A Semiparametric Approach To Dual Modeling with Unreplicated Data

B. A. Starnes (C-N),

Jeffrey B. Birch (VPI&SU),

Tim Robinson (UWYO)

16 May 2008

Forms of Regression

- *q,* the *true* mean function is based on an unknown parameter vector $\boldsymbol{\beta}$, and estimated "globally"
- parametric estimate designated as y^{p}
- *q,* the *true* mean function is of uncertain form and is estimated locally
- nonparametric estimate designated as y^{np}

Parametric Regression

Generally the model is written as

$$y_i = f(\mathbf{x}_i;\boldsymbol{\beta}) + \varepsilon_i,\ \varepsilon_i \sim^{iid} N(0,\sigma),\ i = 1, 2, \ldots n$$

with f being any function of regressors and parameters,

and the $\boldsymbol{\beta}$ estimate selected to minimize

$$SSE(\mathbf{b}) = \sum^{n} (y^{p}_{i} - f(\mathbf{x}_i;\mathbf{b}))^2$$

Note that the error terms are assumed to have a constant variance.

What if, in fact, that is not the case?

What if the proposed model is only accurate for part of the prediction domain?

Nonparametric Regression

$$y_i = f(\mathbf{x}_i;\boldsymbol{\beta}_i, b_i) + \varepsilon_i,\ \varepsilon_i \overset{iid}{\sim} N(0,\sigma),\ i = 1, 2, \ldots n$$

where, again, f can be any function of the regressors and parameters, which is now determined locally, and b_i is the bandwidth (either global or local) which determines the “weight” given to nearby observations.

The advantage of nonparametric modeling is that the model captures unusual trends in the data set which may not be possible with a parametric model.

Nonparametric Modeling (continued)

Local polynomial regression (LLR, LQR, etc.) essentially involves locally weighted least squares in which the weights are "kernel weights".

Kernel functions weight data according to its proximity to a target location.

For example the Normal kernel (for one "x_0") is given by

$$K\left((x_j - x_0)/b\right) = \exp[-\,(x_j - x_0)^{\wedge}2\ /b]$$

The $\boldsymbol{\beta}$ estimate (not b above) in the linear predictor, of course, will have dimension

D = degree + 1,

and will be determined locally.

The next slide illustrates possible advantages of a nonparametric fit to a data set.

An example of Nonparametric modeling

Figure 1(a)

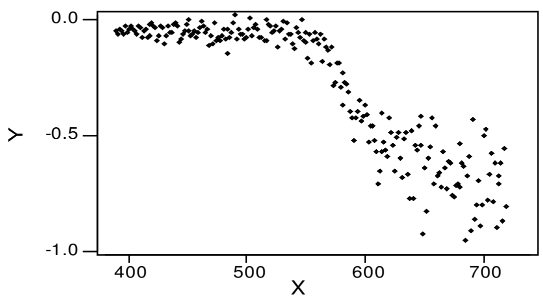
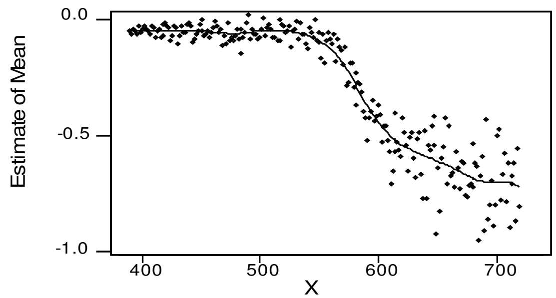
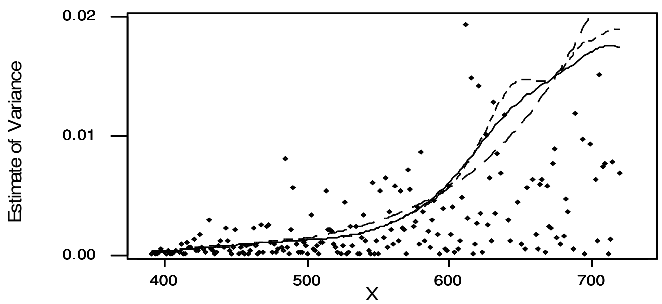


Figure 1(b)



MMRR (Solid)

Figure 1(c)



VMMR (Solid), Parametric (Long Dash), Nonparametric (Short Dash)

This is the famous LIDAR data set that illustrates the value of nonparametric regression. Notice the uncertain pattern for the mean and variance of the "dependent" variable. []

Model Robust Regression (MRR)

- Improves regression function estimate of q by combining parametric and non-parametric estimates with a mixing parameter
- Two initial forms of Model Robust Regression: MRR1 and MRR2 (developed at Virginia Tech)
- Ameliorates the disadvantages of both estimates
- MRR1 developed in 1989 by Rich Einsporn and Jeffrey B. Birch
- MRR2 developed in 1995 by James Mays and Jeffrey B. Birch

The MRR1 Estimate

- Form: $\hat{q}(\mathbf{x}_i) = (1 - \lambda)y_i^{p} + \lambda\, y_i^{np}$
- Mixing parameter: $0 \leq \lambda \leq 1$
  - If $\lambda = 0$, parametric model is correct
  - If $\lambda = 1$, nonparametric model is correct.
  - λ represents a weighting constant for the two estimates and is determined by the relative value of each model.

The MRR2 Estimate

- Form: $\hat{q}(\mathbf{x}_i) = y_i^{p} + \lambda\, y_i^{np}$
- Mixing parameter: $0 \leq \lambda \leq 1$
  - If $\lambda = 0$, parametric model is correct
  - If $\lambda = 1$, nonparametric estimate is added since parametric model is inadequate.
  - λ represents the degree to which the parametric estimate is improved
  - The optimal mixing parameter obtained via simple calculus is given by

$$\lambda^* = \langle \mathbf{y}^{np}, \boldsymbol{q} - \mathbf{y}^{p} \rangle / ||\mathbf{y}^{np}||^2$$

  - The optimal data driven mixing parameter is given by

$$\hat{\lambda}^* = \langle \mathbf{y}^{np}, \mathbf{y} - \mathbf{y}^{p} \rangle / ||\mathbf{y}^{np}||^2$$

  - The parametric estimate is always involved
  - Nonparametrically smoothed residuals

Dual Modeling

- Dual modeling is the phrase describing situations in which there is interest in modeling both the mean and variance of some response
- Research has been done with both replicated and unreplicated data scenarios … we focus on the unreplicated data situation
- Typically the data sets are reasonably large… we focus here on smaller data sets found, for example, in engineering applications where cost of each experimental run is a premium
- Pickle, Robinson, Birch, Anderson-Cook (2008) considered dual modeling in small sample settings where replication was present

Variance Modeling

Bartlett and Kendall (1946)

$$\ln(\sigma_i^2) = \mathbf{x}_i^{*T}\boldsymbol{\gamma} + \varepsilon_i \ , \ \varepsilon_i \overset{iid}{\sim} N(0,\varsigma), \ i = 1, 2, \ldots n$$

Myers and Montgomery (2002) and others suggest GLM

$$g(\sigma_i^2) = \mathbf{x}_i^{*T}\boldsymbol{\gamma} + \varepsilon_i \ , \ \varepsilon_i \overset{iid}{\sim} N(0,\varsigma), \ i = 1, 2, \ldots n$$

Aitken (1987) suggested GLM for modeling squared means residuals in the absence of replication

$$g(\sigma_i^2) = \mathbf{x}_i^{*T}\boldsymbol{\gamma} + \varepsilon_i \ , \ \varepsilon_i \overset{iid}{\sim} N(0,\varsigma), \ i = 1, 2, \ldots n$$

Model Robust Dual Modeling

This is a form of joint estimation in which

two models are estimated "simultaneously".

The models given here, are the "means" model given by:

$$y_i = \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta} + f(\mathbf{x}_i) + \varepsilon_i,\ \varepsilon_i \sim^{\text{id}} N(0,\sigma_{\text{i}}),\ i = 1, 2, \ldots n,$$

and the "variance" model given by:

$$e^2{}_i = g^{-1}(\mathbf{x}_i^{*\mathsf{T}}\boldsymbol{\beta}) + l(\mathbf{x}_i^{*}) + \eta_i,\ \eta_i \sim^{\text{iid}} N(0,\varsigma),\ i = 1, 2, \ldots n.$$

Model Robust Dual Modeling (cont.)

The two models are estimated simultaneously via MRR, AND involve parametric and nonparametric components. So the resulting DMRR estimates are:

$$\hat{q}_{\mu}(\mathbf{x}_i) = y_i^{p(EWLS)} + \lambda_{\mu}\, y_i^{np(LLR)}$$

$$\hat{q}_{\sigma}(\mathbf{x}_i) = (1-\lambda_{\sigma})\, y_i^{p(GLM)} + \lambda_{\sigma}\, y_i^{np(LLR)}$$

Note that we are utilizing local linear regression for the nonparametric components of both the mean and variance estimates.

Note also that MRR1 is used in the variance estimate while MRR2 is used for the means estimate.

The process of obtaining these estimates is via a 10 step algorithm.

Dual Model Algorithm for Unreplicated data

1. Let $\hat{V} = \text{diag}\{s_1^2, s_2^2, \ldots, s_n^2\}$ where $s_i^2 = \hat{\sigma}_i^2 = 1$, $i=1, 2, \ldots, n$.

2. Then, obtain (EWLS) the parametric estimate of the means model :

$$\hat{y}_i^{(EWLS)} = \mathbf{x}_i^T \mathbf{b}^{(EWLS)} = \mathbf{x}_i^T(\mathbf{X}^T \hat{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T \hat{V}^{-1}\mathbf{y},\ i=1, 2, \ldots, n,$$

and let $\hat{\mathbf{y}}^{(EWLS)}$ denote the nx1 vector of EWLS fits.

3. Form the residuals from the fit found in Step 2 , $\mathbf{e}^{(EWLS)} = (\mathbf{y} - \hat{\mathbf{y}}^{(EWLS)})$, 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 $\mathbf{H}_{b_\mu}^{(LLR)}$ and $\mathbf{e}^{(EWLS)}$ is the $nx1$ vector of EWLS residuals.

4. Obtain the MMRR fit to the means model, written as: $\hat{\mathbf{y}}^{(MMRR)} = \hat{\mathbf{y}}^{(EWLS)} + \lambda_\mu \hat{\mathbf{r}}_\mu$ where the i^{th} element of $\hat{\mathbf{r}}_\mu$ is $\hat{r}_{i_\mu}$ from Step 3, and $\lambda_\mu \in [0, 1]$ is the means model mixing parameter.

5. Form the squared residuals from the MMRR fit to the mean, obtaining : $e_i^{2(MMRR)} = (y_i - \hat{y}_i^{(MMRR)})^2$, $i=1,\ldots,n$, and let $\mathbf{e}^{2(MMRR)}$ denote the nx1 vector of squared MMRR residuals.

Dual Model Algorithm (cont.)

6. The GLM model for estimating the variance is: $\phi\{e_i^{2(\text{MMRR})}\} = \mathbf{z}_i^{\text{T}}\boldsymbol{\theta}$ where $\phi(.)$ is the log link function.

7. The fitted values are then given by: $\hat{\sigma}_i^{2(\text{GLM})} = \exp\left\{\mathbf{z}_i\hat{\boldsymbol{\theta}}^{(\text{GLM})}\right\}$, i=1,2, … ,n.

8. Perform local linear regression on the set of squared MMRR residuals, obtaining

$$\hat{\sigma}_i^{2(\text{LLR})} = \mathbf{h}_{\text{ib}\sigma}^{(\text{llr})\,\text{T}}\,\mathbf{e}^{2(\text{MMRR})}$$ where $\mathbf{h}_{\text{ib}\sigma}^{(\text{llr})\,\text{T}}$ is the i^{th} row of $\mathbf{H}_{\text{b}\sigma}^{(\text{LLR})}$, i=1,2, … ,n.

9. Obtain the VMRR estimates of variance which are written as:

$$\hat{\sigma}_i^{2\,(\text{VMRR})} = \lambda_\sigma\ \hat{\sigma}_i^{2\,(\text{LLR})} + \left(1-\lambda_\sigma\right)\hat{\sigma}_i^{2\,(\text{GLM})}, \ i=1,\ldots,n,$$

and where $\lambda_\sigma \in [0, 1]$ is the variance model mixing parameter.

10. Return to step 2 with $\hat{V} = diag\left(\hat{\sigma}_1^{2(\text{VMRR})}, \cdots, \hat{\sigma}_n^{2(\text{VMRR})}\right)$. Continue steps 2 - 9 until means model parameter estimates converge.

Simulation Results

Here we have several simulations of observations generated for a "perturbed" quadratic function:

$$q_{\mu}(x) + \delta = \{2(x-5)^2 + 5x + \kappa \sin(\pi(x-1)/2.25) + (q_{\sigma}(x))^{.5}\varepsilon\} I_x(0, 10) \text{ where } \varepsilon \sim^{iid} N(0, 1), \text{ and}$$

$$q_{\sigma}(x) + \eta = \{e^{(3.125 - 1.25x + .125x^2)} + \eta\} I_x(0, 10) \text{ where } \eta \sim^{iid} N(0, \varsigma).$$

Figure 2(a)

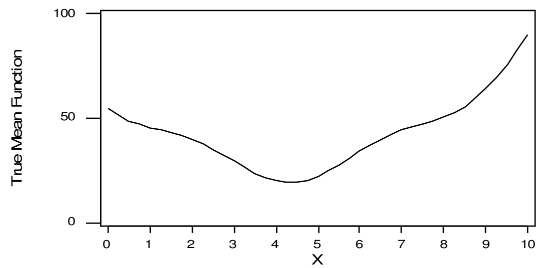
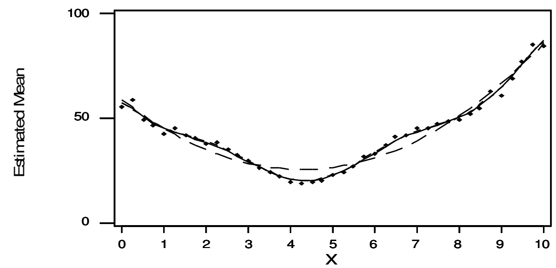
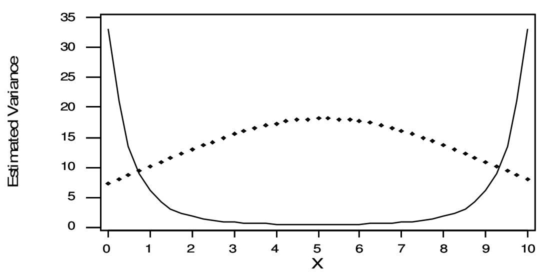


Figure 2(b)



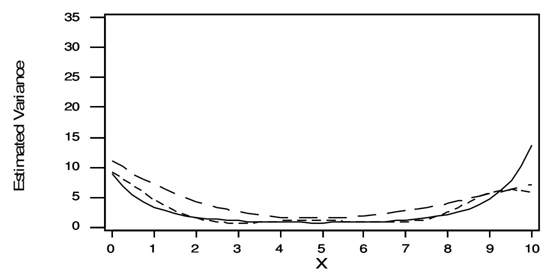
MMRR (Solid), EWLS (Dash), LLR Fit (Dotted)

Figure 2(c)



True Variance Function (Solid), GLM Estimated Variance (Dotted)

Figure 2(d)



VMRR (Solid), Diff (Long Dash), Res (Short Dash)

Here we consider κ as the degree of perturbation.

Simulation Results (cont.)

Here we have results from the situation in which the means model is misspecified, but the variance model is correctly specified. Simulated integrated mean squared error values for the five methods are given with the SIMSE(M) on top, and the SIMSE(V) on bottom. The Best values over the five methods are in bold for each κ. Additional simulations were done with n = 40 and 60 with similar results.

$n=20,\ \kappa=$	PAR	NPAR1	NPAR2	NPAR3	**DMRR**
0	**0.765** ***28.376***	2.788 *304.812*	2.788 *33.161*	2.788 *41.829*	1.563 *29.362*
2.5	3.908 *37.538*	2.974 *369.447*	2.974 *33.689*	2.974 *42.283*	**2.529** ***32.039***
5.0	14.726 *305.975*	3.397 *542.214*	3.397 *36.181*	3.397 *51.674*	**3.069** ***33.667***
7.5	26.450 *928.407*	3.648 *680.427*	3.648 *35.704*	3.648 *54.310*	**3.446** ***35.272***
10	45.813 *2615.06*	3.791 *959.97*	3.791 *34.430*	3.791 *53.776*	**3.600** ***33.951***

Simulation Results (cont.)

Below are the simulation results in the case where the Means and Variance Models are both misspecified. Simulated integrated mean squared error values are given for the five methods with the SIMSE(M) on top, and the SIMSE(V) on bottom. Best values over the five methods in bold for each κ. Additional simulations were done with n = 40, and 60 with similar results.

$n=20,\ \kappa=$	PAR	NPAR1	NPAR2	NPAR3	**DMRR**
0	**1.029** ***30.177***	2.653 *280.804*	2.653 *33.801*	2.653 *43.407*	1.480 *31.553*
2.5	3.945 *38.392*	2.977 *345.403*	2.977 *34.347*	2.977 *45.639*	**2.525** ***33.538***
5.0	14.512 *233.914*	3.319 *518.584*	3.319 *33.795*	3.319 *49.730*	**3.060** ***33.711***
7.5	25.991 *705.696*	3.621 *616.949*	3.621 *35.713*	3.621 *54.640*	**3.361** ***33.843***
10	45.192 *2116.77*	3.573 *874.973*	3.573 *35.667*	3.573 *56.956*	**3.360** ***34.287***

[]

Distance Measures and Asymptotics

- For $i = 1, 2, \ldots n$, and functions of $\mathbf{x}_i$
  - $<h_a, h_b> = n^{-1}\sum(h_a(\mathbf{x}_i)h_b(\mathbf{x}_i))$
  - $<h_a, h_a> = ||h_a||^2$
  - $\delta_n = \inf\{||\boldsymbol{q} - \mathbf{y}^p(\boldsymbol{\beta}^*)||: \boldsymbol{\beta}^* \in \mathbb{R}^d\}$
    - Parametric model correct ($\delta_n = 0$)
    - Parametric model incorrect ($\mathcal{L}\delta_n > 0$)
  - $\gamma_n^2 = E(||\mathbf{y}^{np} - \boldsymbol{q}||)^2$ = AVEMSE
  - For MRR2,

$$\gamma_n^2 = E(||\mathbf{y}^{np} - (\boldsymbol{q} - \mathbf{y}^p(\boldsymbol{\beta}^*))||)^2,\ \boldsymbol{\beta}^* \in \mathbb{R}^d$$

Distance Measures and Asymptotics (cont.)

- A stochastic sequence $\{X_n\}$ is said to be $O_P(1)$ if for every $0 < \eta < 1$, there exist constants $K(\eta)$, $n(\eta) \in \mathbb{R}$, such that for $n \geq n(\eta)$,

$$P\{|X_n| \leq K(\eta)\} \geq 1 - \eta$$

- A stochastic sequence $\{X_n\}$ is said to be $O_P(b_n)$ if

$$\{X_n / b_n\} = O_P(1)$$

- Stochastic parallel to "O *notation*" in the Analysis framework.

The DMRR Estimate

- Form: $\hat{q}_{\mu}(\mathbf{x}_i) = y_i^{p(EWLS)} + \lambda_{\mu}\, y_i^{np(LLR)}$
- Form: $\hat{q}_{\sigma}(\mathbf{x}_i) = (1-\lambda_{\sigma})\, y_i^{p(GLM)} + \lambda_{\sigma}\, y_i^{np(LLR)}$
- Mixing parameter: $0 \le \lambda \le 1$
  - The optimal data driven mixing parameter for the mean estimate is given by

$$\hat{\lambda}_{\mu}^{*} = \langle \mathbf{y}^{np(LLR)}, \mathbf{y} - \mathbf{y}^{p(EWLS)} \rangle / ||\mathbf{y}^{np(LLR)}||^2$$

  - The optimal data driven mixing parameter for the variance estimate is given by

$$\hat{\lambda}_{\sigma}^{*} = \frac{\langle \mathbf{y}^{p(GLM)} - \mathbf{y}^{np(LLR)}, \mathbf{y}^{p(GLM)} - \mathbf{e}^2 \rangle}{||\mathbf{y}^{p(GLM)} - \mathbf{y}^{np(LLR)}||^2}$$

  where $\mathbf{e}^2$ is the vector of squared residuals from the means estimate

Means Model Theorem

Given the model and affiliated assumptions A1 – A6 (references) ...

$$\| \hat{\lambda}_{\mu}^{*}\mathbf{y}^{\mathrm{np(LLR)}} + \mathbf{y}^{\mathrm{p(EWLS)}} - \boldsymbol{q}_{\mu} \| = O_P(\gamma_n)$$

if $\mathcal{L}\ \delta_n > 0$

and

$$\| \hat{\lambda}_{\mu}^{*}\mathbf{y}^{\mathrm{np(LLR)}} + \mathbf{y}^{\mathrm{p(EWLS)}} - \boldsymbol{q}_{\mu} \| = O_P(n^{-.5})$$

if $\delta_n = 0$.

Variance Model Theorem

Given the model and affiliated assumptions A1 – A6 and requirements R1 - R4 (references) ...

$$\|\hat{\lambda}_{\sigma} * \mathbf{y}^{np(LLR)} + (1 - \hat{\lambda}_{\sigma})\mathbf{y}^{p(GLM)} - \boldsymbol{q}_{\sigma}\| = O_P(\gamma_n)$$

if $\mathcal{L}\ \delta_n > 0$

and

$$\|\hat{\lambda}_{\sigma} * \mathbf{y}^{np(LLR)} + (1 - \hat{\lambda}_{\sigma})\mathbf{y}^{p(GLM)} - \boldsymbol{q}_{\sigma}\| = O_P(n^{-.5})$$

if $\delta_n = 0$.

Thoughts

- A correct parametric model in either the mean or variance estimation results in the DMRR estimate converging asymptotically at the same rate as a parametric estimate ... otherwise the estimate converges at the rate of the nonparametric estimate.
- Same asymptotic results for the data driven and theoretical asymptotically optimal λ.
- DMRR achieves the "Golden Result of Model Robust Regression" as in the MRR1 and MRR2 cases.
- Asymptotic results for DMRR are not based on the number of iterations in the IRLS algorithm (above), but rather the number of observations.

References

Aitkin, M. (1987), “Modelling variance heterogeneity in Normal regression using GLIM”, *Appl. Statist.* **36**, 332-339.

Bartlett, M. S. and Kendal, D. G. (1946), “The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformations”, *Journal of the Royal Statistical Society, Series B*, **8**, 128–150.

Bishop, Y. M. M., Feinberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis,Theory and Practice*. Cambridge, MA: MIT Press.

Burman, P., and Chaudhuri, P. (1992), “A Hybrid Approach to Parametric and Nonparametric Regression”, *Technical Report No. 243*, Division of Statistics, University of California-Davis, Davis, CA, USA.

Mays, J., Birch, J., and Starnes, B. (2001), “Model Robust Regression: Combining Parametric, Nonparametric and Semiparametric Methods”, *Journal of Nonparametric Statistics*, **13**, 245-277.

References (cont.)

Myers, R. H., and Montgomery, D. C. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. New York, NY: John Wiley and Sons, Inc.

Pickle, S., Robinson, T., Birch, J., and Anderson-Cook, C. (2008), "Dual Modelling in Small Sample Settings", *Journal of Statistical Inference and Planning*, article to appear.

Robinson, T., Birch, J., and Starnes, B. (2008), "A Semiparametric Approach to Dual Modeling", *Journal of Statistical Inference and Planning*, article submitted for review.

Starnes, B. (1999), "Asymptotic Results for Model Robust Regression", Unpublished Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

Starnes, B., and Birch, J. (2000), "Asymptotic Results for Model Robust Regression", *Journal of Statistical Computation and Simulation*, **66**, 19-33.