# Chapter 6: Examples

## 6.A Introduction

In Chapter 4, several approaches to the dual model regression problem were described and Chapter 5 provided expressions enabling one to compute the MSE of the mean and variance estimates for each of the procedures. The question still remains regarding which method is the most appropriate for a given data set. The answer to this question depends on how closely the procedural assumptions are to being matched by the process at hand. Procedural assumptions can be categorized into error distributions assumptions and assumptions regarding the form of the underlying models. In this research, attention is focused on functional form assumptions. Traditionally, the underlying models are assumed to have a functional form which is either *known* or *unknown*. When functional form is considered to be known, the user explicitly describes the process in terms of a mathematical model and data is then used to estimate model parameters (the parametric approach). If the user has indeed specified models which accurately depict the true state of nature, this parametric approach to understanding a given phenomena is quite effective. However, if the specified models are incorrect (for even part of the data), a parametric approach can lead to faulty inferences regarding the underlying process. Thus, a parametric approach (which depends solely on the researcher's specifications) is only as good as the precision of the researcher's knowledge.

Another approach to data analysis is to assume that the researcher has no knowledge regarding the form of the underlying functions. Under this assumption, the researcher allows the data to determine its own model (nonparametric regression). Although nonparametric procedures often provide more variable fits than their parametric competitors, the concession of more variance is often worth the price of dramatically reducing the bias that could possibly be present in the parametric functional estimate. Thus, if the researcher is incapable of prescribing a model which accurately depicts the true state of nature, nonparametric procedures become the preferred choice for analysis.

Practically speaking, the researcher often has a "feel" for the structure of the underlying models but the specified models are not quite appropriate for the entire range of the data. In such cases, the researcher is presented with a dilemma. If a parametric approach to analysis is taken, the researcher runs the risk of encountering bias problems in the analysis. As we will soon see, bias can be extremely detrimental within the context of an iterative algorithm. If nonparametric approaches are taken, important knowledge that the user may possess regarding the process is ignored and the result is often an estimate that is more variable than what is desired.

It is the contention of this research that the potential risks encountered by the traditional parametric and nonparametric procedures can be ameliorated by basing the analysis on a different philosophy about functional form. Instead of assuming that the forms of the underlying functions are *known* or *unknown*, we take that position that there is always at least some information about the underlying process which can be extracted via a parametric specification (and thus analyzed parametrically). Then, whatever is leftover in the underlying functions can be analyzed nonparametrically. By combining parametric and nonparametric estimates, we hope to obtain a "hybrid" estimate for the mean and variance which borrows the strengths from the traditional approaches. The parametric component brings with it a degree of stability (low variance) and the nonparametric portion brings just enough variation to capture additional trends in the data that the parametric portion could not capture.

Since the forms of the underlying functions are never known in practice, it is important to decide on a method of analysis which is robust to true functional form. By "robust to true functional form" we mean that the method should perform at a high level regardless of the true form of the underlying models. In this chapter we will consider examples which illustrate the performances of the discussed dual modeling procedures in various simulated environments. These environments include situations in which the researcher is capable of specifying the functional form of both the mean and variance to situations in which the parametric specification is largely inadequate.

The performances of the dual modeling procedures (parametric, residual-based nonparametric, difference-based nonparametric and DMRR) are compared in each of the examples based on their theoretical IMMSE and IVMSE values. In all examples, the x-data (and thus the z-data, since x = z in this research) is scaled to be between 0 and 1. The scaling is done to force the bandwidths to be between 0 and 1 for the nonparametric and model robust procedures so that bandwidth magnitudes can be compared across data sets which may be on different scales. Keep in mind that the bandwidths and mixing parameters used for the computation of the IMMSE and IVMSE are chosen to be the optimal values based on minimizing the AVEMMSE and AVEVMSE for the means and variance estimates respectively. For each example, tables will be provided which display the optimal bandwidths and mixing parameters for various sample sizes and degrees of model misspecification.

## 6.B Example 1 (Means Model Misspecification)

In the first example, we will consider the situation where the user has misspecified the means model, but correctly specified the variance model. In this example, data is generated from the dual model

$$y_i = 2\left(x_i - 5.5\right)^2 + 5\,x_i + \gamma \sin\left(\frac{\pi\left(x_i - 1\right)}{2.25}\right) + g^{1/2}\left(\mathbf{z}_i\right)\varepsilon_i, \qquad (6.B.1)$$

$$\sigma_i^2 = g^{1/2}\left(\mathbf{z}_i\right) = \exp\left\{3.0 - 1.9\,z_i - 0.19\,z_i^2\right\} \qquad (6.B.2)$$

at evenly spaced x-values (recall x = z), from 0 to 10 where $\varepsilon \sim N(0,1)$. In the means function above, $\gamma$ can be thought of as a "misspecification" parameter. As the value of $\gamma$ gets large ($\gamma > 0$), the means function deviates from a quadratic model in $x_i$. Figure 6.B.1 shows the means function plotted for various values of $\gamma$ and Figure 6.B.2 displays the underlying variance function, which is independent of the value of $\gamma$. Model 6.B1 was introduced by Einsporn (1987), who studied fitting techniques for values of $\gamma$ in increasing magnitude. We will assume throughout this section that the researcher has specified a quadratic means model, thereby misspecifying model (6.B.1) when $\gamma$>0. We also assume that the variance model has been correctly specified as a quadratic experimental model. The researcher's dual model is then written as follows:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + g^{1/2}(z_i;\theta)\,\varepsilon_i \tag{6.B.3}$$

$$\sigma_i^2 = g^{1/2}(z_i;\theta) = \exp\left\{\theta_0 + \theta_1 z_i + \theta_2 z_i^2\right\}. \tag{6.B.4}$$

When $\gamma = 0$, a parametric analysis of the model in (6.B.3) and (6.B.4), via the generalized least squares algorithm discussed in Section 3.C.1, would yield maximum likelihood estimates for $\beta$ and $\theta$. As a result, this approach is widely accepted as the best approach. However, for $\gamma > 0$, a parametric analysis of this model would produce bias estimates for the mean and variance due to the fact that the specified quadratic model does not adequately describe the true form of the underlying means function.

Although a quadratic specification for the mean seems totally inappropriate when one views Figure 6.B.1 (where $\gamma > 0$), the quadratic specification seems somewhat less ridiculous when we consider that in practice, functional form is often specified upon viewing a scatter plot of the data. To illustrate this, a scatter plot for a random data set (*n* = 41) generated from the dual model in (6.B.1) and (6.B.2), with $\gamma = 5.0$, is provided in Figure 6.B.3. Notice that the scatter plot makes a quadratic specification seem plausible even though the true underlying function deviates substantially from a quadratic model when $\gamma = 5.0$. In Figure 6.B.4, the true means function and raw data are shown along with the EWLS means fit. Figure 6.B.5 displays the true variance function, the squared EWLS residuals, and the corresponding parametric, GLM variance estimate. Note in Figure 6.B.4 that the EWLS fit, as suspected, captures the general trend of the data but is unable to capture the "dips" which are present in the true underlying means function.

The bias in the EWLS means fit translates into "biased data" being used to estimate the variance. Although the user has specified the correct variance function, the "wrong" data is used to estimate the variance function. By using the squared EWLS

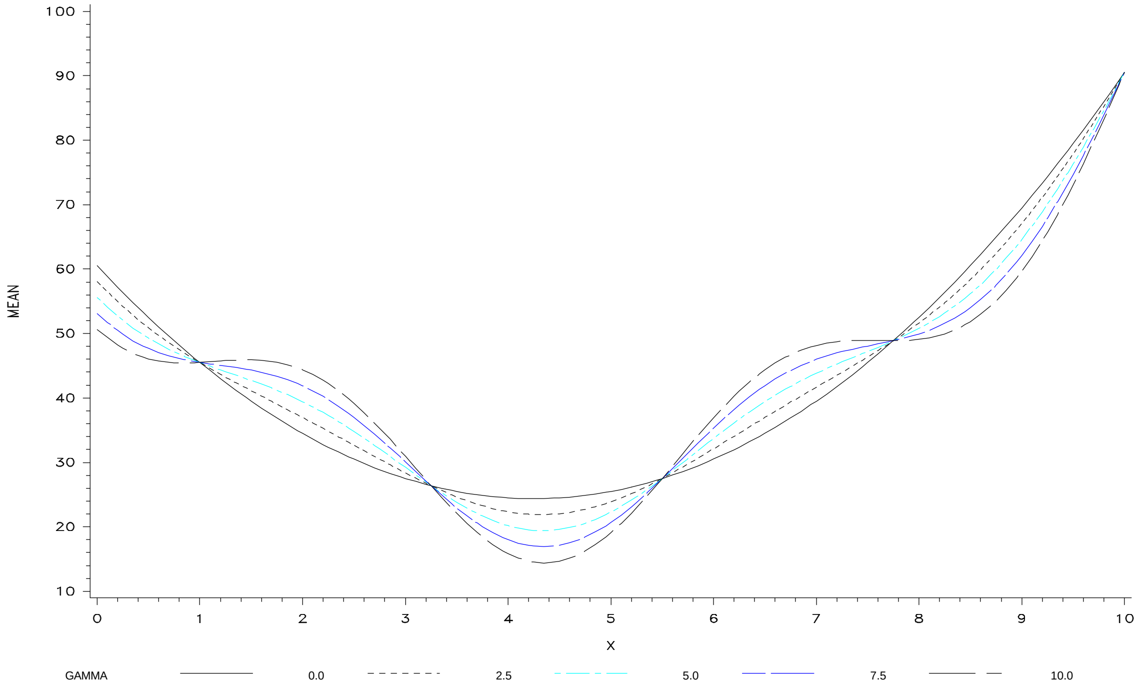

**Figure 6.B.1 True underlying mean function from Equation 6.B.1 for various values of $\gamma$ .**

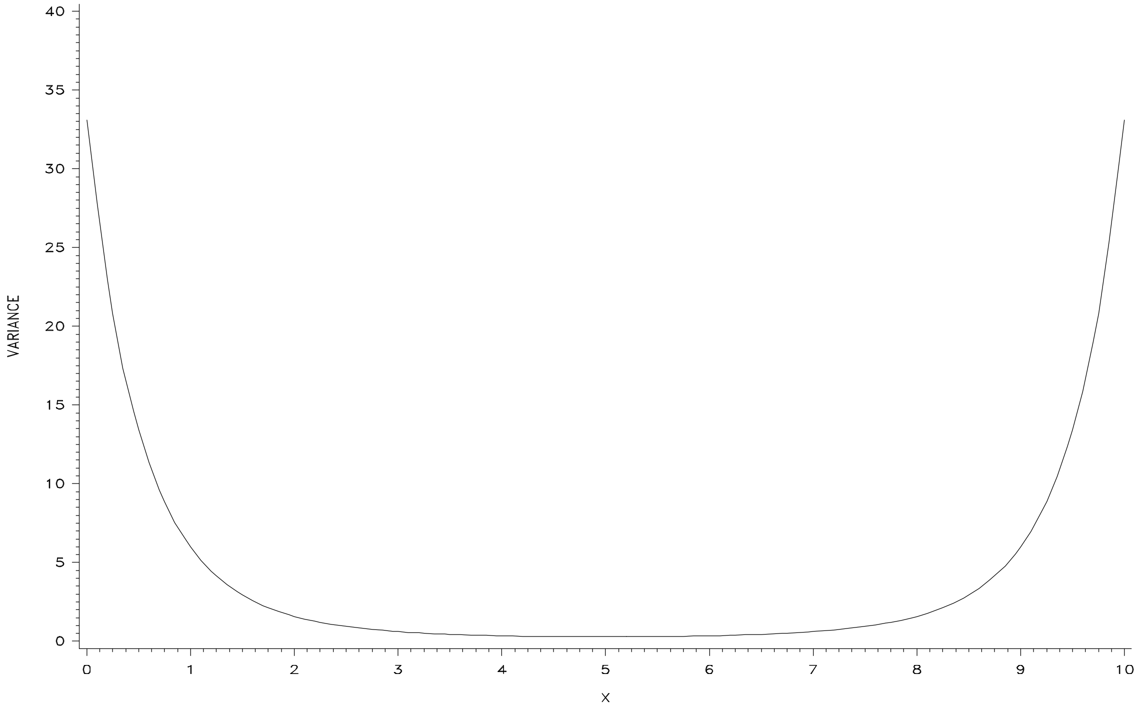

**Figure 6.B.2 True underlying variance function from Equation 6.B.2.**

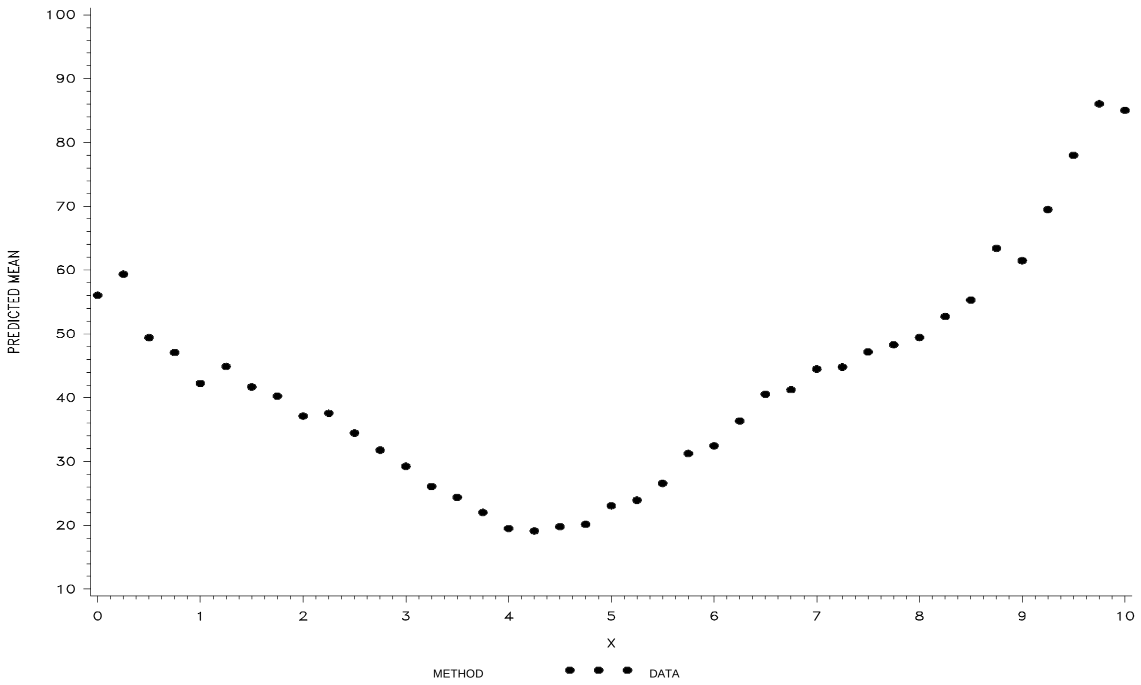

**Figure 6.B.3 Scatter plot of a random data set generated from the dual model in (6.B.1) and (6.B.2) where $\gamma$ = 5.0.**

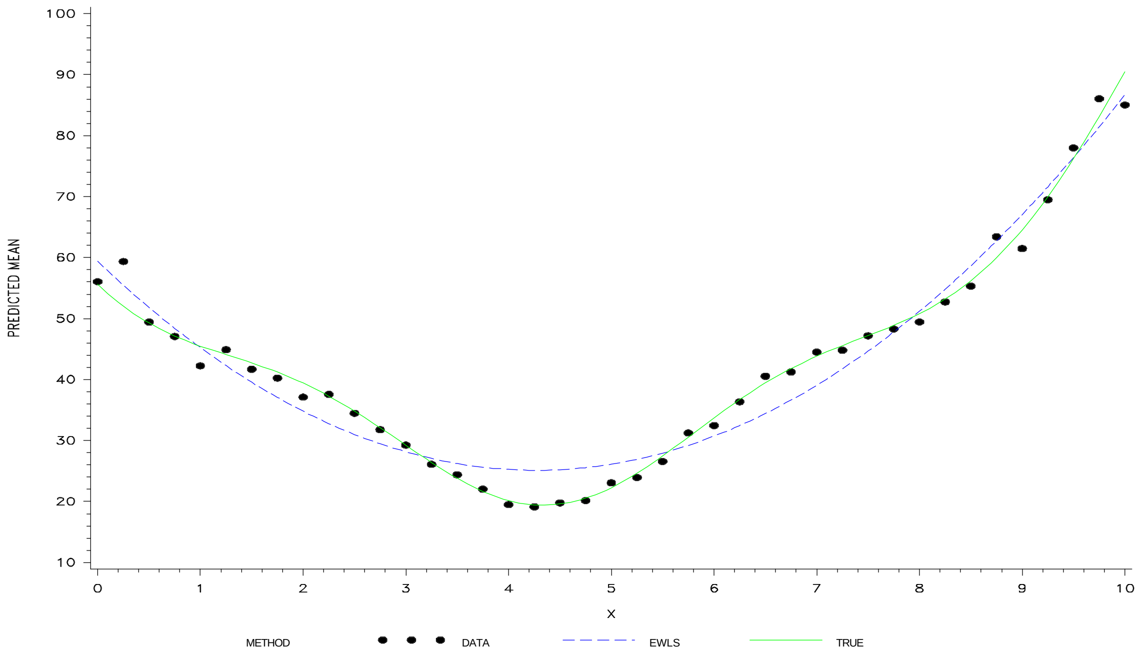

**Figure 6.B.4. Plot of true means function in Example 1 along with the raw data and EWLS means fit.**

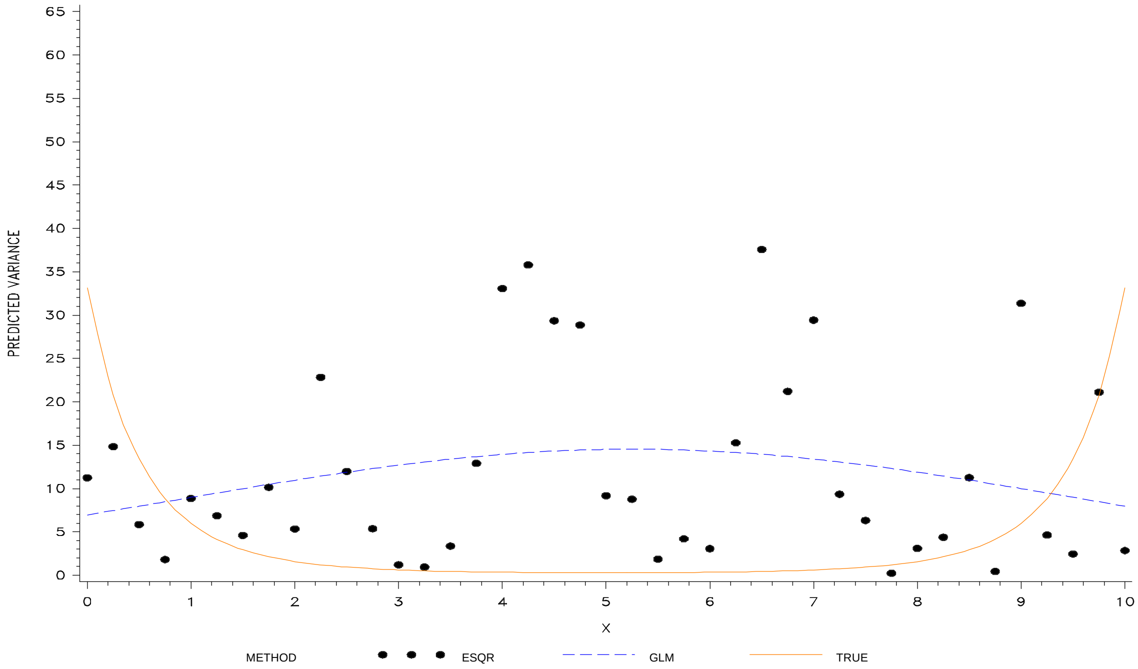

**Figure 6.B.5. Plot of parametric, GLM variance estimate using squared EWLS from the means fit in Figure 6.B.4.**

residuals as variance model data, the user implicitly assumes that $E\left(e_i^{2\,(\text{ewls})}\right) \approx \sigma_i^2$. However, as shown in Appendix C, as bias enters the EWLS fit, this approximation no longer holds true. Using this "biased data" to model the variance causes the user to conclude that the process variance is lower at the periphery of the z-space than at the interior of the z-space, when, in fact, the opposite is true.

Upon observing the scatter plot of the raw data, the researcher may have elected to disregard the general quadratic trend in the data and opted for a purely nonparametric method of analysis. Figure 6.B.6 displays the local linear (LLR) means estimate (based on $b_o = 0.0543$). Notice that the LLR means estimate is much more appealing than the EWLS estimate due to the reduction in bias. The residual-based variance estimate (based on $b_{e_o} = 0.08249$) is plotted along with the true variance function and the squared LLR residuals in Figure 6.B.7. The difference-based variance estimate.

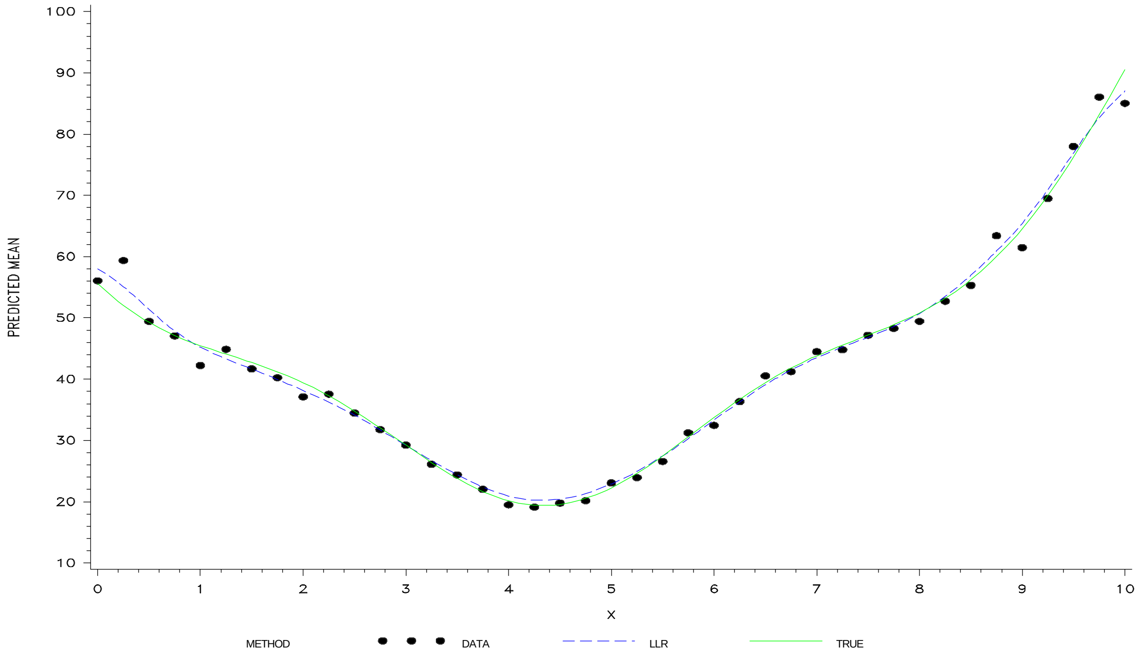

**Figure 6.B.6. Plot of true means function in Example 1 along with the raw data and LLR means fit.**

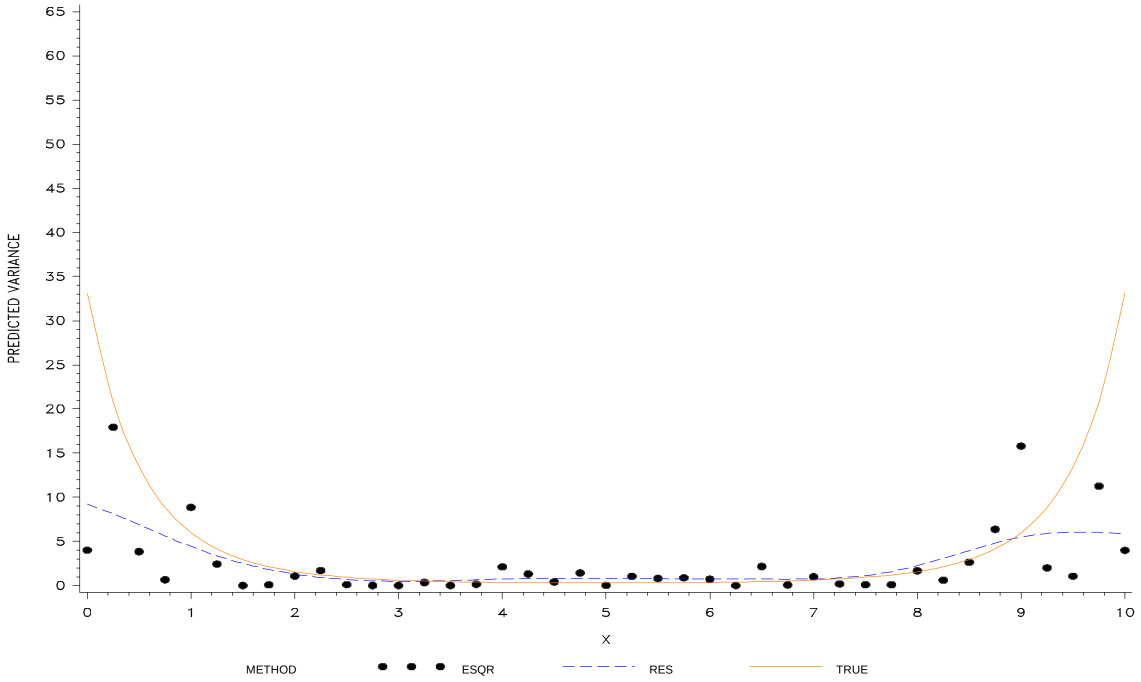

**Figure 6.B.7. Plot of non-parametric, residual-based variance estimate using the squared LLR residuals from the means fit in Figure 6.B.6.**

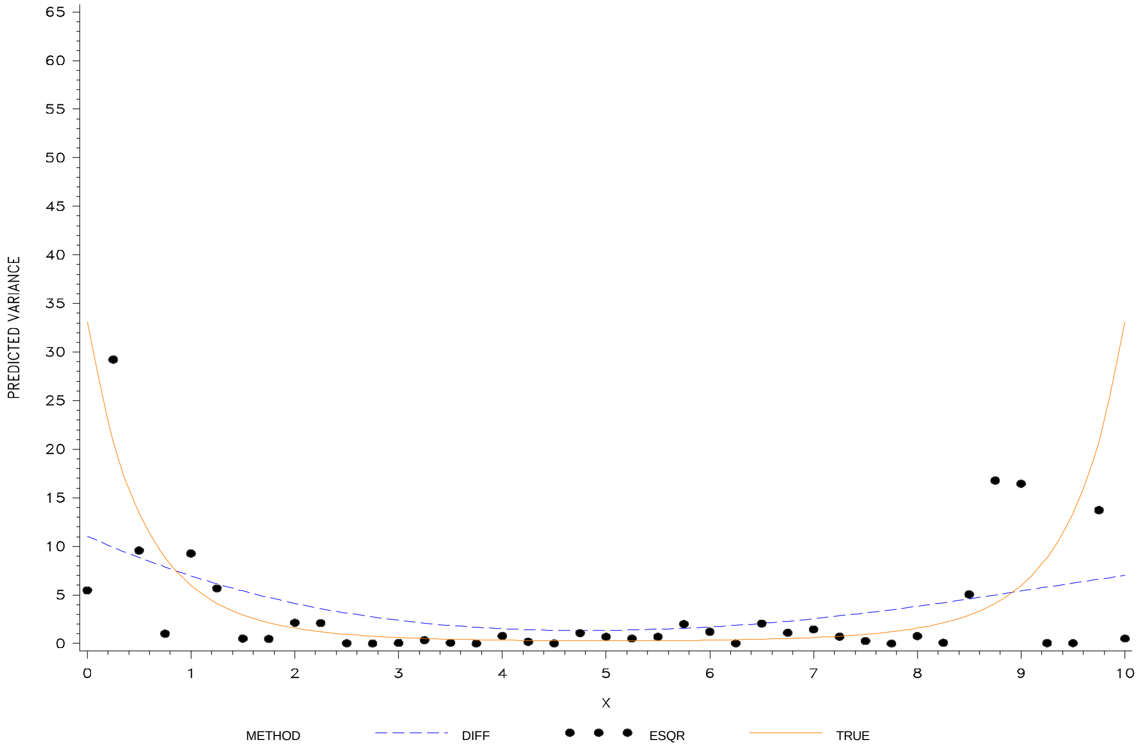

**Figure 6.B.8. Plot of non-parametric, difference-based variance estimate using the squared pseudo-residuals formed from the raw data.**

(based on $b_{\tilde{e}_o} = 0.2969$) is plotted along with the true variance function and the squared pseudo-residuals in Figure 6.B.8. Notice that both nonparametric variance estimates capture the trend of the underlying variance function. The residual-based variance estimate is not as smooth (more variable) than is desired and the difference-based estimate seems more quadratic than exponential in its form.

Figures 6.B.9 and 6.B.10 show the MMRR and VMRR mean and variance fits, respectively. The MMRR estimate of the mean (using $b_{\mu_o} = 0.0594$ and $\lambda_{\mu_o} = 1.0$), provides an estimate of the mean that is nearly identical to the LLR means estimate. For MMRR, recall that the first step is to obtain a parametric, EWLS fit to the raw data (where the weights are one initially); this the fit pictured in 6.B.11. Then, the residuals from this EWLS fit are fit using local linear regression; this is the LLR fit plotted in Figure 6.B.12. A certain portion ($\lambda_{\mu_o} = 1.0$ here) of the fit to the residuals is then added back to the EWLS fit to give the final MMRR means estimate. In Figure 6.B.13 the final MMRR means estimate is plotted along with the EWLS and LLR means estimates.

The most noticeable improvement that dual model robust regression offers in this example comes in the VMRR variance estimate. Figure 6.B.14 shows the VMRR variance estimate (based on $b_{\sigma_o} = 0.7325$ and $\lambda_{\sigma_o} = 0.0$) plotted along with the parametric GLM fit, the difference-based fit, and the residual-based fit. The VMRR variance estimate is far superior to the parametric GLM estimate due to the fact that the VMRR estimate is obtained by using better data in the variance model than was used in the strictly parametric approach. The data used in VMRR is the squared residuals from the MMRR fit whereas the data used in the parametric GLM estimate is the squared residuals from the EWLS means fit. The MMRR estimate is designed to account for the lack of fit that is present in the EWLS fit. Consequently, the squared residuals from the MMRR fit are essentially free from lack of fit and more closely resemble the pattern of the true underlying variance function than do the squared EWLS residuals.

A criticism that could be made of all the variance estimates is their poor performance at the boundaries of the data set. In the difference-based procedure this is due to the fact that the pseudo-residuals are composed using a neighborhood of only two points at the boundaries instead of the three point neighborhoods used for the interior points (see Section 3.C.3). Consequently, the information regarding the process variance at the endpoints is not as complete as it is throughout the rest of the data set. In the VMRR and residual-based methodologies, variance information is also incomplete at the boundaries, but for a different reason. Both of these methodologies are based on the premise that regions of the data set that have large (in an absolute value sense) residuals from the means fit represent areas in which process variance is high. Likewise, regions with small residuals represent areas in which process variance is low. This assumption, however, tends to break down at the endpoints of the data set whenever LLR is used to estimate the mean. Inherent to LLR is the tendency to fit closely to data located at the boundaries. Thus, small residuals are typically found at the boundaries and consequently, low variance is implied there regardless of the true underlying variance function.

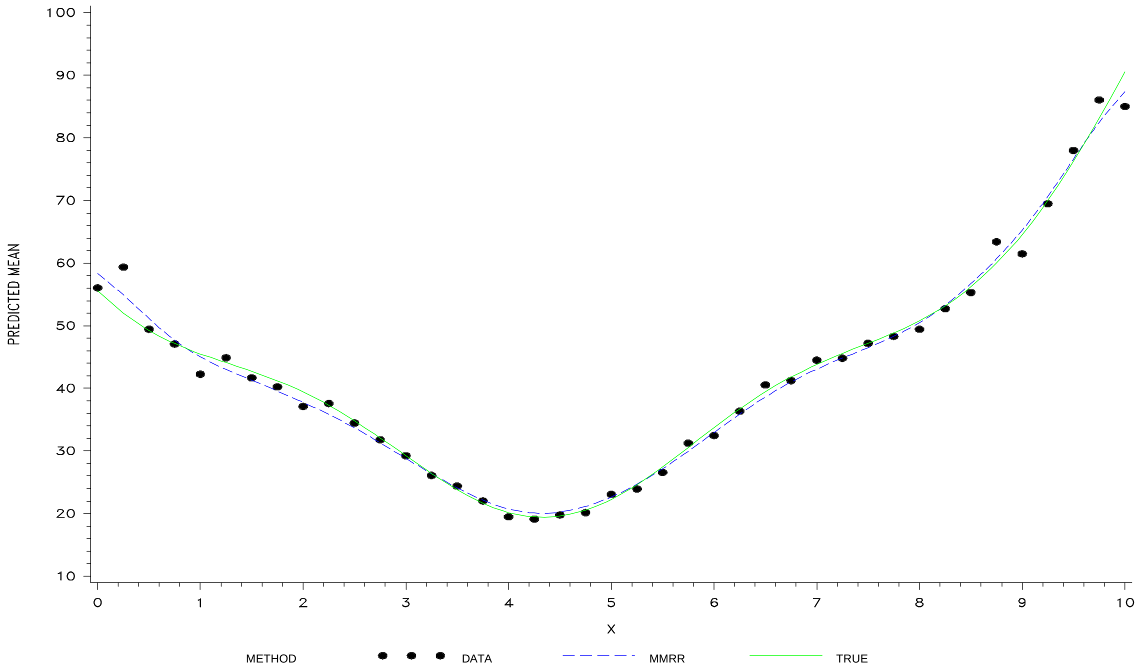

**Figure 6.B.9. Plot of true means function in Example 1 along with the raw data and MMRR means fit.**

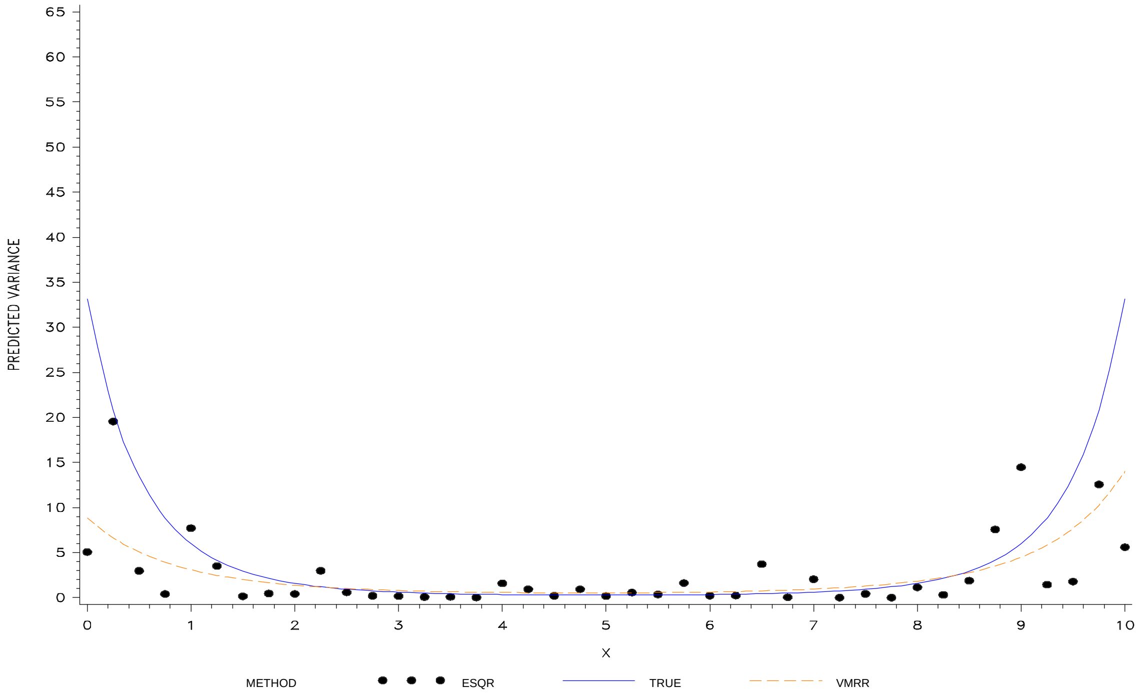

**Figure 6.B.10. Plot of the VMRR variance estimate using the squared MMRR residuals from the fit shown in Figure 6.B.9.**

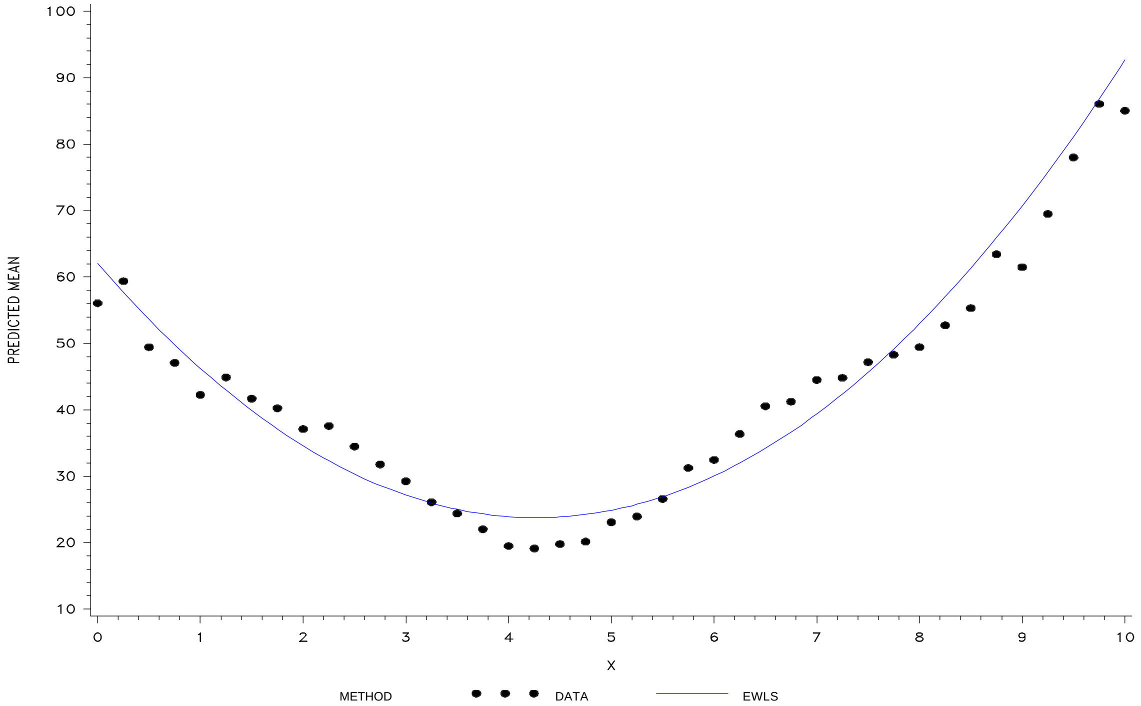

**Figure 6.B.11 OLS fit to the raw data from Example 1.**

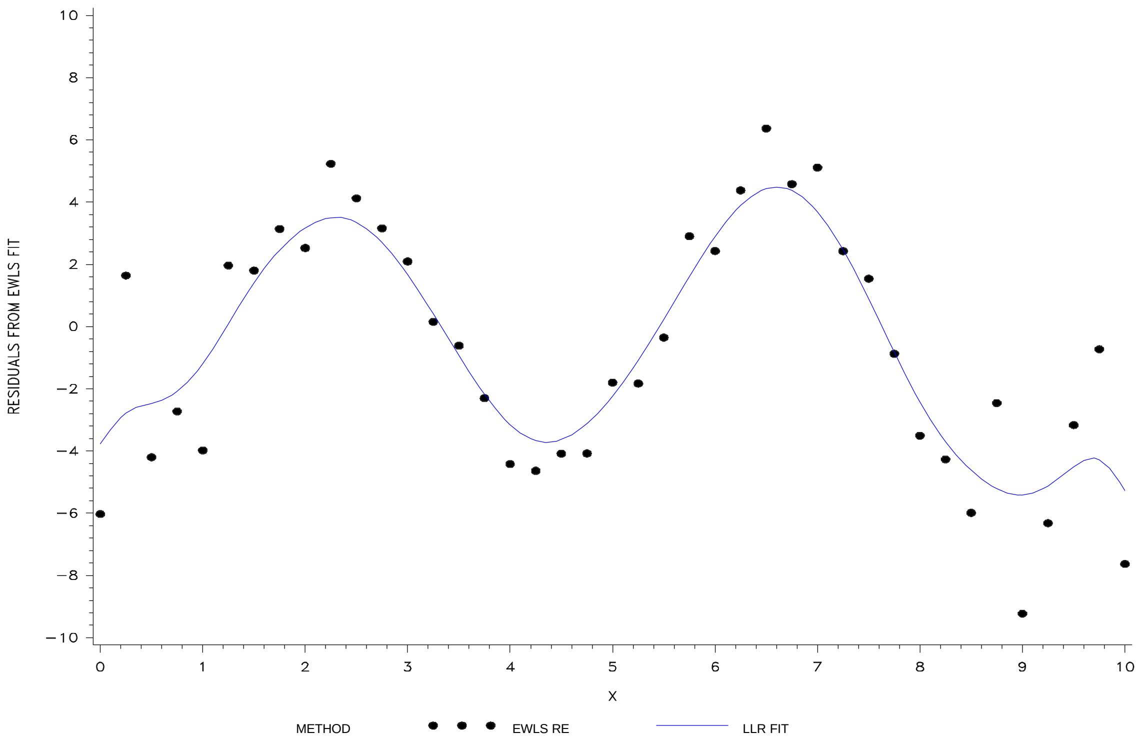

**Figure 6.B.12 LLR fit to the residuals from the OLS fit above.**

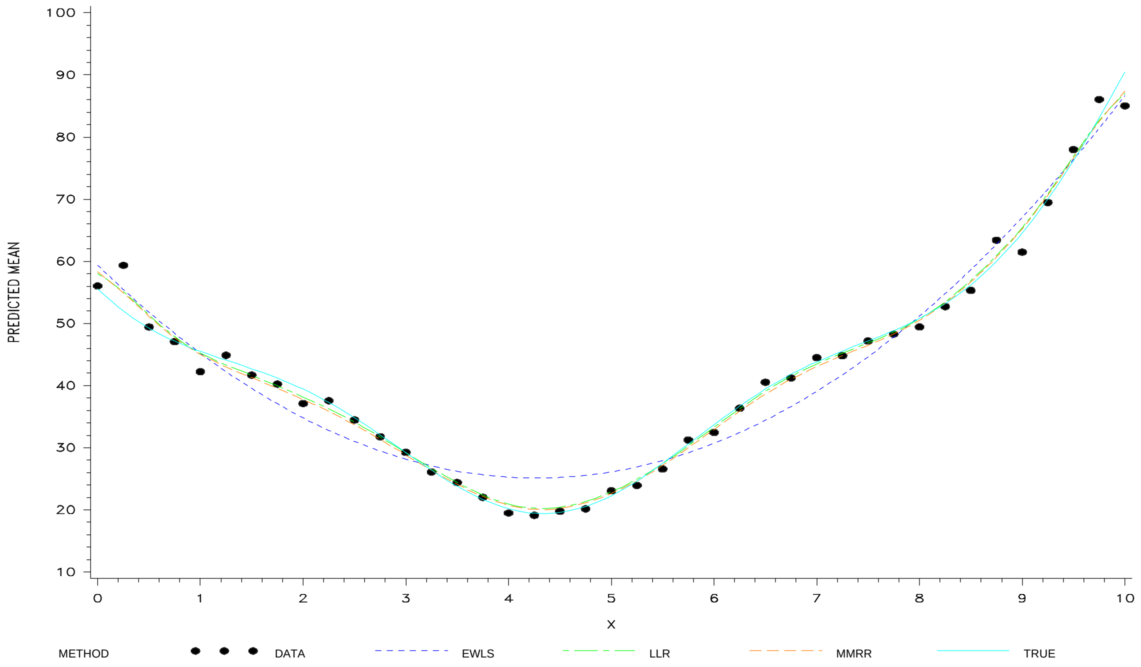

**Figure 6.B.13. Plot of the true mean function in Example 1 along with the EWLS, LLR, and MMRR estimates of the mean.**

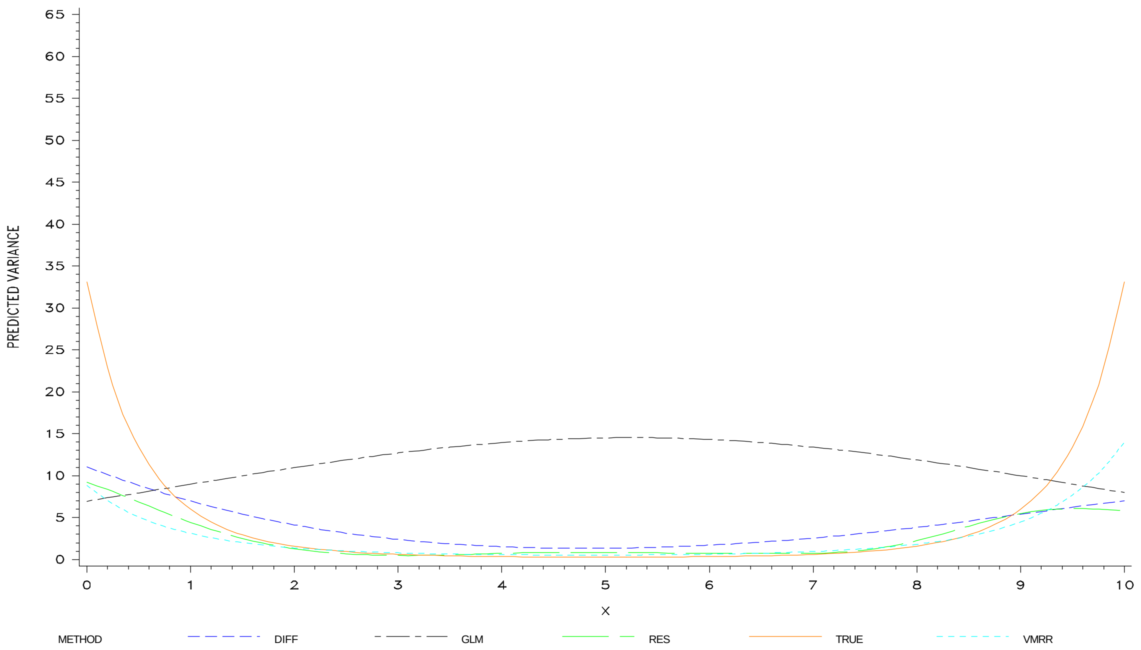

**Figure 6.B.14. Plot of the true variance function in Example 1 along with the GLM, Difference**

Numerical Comparisons

The observations made so far in Example 1 have been informative but they can only be considered to be preliminary comparisons since they have been based on only one of infinitely many data sets that can be generated from the models given by 6.B.1 and 6.B.2. For a more complete comparison it is necessary to quantify the performances of the procedures through the MMIMSE and VMIMSE values for various sample sizes and different degrees of means model misspecification. Recall that the MMIMSE and VMIMSE values are not data dependent but depend only on the true underlying models. In this example three sample sizes are considered ( $n$ = 21, 41 and 61) and means model misspecifcation is considered for five different levels of $\gamma$ ( $\gamma$ = 0, 2.5, 5.0, 7.5 and 10.0).

During the course of comparing procedural performances using the MMIMSE and VMIMSE criteria, we will also provide a check for the accuracy of these values. Recall that the MMIMSE and VMIMSE values are only obtained via asymptotic formulas. Thus, Monte Carlo simulations are performed to check the appropriateness of the asymptotic formulas for small samples (For all Monte Carlo simulations presented in this research, MC = 300 data sets.). The simulated integrated mean square error for a particular estimate is calculated using the following steps. For each of the MC = 300 data sets, the mean and variance fits will be determined at 1000 $x_o$ (recall $x_o$ = $z_o$ ) locations. After obtaining these fits ( the $\hat{y}_i$ 's and the $\hat{\sigma}_i$ 's), compute the average squared error in the means fit (asem) and the average squared error in the variance fit (asev) as

$$\text{asem} = \frac{\sum_{i=1}^{n} \left(\mathrm{E}\left(\mathrm{y}_i\right) - \hat{\mathrm{y}}_i\right)^2}{1000} \tag{6.B.6}$$

$$\text{asev} = \frac{\sum_{i=1}^{n} \left(\sigma_i^2 - \hat{\sigma}_i^2\right)^2}{1000} \tag{6.B.7}$$

for each of the 300 simulated data sets (i.e. get $\text{asem}_j$ and $\text{asev}_j$ for $j = 1,2, \cdots, 300)$. The Monte Carlo simulated integrated mean square error for the mean and variance estimates is then given by

$$\text{SIMMSE}_\mu = \frac{\sum_{j=1}^{300} \text{asem}_j}{300} \tag{6.B.8}$$

$$\text{SIVMSE}_\sigma = \frac{\sum_{j=1}^{300} \text{asev}_j}{300}. \tag{6.B.9}$$

It is important to note that in the nonparametric and model robust procedures, the values of *b* and $\lambda$ (when applicable) are fixed at their optimal values during the simulations.

Table 6.B.1 gives the results for the simulated MSE values along with the theoretical integrated mean square error values. Table 6.B.2 provides the optimal values of *b* and $\lambda$, upon which the simulated values are based. Notice that, in many cases, the simulated integrated mean square error values are quite close to the theoretical integrated mean square error values. It should be noted that the theoretical formulas are asymptotic and thus, their accuracy tends to improve for larger sample sizes.

Several observations can be made from Table 6.B.1. When there is no means model misspecification ($\gamma = 0$), the MMRR means estimate is close to the EWLS means estimate, and both are substantially better than the LLR means estimate. However, as the means model becomes misspecified ($\gamma > 0$), the MMRR means estimate is dramatically better than the EWLS fit. Notice that as the means model becomes more and more misspecified, the means model optimal bandwidth (from Table 6.B.2) becomes smaller and the corresponding means model mixing parameter gets closer to 1. This is necessary because the data being fit is becoming more complex and thus, more structure exists in the EWLS residuals. To capture this structure, a small bandwidth is needed and subsequently, a large mixing parameter is required to add this structure back to the EWLS fit.

The same sort of comparisons also hold true when observing the variance estimates. For $\gamma = 0$, the VMRR variance estimate is identical to the parametric GLM estimate. However, the VMRR variance estimate is dramatically better than the GLM estimate when there is misspecification in the means model. This is due to the fact that the VMRR procedure relies on data (the squared MMRR residuals) that is essentially free from bias whereas the GLM variance estimate utilizes data (the squared EWLS residuals) which is plagued with bias from the means model fit. It is useful to provide a plot which shows the difference in the variance model data used by the two procedures. Figure 6.B.15 shows a plot of the expected values of the squared MMRR residuals and the squared EWLS residuals. The expected values were calculated from the formulas derived in Appendix C where the underlying dual model was assumed to be the one given in (6.B.1) and (6.B.2) with $\gamma = 5.0$ and $n = 41$. Notice that this plot shows what the data looks like, on the average, for the GLM and VMRR procedures. The squared MMRR residuals clearly give a more accurate representation of the underlying variance function than do the squared EWLS residuals.

This plot also illustrates the point brought up earlier regarding the poor performance of the VMRR estimate at the boundaries. Since MMRR relies partially on local linear regression in its estimate of the mean, and local linear regression tends to fit closely to data located at the boundaries, we expect small residuals at those boundary points. In this plot we see that the squared residuals at the boundaries are quite small compared to the trend in the squared residuals just before the boundaries. Although the VMRR estimate could be better at the boundaries, it is still by far the best estimate of variance in terms of IVMSE.

Another important point that should be addressed concerns the accuracy of the asymptotic formulas for the difference-based variance estimate. Notice that the integrated mean squared error for the variance estimate, as calculated from the expressions for mean squared error given by Müller and Stadtmüller, grossly underestimates the true integrated mean squared error (represented by the simulated IVMSE) of the difference-based variance estimate. Notice that the theoretical IVMSE values become closer to the true IVMSE values for larger sample sizes, but even for the largest sample size considered ($n$=61), the theoretical values are far from accurate.

**Table 6.B.1 Simulated mean square error values for optimal mean and variance fits from 300 Monte Carlo runs (means model misspecified, variance model correctly specified). Theoretical MMIMSE and VMIMSE values are in bold.**

| γ | OLS | PARAMETRIC | | DIFF-BASED | | RES-BASED | | DMRR | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Mean | VAR | Mean | VAR | Mean | VAR | Mean | VAR |
| *n* = 21 | | | | | | | | | |
| | | | | | | | | | |
| 0.0 | **1.35398**<br>1.37878 | **.39407**<br>.45808 | **29.9596**<br>61.4168 | **2.43217**<br>2.46754 | **75.3459**<br>233.875 | **2.43217**<br>2.46754 | **42.5676**<br>44.9409 | **.39407**<br>.45808 | **30.2734**<br>61.4168 |
| 2.5 | **4.10819**<br>4.11166 | **3.58887**<br>3.93084 | **41.3451**<br>54.6749 | **2.50996**<br>2.52663 | **75.3459**<br>278.268 | **2.50996**<br>2.52663 | **47.2060**<br>45.8735 | **2.51587**<br>2.64742 | **28.0423**<br>33.8994 |
| 5.0 | **12.3708**<br>12.3530 | **12.2346**<br>12.6707 | **194.745**<br>246.660 | **2.61931**<br>2.62218 | **75.3459**<br>344.458 | **2.61931**<br>2.62218 | **54.3095**<br>48.4562 | **2.74212**<br>2.77465 | **38.0195**<br>39.9176 |
| 7.5 | **26.1419**<br>26.1027 | **26.3499**<br>26.5150 | **796.099**<br>788.744 | **2.71968**<br>2.71881 | **75.3459**<br>437.515 | **2.71968**<br>2.71881 | **56.6704**<br>50.8455 | **2.72568**<br>2.74096 | **44.4470**<br>45.2852 |
| 10.0 | **45.4215**<br>45.5090 | **45.9967**<br>45.9239 | **2369.71**<br>2695.17 | **2.83772**<br>2.83195 | **75.3459**<br>497.330 | **2.83772**<br>2.83195 | **55.6576**<br>53.0691 | **2.84135**<br>2.84396 | **49.5122**<br>49.9357 |
| | | | | | | | | | |
| *n* = 41 | | | | | | | | | |
| | | | | | | | | | |
| 0.0 | **0.64147**<br>0.67893 | **0.19973**<br>0.21532 | **18.5019**<br>29.8169 | **1.44325**<br>1.50883 | **42.3447**<br>75.0863 | **1.44325**<br>1.50883 | **27.5035**<br>32.3668 | **0.19973**<br>0.21532 | **18.5310**<br>29.8169 |
| 2.5 | **3.37926**<br>3.41761 | **3.24663**<br>3.41848 | **31.4967**<br>38.0665 | **1.50893**<br>1.56864 | **42.3447**<br>79.2287 | **1.50893**<br>1.56864 | **28.8624**<br>32.9330 | **1.40214**<br>1.46651 | **17.9955**<br>19.3916 |
| 5.0 | **11.5926**<br>11.6319 | **11.5290**<br>11.6595 | **179.027**<br>189.171 | **1.62415**<br>1.67862 | **42.3447**<br>84.3164 | **1.62415**<br>1.67862 | **31.3140**<br>34.3424 | **1.57524**<br>1.64514 | **21.2146**<br>22.0281 |
| 7.5 | **25.2816**<br>25.3217 | **25.3762**<br>25.4427 | **758.913**<br>788.744 | **1.72866**<br>1.78150 | **42.3447**<br>90.9825 | **1.72866**<br>1.78150 | **33.7485**<br>36.0487 | **1.70066**<br>1.78083 | **24.1884**<br>24.9416 |
| 10.0 | **44.4461**<br>44.4871 | **44.7468**<br>44.7503 | **2280.87**<br>2352.73 | **1.82160**<br>1.87462 | **42.3447**<br>97.8776 | **1.82160**<br>1.87462 | **35.9420**<br>37.6622 | **1.80137**<br>1.88952 | **26.8342**<br>27.5846 |
| | | | | | | | | | |
| *n* = 61 | | | | | | | | | |
| | | | | | | | | | |
| 0.0 | **0.41819**<br>0.42123 | **0.13381**<br>0.13910 | **13.2305**<br>18.1739 | **1.04773**<br>1.08708 | **31.2596**<br>45.8120 | **1.04773**<br>1.08708 | **22.5681**<br>26.6730 | **0.13381**<br>0.13910 | **13.2303**<br>18.1739 |
| 2.5 | **3.15293**<br>3.15488 | **3.14436**<br>3.23887 | **15.5038**<br>31.7697 | **1.10297**<br>1.14175 | **31.2596**<br>46.4028 | **1.10297**<br>1.14175 | **23.3054**<br>27.0620 | **1.00613**<br>1.00727 | **12.9956**<br>13.3556 |
| 5.0 | **11.3571**<br>11.3580 | **11.3207**<br>11.4093 | **150.506**<br>178.926 | **1.19882**<br>1.23687 | **31.2596**<br>47.1369 | **1.19882**<br>1.23687 | **24.6408**<br>28.1210 | **1.15793**<br>1.16346 | **14.7046**<br>16.1578 |
| 7.5 | **25.0308**<br>25.0306 | **25.0991**<br>25.1816 | **699.534**<br>776.036 | **1.28622**<br>1.32399 | **31.2596**<br>48.0289 | **1.28622**<br>1.32399 | **26.0235**<br>29.3105 | **1.26242**<br>1.35533 | **16.4396**<br>16.7218 |
| 10.0 | **44.1740**<br>44.1727 | **44.4087**<br>44.4565 | **2169.89**<br>2325.19 | **1.36227**<br>1.39997 | **31.2596**<br>49.0952 | **1.36227**<br>1.39997 | **27.3444**<br>30.4471 | **1.34523**<br>1.44020 | **18.0359**<br>18.3402 |
| | | | | | | | | | |

**Table 6.B.2 Optimal bandwidths and mixing parameters chosen by minimizing the AVEMMSE and AVEVMSE values for the nonparametric and model robust procedures (means model misspecified, variance model correctly specified).**

| $\gamma$ | DMRR | | | | Diff-Based | | Res-Based | |
|---|---|---|---|---|---|---|---|---|
| | Mean | | VAR | | Mean | VAR | Mean | VAR |
| *n* = 21 | $b_{\mu_o}$ | $\lambda_{\mu_o}$ | $b_{\sigma_o}$ | $\lambda_{\sigma_o}$ | $b_{\mu_o}$ | $b_{\sigma_o}$ | $b_{\mu_o}$ | $b_{\sigma_o}$ |
| 0.0 | 1.0 | 0.0 | 0.9356 | 0.0 | 0.0839 | 0.4530 | 0.0839 | 0.1266 |
| 2.5 | 0.11477 | 0.55107 | 0.76875 | 0.0 | 0.07637 | 0.4530 | 0.07637 | 0.12317 |
| 5.0 | 0.07024 | 0.92378 | 0.10977 | 0.0 | 0.06137 | 0.4530 | 0.06137 | 0.10423 |
| 7.5 | 0.05449 | 0.99402 | 0.09984 | 0.0 | 0.05025 | 0.4530 | 0.05025 | 0.08554 |
| 10.0 | 0.04613 | 1.0 | 0.09586 | 0.0 | 0.04371 | 0.4530 | 0.04371 | 0.07148 |
| | | | | | | | | |
| *n* = 41 | | | | | | | | |
| | | | | | | | | |
| 0.0 | 1.0 | 0.0 | 0.91130 | 0.0 | 0.0722 | 0.2969 | 0.0722 | 0.1008 |
| 2.5 | 0.08828 | 0.83774 | 0.82625 | 0.0 | 0.06574 | 0.2969 | 0.06574 | 0.09562 |
| 5.0 | 0.05941 | 1.0 | 0.73250 | 0.0 | 0.05434 | 0.2969 | 0.05434 | 0.08249 |
| 7.5 | 0.04842 | 1.0 | 0.70125 | 0.0 | 0.04629 | 0.2969 | 0.04629 | 0.07196 |
| 10.0 | 0.04213 | 1.0 | 0.05707 | 0.0 | 0.04099 | 0.2969 | 0.04099 | 0.06423 |
| | | | | | | | | |
| *n* = 61 | | | | | | | | |
| | | | | | | | | |
| 0.0 | 1.0 | 0.0 | 0.90375 | 0.0 | 0.06551 | 0.22324 | 0.06551 | 0.08861 |
| 2.5 | 0.07766 | 1.0 | 0.85125 | 0.0 | 0.05961 | 0.22324 | 0.05961 | 0.08272 |
| 5.0 | 0.05367 | 1.0 | 0.8125 | 0.0 | 0.04941 | 0.22324 | 0.04941 | 0.07175 |
| 7.5 | 0.04410 | 1.0 | 0.80125 | 0.0 | 0.04238 | 0.22324 | 0.04238 | 0.06335 |
| 10.0 | 0.03857 | 1.0 | 0.7925 | 0.0 | 0.03769 | 0.22324 | 0.03769 | 0.05716 |
| | | | | | | | | |

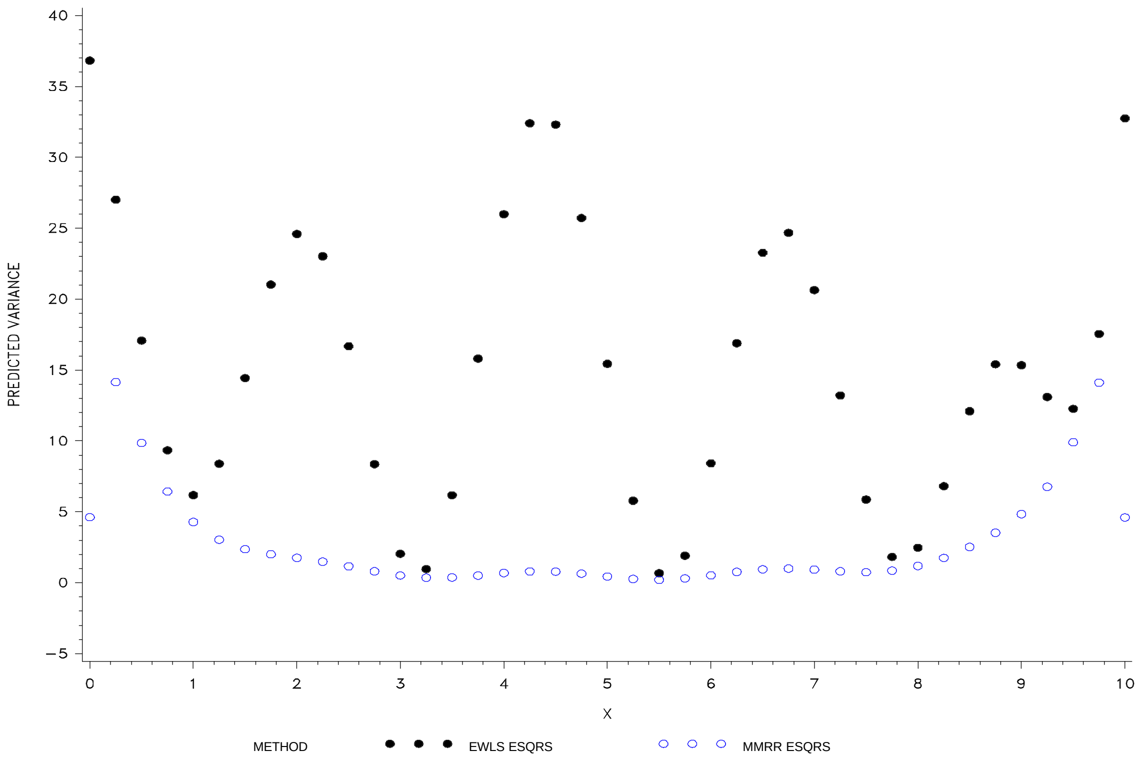

**Figure 6.B.15. Plot of $E\left(e_i^{2\,(ewls)}\right)$ and $E\left(e_i^{2\,(mmrr)}\right)$ for Example 1 when $\gamma = 5.0$ and $n = 41$.**

## 6.C Example 2 (Variance Model Misspecification)

Introduction

In Example 1, dual model robust regression (DMRR) was observed to be robust to the user's specification of the form of the true underlying means function. The parametric procedure did well when the researcher assumed the correct functional form for the means model but the quality of the parametric performance suffered when the researcher's assumed form became inadequate across the range of the data (when $\gamma > 0$ ). The nonparametric procedures were far superior to the parametric methodology when the true underlying function became more complex, but for $\gamma$ = 0.0, they were far less appealing than their parametric competitor. The DMRR procedure however, was observed to be superior in terms of IMMSE and IVMSE for the entire range of $\gamma$ considered.

Since variance model specification often relies on scatter plots of the residuals from the means fit, it is quite likely that misspecification of the mean will also lead to variance model misspecification. In this section, we will explore the impact of variance model mispecification on the dual model analysis. This study will be conducted in two phases. First, we will assume that the user has correctly specified the means function but misspecified the variance model. Next, we will consider the case in which the user has misspecified both the mean and variance models. Discussion will follow the same type of format as that used in Example 1 where the dual modeling procedures will first be compared graphically for a given data set. Then the procedures will be compared in a more general sense based on their theoretical and simulated integrated mean square error values. The underlying dual model used in this study will be the same as the one used in Example 1, written again here as

$$y_i = 2\left(x_i - 5.5\right)^2 + 5\,x_i + \gamma \sin\left(\frac{\pi\left(x_i - 1\right)}{2.25}\right) + g^{1/2}\left(\mathbf{z}_i\right)\varepsilon_i, \qquad (6.C.1)$$

$$\sigma_i^2 = g^{1/2}\left(\mathbf{z}_i\right) = \exp\left\{3.0 - 1.9\,z_i - 0.19\,z_i^2\right\}. \qquad (6.C.2)$$

where $\gamma \in [0,1]$ and $\varepsilon_i \sim N(0,1)$.

Variance Model Misspecification

The first phase of this study concentrates on the scenario in which the researcher has correctly specified the mean's model but misspecified the variance model. Assuming the underlying models in (6.C.1) and (6.C.2) with $\gamma = 0.0$, suppose that the researcher prescribes the following dual model for the process:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + g^{1/2}\left(z_i;\boldsymbol{\theta}\right)\varepsilon_i \qquad (6.C.2)$$

$$\sigma_i^2 = g^{1/2}\left(z_i;\boldsymbol{\theta}\right) = \exp\left\{\theta_0 + \theta_1 z_i\right\}. \qquad (6.C.3)$$

Notice that the user has correctly specified a quadratic means model and the functional form of the variance model has also been correctly specified. However, the user's variance misspecification is due to failure to include the quadratic term in the argument of the exponential function (the $z_i^2$ term). It is important to note that the nonparametric dual modeling procedures will be unaffected by this mistake on the user's part because they, in no way, rely on user-supplied knowledge to form their estimate. Thus, the key issue in this study will be how well the parametric and DMRR procedures perform under this misspecification and how they compare with the two nonparametric procedures. As we will see, DMRR is robust to this misspecification on the user's behalf and its performance is still superior to that of its competitors.

To begin this study, consider a random data set of 21 equally spaced points from 0 to 10, generated from the models in (6.C.1) and (6.C.2) where $\gamma = 0.0$. A scatter plot of the raw data is presented in Figure 6.C.1. The EWLS means fit is shown in Figure 6.C.2 (a) along with the true underlying means function and raw data. The corresponding parametric variance estimate is pictured in Figure 6.C.3 along with the squared EWLS residuals and true underlying variance function. Notice that the EWLS means fit captures the true structure of the process mean but the parametric variance estimate is unable to pick up the quadratic structure that is present in the true underlying variance function. Instead, the parametric procedure offers an estimate of variance that is constant across the range of the data. This is clearly due to the omission of the quadratic term in the user's variance model. Without any structure in the variance estimate, EWLS provides essentially the same fit as ordinary least squares (OLS) , which ignores variance heterogeneity.

The two nonparametric dual modeling procedures being considered use local linear regression to estimate the underlying means function. Figure 6.C.4 displays the LLR means fit (based on $b_{\mu_o} = 0.0839$) along with the true underlying means function and the raw data. Notice that LLR fits too close to the data located at the boundaries of the data set. The nonparametric difference-based variance estimate (based on $b_{\sigma_o} = 0.453$) is pictured in Figure 6.C.5 along with the squared pseudo-residuals and underlying variance function. Notice that the difference-based variance estimate follows the trend of the underlying variance function, and, if the whole estimated curve were shifted down, it would match closely with the true underlying variance function.

The plot in Figure 6.C.5 however, illustrates a flaw that is associated with difference-based variance estimation. Notice that the only squared pseudo-residuals that are not pictured in Figure 6.C.4 are those located at the endpoints of the data. This is due to the large magnitude of those squared pseudo-residuals. Recall that the pseudo-residuals at the endpoints are only based on a neighborhood of two points as opposed to the three point basis used for the data's interior (see Section 3.C.1). As a consequence, these endpoint pseudo-residuals are often highly unstable, and consequently, they often adversely effect the overall variance estimate.

The residual-based variance estimate (based on $b_{\sigma_o} = 0.12663$) is given in Figure 6.C.7 . This estimate also suffers at the boundaries of the data set. As mentioned in Example 1, the problem with residual-based variance estimation is that it depends on the assumption that variance information lies within the residuals from the LLR means fit. That is, large residuals are assumed to imply large variance and small residuals imply small variance. However, inherent to LLR is its tendency to fit too closely to data located at the boundaries. Thus, the residuals at the endpoints are often quite small, suggesting low variability regardless of what the true variance is at those points.

The MMRR means estimate (based on $b_{\mu_o} = 0.7225$ and $\lambda_{\mu_o} = 0.0$) is pictured in Figure 6.C.7 along with the true means function and raw data. In Figure 6.C.8 all of the competing estimates of the mean (EWLS, LLR, and MMRR) are plotted along with the raw data and underlying means function. Notice that like its parametric and nonparametric counterparts, MMRR captures the general structure of the true process mean function. However there are key differences which exist between the estimates. Observe in Figure 6.C.8 that, unlike the LLR fit, MMRR is does not fit too closely to data located at the boundaries. In comparing the MMRR fit to the EWLS fit it is first

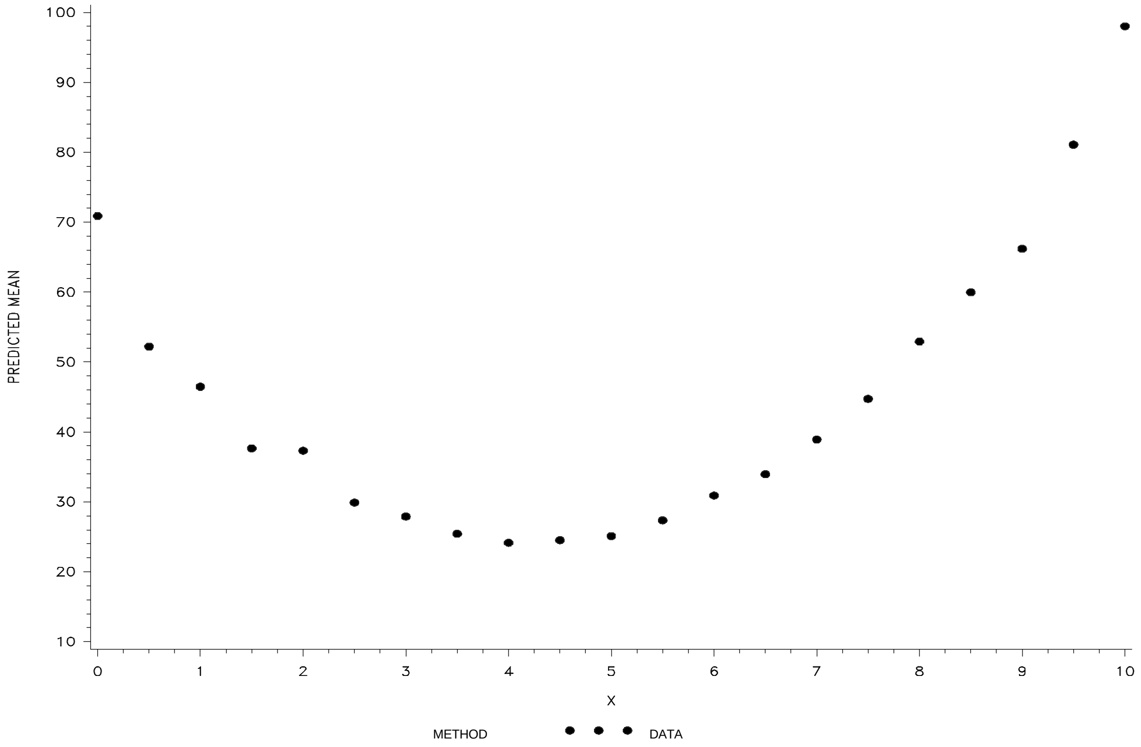

**Figure 6.C.1 Scatter plot of a random data set generated from the dual model given in equations (6.C.1) and (6.C.2) where $\gamma$ = 0.0.**

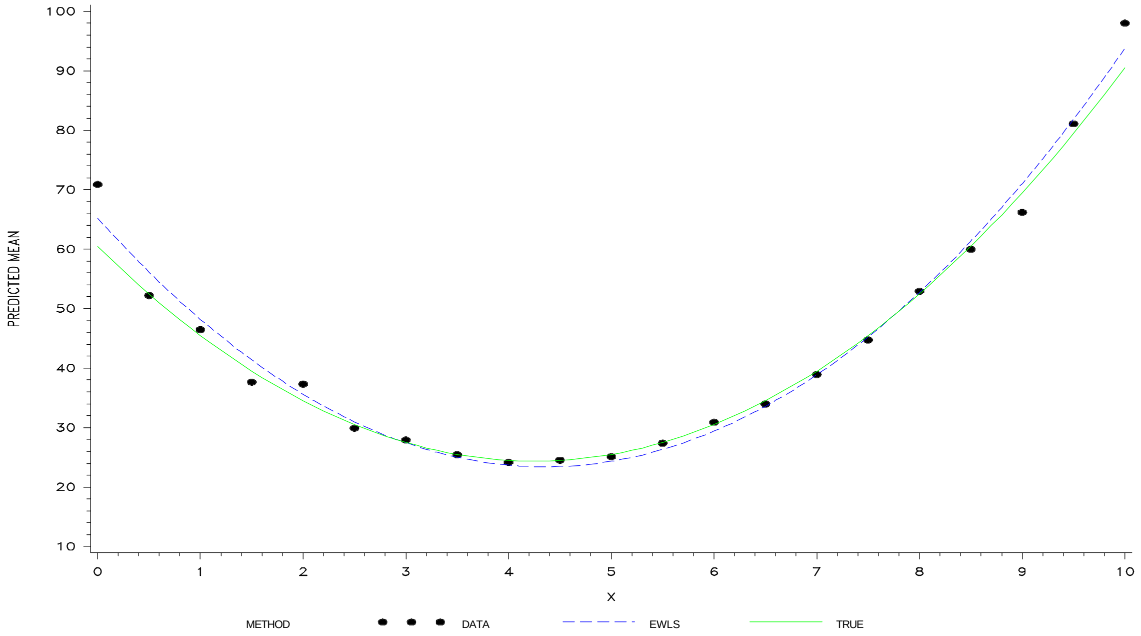

**Figure 6.C.2. Plot showing generated data for Example 2 along with EWLS fit and the true means function.**

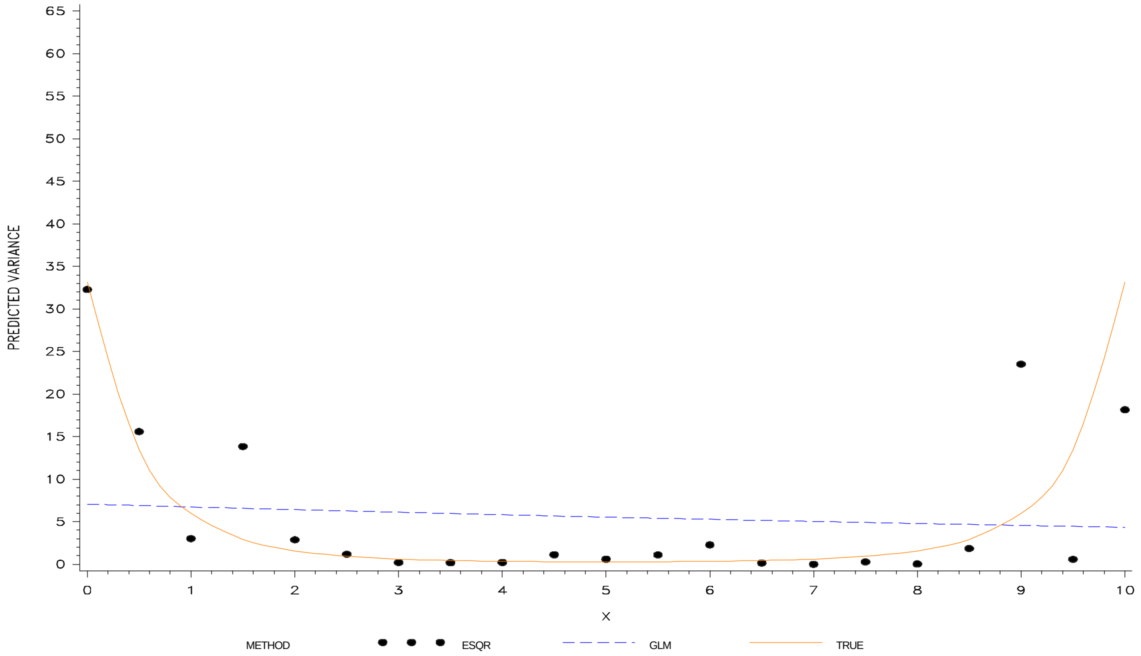

**Figure 6.C.3. Plot showing the EWLS squared residuals, the GLM variance estimate and the true underlying variance function for Example 2.**

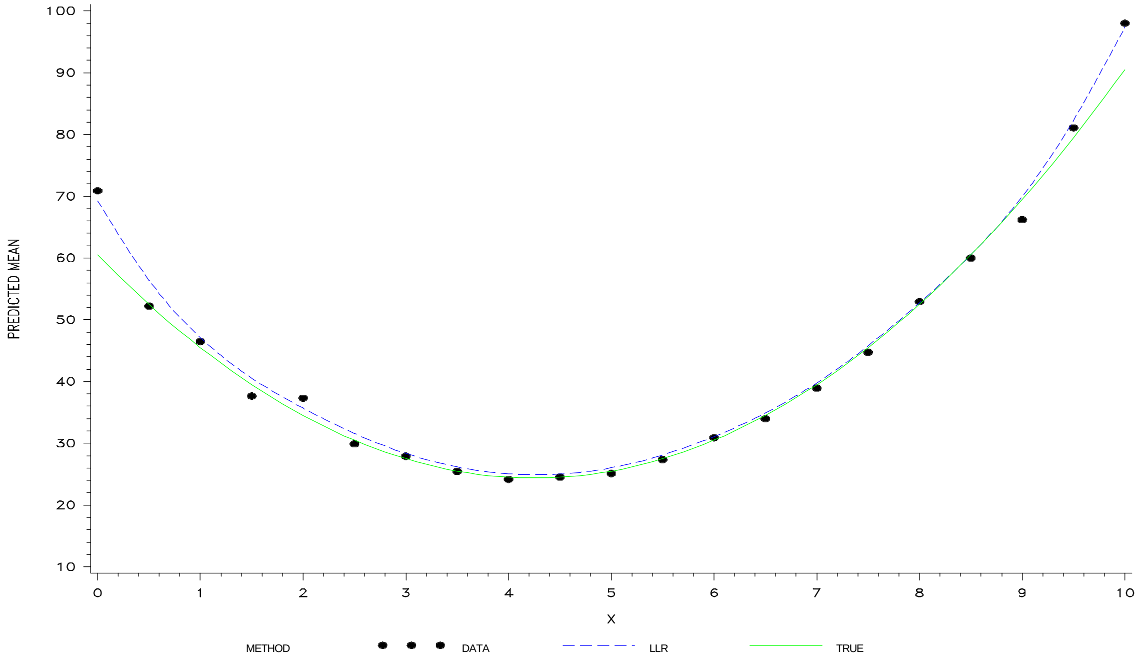

**Figure 6.C.4. Plot showing generated data for Example 2 along with LLR fit and the true means function.**

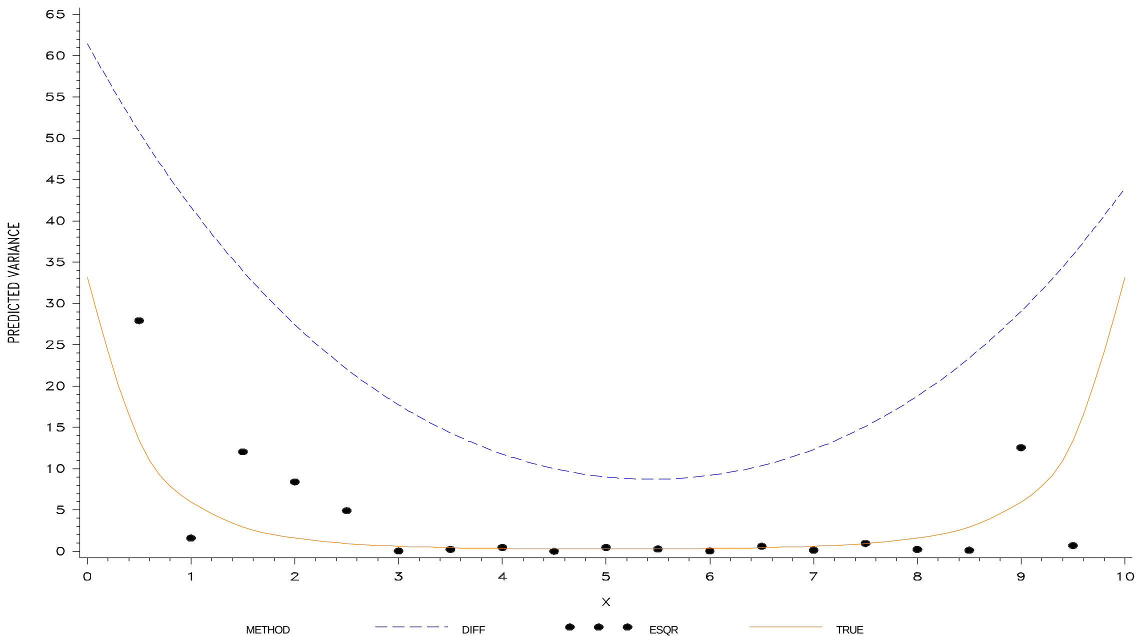

**Figure 6.C.5. Plot showing the squared pseudo-residuals, the difference-based estimate and true underlying variance function for Example 2.**

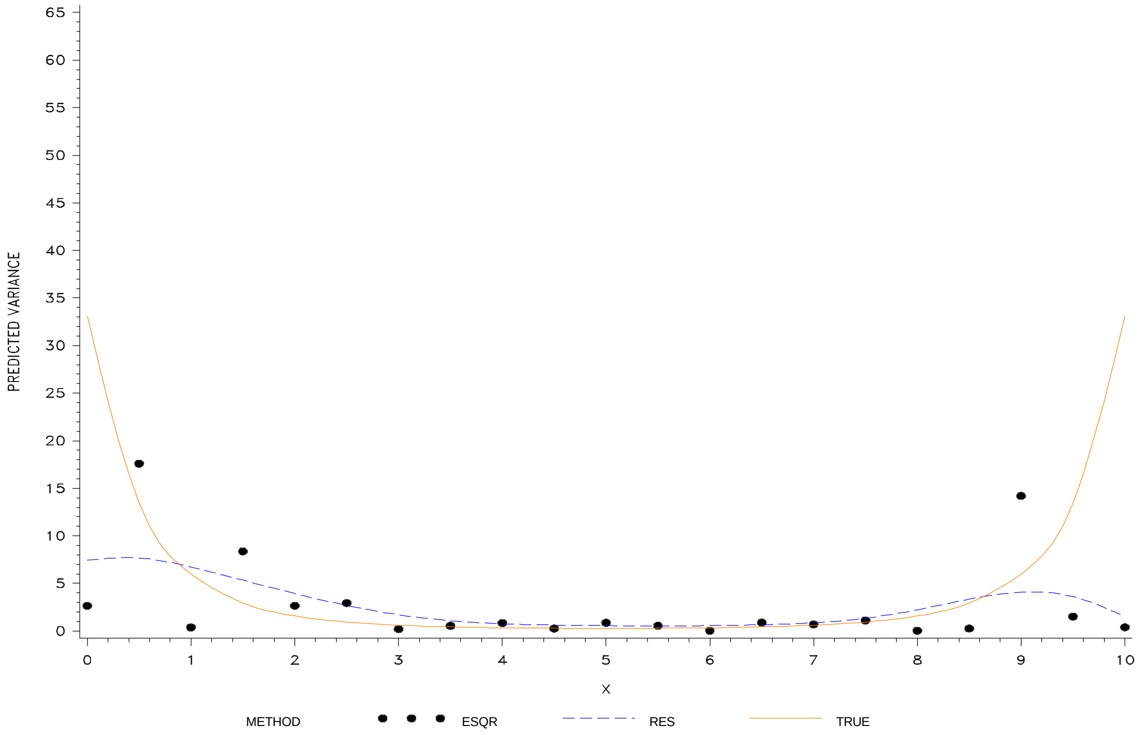

**Figure 6.C.6. Plot showing the squared local linear residuals, the residual-based estimate and true underlying variance function for Example 2.**

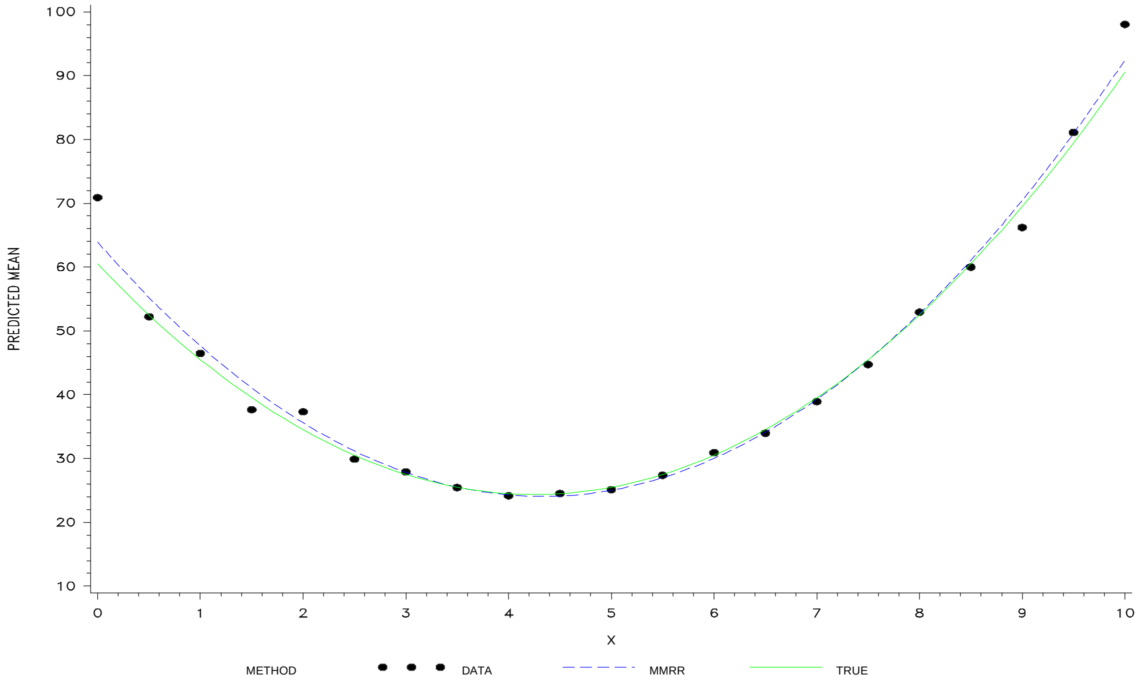

**Figure 6.C.7. Plot showing generated data for Example 2 along with MMRR fit and the true means function.**

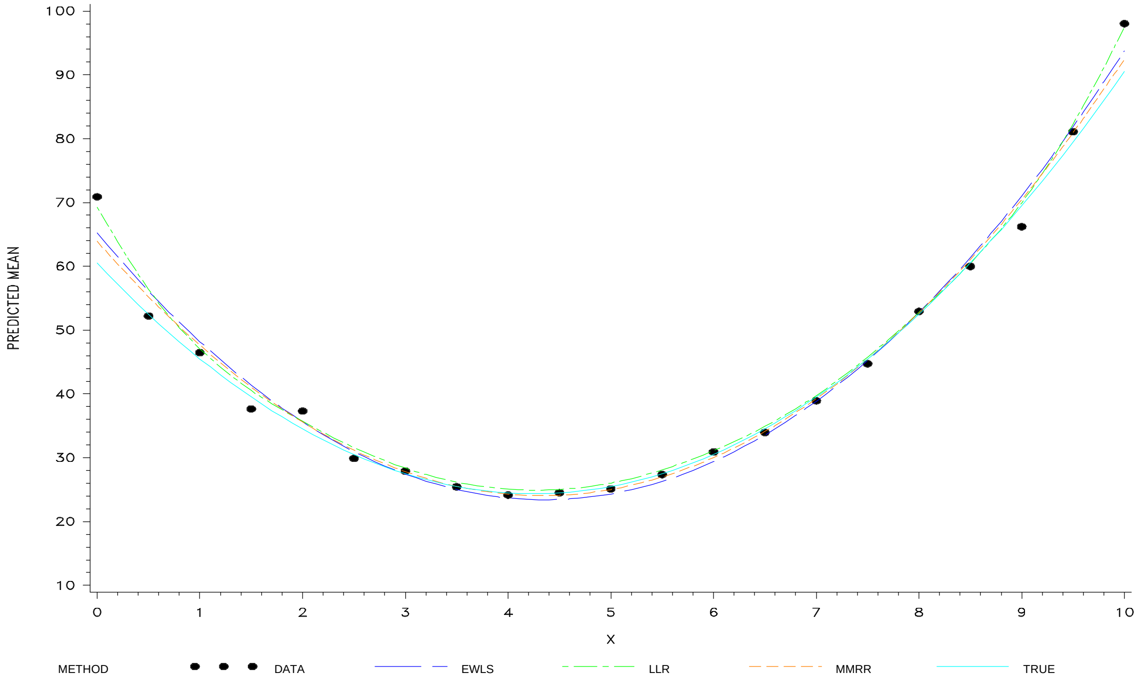

**Figure 6.C.8. Plot of true means function in Example 2 along with the EWLS, LLR, and MMRR estimates of the mean.**

important to recognize that with $\lambda_{\mu_o} = 0.0$, MMRR is simply an estimated weighted least squares mean estimate. However, the EWLS fit obtained in MMRR is based on the weights obtained from the model robust variance estimate (VMRR) whereas the EWLS fit in parametric dual modeling uses weights based on the parametrically estimated variance function.

The VMRR variance estimate (based on $b_{\sigma_o} = 0.35570$ and $\lambda_{\sigma_o} = 1.0$) is shown in Figure 6.C.9 along with the true variance function and the squared MMRR residuals. Notice that the quadratic structure, which was not specified by the researcher, has been captured by the VMRR estimate. For VMRR, recall that the first step is to obtain a parametric fit to the MMRR squared residuals; this is the fit pictured in Figure 6.C.10. Then, the residuals from this parametric fit are fit using local linear regression; this is the LLR fit plotted in Figure 6.C.10.. A certain portion ($\lambda_{\sigma_o} = 1.0$ here) of the fit to the residuals is then added back to the parametric fit to the data to give the final VMRR variance estimate. The final VMRR variance estimate is pictured along with the parametric, difference-based, and residual based variance estimates in Figure 6.C.12.

Comparing the VMRR fit of the variance function to the parametric GLM fit, it is now easy to understand why the EWLS portion of MMRR is a better fit to the means model than the EWLS fit obtained in parametric dual modeling. This difference is essentially the difference between ordinary least squares (OLS) and weighted least squares (WLS) when heteroscedasticity is present in a data set. When the model errors have non-constant variance, WLS is preferred over OLS because the WLS fit is obtained by giving less emphasis to those points which are associated with high variance. The result is a fit which is more precise than the OLS fit which ignores heteroscedasticity (Myers (1990)). In this example, the EWLS fit associated with parametric dual modeling uses the estimated variance function pictured in Figure 6.C.1 to determine it weights. Since the function pictured in Figure 6.C.1 is basically constant over the range of the data, the parametric means fit is essentially an OLS fit to the mean. In contrast, the EWLS fit in MMRR uses weights (based on the VMRR estimate) which closely resemble the true variances.

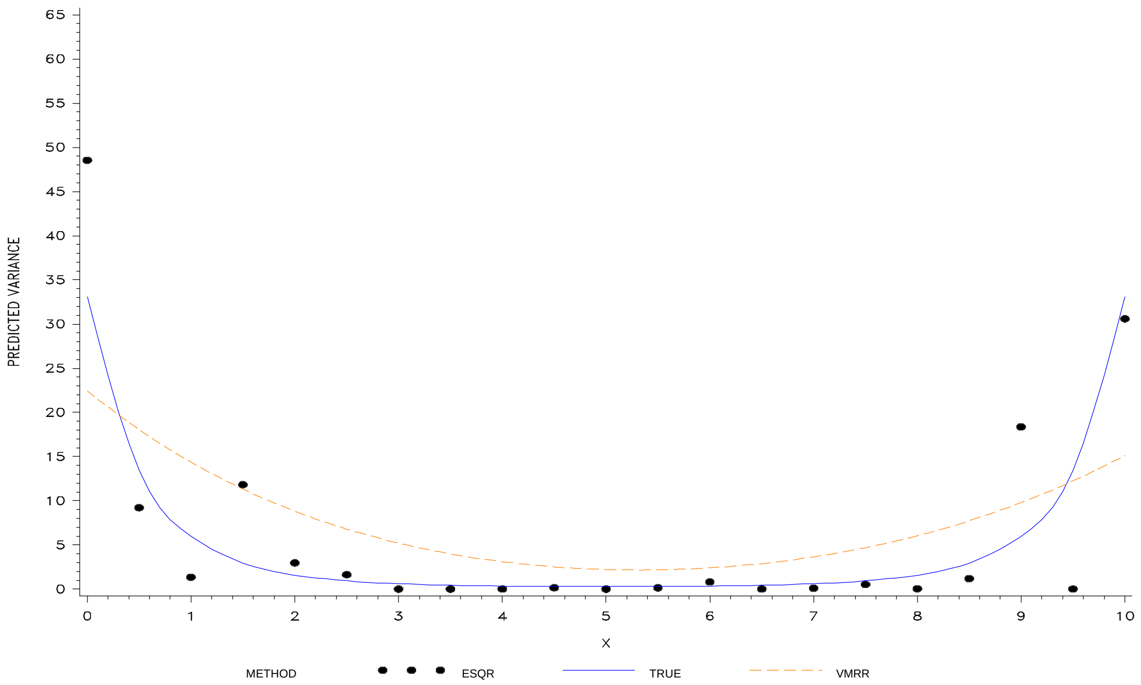

**Figure 6.C.9. Plot showing the MMRR squared residuals along with the VMRR estimate and true variance function for Example 2.**

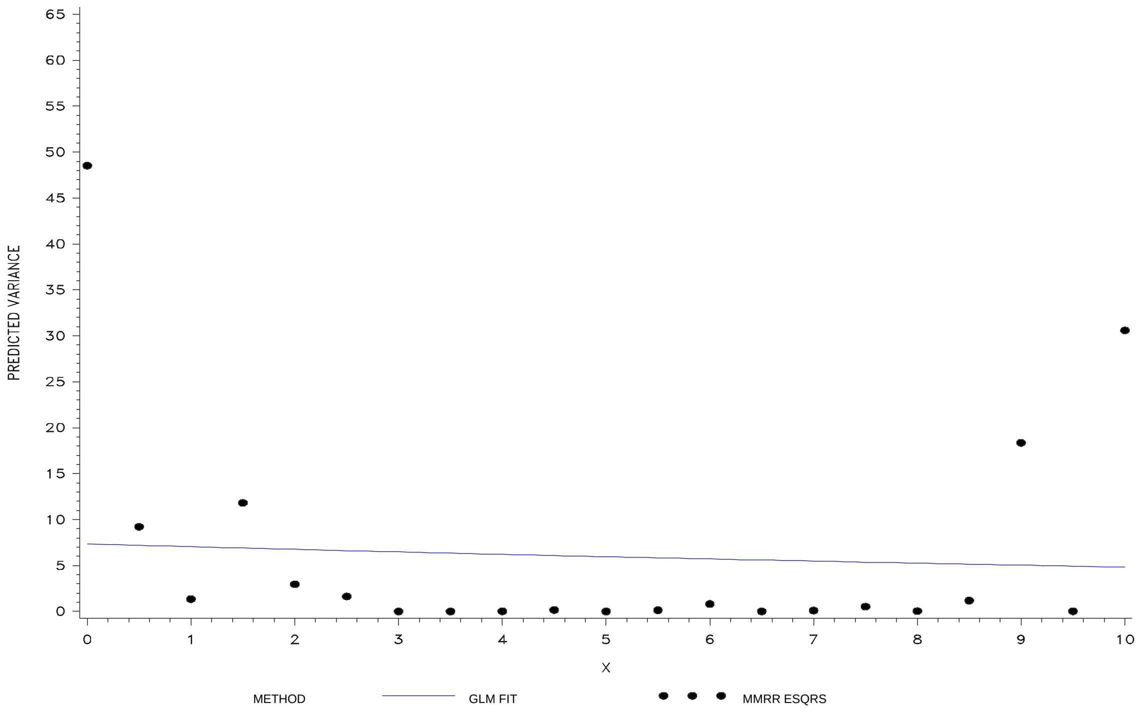

**Figure 6.C.10 Parametric GLM fit to the squared MMRR residuals from Example 2.**

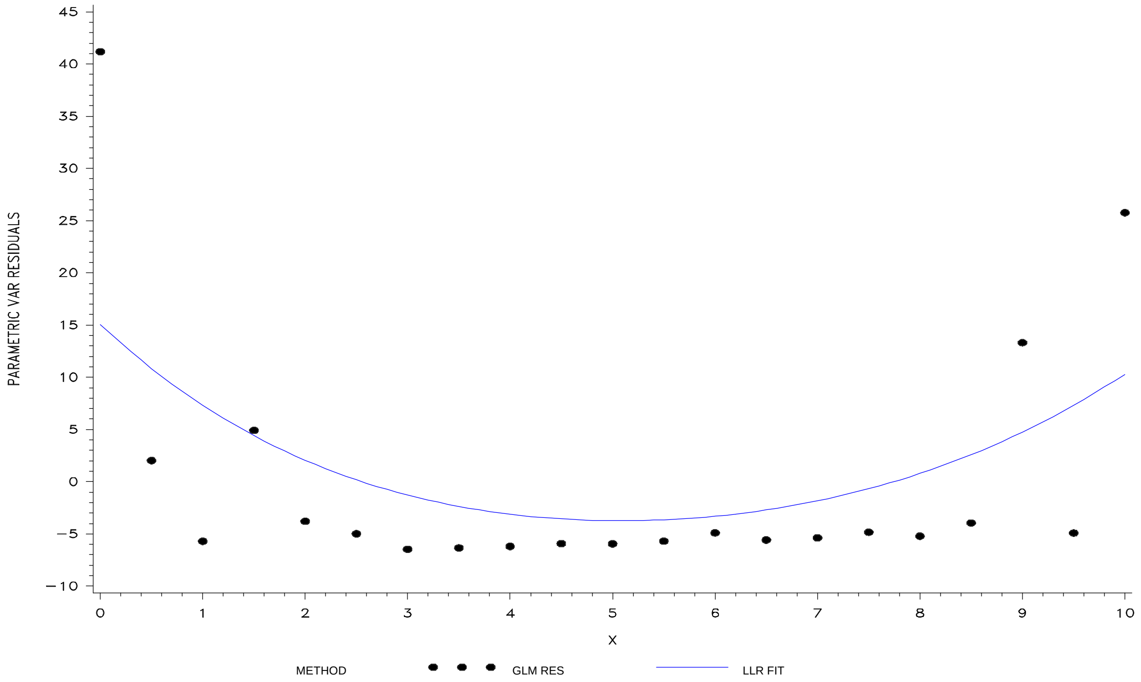

**Figure 6.C.11 LLR fit to residuals from the parametric GLM estimate displayed in Figure 6.C.10.**

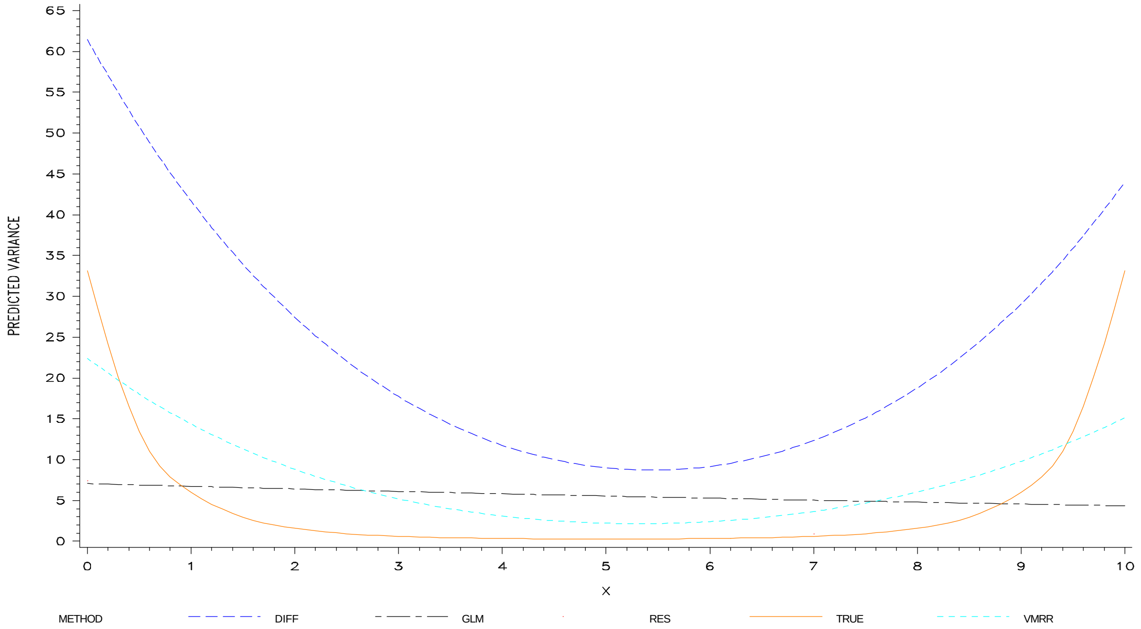

**Figure 6.C.12 Plot of the true variance function in Example 1 along with the GLM, Difference-Based, Residual-Based, and VMRR variance estimates ($\gamma = 0.0$).**

Tables 6.C.1 and 6.C.2 provides the numerical results of interest for this example. The rows of interest are the three rows that correspond to $\gamma = 0$ because when $\gamma = 0$ the user has correctly specified the means model. Notice from Table 6.C.2 that the optimal mixing parameter used in the VMRR variance estimate ($\lambda_{\sigma_o}$) is equal to 1. Recall from Table 6.B.2 that when the variance model was not misspecified the optimal variance model mixing parameter was $\lambda_{\sigma_o}$ = 0.0. This is a promising observation because it shows that the VMRR procedure mixes the parametric and nonparametric techniques when model misspecification is present and it relies only on the parametric variance estimate when the variance model is correctly specified.

By comparing the rows of Table 6.C.1 and Table 6.B.1 where $\gamma = 0$, we can observe the influence of variance model misspecification on each of the dual modeling procedures. Notice that misspecification of the variance function has a direct influence on the quality of the means estimate for both parametric and dual model robust estimation. The parametric IMMSE is approximately three times larger when the variance model has been misspecified (Table 6.C.1) than when it is not misspecified (Table 6.B.1). The IMMSE for DMRR is also larger when the variance function has been misspecified than when it is not misspecified. As expected, the IMMSE values for the nonparametric dual modeling procedures are the same regardless of what variance model the user has specified because these procedures do not rely on a user-specified model to estimate the variance. However, when comparing the IMMSE across all procedures it can be concluded that DMRR provides the best estimate of the mean. When comparing the IVMSE values for each of the dual modeling procedures, DMRR also performs better than the other procedures. It is important to note however that DMRR performs better when the user specifies the correct variance function than if the user misspecifies the underlying variance function. This can be observed by comparing the IVMSE values in Table 6.C.1 (variance misspecified) to those in Table 6.B.1 (variance correctly specified). As a final note, it can be concluded that although DMRR performs better when the user specifies the correct variance function, its performance is still better, and dramatically better in some cases, than the traditional procedures even when the user misspecfies the variance function.

Mean and Variance Model Misspecification

Now that the dual modeling procedures have been compared in situations where there is only means model misspecification or only variance model misspecification, it is important to look at the scenario in which the user misspecifies both functions. As mentioned earlier, this scenario is an important one because if the user misspecifies the means function, more than likely the variance function will also be misspecified since the residuals from the means fit, based on EWLS, often serve as building blocks for constructing a variance model. The underlying dual model is again taken to be the one given in equations (6.C.1) and (6.C.2) and we will assume that the user specifies the models given in (6.C.3) and (6.C.4).

For making graphical comparisons of the dual modeling procedures, a random data set of 41 equally spaced points from 0 to 10 will be generated from the models given in equations (6.C.1) and (6.C.2) with $\gamma = 2.5$. A scatter plot of the data set is pictured in Figure 6.C.13. It is useful to compare the scatter plot of this data set to the scatter plots in Figures 6.B.3 and 6.C.1.

Figure 6.B.3 represents a data set generated from a moderately contaminated quadratic function ($\gamma = 5.0$) and the data set represented in Figure 6.C.1 was generated from a non-contaminated quadratic function ($\gamma = 0.0$). The data set presently being considered was generated from a quadratic model that is only slightly ($\gamma = 2.5$) contaminated.

The EWLS fit to this data, as well as a plot of the true underlying means function are given in Figure 6.C.14. Notice that the EWLS fit captures the general trend of the underlying means function but it fails to capture the "dip" between x = 3.0 and x = 6.0. The corresponding parametric variance estimate is pictured in Figure 6.C.15 along with the squared EWLS residuals and true underlying variance function. As expected, the parametric variance estimate is poor in quality due to the user's omission of the quadratic term (the $z_i^2$ term). Note in the large squared EWLS residuals in the region from x = 3 to x = 6 in Figure 6.C.5. Recall that the EWLS means fit was unable to capture the "dip" in this region and as a result, large residuals are present there. Consequently, even with a correctly specified variance function, these large squared residuals in the data's interior could have had an adverse affect on the parametric variance estimate. This "bad data" phenomenon was discussed in Example 1.

The LLR estimate of the mean (based on $b_{\mu_o} = 0.06574$) is given in Figure 6.C.16. As mentioned earlier, LLR fits close to the data points but the fit is not very smooth (low bias, high variance), especially at the data's endpoints. The nonparametric variance estimates are provided in Figure 6.C.17 (the difference-based estimate, based on $b_{\sigma_o} = 0.29695$) and Figure 6.C.18 (the residual-based estimate, based on $b_{\sigma_o} = 0.08249$). Figure 6.C.17 illustrates a point made earlier concerning the instability of the squared pseudo-residuals at the boundary points. Even though the process variance is high at *both* endpoints of the data set, the squared pseudo-residuals only suggest high variance at the *left* endpoint of the data. In Figure 6.C.18 the residual-based variance estimate is observed to have the same boundary problems that were brought up earlier in this section.

The DMRR means estimate (based on $b_{\mu_o} = 0.08898$ and $\lambda_{\mu_o} = 0.85439$) and variance estimate (based on $b_{\sigma_o} = 0.27672$ and $\lambda_{\sigma_o} = 1.0$) are pictured in Figures 6.C.19 and 6.C.20, respectively. The DMRR means fit and variance fits are then plotted with their respective parametric and nonparametric counterparts in Figures 6.C.21 and 6.C.22. As expected, the robust means estimate captures the dip between x = 3 and x = 6 that EWLS was unable to capture and DMRR provides a slightly smoother (less variable) fit than LLR. While the model robust variance estimate (VMRR) is not outstanding at the data's endpoints, its overall fit is better than its competitors.

The numerical results of interest for this example are found in Tables 6.C.1 and 6.C.2. Notice in Table 6.C.2 that the optimal means model mixing parameter $(\lambda_{\mu_o})$ increases from 0 to 1 as the means model becomes more misspecified ($\gamma > 0$). This demonstrates that DMRR is mixing the parametric and nonparametric techniques as
the specified means function deviates more and more from the true underlying means function. Also, notice that the same observation can be made regarding the optimal variance model mixing parameter $(\lambda_{\sigma_o})$. Except for the small sample ($n$ = 21), highly

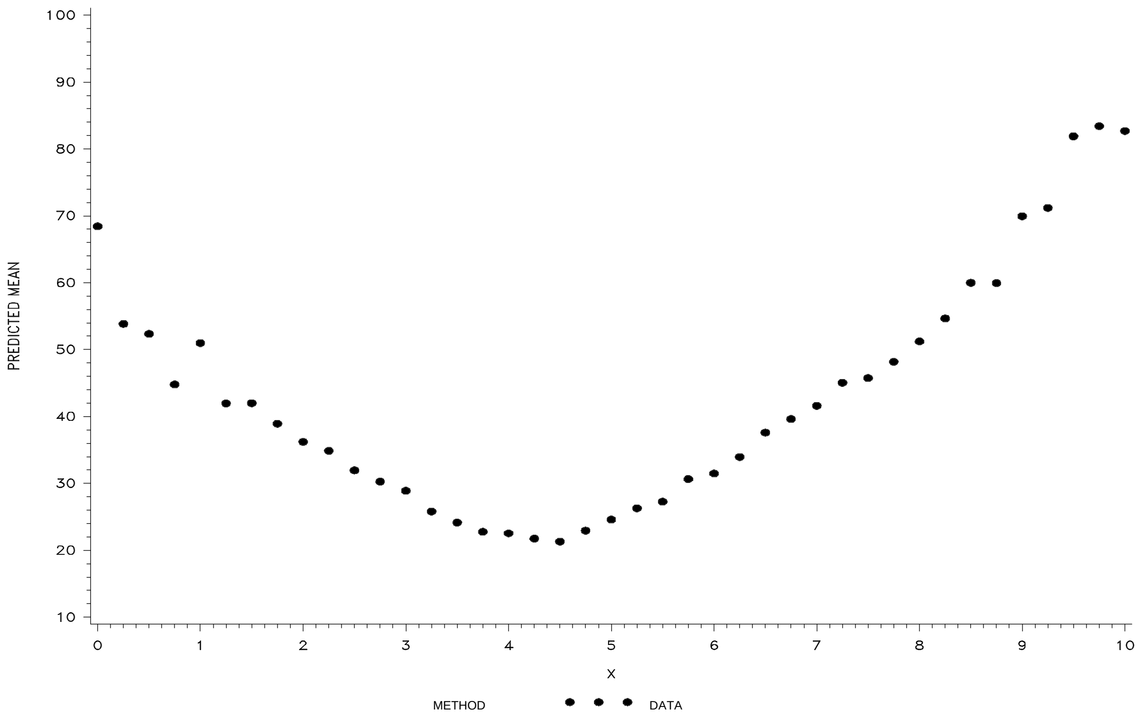

**Figure 6.C.13. Scatter plot of a random data set generated from the dual model given in equations (6.C.1) and (6.C.2) where $\gamma$ = 2.5.**

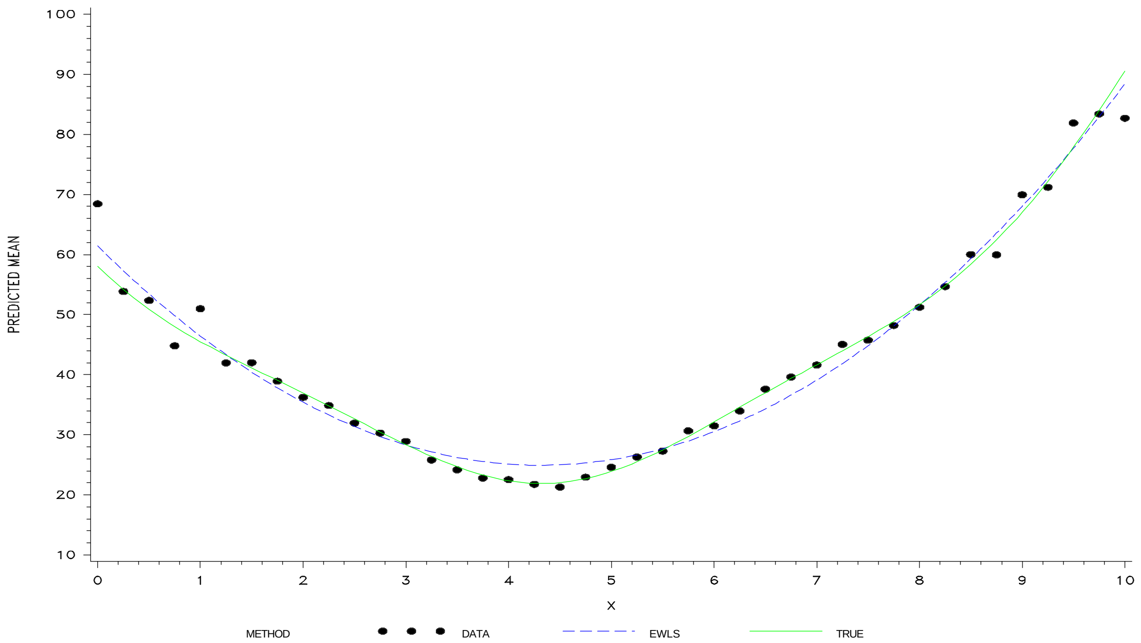

**Figure 6.C.14. Plot showing generated data for Example 2 ($\gamma$ = 2.5) along with the EWLS fit and true underlying means function.**

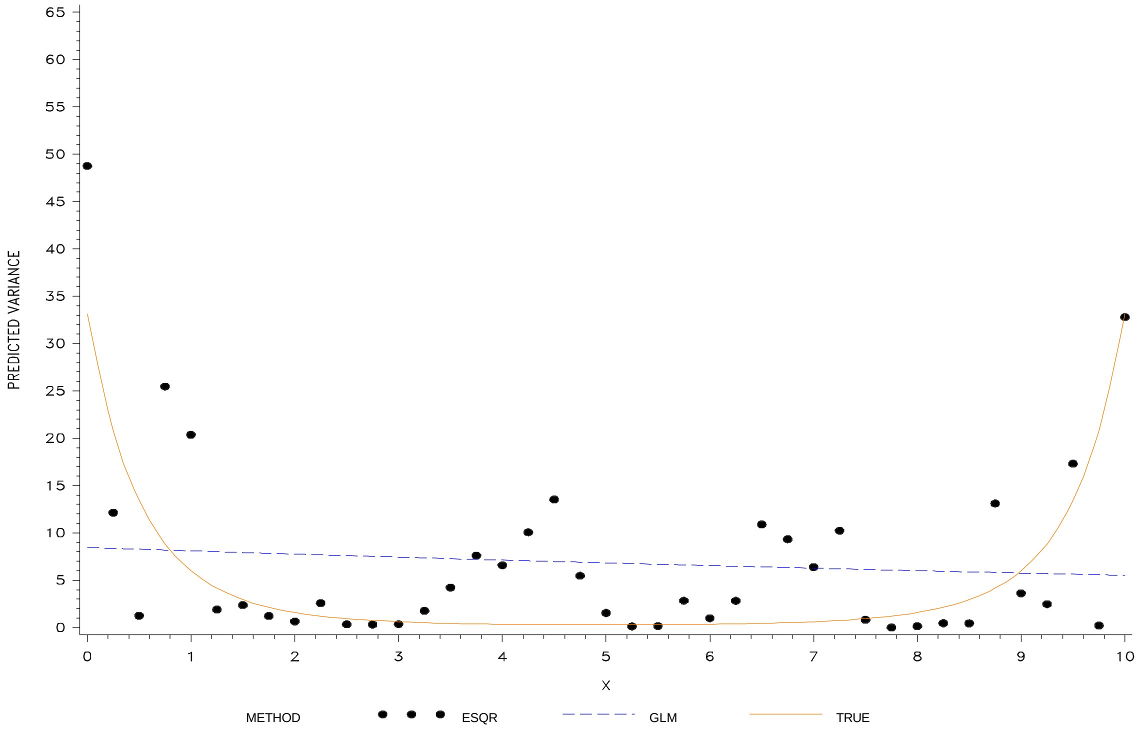

**Figure 6.C.15. Plot showing the EWLS squared residuals, the parametric variance estimate and the true underlying variance function for Example 2 (γ = 2.5).**

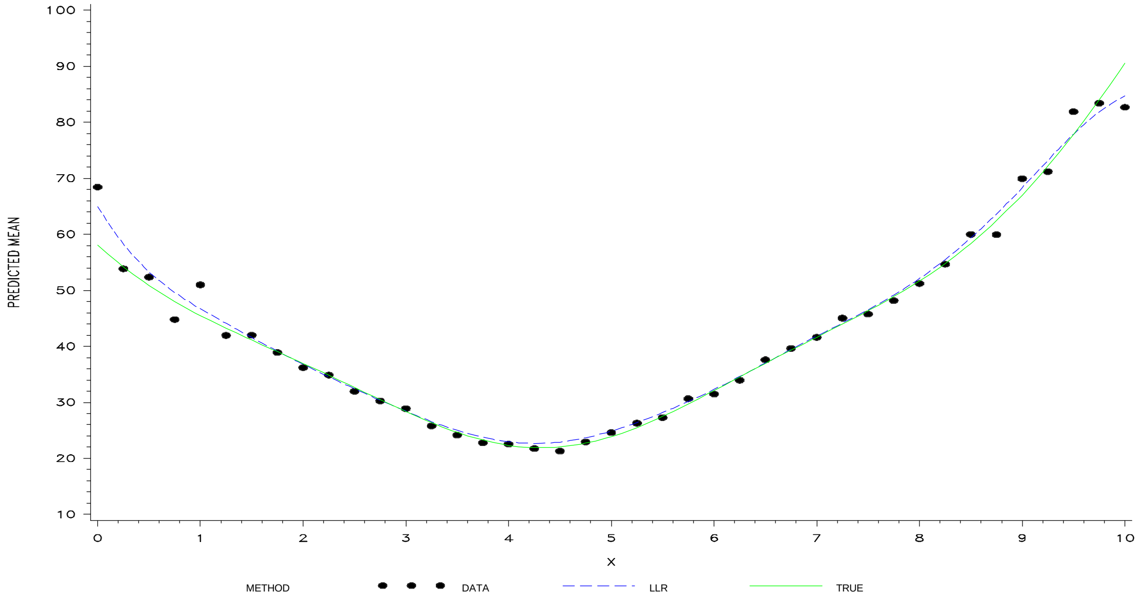

**Figure 6.C.16. Plot showing generated data for Example 2 (γ = 2.5) along with the LLR fit and true underlying means function.**

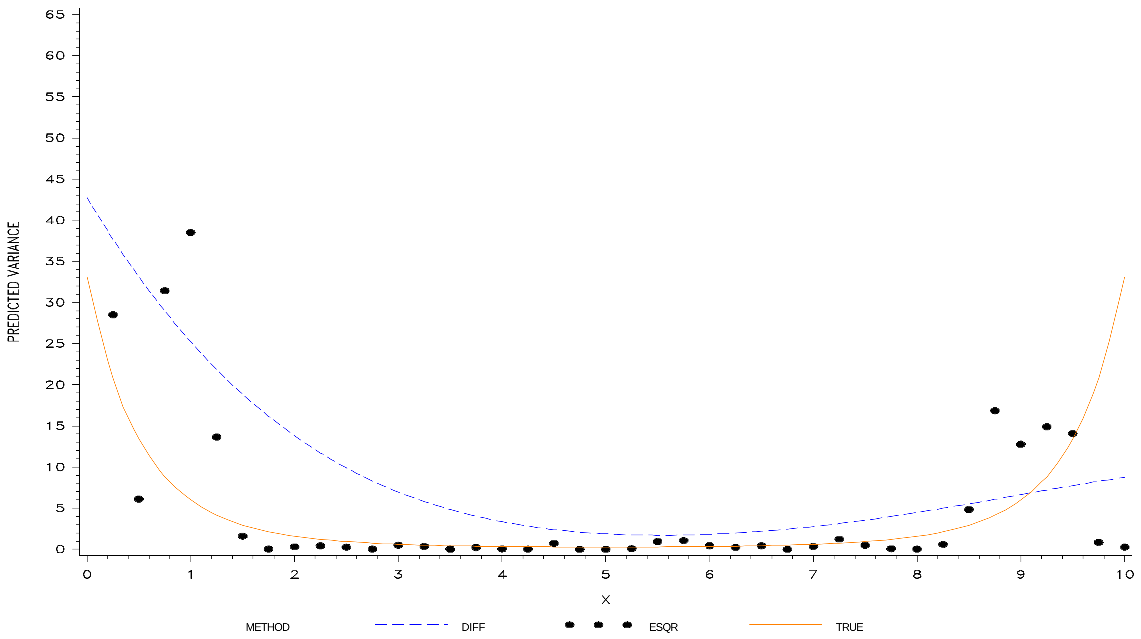

**Figure 6.C.17. Plot showing the squared pseudo-residuals, difference-based variance estimate, and true underlying variance function for Example 2 (γ = 2.5).**

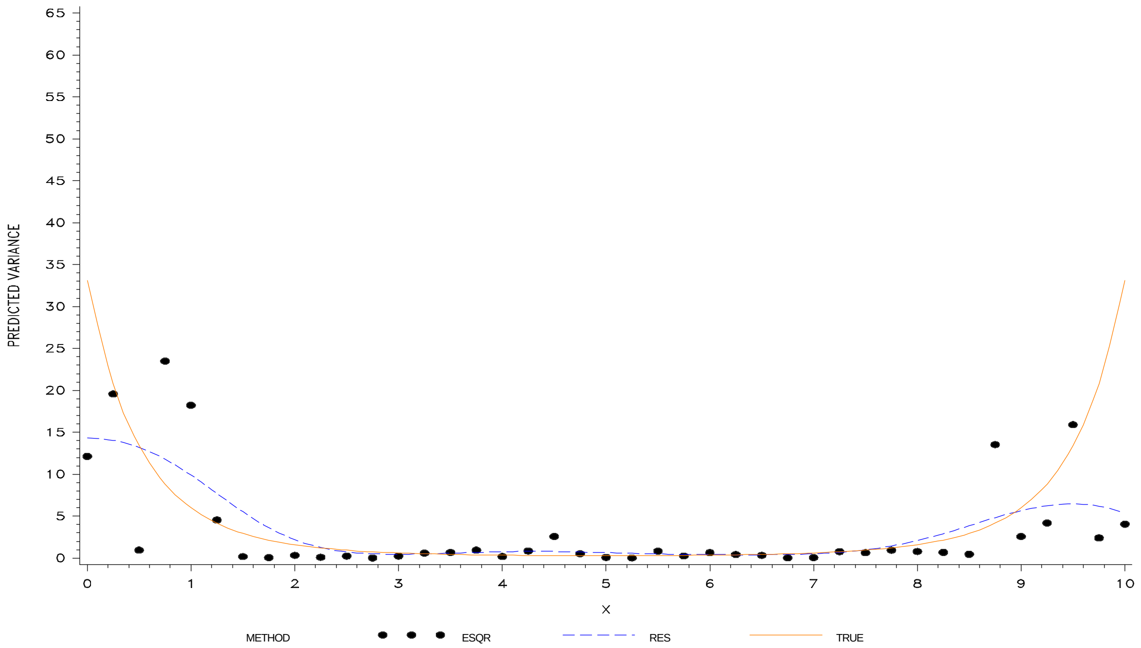

**Figure 6.C.18. Plot showing the squared LLR residuals, residual-base variance estimate, and true underlying variance function for Example 2 (γ = 2.5).**

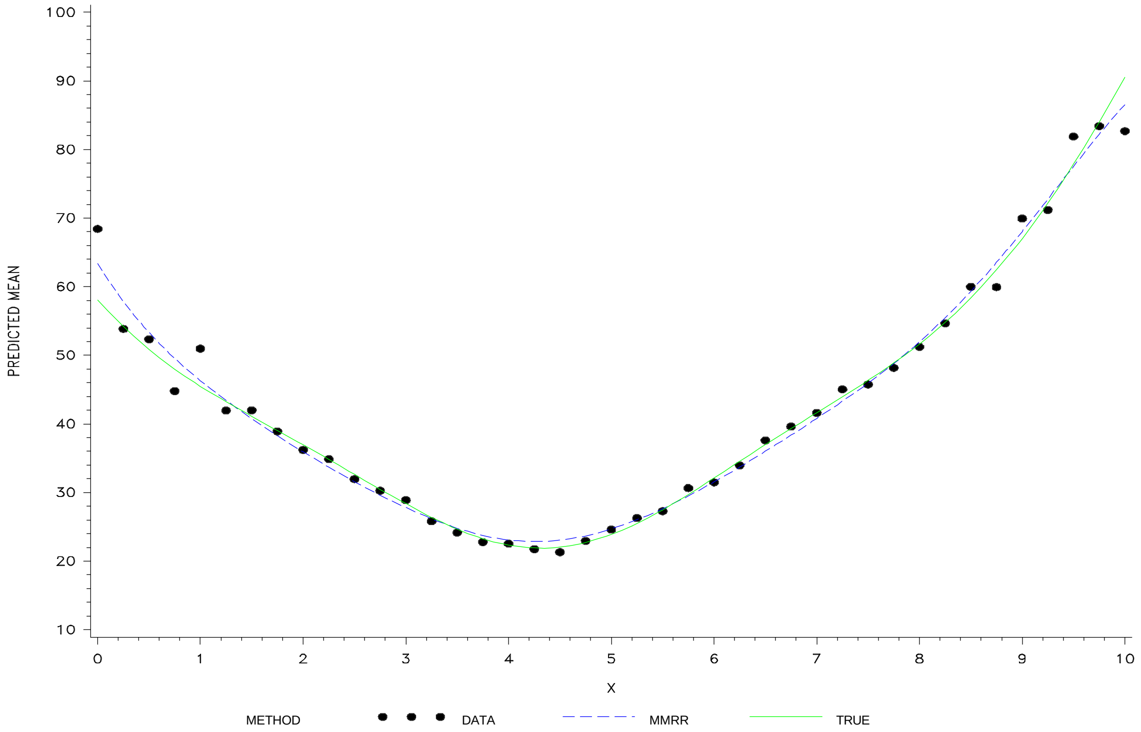

**Figure 6.C.19. Plot showing the generated data for data for Example 2 (γ = 2.5) along with the MMRR fit and true underlying means function.**

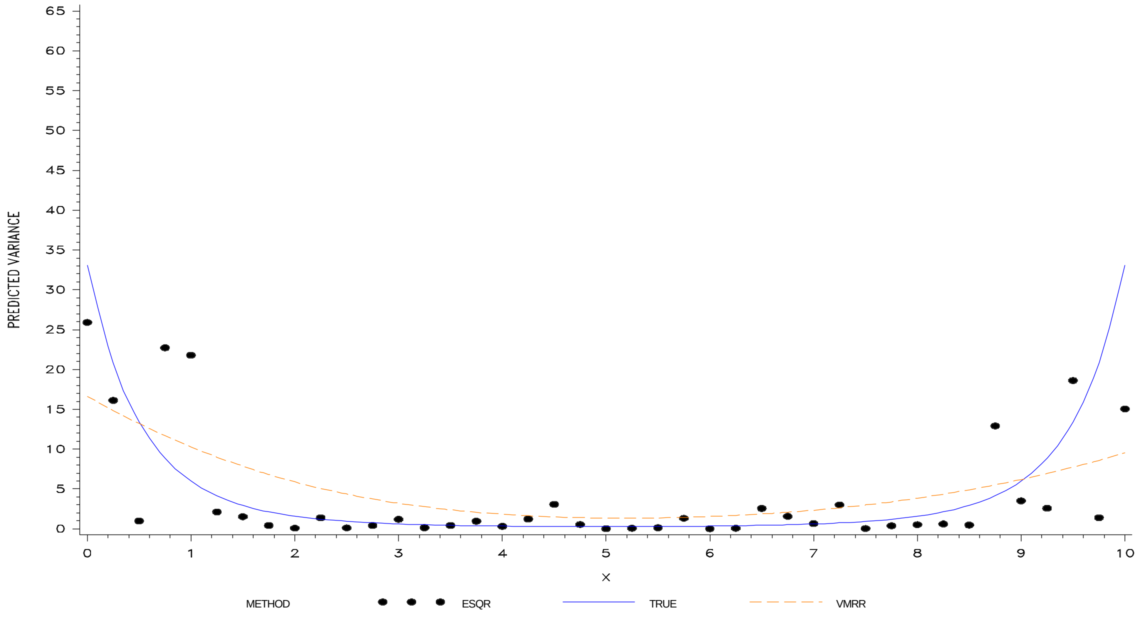

**Figure 6.C.20 Plot showing the MMRR squared residuals along with the VMRR estimate and true variance function for Example 2 (γ = 2.5).**

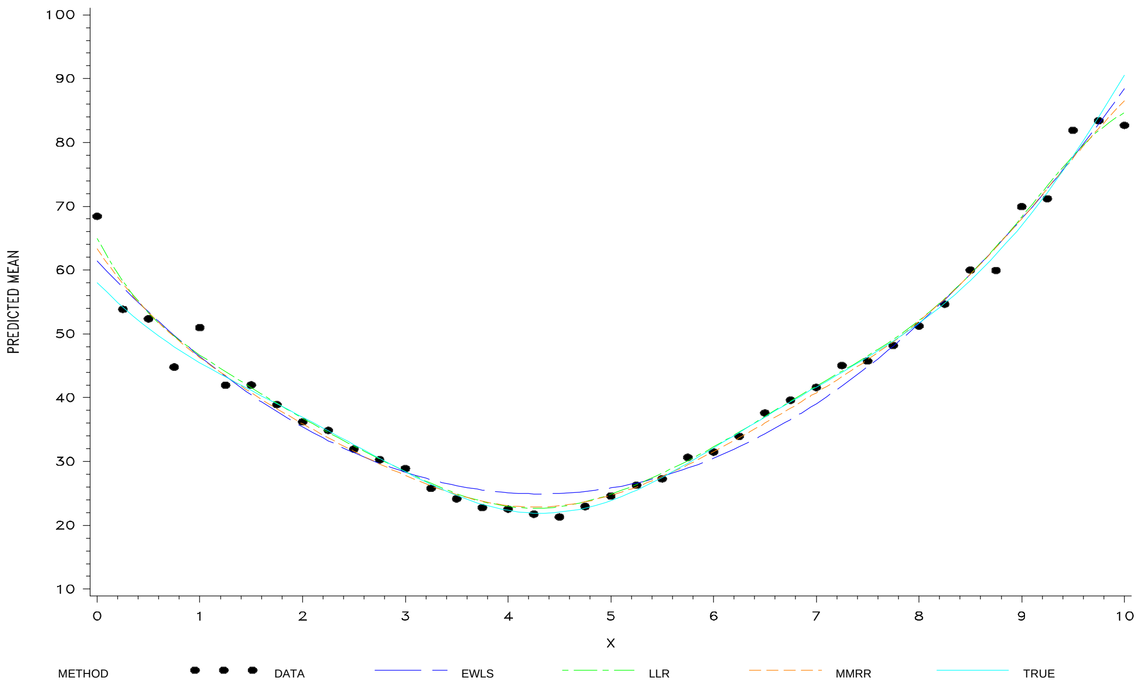

**Figure 6.C.21. Plot of the true means function in Example 2 (γ = 2.5) along with the EWLS, LLR, and MMRR estimates of the mean.**

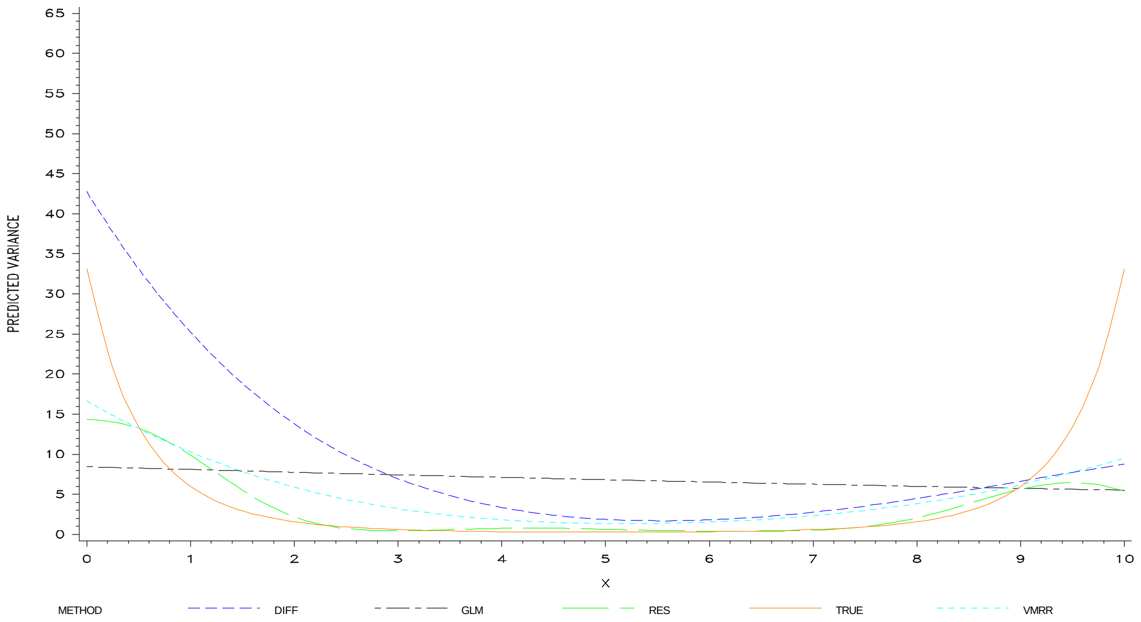

**Figure 6.C.22. Plot of the true variance function in Example 2 (γ = 2.5) along with the GLM, Difference-Based, Residual-Based, and VMRR variance estimates.**

misspecified ($\gamma \geq 7.5$) means cases, the optimal variance model mixing parameter is equal to 1. Contrast this to numerical results found in Table 6.B.2 where there was no variance model misspecification and $\lambda_{\sigma_o}$ was consistently 0.

There are several key results from Table 6.C.1 that are worth mentioning. First, the IMMSE and IVMSE values for the DMRR procedure are uniformly smaller than those values of the parametric and nonparametric procedures. The only exception to this is in the case in which $n = 21$ and $\gamma > 0$ where the nonparametric residual-based estimate provides slightly better IMMSE values than MMRR. The most notable improvements to dual model analysis offered by DMRR exist in small samples when there is little or no misspecification of the means model ($\gamma \leq 2.5$). The IMMSE and IVMSE values for the nonparametric residual-based approach become more competitive with DMRR as the sample size increases and as the specified means function begins to deviate more and more from the true means function. A couple of other interesting notes concern the performance of the parametric dual modeling procedure. Notice that for $\gamma = 0.0$ (no means misspecification) the IMMSE for EWLS is the same as that for OLS, reflecting the constant variance function estimated by the parametric dual modeling procedure. Also observe that for $\gamma = 0.0$, the IMMSE for EWLS is still better than the IMMSE for LLR, even though the heteroscedasticity is ignored. However, for even slight misspecification ($\gamma = 2.5$) of the mean, parametric dual modeling performs noticeably worse than all other procedures. This indicates that parametric dual modeling is very sensitive to misspecification of the means model and thus, should only be used as a means of analysis when the user is extremely confident in the accuracy of the specified means function.

**Table 6.C.1 Simulated mean square error values for optimal mean and variance fits from 300 Monte Carlo runs (mean and variance models misspecified). Theoretical MMIMSE and VMIMSE values are in bold.**

| γ | OLS | PARAMETRIC | | DIFF-BASED | | RES-BASED | | DMRR | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Mean | VAR | Mean | VAR | Mean | VAR | Mean | VAR |
| *n* = 21 | | | | | | | | | |
| | | | | | | | | | |
| 0.0 | **1.35398**<br>1.37878 | **1.35398**<br>1.24465 | **58.1300**<br>49.9427 | **2.43217**<br>2.46754 | **75.3459**<br>233.875 | **2.43217**<br>2.46754 | **42.5676**<br>44.9409 | **0.78090**<br>0.80219 | **37.8466**<br>57.1043 |
| 2.5 | **4.10819**<br>4.11166 | **4.10860**<br>4.00680 | **69.8223**<br>63.0723 | **2.50996**<br>2.52663 | **75.3459**<br>278.268 | **2.50996**<br>2.52663 | **47.2060**<br>45.8735 | **2.58538**<br>2.59764 | **38.2848**<br>43.9164 |
| 5.0 | **12.3708**<br>12.3530 | **12.3737**<br>12.3087 | **197.055**<br>202.936 | **2.61931**<br>2.62218 | **75.3459**<br>344.458 | **2.61931**<br>2.62218 | **54.3095**<br>48.4562 | **2.60261**<br>2.58685 | **43.5269**<br>46.5486 |
| 7.5 | **26.1419**<br>26.1027 | **26.1487**<br>26.0619 | **718.417**<br>761.266 | **2.71968**<br>2.71881 | **75.3459**<br>437.515 | **2.71968**<br>2.71881 | **56.6704**<br>50.8455 | **2.75114**<br>2.73585 | **48.7496**<br>49.1884 |
| 10.0 | **45.4215**<br>49.4811 | **45.4328**<br>49.4613 | **2100.18**<br>2413.79 | **2.83772**<br>2.83195 | **75.3459**<br>497.330 | **2.83772**<br>2.83195 | **55.6576**<br>53.0691 | **2.86908**<br>2.84567 | **52.6728**<br>50.9980 |
| | | | | | | | | | |
| *n* = 41 | | | | | | | | | |
| | | | | | | | | | |
| 0.0 | **0.64147**<br>0.67893 | **0.64147**<br>0.66079 | **54.5872**<br>48.5400 | **1.44325**<br>1.50883 | **42.3447**<br>75.0863 | **1.44325**<br>1.50883 | **27.5035**<br>32.3668 | **0.30253**<br>0.33477 | **24.3792**<br>40.0198 |
| 2.5 | **3.37926**<br>3.41761 | **3.37945**<br>3.39814 | **64.6027**<br>58.7028 | **1.50893**<br>1.56864 | **42.3447**<br>79.2287 | **1.50893**<br>1.56864 | **28.8624**<br>32.9330 | **1.42027**<br>1.46850 | **25.8844**<br>32.2937 |
| 5.0 | **11.5926**<br>11.6319 | **11.5939**<br>11.6125 | **185.864**<br>187.366 | **1.62415**<br>1.67862 | **42.3447**<br>84.3164 | **1.62415**<br>1.67862 | **31.3140**<br>34.3424 | **1.58074**<br>1.62092 | **29.5449**<br>33.2461 |
| 7.5 | **25.2816**<br>25.3217 | **25.2844**<br>25.3034 | **695.577**<br>717.792 | **1.72866**<br>1.78150 | **42.3447**<br>90.9825 | **1.72866**<br>1.78150 | **33.7485**<br>36.0487 | **1.70569**<br>1.74802 | **32.3756**<br>34.9380 |
| 10.0 | **44.4461**<br>44.4871 | **44.4502**<br>44.4706 | **2059.65**<br>2118.39 | **1.82160**<br>1.87462 | **42.3447**<br>97.8776 | **1.82160**<br>1.87462 | **35.9420**<br>37.6622 | **1.80606**<br>1.85094 | **34.6575**<br>36.5402 |
| | | | | | | | | | |
| *n* = 61 | | | | | | | | | |
| | | | | | | | | | |
| 0.0 | **0.41819**<br>0.42123 | **0.41819**<br>0.40899 | **52.2169**<br>47.4726 | **1.04773**<br>1.08708 | **31.2596**<br>45.8120 | **1.04773**<br>1.08708 | **22.5681**<br>26.6730 | **0.18016**<br>0.18972 | **18.0094**<br>30.1530 |
| 2.5 | **3.15293**<br>3.15488 | **3.15305**<br>3.14202 | **61.4185**<br>56.0406 | **1.10297**<br>1.14175 | **31.2596**<br>46.4028 | **1.10297**<br>1.14175 | **23.3054**<br>27.0620 | **1.00945**<br>1.03573 | **19.3288**<br>26.3172 |
| 5.0 | **11.3571**<br>11.3580 | **11.3580**<br>11.3487 | **179.950**<br>176.772 | **1.19882**<br>1.23687 | **31.2596**<br>47.1369 | **1.19882**<br>1.23687 | **24.6408**<br>28.1210 | **1.15929**<br>1.18883 | **21.4640**<br>27.0527 |
| 7.5 | **25.0308**<br>25.0306 | **25.0326**<br>25.0259 | **684.700**<br>690.793 | **1.28622**<br>1.32399 | **31.2596**<br>48.0289 | **1.28622**<br>1.32399 | **26.0235**<br>29.3105 | **1.26426**<br>1.29592 | **23.3075**<br>28.0722 |
| 10.0 | **44.1740**<br>44.1727 | **44.1765**<br>44.1709 | **2041.89**<br>2065.66 | **1.36227**<br>1.39997 | **31.2596**<br>49.0952 | **1.36227**<br>1.39997 | **27.3444**<br>30.4471 | **1.34700**<br>1.37999 | **24.8952**<br>29.0817 |
| | | | | | | | | | |

**Table 6.C.2 Optimal bandwidths and mixing parameters chosen by minimizing the AVEMMSE and AVEMMSE values for the nonparametric and model robust procedures (mean and variance models misspecified).**

| $\gamma$ | DMRR | | | | Diff-Based | | Res-Based | |
|---|---|---|---|---|---|---|---|---|
| | Mean | | VAR | | Mean | VAR | Mean | VAR |
| $n = 21$ | $b_{\mu_o}$ | $\lambda_{\mu_o}$ | $b_{\sigma_o}$ | $\lambda_{\sigma_o}$ | $b_{\mu_o}$ | $b_{\sigma_o}$ | $b_{\mu_o}$ | $b_{\sigma_o}$ |
| 0.0 | 0.7225 | 0.0 | 0.35570 | 1.0 | 0.08387 | 0.45305 | 0.08387 | 0.12663 |
| 2.5 | 0.11383 | 0.62139 | 0.40609 | 1.0 | 0.07637 | 0.45305 | 0.07637 | 0.12317 |
| 5.0 | 0.07066 | 1.0 | 0.08184 | 1.0 | 0.06137 | 0.45305 | 0.06137 | 0.10423 |
| 7.5 | 0.05666 | 1.0 | 0.08019 | 0.78389 | 0.05022 | 0.45305 | 0.05022 | 0.08554 |
| 10.0 | 0.04838 | 1.0 | 0.07945 | 0.47402 | 0.04371 | 0.45305 | 0.04371 | 0.07148 |
| | | | | | | | | |
| $n = 41$ | | | | | | | | |
| | | | | | | | | |
| 0.0 | 0.755 | 0.0 | 0.25840 | 1.0 | 0.07227 | 0.29695 | 0.07227 | 0.10084 |
| 2.5 | 0.08898 | 0.85439 | 0.27672 | 1.0 | 0.06574 | 0.29695 | 0.06574 | 0.09562 |
| 5.0 | 0.06012 | 1.0 | 0.29078 | 1.0 | 0.05434 | 0.29695 | 0.05434 | 0.08249 |
| 7.5 | 0.04887 | 1.0 | 0.30391 | 1.0 | 0.04629 | 0.29695 | 0.04629 | 0.07196 |
| 10.0 | 0.04246 | 1.0 | 0.31391 | 1.0 | 0.04099 | 0.29695 | 0.04099 | 0.06422 |
| | | | | | | | | |
| $n = 61$ | | | | | | | | |
| | | | | | | | | |
| 0.0 | 0.765 | 0.0 | 0.20852 | 0.94926 | 0.06551 | 0.22324 | 0.06551 | 0.08861 |
| 2.5 | 0.07805 | 0.94575 | 0.22203 | 1.0 | 0.05961 | 0.22324 | 0.05961 | 0.08272 |
| 5.0 | 0.05402 | 1.0 | 0.22828 | 1.0 | 0.04941 | 0.22324 | 0.04941 | 0.07175 |
| 7.5 | 0.04439 | 1.0 | 0.23492 | 1.0 | 0.04238 | 0.22324 | 0.04238 | 0.06335 |
| 10.0 | 0.03879 | 1.0 | 0.24047 | 1.0 | 0.03769 | 0.22324 | 0.03769 | 0.05716 |
| | | | | | | | | |