

Appendix A: Technical Assumptions

The following list contains the assumptions necessary for the development of the asymptotic formulas in chapter 7 and to follow in appendix B (from Huber (1981)).

- A.1 \mathbf{X}^P , the parametric model matrix, is of full rank $R = k+1$
- A.2 $\max_i h_{ii} \ll 1$, that is, the diagonals of the ols HAT matrix are uniformly small (no leverage points).
- A.3 $\psi(x) = \frac{d}{dx} \rho(x)$ is continuous and bounded
- A.4 $\varepsilon_i \stackrel{iid}{\sim} \varepsilon: E[\psi(\varepsilon_i)] = 0$

Note that the derivations for M-Regression and Least Squares are left out of Appendix B, as they are well documented.

Appendix B: Asymptotic Bias and Variance Derivations

Appendix B.1: Introduction

The asymptotic bias and variance formulas for each procedure considered in this research are derived in this appendix. The results are based on the assumed underlying model of the form

$$y_i = m(x_i) + \varepsilon_i,$$

This can be rewritten as the following, in terms of the user's specified model:

$$y_i = \mathbf{x}_i^p \beta + f_i + \varepsilon_i,$$

where $\mathbf{x}_i^p = (1 \ x_i \ \dots \ x_i^k)$ is the i^{th} row of the user's model matrix \mathbf{X}^p . Based on this representation of the true underlying model, the user's misspecified model can be expressed as

$$\mathbf{y} = \mathbf{X}^p \beta + \boldsymbol{\varepsilon}^*,$$

where

$$E(\boldsymbol{\varepsilon}^*) = \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}.$$

We also assume that the errors are distributed according to the following probability distribution:

$$\boldsymbol{\varepsilon} \sim \text{CN}(\boldsymbol{\pi}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2).$$

This assumption is non-restrictive in that it neither makes assumptions about the exact number of outliers, nor the locations (in the regressor space) in which they may occur. Note that if $\boldsymbol{\varepsilon} \sim \text{CN}(\boldsymbol{\pi}, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$, then $\text{Var}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}^2)$ since we assume that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$. Hence we have that

$$\begin{aligned} \text{Var}(\boldsymbol{\varepsilon}) &= (1 - \pi) \int \boldsymbol{\varepsilon}^2 dN(0, \boldsymbol{\sigma}_1) + \pi \int \boldsymbol{\varepsilon}^2 dN(0, \boldsymbol{\sigma}_2) \\ &= (1 - \pi) \boldsymbol{\sigma}_1^2 + \pi \boldsymbol{\sigma}_2^2 = \boldsymbol{\sigma}^2, \end{aligned}$$

which leads to constant variance for the vector of errors: $\text{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\sigma}^2 \cdot \mathbf{I}$.

Appendix B.2: Loess and RLLR

RLLR and Loess, as used in this research, are unique in only two respects: the neighborhood weighting functions, and the iteration stopping criterion (RLLR iterates the residual-based weights until convergence, while Loess utilizes only 2 iterations). The generic label NP will be used to denote either of these robust nonparametric techniques in this section, as the formulas apply to both (note that when calculations are performed using these formulas, they result in distinct values for RLLR and Loess because of the differences of the two procedures mentioned above). In general, for both Loess and RLLR (which are special cases of local polynomial regression), the i^{th} predicted value is given by

$$\begin{aligned}\hat{y}_i^{\text{NP}} &= \mathbf{h}_i^{\text{NP}'} \mathbf{y} \\ &= \mathbf{x}_i^{\text{NP}'} (\mathbf{X}^{\text{NP}'} \mathbf{W}_i^{\text{NP}} \mathbf{X}^{\text{NP}})^{-1} \mathbf{X}^{\text{NP}'} \mathbf{W}_i^{\text{NP}} \mathbf{y},\end{aligned}$$

where $\mathbf{h}_i^{\text{NP}'}$ is the i^{th} row of the Hat matrix for the respective nonparametric technique, and \mathbf{W}_i is the weight matrix that results when predicting at the i^{th} location:

- $\mathbf{W}_i^{\text{NP}} = \mathbf{W}_i^{\text{Loess}} = \text{diagonal} (h_{ij}^{\text{KER-L}} \cdot \delta_j)$ for Loess, a diagonal matrix of the product of neighborhood weights for x_1, \dots, x_n calculated according to the kernel function used in the Loess algorithm and residual-based weights resulting from the use of the ψ function.

- $\mathbf{W}_i^{\text{NP}} = \mathbf{W}_i^{\text{RLLR}} = \text{diagonal} (h_{ij}^{\text{KER-R}} \cdot \delta_j)$ for RLLR, a diagonal matrix of the product of neighborhood weights for x_1, \dots, x_n calculated according to the kernel function used in the RLLR algorithm and residual-based weights resulting from the use of the ψ function.

When applying either the asymptotic expectation or variance operators, note that the residual based weights become unity (see bias and variance calculations for ORMRR in Appendix

A.2 for further explanation). Hence, the weight matrix \mathbf{W}_i^{NP} becomes just $\mathbf{W}_i^{\text{KER}}$ ($= \mathbf{W}_i^{\text{KER-R}}$ or $\mathbf{W}_i^{\text{KER-L}}$, depending on the method being considered) as in equation (4.B.1).

Note that this matrix is merely a diagonal matrix of neighborhood weights. In addition, the Hat matrix \mathbf{H}^{NP} becomes \mathbf{H}^{LLR} since it is identical to the local linear regression hat matrix. Note that this matrix depends on the weight function that is used to determine the neighborhood weights, and hence the notation $\mathbf{H}^{\text{LLR-L}}$ and $\mathbf{H}^{\text{LLR-R}}$ will be used to distinguish between the weight functions utilized by Loess and RLLR, respectively. The general notation that will be used in the derivations is $\mathbf{H}^{\text{LLR-}\bullet}$.

Based on these assumptions and results, the asymptotic bias for the vector of predicted values resulting from either RLLR or Loess is given by

$$\begin{aligned} \text{Bias}(\hat{\mathbf{y}}^{\text{NP}}) &= \text{E}(\hat{\mathbf{y}}^{\text{NP}}) - \text{E}(\mathbf{y}) \\ &= \text{E}(\mathbf{H}^{\text{NP}} \mathbf{y}) - \mathbf{m} \\ &= \mathbf{H}^{\text{LLR-}\bullet} \mathbf{m} - \mathbf{m} \\ &= -(\mathbf{I} - \mathbf{H}^{\text{LLR-}\bullet}) \mathbf{m} \end{aligned}$$

Similarly, the asymptotic variance of the vector of predicted values calculated by using either RLLR or Loess is given by

$$\begin{aligned} \text{Var}(\hat{\mathbf{y}}^{\text{NP}}) &= \text{Var}(\mathbf{H}^{\text{NP}} \mathbf{y}) \\ &= \mathbf{H}^{\text{LLR-}\bullet} \text{Var}(\mathbf{y}) \mathbf{H}^{\text{LLR-}\bullet} \\ &= \mathbf{V}^2 \mathbf{H}^{\text{LLR-}\bullet} \mathbf{H}^{\text{LLR-}\bullet'} \end{aligned}$$

where

$$\mathbf{V}^2 = \sigma^2 \cdot \frac{\text{E}\{\psi(\varepsilon_i/\sigma)^2\}}{\text{E}\{\psi'(\varepsilon_i/\sigma)\}^2}.$$

Appendix B.3: ORMRR

Recall that the vector of fitted values for ORMRR is given by:

$$\begin{aligned}\hat{\mathbf{y}}^{\text{ORMRR}} &= \mathbf{H}^{\text{ORMRR}} \mathbf{y} \\ &= \mathbf{H}^{\text{M}} \mathbf{y} + \lambda \mathbf{H}^{\text{RLLR}} \mathbf{r} \\ &= [\mathbf{H}^{\text{M}} + \lambda \mathbf{H}^{\text{RLLR}} (\mathbf{I} - \mathbf{H}^{\text{M}})] \mathbf{y}\end{aligned}$$

where \mathbf{H}^{M} is the Hat matrix for the M-regression fit to the raw data, and \mathbf{H}^{RLLR} is the Hat matrix for the robust local linear regression fit to the residuals formed from M-regression. Note that the matrix \mathbf{H}^{RLLR} depends implicitly on the bandwidth b , and the overall fit depends on the mixing parameter λ . In addition, both \mathbf{H}^{M} and \mathbf{H}^{RLLR} are functions of residual based weight matrices (unique to each) that depend on their corresponding ψ function.

The following identity is used in order to obtain the simplified bias and variance calculations for the estimator:

$$\mathbf{H}^{\text{M}} \mathbf{X}^{\text{P}} = \mathbf{X}^{\text{P}} (\mathbf{X}^{\text{P}'} \mathbf{W} \mathbf{X}^{\text{P}})^{-1} \mathbf{X}^{\text{P}'} \mathbf{W} \mathbf{X}^{\text{P}} = \mathbf{X}^{\text{P}}.$$

The derivation, then, of the asymptotic bias for the vector of predicted values that result from ORMRR is given by

$$\text{Bias}(\hat{\mathbf{y}}^{\text{ORMRR}}) = E(\hat{\mathbf{y}}^{\text{ORMRR}}) - E(\mathbf{y}) \quad (1)$$

$$= E(\mathbf{H}^{\text{ORMRR}} \mathbf{y}) - (\mathbf{X}^{\text{P}} \boldsymbol{\beta} + \mathbf{f}) \quad (2)$$

$$= E(\mathbf{H}^{\text{ORMRR}}) \cdot E(\mathbf{y}) - (\mathbf{X}^{\text{P}} \boldsymbol{\beta} + \mathbf{f}) \quad (3)$$

$$= E[\mathbf{H}^{\text{M}} + \lambda \mathbf{H}^{\text{RLLR}} (\mathbf{I} - \mathbf{H}^{\text{M}})] (\mathbf{X}^{\text{P}} \boldsymbol{\beta} + \mathbf{f}) - (\mathbf{X}^{\text{P}} \boldsymbol{\beta} + \mathbf{f}) \quad (4)$$

$$= [\mathbf{H}^{\text{OLS}} + \lambda \mathbf{H}^{\text{LLR}} (\mathbf{I} - \mathbf{H}^{\text{OLS}})] (\mathbf{X}^{\text{P}} \boldsymbol{\beta} + \mathbf{f}) - (\mathbf{X}^{\text{P}} \boldsymbol{\beta} + \mathbf{f}) \quad (5)$$

$$= \mathbf{X}^{\text{P}} \boldsymbol{\beta} + 0 - \mathbf{X}^{\text{P}} \boldsymbol{\beta} + \mathbf{H}_{\text{Asym}}^{\text{ORMRR}} \mathbf{f} - \mathbf{f} \quad (6)$$

$$= -(\mathbf{I} - \mathbf{H}_{\text{Asym}}^{\text{ORMRR}}) \mathbf{f}, \quad (7)$$

where $\mathbf{H}_{\text{Asym}}^{\text{ORMRR}} = \mathbf{H}^{\text{OLS}} + \lambda \mathbf{H}^{\text{LLR-R}} (\mathbf{I} - \mathbf{H}^{\text{OLS}})$.

The justification for step 5 follows from the asymptotic variance calculations for M-Regression developed by Huber (1981). This leads to the asymptotic result $E(\mathbf{H}^{\text{M}}) = \mathbf{H}^{\text{OLS}}$ and

$E(\mathbf{H}^{\text{RLLR}}) = \mathbf{H}^{\text{LLR-R}}$, or in other words, $\mathbf{H}_{\text{Asym}}^{\text{M}} = \mathbf{H}^{\text{OLS}}$ and $\mathbf{H}_{\text{Asym}}^{\text{RLLR}} = \mathbf{H}^{\text{LLR-R}}$, since the residual-based weights in \mathbf{W} for M-regression in \mathbf{H}^{M} and that form \mathbf{W}_i^{NP} in \mathbf{H}^{RLLR} become ones.

The derivation of the variance for the ORMRR vector of fitted values is similar to that of the bias derivation, in that $\mathbf{H}^{\text{ORMRR}}$ becomes $\mathbf{H}_{\text{Asym}}^{\text{ORMRR}}$ when the asymptotic variance operator is applied. The derivation also depends on the assumption of constant variance (which implies that $\text{Var}(\mathbf{y}) = V^2\mathbf{I}$). Based on these results and assumptions, the asymptotic variance of the vector of predicted values that result from ORMRR is given by

$$\begin{aligned}\text{Var}(\hat{\mathbf{y}}^{\text{ORMRR}}) &= \text{Var}(\mathbf{H}^{\text{ORMRR}}\mathbf{y}) \\ &= \mathbf{H}_{\text{Asym}}^{\text{ORMRR}}\text{Var}(\mathbf{y})\mathbf{H}_{\text{Asym}}^{\text{ORMRR}'} \\ &= V^2\mathbf{H}_{\text{Asym}}^{\text{ORMRR}}\mathbf{H}_{\text{Asym}}^{\text{ORMRR}'} \\ &= V^2[\mathbf{H}^{\text{OLS}} + \lambda\mathbf{H}^{\text{LLR-R}}(\mathbf{I} - \mathbf{H}^{\text{OLS}})][\mathbf{H}^{\text{OLS}} + \lambda\mathbf{H}^{\text{LLR-R}}(\mathbf{I} - \mathbf{H}^{\text{OLS}})]'\end{aligned}$$

where

$$V^2 = \sigma^2 \cdot \frac{E\{\psi(\varepsilon_i/\sigma)^2\}}{E\{\psi(\varepsilon_i/\sigma)\}^2}$$

The properties of the vector of smoothed residuals (residuals from the parametric fit, smoothed by way of RLLR) are conditioned on the parametric fit to the raw data. The residuals are represented by $\mathbf{r} = \mathbf{y} - \mathbf{X}^{\text{P}}\boldsymbol{\beta}^{\text{M}}$, and the smoothed residuals are given by $\hat{\mathbf{r}}^{\text{RLLR}} = \mathbf{H}^{\text{RLLR}}\mathbf{r}$. The bias and variance derivations are given below:

$$\text{Bias}(\hat{\mathbf{r}}^{\text{RLLR}}) = E(\hat{\mathbf{r}}^{\text{RLLR}}) - E(\mathbf{r}) \quad (1)$$

$$= E(\mathbf{H}^{\text{RLLR}}\mathbf{r}) - E(\mathbf{r}) \quad (2)$$

$$= E(-(\mathbf{I} - \mathbf{H}^{\text{RLLR}})\mathbf{r}) \quad (3)$$

$$= -(\mathbf{I} - \mathbf{H}^{\text{LLR-R}})E(\mathbf{r}) \quad (4)$$

$$= -(\mathbf{I} - \mathbf{H}^{\text{LLR-R}})E(\mathbf{y} - \hat{\mathbf{y}}^{\text{M}}) \quad (5)$$

$$= -(\mathbf{I} - \mathbf{H}^{\text{LLR-R}}) \mathbf{E}((\mathbf{I} - \mathbf{H}^{\text{M}}) \mathbf{y}) \quad (6)$$

$$= -(\mathbf{I} - \mathbf{H}^{\text{LLR-R}}) (\mathbf{I} - \mathbf{H}^{\text{OLS}}) (\mathbf{X}^{\text{P}} \boldsymbol{\beta} + \mathbf{f}) \quad (7)$$

$$= -(\mathbf{I} - \mathbf{H}^{\text{LLR-R}}) (\mathbf{I} - \mathbf{H}^{\text{OLS}}) \mathbf{f} \quad (8)$$

Note that step 4 can be justified for the same reasons as those given for step 5 in the derivation of the bias of the overall ORMRR fit. Step 8 follows from the fact that $\mathbf{H}^{\text{OLS}} \mathbf{X}^{\text{P}} \boldsymbol{\beta} = \mathbf{X}^{\text{P}} \boldsymbol{\beta}$.

Based on the same results and properties used in determining the variance properties for the global ORMRR fit, the asymptotic variance of the smoothed residuals using RLLR is given by

$$\begin{aligned} \text{Var}(\hat{\mathbf{r}}^{\text{RLLR}}) &= \text{Var}(\mathbf{H}^{\text{RLLR}} \mathbf{r}) = \text{Var}(\mathbf{H}^{\text{RLLR}} (\mathbf{I} - \mathbf{H}^{\text{M}}) \mathbf{y}) \\ &= [\mathbf{H}^{\text{LLR-R}} (\mathbf{I} - \mathbf{H}^{\text{OLS}})] \text{Var}(\mathbf{y}) [\mathbf{H}^{\text{LLR-R}} (\mathbf{I} - \mathbf{H}^{\text{OLS}})]' \\ &= \mathbf{V}^2 [\mathbf{H}^{\text{LLR-R}} (\mathbf{I} - \mathbf{H}^{\text{M}})] [\mathbf{H}^{\text{LLR-R}} (\mathbf{I} - \mathbf{H}^{\text{M}})]' \\ &= \mathbf{V}^2 \mathbf{H}^{\text{LLR-R}} (\mathbf{I} - \mathbf{H}^{\text{OLS}}) \mathbf{H}^{\text{LLR-R}'} \end{aligned}$$

Vita

Christopher Ashley Assaid

Christopher A. Assaid was born February 11, 1970 to Don Alan and Ruth Ann Longanacre Assaid of Roanoke, Virginia. He graduated fourth in his class of 42 from Roanoke Valley Christian School in Roanoke, VA on June 10, 1988. On May 2, 1992 he graduated Magna Cum Laude from Roanoke College in Salem, VA with a B.S. degree in Mathematics and as an Honors Program graduate.

On December 18, 1993, and May 9, 1997, he received his M.S. degree in Statistics and his Ph.D. degree in Statistics, respectively from Virginia Polytechnic Institute and State University. He was awarded the Klaus Hinkelmann award for Outstanding Graduate Assistant for 1995-1996.

The author is a member of the honor societies Pi Mu Epsilon, Alpha Chi, and Mu Sigma Rho. He is also a member of the American Statistical Association.

Upon completion of his Ph.D., he was employed as a Senior Project Biostatistician with the DuPont Merck Pharmaceutical Company in Wilmington, Delaware.