathbf{E}(\mathbf{e}^{2(\text{mmrr})}) - \exp\{\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\} \right), \quad (5.B.35)$$

$$\text{Var}(\hat{\mathbf{r}}_{\sigma}) \approx \mathbf{H}_{b_{\sigma}}^{(\text{llr})} \left(\mathbf{I} - \mathbf{H}^{(\text{glm})} \right) \mathbf{V}_{\mathbf{e}^2(\text{mmrr})} \left(\mathbf{I} - \mathbf{H}^{(\text{glm})} \right)' \mathbf{H}_{b_{\sigma}}^{(\text{llr})}', \quad (5.B.36)$$

where $\mathbf{E}(\mathbf{e}^{2(\text{mmrr})})$ and $\mathbf{V}_{\mathbf{e}^2(\text{mmrr})}$ are defined by (5.B.30) and (5.B.32) respectively.

The bias and variance of $\hat{\sigma}_0^2(\text{vmrr})$ at any location \mathbf{z}_0' can be expressed as

$$\text{Bias}(\hat{\sigma}_0^2(\text{vmrr})) \approx \exp\{\mathbf{z}_0' \boldsymbol{\theta}^{(\text{glm})}\} - \left[\exp\{\mathbf{z}_0' \boldsymbol{\theta}\} + l(z_0) \right] + \mathbf{h}_0'(\text{vmrr}) \left(\mathbf{E}(\mathbf{e}^{2(\text{mmrr})}) - \exp\{\mathbf{Z}\boldsymbol{\theta}^{(\text{glm})}\} \right), \quad (5.B.37)$$

$$\text{Var}(\hat{\sigma}_0^2(\text{vmrr})) \approx \mathbf{h}_0'(\text{vmrr}) \mathbf{V}_{\mathbf{e}^2(\text{mmrr})} \mathbf{h}_0(\text{vmrr}), \quad (5.B.38)$$

where $\mathbf{h}_0'(\text{vmrr}) = \mathbf{h}_0'(\text{glm}) + \lambda_{\sigma} \mathbf{h}_{b_{\sigma},0}'(\text{llr}) (\mathbf{I} - \mathbf{H}^{(\text{glm})})$. The term $\mathbf{h}_{b_{\sigma},0}'(\text{llr})$ refers to the row of the local linear hat matrix determined by \mathbf{x}_0' when fitting the residuals from the GLM variance estimate.

Now that the expressions for the bias and variance of the mean and variance estimates (expressions (5.B.1 - 5.B.38)) have been given for each of the dual modeling techniques of interest, the MSE values are simple to obtain. The AVEMMSE and AVEVMSE for the various procedures are found by averaging the MSE's across n the data points in the means and variance models respectively. These are the statistics which are used to find the appropriate smoothing parameter, b , and mixing parameter, λ , (when applicable) for the nonparametric and model robust dual modeling procedures. The means model integrated mean squared error (MIMSE) and variance model mean squared error (VIMSE) are approximated by averaging the MSE's over a fine grid (1000) of points within the range of the data for each of the methods. These two diagnostics will be the ones which will be used to compare performances of the dual modeling techniques. In the next chapter, examples will be provided which illustrate the performance of the various techniques.