Simultaneous Confidence Bands for Linear Regression With Heteroscedastic Errors

Julian J. FARAWAY and Jiayang SUN*

The Scheffé method may be used to construct simultaneous confidence bands for a regression surface for the whole predictor space. When the bands need only hold for a subset of that space, previous authors have described how the bands can be appropriately narrowed while still maintaining the desired level of confidence. Data with heteroscedastic errors occur often, and unless some transformation is feasible, there is no obvious way to construct bands using the current methods. This article shows how to construct approximate simultaneous confidence bands when the errors are heteroscedastic and symmetric. The method works when the weights are known or unknown and have to be estimated. The region in which the bands must hold can be quite general and will work for any linear unbiased estimate of the regression surface. The method can even be extended to linear estimates with a small amount of bias such as nonparametric kernel regression smoothers.

KEY WORDS: Scheffé confidence interval; Tube formula.

1. INTRODUCTION

Consider the linear regression model

$$Y_i = \beta_0 + \sum_{i=1}^{p} \beta_j x_{ij} + \varepsilon_i \qquad i = 1, \ldots, n$$

$$= f(x_{i1}, \ldots, x_{ip}) + \varepsilon_i,$$

where for simplicity the errors, ε_i, are normally distributed with mean zero and variance $\sigma_i^2 = \sigma^2/w_i$. We may also express this in the usual matrix form, with $\mathbf{X}$ an $n \times (p + 1)$ matrix, $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ as

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon.$$

Scheffé's (1959) S method may be used to obtain simultaneous confidence bands for the regression surface $f(x_1, \ldots, x_p)$ for $x_1, \ldots, x_p \in \mathcal{X}$, when $\mathcal{X} \equiv \mathbb{R}^p$ and when the variances, σ_i^2, are equal. The method may be adapted if the variances are unequal but known. In practice we will not need simultaneous confidence bands over all $\mathbb{R}^p$, so the Scheffé bands will be wider than necessary for the desired level of confidence.

Wynn and Bloomfield (1971) described a straightforward adjustment of Scheffé's method that works for simple linear regression ($p = 1$) for a finite interval. Wynn (1984) extended this to the one-dimensional polynomial regression case, where Uusipaikka (1983) allowed for bands on an arbitrary finite union of intervals and points. Naiman (1987) described the construction of bands in the multiple regression case over convex polyhedral sets. Seber (1977) provided a survey of earlier methods. Sun and Loader (1994) generalized further to linear estimates over general regions and showed how to adjust for bias when a nonlinear estimate is used.

This article addresses the heteroscedastic error case where the form of the weights, w_i, is known or unknown. When the weights are known, we can construct simultaneous confidence bands by using an extension of the method described by Sun and Loader (1994). But if the weights are unknown and are estimated, then we must make further adjustments to account for the extra variation introduced by the estimation of the weights. We outline the known weights case by way of introduction. Theoretical justification has been provided by Sun and Loader (1994).

The least squares estimate of $f(\mathbf{x})$, where $\mathbf{x} = (1, x_1, \ldots, x_p)^T$, is

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{n} l_i(\mathbf{x}) Y_i = \mathbf{l}(\mathbf{x})^T \mathbf{Y},$$

where

$$\mathbf{l}(\mathbf{x})^T = \mathbf{x}^T (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1}$$

and $\mathbf{\Sigma}^{-1} = \operatorname{diag}(w_i)$. Now, because

$$\operatorname{var} \hat{f}(\mathbf{x}) = \sum_{i=1}^{n} l_i(\mathbf{x})^2 \sigma_i^2$$

$$= \sum_{i=1}^{n} l_i^w(\mathbf{x})^2 \sigma^2,$$

where $l_i^w(\mathbf{x}) = l_i(\mathbf{x})/w_i^{1/2}$, reasonable simultaneous confidence bands are of the form

$$|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \le c\hat{\sigma} \|\mathbf{l}^w(\mathbf{x})\|,$$

where $\|\cdot\|$ denotes the L_2 norm, $\hat{\sigma}^2 = \nu^{-1} \|(\mathbf{I} - \mathbf{L})\mathbf{Y}\|^2$, $\mathbf{L} = \mathbf{X}(\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1}$, $\mathbf{l}^w = (l_1^w, \ldots, l_n^w)^T$, and ν is the residual degrees of freedom, computed as

$$\nu = \operatorname{tr}((\mathbf{I} - \mathbf{L})(\mathbf{I} - \mathbf{L})^T).$$

Because $|\hat{f}(\mathbf{x}) - f(\mathbf{x})| = |\mathbf{l}(\mathbf{x})^T \varepsilon|$, for a $1 - \alpha$ simultaneous confidence band, we must choose c such that

$$\alpha = P(|\mathbf{l}(\mathbf{x})^T \varepsilon| > c\hat{\sigma} \|\mathbf{l}^w(\mathbf{x})\|, \text{ for some } \mathbf{x} \in \mathcal{X})$$

$$= P\left(\sup_{\mathbf{x} \in \mathcal{X}} \left| \left\langle \mathbf{T}(\mathbf{x}), \frac{\varepsilon_w}{\sigma} \right\rangle \right| > \frac{c\hat{\sigma}}{\sigma} \right),$$

where $\mathbf{T}(\mathbf{x}) = \mathbf{l}^w(\mathbf{x}) / \|\mathbf{l}^w(\mathbf{x})\|$, $\langle , \rangle$ denotes the inner product, and $\varepsilon_w = (\varepsilon_1 \sqrt{w_1}, \ldots, \varepsilon_n \sqrt{w_n})^T \sim N(0, \sigma^2 \mathbf{I})$. Because $\langle \mathbf{T}(\mathbf{x}), \varepsilon_w/\sigma \rangle$ is a Gaussian random field and independent of $\hat{\sigma}/\sigma$,

* Julian J. Faraway is Associate Professor and Jiayang Sun is Assistant Professor, Department of Statistics, University of Michigan, Ann Arbor, MI 48109. Sun's research is supported by National Science Foundation Grant DMS-09207357. The authors acknowledge the use of the University of Michigan Center for Parallel Computing, partially funded under National Science Foundation Grant CDA-92-14296.

Journal of the American Statistical Association
September 1995, Vol. 90, No. 431, Theory and Methods

Sun's (1993) tube formula–based approximation for the tail probability may be applied to find c. When $\dim(\mathcal{X}) = 1$ the approximation yields

$$\alpha \approx \frac{\kappa_0}{\pi}\left(1 + \frac{c^2}{\nu}\right)^{-\nu/2} + E \cdot P(|t_\nu| > c), \qquad (1)$$

where t_ν denotes a t-distributed random variable with ν degrees of freedom and $\kappa_0 = \int_{\mathbf{x}\in\mathcal{X}} \|T'(\mathbf{x})\| \, d\mathbf{x}$. The $P(|t_\nu| > c)$ term is a boundary correction. When $\dim(\mathcal{X}) > 1$, a similar approximation may be obtained using the same tube formula. When $\mathcal{X}$ is a finite interval or rectangular region, E is 1; each extra disjoint region or point adds 1 to E (for other $\mathcal{X}$, consult Sun and Loader (1994) for details). An easier way to compute κ_0 when $\mathcal{X} = [a, b]$ is to partition $[a, b]$ into $a = z_0 < \cdots < z_m = b$; then

$$\kappa_0 = \sum_{i=1}^{m} \int_{z_{i-1}}^{z_i} \|T'(x)\| \, dx \approx \sum_{i=1}^{m} \|T(z_i) - T(z_{i-1})\|.$$

This method can be extended to work with other linear estimates of $\hat{f}(\mathbf{x})$ (just define $l_i(\mathbf{x})$ appropriately) and for nonnormal errors (the same approximation for the tail probability of the maximum of the Gaussian random field applies). When the weights are unknown and need to be estimated, some complex adjustments are necessary; these are detailed in the next section.

2. UNKNOWN WEIGHTS

It would be nice to simply estimate the unknown weights using any consistent parametric or nonparametric estimator and then just apply the method described in Section 1. Unfortunately, this is insufficient, because the bands obtained have rather less than the desired level of confidence. Extra variation is introduced by the estimation of the weights; adjustments to account for this are detailed next.

Denote $\hat{\Sigma}^{-1} = \operatorname{diag}(\hat{w}_i)$, where $\hat{w}_i$ are consistent estimates of weights w_i. Accordingly, write

$$\hat{\mathbf{l}}(\mathbf{x})^T = \mathbf{x}^T(\mathbf{X}^T\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\Sigma}^{-1},$$
$$\hat{l}_i^w(\mathbf{x}) = \hat{l}_i(\mathbf{x})/\hat{w}_i^{1/2},$$
$$\hat{\mathbf{L}} = \mathbf{X}(\mathbf{X}^T\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\Sigma}^{-1},$$

and

$$\tilde{\sigma}^2 = \nu^{-1}\|(\mathbf{I} - \hat{\mathbf{L}})\mathbf{Y}\|^2.$$

Thus in the unknown weights case, $\hat{f}(\mathbf{x}) = \langle \hat{\mathbf{l}}(\mathbf{x}), \mathbf{Y} \rangle$, and the simultaneous confidence bands are of the form

$$|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \le c\tilde{\sigma}\|\hat{\mathbf{l}}^w(\mathbf{x})\|,$$

where $\hat{\mathbf{l}}^w(\mathbf{x}) = (\hat{l}_1^w(\mathbf{x}), \ldots, \hat{l}_n^w(\mathbf{x}))^T$. The constant c is chosen so that the coverage probability for all $\mathbf{x}$ simultaneously is a predetermined level, $1 - \alpha$. A (conservative) approximation formula for determining c for this coverage probability is

$$\alpha \approx E \cdot P(|t_\nu| > c) + \frac{\kappa_0}{\pi}\left\{\left(1 + \frac{c^2}{\nu}\right)^{-\nu/2} + c(\delta' + \gamma' c \nu^{-1/2}) \times \frac{2^{1/2}\Gamma\left(\frac{\nu+1}{2}\right)}{\nu^{1/2}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{c^2}{\nu}\right)^{-(\nu+1)/2}\right\}, \qquad (2)$$

where E is as in Section 1 and γ' and δ' are two constants that can be estimated by estimates of

$$\gamma = \|(\mathbf{L} - \hat{\mathbf{L}})\mathbf{Y}\|^2/\sigma^2, \qquad \delta = \sup_{\mathbf{x}\in\mathcal{X}} \frac{|\langle \mathbf{l}(\mathbf{x}) - \hat{\mathbf{l}}(\mathbf{x}), \mathbf{Y} \rangle|}{\sigma\|\mathbf{l}^w(\mathbf{x})\|}.$$

Note that $2^{1/2}\Gamma[(\nu + 1)/2]/\Gamma(\nu/2) \approx \sqrt{\nu}(4\nu - 1)/(4\nu)$. See the Appendix for details on how formula (2) is obtained. A simultaneous confidence interval based on c calculated from (2) should be conservative and close to the nominal level $1 - \alpha$ when n is large and the errors are symmetric.

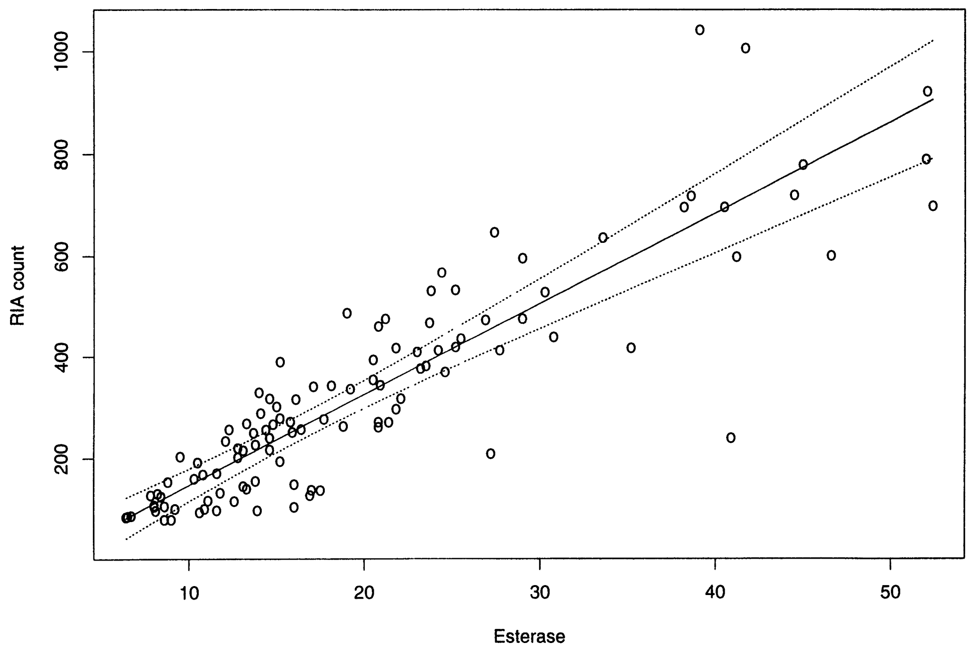

Figure 1. 99% Simultaneous Confidence Bands for the Mean Response. Notice how the bands are wider where the response is more variable.

Table 1. Known Weights

Weights	Constant		Linear		Quadratic		Valley	
[a, b]	[.25, .75]	[−1, 2]	[.25, .75]	[−1, 2]	[.25, .75]	[−1, 2]	[.25, .75]	[−1, 2]
				n = 25, Normal error				
$1-\alpha=.9$	.901	.921	.896	.915	.881	.897	.900	.920
$1-\alpha=.95$	.950	.959	.946	.955	.934	.943	.950	.958
$1-\alpha=.99$	.990	.992	.988	.990	.984	.987	.990	.991
				n = 50, Normal error				
$1-\alpha=.9$	.901	.917	.896	.915	.887	.907	.900	.915
$1-\alpha=.95$	.950	.958	.948	.956	.941	.950	.950	.958
$1-\alpha=.99$	.990	.991	.989	.991	.987	.989	.990	.991
				n = 50, Contaminated normal error				
$1-\alpha=.9$	.902	.919	.898	.916	.886	.907	.900	.915
$1-\alpha=.95$	.952	.960	.949	.958	.943	.951	.949	.957
$1-\alpha=.99$	.991	.993	.991	.992	.989	.990	.990	.991

When the bias is not zero but is sufficiently small, the same argument holds approximately. We simply replace δ' in (2) by $\delta' + b'$, where b' is an estimate of

$$b = \sup_{\mathbf{x}\in\mathcal{X}} \frac{|f(\mathbf{x}) - \langle \mathbf{l}(\mathbf{x}), \mu\rangle|}{\sigma\|\mathbf{l}^w(\mathbf{x})\|}.$$

To make this work in practice, we need to find δ', γ', and possibly b' (if we are using a biased estimator).

Often, the statistician has no hard information about the form of the weight function, so parametric methods may be inapplicable. It is usually reasonable to assume that the weight function is smooth, and so a nonparametric method of estimation may be used. We use the one described by Müller (1988, p. 153). The "raw variance" at x_i is given by

$$\tau_i^2 = \frac{2}{3}\left[Y_i - \frac{1}{2}(Y_{i-1} + Y_{i+1})\right]^2.$$

These τ_i^2 are then smoothed using a kernel-based nonparametric estimator, and estimates of the weights can then be derived. If possible, the statistician is advised to plot these τ_i^2 and then choose the bandwidth based on knowledge of the smoothness of the weight function. Automatic methods of bandwidth selection, many of which are available, could also be used. Other methods, such as smoothing the $\log \tau_i^2$, were tried, but these were no better.

Regardless of which method of estimating the weights is chosen, the formula (2) cannot be applied without some values for δ' and γ'. This in turn requires estimates of $\mathbf{l}(\mathbf{x})$ and $\mathbf{L}$, called $\tilde{\mathbf{l}}$ and $\tilde{\mathbf{L}}$, which depend on the unknown true weights. We use the nonparametric estimator described earlier to estimate these weights solely for the purpose of finding δ and γ. The choice of bandwidth used for this estimation is problematic. We found, from simulation experience, that using a bandwidth about two-thirds the size of the one that we actually use to estimate the weight function produced acceptable results. So now

$$\hat{\gamma}' = \|(\tilde{\mathbf{L}} - \hat{\mathbf{L}})\mathbf{Y}\|^2/\hat{\sigma}^2, \qquad \hat{\delta}' = \sup_{\mathbf{x}\in\mathcal{X}} \frac{|\langle \tilde{\mathbf{l}}(\mathbf{x}) - \hat{\mathbf{l}}(\mathbf{x}), \mathbf{Y}\rangle|}{\hat{\sigma}\|\tilde{\mathbf{l}}^w(\mathbf{x})\|}.$$

3. EXAMPLE

Carroll and Ruppert (1988, p. 48) presented data from an assay for the concentration of an enzyme esterase. The observed concentration of esterase is the predictor, and the number of bindings counted is the response. We estimate the weight function nonparametrically, choosing the bandwidth by eye to produce a smooth weight function. We set $\mathcal{X}$ as the range of the predictor. Figure 1 shows 99% simultaneous confidence bands for the mean response.

4. SIMULATION

The design used throughout was

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where $\varepsilon_i \sim N(0, w_i^{-1})$, normal or $\varepsilon_i \sim \frac{2}{3}N(0, w_i^{-1}) + \frac{1}{3}N(0, 3^2 w_i^{-1})$, contaminated normal and $x_i = (2i-1)/2n$. We set $\beta_0 = \beta_1 = 0$ without loss of generality.

First, we investigate the performance of our bands when the weights are known. We try four weight functions: con-

Table 2. Parametrically Estimated Weights

Weights [a, b]	$\eta = 0$		$\eta = 1$	
	[2.25, .75]	[−1, 2]	[.25, .75]	[−1, 2]
		n = 25, Normal error		
$1-\alpha=.9$	.909	.936	.904	.933
$1-\alpha=.95$	.956	.968	.950	.966
$1-\alpha=.99$	.992	.994	.989	.993
		n = 50, Normal error		
$1-\alpha=.9$	.911	.936	.908	.936
$1-\alpha=.95$	.956	.968	.954	.968
$1-\alpha=.99$	.992	.994	.990	.993
		n = 50, Contaminated normal error		
$1-\alpha=.9$	.912	.936	.918	.942
$1-\alpha=.95$	.959	.970	.962	.973
$1-\alpha=.99$	.994	.995	.994	.996

Table 3. Nonparametrically Estimated Weights

Weights	Constant		Linear		Quadratic		Valley	
[a, b]	[.25, .75]	[−1, 2]	[.25, .75]	[−1, 2]	[.25, .75]	[−1, 2]	[.25, .75]	[−1, 2]
			n = 25, Normal error					
$1-\alpha=.9$	.897	.924	.876	.903	.893	.933	.953	.978
$1-\alpha=.95$	.946	.959	.930	.944	.943	.964	.979	.991
$1-\alpha=.99$	.988	.991	.981	.985	.986	.991	.997	.999
			n = 50, Normal error					
$1-\alpha=.9$	.896	.921	.886	.907	.921	.940	.960	.983
$1-\alpha=.95$	.946	.958	.937	.948	.960	.968	.983	.993
$1-\alpha=.99$	.988	.991	.984	.986	.992	.991	.998	.999
			n = 50, Contaminated normal error					
$1-\alpha=.9$	.914	.939	.903	.925	.922	.944	.963	.983
$1-\alpha=.95$	.959	.971	.951	.961	.961	.969	.984	.994
$1-\alpha=.99$	.993	.995	.990	.992	.992	.992	.998	.999

stant weights, $w_i = 1$; linear weights, $w_i = i$; quadratic weights, $w_i = i^2$; and valley weights, $w_i = |(n+1)/2 - i| + .5$. The results are shown in Table 1, where the entries are the estimated true confidence probabilities.

Pseudorandom normal numbers were generated using a linear congruential generator with large period and the polar method; 100,000 replications were used throughout. The entries in Table 1 are the simulated coverage probabilities of the confidence bands using the formula (1).

Simulation standard error is at most ± 1 in the third digit. The performance is better for high confidence levels, because the tail approximation used is more accurate for smaller α. The performance is reasonable even for $n = 25$, and the presence of outliers in the error distribution has little impact.

When the weight function is of the form $w_i = (1 + \eta i)^{-2}$, η may be simply estimated by iteratively regressing the absolute vale of the residuals, $\hat{\varepsilon}_i$, on x_i. Of course there are many conceivable parametric forms for the weight function and ways of estimating it (see Carroll and Ruppert 1988, for example), and because our method demands only the estimated weights, any of these methods could be used in conjunction with ours to produce the simultaneous confidence bands. The weights must be estimated well for the confidence bands to be accurate, so naturally the user must choose an appropriate method of estimation. The results are shown in Table 2. The entries in the table are the estimated coverage probabilities of the confidence bands using the formula (2).

The results tend to the conservative side, which is not unexpected given the approximations made. The effect of a contaminated normal error is to give wider confidence intervals. In this circumstance a more robust estimation of the weights may produce a better result. When $n = 25$, we found that our method of estimating the weights had to be modified, because, in some instances, the estimated weights were too large. Our solution was to truncate these large weights. Be clear that this modification involves only the estimation of the weights and not the construction of the confidence bands themselves. Furthermore, such truncation is standard practice in weight estimation (see Carroll and Ruppert 1988). This merely amplifies our prior point that the bands are only as good as the estimated weights.

When we are uncertain as to the parametric form of the weight function, we can always use a nonparametric estimator as described previously. The results are shown in Table 3.

We used the same bandwidth within each sample size. Relatively large bandwidths were used, as this gave the best overall results, although better results could have been obtained had we chosen the bandwidth individually for each case. In practice, we recommend that the user plot the estimated weight function and select the bandwidth using any ancillary knowledge about the smoothness of the true weight function. A long-tailed error distribution results in intervals that are more conservative, although some experimentation with more robust smoothing methods revealed that this could be corrected. In any case, it would be better if the data were examined for outliers before blindly applying this method.

5. CONCLUSION

We have described a practical method for constructing simultaneous confidence bands for linear regression estimates when there is heteroscedastic error. We have seen that the method works well enough for small sample sizes and when the errors are not necessarily normally distributed. But users of this method must take responsibility for estimating the weights sensibly as well as taking all the other precautions of a prudent regression analyst. Given this, the user will be rewarded with simultaneous confidence bands narrower than those obtainable using the previously available Scheffé method but that still maintain the desired level of confidence.

APPENDIX: DERIVATION

To derive the approximation formula for the simultaneous confidence interval, assume for the moment that the image of $\mathbf{T}(\mathbf{x}) = \mathbf{l}^w(\mathbf{x})/\|\mathbf{l}^w(\mathbf{x})\|$ on $\mathcal{X}$ is connected and so no boundary correction is necessary.

If we let $\mu = E\mathbf{Y}$, so that $\mathbf{Y} = \mu + \varepsilon$, then we have

$$\begin{aligned} |f(\mathbf{x}) - \hat{f}(\mathbf{x})| &= |f(\mathbf{x}) - \langle \hat{\mathbf{l}}(\mathbf{x}), \mathbf{Y}\rangle| \\ &= |f(\mathbf{x}) - \langle \mathbf{l}(\mathbf{x}), \mu\rangle - \langle \mathbf{l}(\mathbf{x}), \varepsilon\rangle \\ &\quad + \langle \mathbf{l}(\mathbf{x}) - \hat{\mathbf{l}}(\mathbf{x}), \mathbf{Y}\rangle| \\ &\le |f(\mathbf{x}) - \langle \mathbf{l}(\mathbf{x}), \mu\rangle| + |\langle \mathbf{l}(\mathbf{x}), \varepsilon\rangle| \\ &\quad + |\langle \mathbf{l}(\mathbf{x}) - \hat{\mathbf{l}}(\mathbf{x}), \mathbf{Y}\rangle|. \end{aligned} \tag{A.1}$$

We must choose c such that

$$\begin{aligned} \alpha &= P\{|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \ge c\tilde{\sigma}\|\hat{\mathbf{l}}^w(\mathbf{x})\|, \text{ for some } \mathbf{x} \in \mathcal{X}\} \\ &\le P\left\{\frac{|f(\mathbf{x}) - \langle \mathbf{l}(\mathbf{x}), \mu\rangle|}{\|\mathbf{l}^w(\mathbf{x})\|} + \frac{|\langle \mathbf{l}(\mathbf{x}), \varepsilon\rangle|}{\|\mathbf{l}^w(\mathbf{x})\|}\right. \\ &\quad \left. + \frac{|\langle \mathbf{l}(\mathbf{x}) - \hat{\mathbf{l}}(\mathbf{x}), \mathbf{Y}\rangle|}{\|\mathbf{l}^w(\mathbf{x})\|} \ge c\tilde{\sigma}\frac{\|\hat{\mathbf{l}}^w(\mathbf{x})\|}{\|\mathbf{l}^w(\mathbf{x})\|}, \text{ for some } \mathbf{x} \in \mathcal{X}\right\} \\ &\le P\left\{\sup_{\mathbf{x}\in\mathcal{X}} \left|\left\langle \mathbf{T}(\mathbf{x}), \frac{\varepsilon_w}{\sigma}\right\rangle\right| \ge c\frac{\tilde{\sigma}}{\sigma}a - b - \delta\right\} \\ &\le 2P\left\{\sup_{\mathbf{x}\in\mathcal{X}} \left\langle \mathbf{T}(\mathbf{x}), \frac{\varepsilon_w}{\sigma}\right\rangle \ge c\frac{\tilde{\sigma}}{\sigma}a - b - \delta\right\}, \end{aligned}$$

where b is the normalized bias, a is the minimum of the ratio of estimated $\mathbf{l}^w$ and the true $\mathbf{l}^w$, and δ is the difference in regression estimate due to the variation in estimating the unknown weights:

$$b = \sup_{\mathbf{x}\in\mathcal{X}} \frac{|f(\mathbf{x}) - \langle \mathbf{l}(\mathbf{x}), \mu\rangle|}{\sigma\|\mathbf{l}^w(\mathbf{x})\|}, \qquad a = \inf_{\mathbf{x}\in\mathcal{X}} \frac{\|\hat{\mathbf{l}}^w(\mathbf{x})\|}{\|\mathbf{l}^w(\mathbf{x})\|},$$

$$\delta = \sup_{\mathbf{x}\in\mathcal{X}} \frac{|\langle \mathbf{l}(\mathbf{x}) - \hat{\mathbf{l}}(\mathbf{x}), \mathbf{Y}\rangle|}{\sigma\|\mathbf{l}^w(\mathbf{x})\|}.$$

Now $\tilde{\sigma}/\sigma \ge \hat{\sigma}/\sigma - \gamma/\sqrt{\nu}$, where $\gamma = \|(\mathbf{L} - \hat{\mathbf{L}})\mathbf{Y}\|/\sigma$ by the triangle inequality, so

$$\frac{\alpha}{2} \le P\left\{\sup_{\mathbf{x}\in\mathcal{X}} \left\langle \mathbf{T}(\mathbf{x}), \frac{\varepsilon_w}{\sigma}\right\rangle \ge \frac{ac}{\nu^{1/2}}\left(\frac{\nu^{1/2}\hat{\sigma}}{\sigma} - \gamma\right) - b - \delta\right\}.$$

Hence γ is the difference in the variance estimate due to the variation in estimating the unknown weights. Note that $b = 0$ if the estimate of the regression surface is unbiased, as it is in least squares. If a nonparametric, linear estimator were used, then confidence bands could be constructed by taking this term into account.

When the estimates of the weights are estimated consistently,

$$a = 1 + o_p(1), \qquad \delta = o_p(1), \quad \text{and} \quad \gamma = o_p(1), \quad \text{as} \quad n \to \infty,$$

so we can bound δ and γ by two positive constants, δ' and γ', such that $\delta \le \delta'$ and $\gamma \le \gamma'$ probability as $n \to \infty$ and

$$\frac{\alpha}{2} \le P\left\{\sup_{\mathbf{x}\in\mathcal{X}} \left\langle \mathbf{T}(\mathbf{x}), \frac{\varepsilon_w}{\sigma}\right\rangle \ge \frac{c}{\nu^{1/2}}\left(\frac{\nu^{1/2}\hat{\sigma}}{\sigma} - \gamma'\right) - b - \delta'\right\} + o(\alpha). \tag{A.2}$$

When $b = 0$, the variable $\nu^{1/2}\hat{\sigma}/\sigma$ has a χ distribution with degrees of freedom ν with probability density function

$$f(y, \nu) = \frac{y^{\nu-1}}{2^{\nu/2-1}\Gamma(\nu/2)} e^{-y^2/2}.$$

Because $\langle \mathbf{T}(\mathbf{x}), \varepsilon_w/\sigma\rangle$ is a Gaussian random field with mean zero and variance 1 and is independent of $\hat{\sigma}$, Sun's (1993) approximation formula gives

$$\begin{aligned} \frac{\alpha}{2} &\le \int_0^\infty P\left\{\sup_{\mathbf{x}\in\mathcal{X}} \left\langle \mathbf{T}(\mathbf{x}), \frac{\varepsilon_w}{\sigma}\right\rangle \ge \frac{c}{\nu^{1/2}}y - d\right\} f(y, \nu)\, dy + o(\alpha) \\ &\approx \int_0^\infty \kappa_0 \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(\frac{c}{\nu^{1/2}}y - d\right)^2\right\} f(y, \nu)\, dy, \end{aligned} \tag{A.3}$$

where $d = (c/\nu^{1/2})\gamma' + b + \delta'$. By using the Taylor expansion,

$$\begin{aligned} \exp\left\{-\frac{1}{2}(g(y) - \delta_0)^2\right\} &= \exp\left\{-\frac{1}{2}g(y)^2\right\} \\ &\quad + \exp\left\{-\frac{1}{2}g(y)^2\right\} g(y)\delta_0 + o(\delta_0), \quad \text{as} \quad \delta_0 \to 0 \end{aligned}$$

for $g(y) = cy/\sqrt{\nu}$, and the identities

$$\int_0^\infty \exp\left\{-\frac{c^2}{2\nu}y^2\right\} f(y, \nu)\, dy = \left(1 + \frac{c^2}{\nu}\right)^{-\nu/2}$$

and

$$\int_0^\infty \exp\left\{-\frac{c^2}{2\nu}y^2\right\} y f(y, \nu)\, dy = \frac{2^{1/2}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left(1 + \frac{c^2}{\nu}\right)^{(\nu+1)/2}},$$

we find, after some calculus from (A.3), that

$$\frac{\alpha}{2} \mathrel{\dot{\le}} \frac{\kappa_0}{2\pi}\left\{\left(1 + \frac{c^2}{\nu}\right)^{-\nu/2} + d \cdot \frac{c 2^{1/2}\Gamma\left(\frac{\nu+1}{2}\right)}{\nu^{1/2}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{c^2}{\nu}\right)^{-(\nu+1)/2}\right\}.$$

Now, if the image of $\mathbf{T}(\mathbf{x}) = l^w(\mathbf{x})/\|l^w(\mathbf{x})\|$ on $\mathcal{X}$ has a boundary, as it usually has when $\mathcal{X}$ is a finite interval, then this formula (A.3) can be corrected as in the case of known weights. So the final approximation formula is given by (2). Strictly speaking, the formula (2) gives a conservative confidence band (although narrower than Scheffé's), because we used upper bounds in the construction. Nevertheless, if the errors are symmetric, δ' and γ' are small, and n is large, then the upper bounds will be close to the values that they are bounding and thus the nominal level will be closely approximated.

[Received September 1993. Revised August 1994.]

REFERENCES

Carroll, R., and Ruppert, D. (1988), *Transformation and Weighting in Regression,* London: Chapman and Hall.

Müller, H-G. (1988), *Nonparametric Regression Analysis of Longitudinal Data,* New York: Springer-Verlag.

Naiman, D. Q. (1987), "Simultaneous Confidence Bands in Multiple Regression Using Predictor Variable Constraints," *Journal of the American Statistical Association,* 82, 214–219.

Scheffé, H. (1959), *The Analysis of Variance,* New York: John Wiley.

Seber, G. (1977), *Linear Regression Analysis,* New York: John Wiley.

Sun, J. (1993), "Tail Probabilities of the Maxima of Gaussian Random Fields," *Annals of Probability,* 21, 34–71.

Sun, J., and Loader, C. (1994), "Simultaneous Confidence Bands for Linear Regression and Smoothing," *The Annals of Statistics,* 22, 1328–1347.

Uusipaikka, E. (1983), "Exact Confidence Bands for Linear Regression Over Intervals," *Journal of the American Statistical Association,* 78, 638–644.

Wynn, H. P. (1984), "An Exact Confidence Band for One-Dimensional Polynomial Regression," *Biometrika,* 71, 375–380.

Wynn, H. P., and Bloomfield, P. (1971), "Simultaneous Confidence Bands in Regression Analysis," *Journal of the Royal Statistical Society,* Ser. B, 33, 202–217.