Confidence sets for a level set in linear regression

Fang Wan1 | Wei Liu2 | Frank Bretz3

1Department of Mathematics and Statistics, Lancaster University, Bilrigg Lane, Lancaster, LA1 4YF, UK
2School of Mathematical Sciences, University of Southampton, Southampton, UK
3Novartis Pharma AG, Basel, Switzerland

Correspondence
Fang Wan, Department of Mathematics and Statistics, Lancaster University, Lancaster, UK.
Email: f.wan@lancaster.ac.uk

Regression modeling is the workhorse of statistics and there is a vast literature on estimation of the regression function. It has been realized in recent years that in regression analysis the ultimate aim may be the estimation of a level set of the regression function, i.e., the set of covariate values for which the regression function exceeds a predefined level, instead of the estimation of the regression function itself. The published work on estimation of the level set has thus far focused mainly on nonparametric regression, especially on point estimation. In this article, the construction of confidence sets for the level set of linear regression is considered. In particular, $1 - \alpha$ level upper, lower and two-sided confidence sets are constructed for the normal-error linear regression. It is shown that these confidence sets can be easily constructed from the corresponding $1 - \alpha$ level simultaneous confidence bands. It is also pointed out that the construction method is readily applicable to other parametric regression models where the mean response depends on a linear predictor through a monotonic link function, which include generalized linear models, linear mixed models and generalized linear mixed models. Therefore, the method proposed in this article is widely applicable. Simulation studies with both linear and generalized linear models are conducted to assess the method and real examples are used to illustrate the method.

KEYWORDS
confidence sets, linear regression, nonparametric regression, parametric regression, simultaneous confidence bands, statistical inference

1 | INTRODUCTION

Decompression sickness (DCS) is an injury caused by rapid change of pressure, such as during or after water dives. Mild DCS involves symptoms such as muscle or joint pain, while serious DCS can cause paralysis or death. It is important to understand the relationship between risk factors, such as the exposure pressure (depth) and exposure duration, and mortality rate, that is, the chance of death due to serious DCS. Since adult sheep have a body mass similar to human they are considered to have a similar DCS susceptibility as humans. A sheep decompression trial is reported and studied in Li et al. In the article, logistic regression is used to model the mortality rate $p$ as a function of the two covariates exposure pressure ($x_1$) and exposure duration ($x_2$): $\log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. Of particular interest are the values of $x = (x_1, x_2)^T$ that correspond to a relatively low mortality rate, say 0.05, that is, the set

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. Statistics in Medicine published by John Wiley & Sons Ltd.

Statistics in Medicine. 2024;43:1103–1118. wileyonlinelibrary.com/journal/sim | 1103
where \( K \in \mathbb{R}^2 \) is a prespecified region of \( x \). How to make inference about this set (that depends on the unknown parameters \( \beta_i \)), especially the construction of confidence sets, is the well-known and well-studied effective-dose problem; See, for example, Li et al., Tompsett et al., and the references therein for an overview.

A set of potentially more interest is

\[
\{ x \in K : \beta_0 + \beta_1 x_1 + \beta_2 x_2 \leq \log(0.05/(1 - 0.05)) \},
\]

since one would probably be more interested in identifying all the combinations of \( x_1 \) and \( x_2 \) for which the mortality rate \( p \) is no more than the threshold 0.05. It becomes clear below that this is a level set. This motivates us to study the construction of confidence sets for a level set.

In general, let \( h \) be a regression function with covariate (vector) \( x \in \mathbb{R}^p \). A level set can be expressed as follows:

\[
G_{\lambda} = G(h) = \{ x \in K : h(x) \geq \lambda \},
\]

where \( \lambda \) is a prespecified number, and \( K \subset \mathbb{R}^p \) is a given covariate \( x \) region of interest. In regression analysis, there is a vast literature on how to estimate the regression function \( h \), based on the observed data \((Y_i, x_i), i = 1, \ldots, n \). In recent years, it has been realized that an important problem in regression is the inference of the level set \( G_{\lambda} \). For instance, Scott and Davenport assert that “In a wide range of regression problems, if it is worthwhile to estimate the regression function \( h \), then it is also worthwhile to estimate certain level sets. Moreover, these level sets may be of ultimate importance. And in many classification problems, labels are obtained by thresholding a continuous variable. Thus, estimating regression level sets may be a more appropriate framework for addressing many problems that are currently envisioned in other ways.”

Other than its application to the DCS problem alluded to above, one can envisage that, when considering a regression model of perinatal mortality rate on birth weight, it is interesting to identify the range of birth weight over which the perinatal mortality rate exceeds a certain \( \lambda \). Further possible applications have been pointed out, for example, in Scott and Davenport and Dau et al. Inference of the level set \( G_{\lambda} \) is an important component of the more general field of subgroup analysis.

In nonparametric regression where \( h \) is not assumed to have a specific form, point estimation of \( G_{\lambda} \) aims to construct \( \hat{G}_{\lambda} \) to approximate \( G_{\lambda} \) using the observed data. This has been considered by Cavalier, Polonik and Wang, Willett and Nowak, Scott and Davenport, Dau et al., and Reeve et al. among others. The main focus of these works is on large sample properties such as consistency and rate of convergence. Related work on estimation of level-sets of a nonparametric density function can be found in Hartigan, Tsybakov, Cadre, Mason and Polonik, Chen et al., and Qiao and Polonik. Confidence-set estimation of \( G_{\lambda} \) aims to construct sets \( \hat{G}_{\lambda} \) to contain or be contained in \( G_{\lambda} \) with a prespecified confidence level \( 1 - \alpha \). Large sample approximate \( 1 - \alpha \) confidence-set estimation of \( G_{\lambda} \) is considered in Mammen and Polonik.

In this article, confidence-set estimation of \( G_{\lambda} \) for both linear and generalized linear regressions is considered. It is shown that lower, upper and two-sided confidence-set estimators of \( G_{\lambda} \) can be easily constructed from the corresponding lower, upper and two-sided simultaneous confidence bands for a linear regression function over the covariate region of interest. Simultaneous confidence bands for linear regression have been considered in Wynn and Bloomfield, Naiman, Piegersch, Sun and Loader, Liu and Hayter, and numerous others; see Liu for an overview. The method also relates to general problems of simultaneous inference for parametric models. The method is then directly extended to the generalized linear regression models (including the logistic regression for the DCS problem), though the confidence-set estimations are of approximate \( 1 - \alpha \) level since the simultaneous confidence bands are of asymptotic \( 1 - \alpha \) level in this case. A related problem is the confidence-set estimation of the maximum (or minimum) point of a linear regression model; see Wan et al. and the references therein.

The layout of the article is as follows. The construction method of confidence-set estimators is given in Section 2. Simulation results for the coverage of the confidence sets in both linear and generalized linear models are given in Section 3. The method is illustrated with real examples including the DCS example in Section 4. Section 5 contains conclusions and a brief discussion. Finally, the appendix sketches the proofs of two theorems in Section 2 and provides the additional simulation results mentioned in Section 3.
2  |  METHOD

2.1  |  The construction of confidence sets for a level set in linear regression

Let a normal-error linear regression model be given by

\[ Y = h(x) + e = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + e, \]

where the independent errors \( e_i = Y_i - h(x_i) \) have distribution \( N(0, \sigma^2) \). From the observed sample of observations \((Y_1, x_1), \ldots, (Y_n, x_n)\), the usual estimator of \( \beta = (\beta_0, \cdots, \beta_p)^T \) is given by \( \hat{\beta} = (X^T X)^{-1}X^T Y \) where \( X \) is the \( n \times (p + 1) \) design matrix and \( Y = (Y_1, \ldots, Y_n)^T \). The estimator of the error variance \( \sigma^2 \) is given by \( \hat{\sigma}^2 \). It is known that \( \hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1}) \), \( \hat{\sigma}^2 \sim \sigma^2 X^T X / v \) with \( v = n - p - 1 \), and \( \hat{\beta} \) and \( \hat{\sigma}^2 \) are independent. In order for both estimators \( \hat{\beta} \) and \( \hat{\sigma}^2 \) to be available, the sample size \( n \) must be at least \( n \geq p + 2 \).

Let \( \tilde{x} = (1, x^T)^T = (1, x_1, \cdots, x_p)^T \) and \( m(x) = \sqrt{\tilde{x}^T (X^T X)^{-1} \tilde{x}} \). Define the \( 1 - \alpha \) confidence sets as follows:

\[
\begin{align*}
    \hat{G}_{1,1u} &= \{ x \in K : \tilde{x}^T \hat{\beta} + c_1^u \hat{\sigma} m(x) \geq \lambda \}, \\
    \hat{G}_{1,1l} &= \{ x \in K : \tilde{x}^T \hat{\beta} - c_1^l \hat{\sigma} m(x) \geq \lambda \}, \\
    \hat{G}_{2,u} &= \{ x \in K : \tilde{x}^T \hat{\beta} + c_2^u \hat{\sigma} m(x) \geq \lambda \}, \\
    \hat{G}_{2,l} &= \{ x \in K : \tilde{x}^T \hat{\beta} - c_2^l \hat{\sigma} m(x) \geq \lambda \},
\end{align*}
\]

where \( c_1^u > 0 \) and \( c_2^u > 0 \) are the critical constants to achieve the exact \( 1 - \alpha \) confidence level and can be solved from

\[
\begin{align*}
    P\{ \tilde{x}^T \hat{\beta} + c_1^u \hat{\sigma} m(x) \leq \lambda | x \in K \} &= 1 - \alpha, \\
    P\{ \tilde{x}^T \hat{\beta} - c_1^l \hat{\sigma} m(x) \leq \lambda | x \in K \} &= 1 - \alpha, \\
    P\{ \tilde{x}^T \hat{\beta} - c_2^l \hat{\sigma} m(x) \leq \lambda | x \in K \} &= 1 - \alpha.
\end{align*}
\]

Note that \( \tilde{x}^T \hat{\beta} + c_1^u \hat{\sigma} m(x), \tilde{x}^T \hat{\beta} - c_1^l \hat{\sigma} m(x) \) and \( \tilde{x}^T \hat{\beta} \pm c_2^u \hat{\sigma} m(x) \) with \( x \in K \) are the upper, lower and two-sided \( 1 - \alpha \) (hyperbolic) simultaneous confidence bands over the covariate region \( x \in K \). Whilst another popular form is \( m(x) = 1 \), corresponding to the constant-width confidence bands, the hyperbolic bands are often better than the constant-width band under various optimality criteria such as with smaller average band width and smaller minimum area (see, e.g., Liu and Hayter,\(^{25}\) and the references therein) and so used throughout this article. The critical constants \( c_1^u \) and \( c_2^u \) can be computed by using the method of Liu et al.\(^{30,31}\), the R code for computing the critical constants is provided as supplementary material.

It is worth emphasizing that the three probabilities in (4)–(6) do not depend on the unknown parameters \( \beta \in \mathbb{R}^{p+1} \) and \( \sigma > 0 \), and that \( c_1^u < c_2^u \).

The following theorem establishes that \( \hat{G}_{1,1u} \) is an upper, and \( \hat{G}_{1,1l} \) is a lower, confidence set for \( G_1 \) of exact \( 1 - \alpha \) level, whilst \( [\hat{G}_{2,l}, \hat{G}_{2,u}] \) is a two-sided confidence set for \( G_1 \) of at least \( 1 - \alpha \) level. A proof is sketched in the appendix.

\textbf{Theorem 1.} We have

\[
\begin{align*}
    \inf_{\beta \in \mathbb{R}^{p+1}, \sigma > 0} P\{ G_1 \subseteq \hat{G}_{1,1u} \} &= 1 - \alpha, \\
    \inf_{\beta \in \mathbb{R}^{p+1}, \sigma > 0} P\{ \hat{G}_{1,1l} \subseteq G_1 \} &= 1 - \alpha, \\
    \inf_{\beta \in \mathbb{R}^{p+1}, \sigma > 0} P\{ \hat{G}_{2,l} \subseteq G_1 \subseteq \hat{G}_{2,u} \} &\geq 1 - \alpha.
\end{align*}
\]

From the definitions in (1)–(3), it is clear that each set \( \hat{G}_* \) is given by all the points in \( K \) at which the corresponding simultaneous confidence band is at least as high as the given threshold \( \lambda \). Note that each set could be an empty set when \( \lambda \) is sufficiently large, and become \( K \) when \( \lambda \) is sufficiently small. Of course, each set cannot be larger than the given covariate set \( K \) from the definition. Since \( c_1^u > 0 \) and \( c_2^u > 0 \), it is clear that \( \hat{G}_{1,1l} \subseteq \hat{G}_{1,1u} \) and \( \hat{G}_{1,2l} \subseteq \hat{G}_{1,2u} \). Since \( c_1^u < c_2^u \), it is clear that \( \hat{G}_{1,1u} \subseteq \hat{G}_{1,2u} \) and \( \hat{G}_{1,2l} \subseteq \hat{G}_{1,1l} \). Hence \( \hat{G}_{1,2l} \subseteq \hat{G}_{1,2u} \subseteq \hat{G}_{1,1l} \subseteq \hat{G}_{1,1u} \subseteq \hat{G}_{1,2u} \).
Intuitively, since the regression function $\hat{x}^T \beta$ is bounded from above by the upper simultaneous confidence band $\hat{x}^T \beta + c_1^q \hat{\sigma}(x)$ over the region $x \in K$, the level set $G_1$ cannot be bigger than the set $\hat{G}_{1,1u}$. Similarly, since the regression function $\hat{x}^T \beta$ is bounded from below by the lower simultaneous confidence band $\hat{x}^T \beta - c_1^q \hat{\sigma}(x)$ over the region $x \in K$, the level set $G_1$ cannot be smaller than the set $\hat{G}_{1,1l}$. Finally, since the regression function $\hat{x}^T \beta$ is bounded, simultaneously, from below by the lower confidence band $\hat{x}^T \beta - c_1^q \hat{\sigma}(x)$, and from above by the upper confidence band $\hat{x}^T \beta + c_1^q \hat{\sigma}(x)$, over the region $x \in K$, the level set $G_1$ must contain the set $\hat{G}_{1,2l}$ and be contained in the set $\hat{G}_{1,2u}$ simultaneously.

Instead of the level set $G_1$, the set

$$M_2 = M_2(h) = \{ x \in K: h(x) \leq \lambda \}$$

may be of interest in some applications. In this case, one can consider the regression of $-Y$ on $x$, given by $-Y = -h(x) + (\gamma)$, and hence $M_2$ becomes the level set $G_2$ of the regression function $-h(x)$.

Now suppose that the value of $\lambda$ is not prespecified, that is, one might be interested in the confidence sets for $G_1$ for several different values of $\lambda$. Of course one can use the results above to construct a confidence set $\hat{G}_{1,1l}$, say, for each given value of $\lambda$. The question is “what is the joint confidence level of the confidence sets $\{ \hat{G}_{1,1l} \subseteq G_1 \}$, $\{ \hat{G}_{1,1l} \subseteq G_1 \}$, … for a sequence of $\lambda$-values $\lambda_1$, $\lambda_2$, …?” The theorem below asserts that the joint confidence level is at least $1 - \alpha$. Intuitively, since the regression function $\hat{x}^T \beta$ is bounded from above by the upper simultaneous confidence band $\hat{x}^T \beta + c_1^q \hat{\sigma}(x)$ over the region $x \in K$, the condition $G_1 \subseteq \hat{G}_{1,1u}$ should hold for all $\lambda$ simultaneously and hence $\bigcap_{\lambda} G_1 \subseteq \hat{G}_{1,1u}$ holds.

Similarly, since the regression function $\hat{x}^T \beta$ is bounded from below by the lower simultaneous confidence band $\hat{x}^T \beta - c_1^q \hat{\sigma}(x)$ over the region $x \in K$, the level set $G_1$ cannot be smaller than the set $\hat{G}_{1,1l}$ simultaneously for all $\lambda$ and hence $\bigcap_{\lambda} G_1 \supseteq \hat{G}_{1,1l}$ holds. Finally, since the regression function $\hat{x}^T \beta$ is bounded, simultaneously, from below by the lower confidence band $\hat{x}^T \beta - c_1^q \hat{\sigma}(x)$, and from above by the upper confidence band $\hat{x}^T \beta + c_1^q \hat{\sigma}(x)$, over the region $x \in K$, the condition $\hat{G}_{1,2l} \subseteq G_1 \subseteq \hat{G}_{1,2u}$ holds for all $\lambda$ simultaneously; hence $\bigcap_{\lambda} \hat{G}_{1,2l} \subseteq G_1 \subseteq \hat{G}_{1,2u}$ holds. The proof of Theorem 2 is also sketched in Appendix A.1.

**Theorem 2.** We have

$$\inf_{\beta \in \mathbb{R}^p, \sigma > 0} \left\{ P \left( \hat{G}_{1,1u} \subseteq G_1 \forall \lambda \in \mathbb{R}^1 \right) = 1 - \alpha \right. \}$$

$$\inf_{\beta \in \mathbb{R}^p, \sigma > 0} \left\{ P \left( \hat{G}_{1,1l} \subseteq G_1 \forall \lambda \in \mathbb{R}^1 \right) = 1 - \alpha \right. \}$$

$$\inf_{\beta \in \mathbb{R}^p, \sigma > 0} \left\{ P \left( \hat{G}_{1,2l} \subseteq G_1 \subseteq \hat{G}_{1,2u} \forall \lambda \in \mathbb{R}^1 \right) \geq 1 - \alpha \right. \}$$

2.2 Extension to generalized linear models

The confidence sets given in (1–3) for the normal-error linear regression can be generalized to other models that involve a linear predictor $\hat{x}^T \beta$. In generalized linear models, linear mixed models and generalized linear mixed models (cf. McCulloch and Searle32 and Faraway33), for example, the mean response $E(Y)$ is often related to a linear predictor $\hat{x}^T \beta$ by a given monotonic link function $L(\cdot)$, that is, $L(E(Y)) = \hat{x}^T \beta$. Since $L(\cdot)$ is monotone, the set of interest $\{ x \in K: E(Y) \geq L_0 \}$, for a given threshold $L_0$, becomes either $\{ x \in K: \hat{x}^T \beta \geq \lambda \}$ or $\{ x \in K: \hat{x}^T \beta \leq \lambda \}$, where $\lambda = L(L_0)$, depending on whether the function $L(\cdot)$ is increasing or decreasing. For example, in the logistic regression scenario $L(\cdot) = \logit(\cdot) = \log \left( \frac{\pi}{1-\pi} \right)$ is an increasing function and hence the set of interest $\{ x \in K: E(Y) \geq L_0 \}$ becomes $\{ x \in K: \hat{x}^T \beta \geq \lambda \}$ where $\lambda = \logit(L_0)$. Note that unlike in the linear regression, the critical constants $c_1^q$ and $c_1^l$ in GLMs depend on the unknown parameters $\beta$ and $\Sigma$ via the estimates $\hat{\beta}$ and $\hat{\Sigma}$. If the distribution of $\hat{\beta}$ is asymptotically normal $N(\beta, \Sigma)$, then the simultaneous confidence bands of the forms in (4–6) are of approximate $1 - \alpha$ level; see, e.g., Liu26 (Chapter 8). As a result, the corresponding confidence sets of the forms in (1–3) are of approximate $1 - \alpha$ level too. The accuracy of this approximation is explored via simulation in the next section.

3 Simulation Results

In this section, we conducted numerical studies to assess the coverage of the upper and lower one-sided confidence sets for the level sets in both linear regression and generalized linear regressions (specifically logistic regression, as in the
motivating example). The simulation studies included is to assess the validity of the (asymptotic) construction methods for a nominal level set, that is, the minimum coverage of the confidence set over the parameter space is $1 - \alpha$. It is proved in Appendix A.1 that the coverage of our one-sided upper confidence sets for the $\lambda$-level set is at least $1 - \alpha$ attained at $\beta = (\lambda, 0, \ldots, 0)$; and the coverage of our one-sided lower confidence set is at least $1 - \alpha$ attained at $\beta = (\lambda^-, 0, \ldots, 0)$, where $\lambda^-$ denote a number that is infinitesimally smaller than $\lambda$. Hence in Tables 1, 2, A1–A4, the one-sided upper/lower confidence sets are assessed at $\beta = (\lambda, 0, \ldots, 0)$ and $\beta = (\lambda^-, 0, \ldots, 0)$. In Table A5, we consider a series of scenarios where the true parameter values are away from $(\lambda, 0, 0)$ and show that the coverage of the 1-sided upper/lower confidence sets are indeed conservative as expected, although the precise value of $\beta$ does not appear to have a large impact on the coverage once it is not $(\lambda, 0, 0)$ or $(\lambda^-, 0, 0)$. In reality, the true parameter is unknown and hence the coverage of our confidence set is guaranteed to be (asymptotically) $1 - \alpha$.

For the linear scenario, we generated data from a Normal distribution with mean $x\beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_5 x_5$ and variance $\sigma^2 = 1$. Three design matrices were used in the data generation, with 3, 5 or 11 data points evenly chosen from $-1$ to $1$ for each covariate (ie increment by $q = 1, 0.5,$ or $0.2$). In Appendix A.1, it is suggested that the coverage of the upper 1-sided confidence sets should be exact at the parameter $\beta = (\lambda, 0, 0, 0, 0)$ and hence the upper confidence set is assessed at $\beta = (\lambda, 0, 0, 0, 0)$ where $\lambda \in \{-0.9, -0.5, 0.5, 0.9\}$. Note that when $\beta = (\lambda, 0, 0, 0, 0)$ the corresponding level set is the covariate space $K$, and hence we check whether the one-sided upper confidence set in each of the 10,000 simulation covers $K$ (ie equal to $K$). Similarly, the lower one-sided 95% confidence sets are assessed each with 10,000 simulation at $\beta = (\lambda^-, 0, 0, 0, 0)$ as suggested, where $\lambda^-$ is a number that is infinitesimally smaller than $\lambda$.

### Table 1

Coverage of 95% confidence sets for the $\lambda$-level set of the linear model $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_5 x_5 + \epsilon$ with $\epsilon \sim N(0, 1)$ and $x_i \in [-1, 1]$, $i = 1, \ldots, 5$.

| Coverage of 95% confidence sets | $\lambda$ | $q$ | $-0.9$ | $-0.5$ | $0$ | $0.5$ | $0.9$ |
|--------------------------------|----------|----|-------|-------|----|------|------|
| One-sided upper               |          |    |       |       |    |      |      |
| Confidence sets               |          |    |       |       |    |      |      |
| 1                             |          |    | 0.9505| 0.9482| 0.9508| 0.9490| 0.9459|
| 0.5                           |          |    | 0.9516| 0.9472| 0.9504| 0.9527| 0.9518|
| 0.2                           |          |    | 0.9492| 0.9518| 0.9508| 0.9507| 0.9471|
| One-sided lower               |          |    |       |       |    |      |      |
| Confidence sets               |          |    |       |       |    |      |      |
| 1                             |          |    | 0.9455| 0.9508| 0.9453| 0.9510| 0.9496|
| 0.5                           |          |    | 0.9493| 0.9501| 0.9499| 0.9509| 0.9506|
| 0.2                           |          |    | 0.9534| 0.9532| 0.9529| 0.9497| 0.9469|

Note: The design points are evenly spread in the square space $[-1, 1]^6$ by $q$. Hence the samples sizes used in the simulation are $n = 3^4$ for $q = 1, n = 5^4$ for $q = 0.5$ and $n = 11^4$ for $q = 0.2$. The coverage of the upper and lower one-sided confidence sets are calculated with 10,000 simulations.

### Table 2

Coverage of 95% confidence sets for the $\lambda$-level set of the logistic model $Y = \text{Bern}(p), \logit(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

| Coverage of 95% confidence sets | $L_0$ | $q$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|--------------------------------|-------|----|-----|-----|-----|-----|-----|
| One-sided upper               |       |    |     |     |     |     |     |
| Confidence sets               |       |    |     |     |     |     |     |
| 0.5                           |       |    | 1   | 1   | 0.9979| 0.9780| 0.9516|
| 0.1                           |       |    | 0.9687| 0.9543| 0.9519| 0.9507| 0.9347|
| 0.01                          |       |    | 0.9547| 0.9526| 0.9512| 0.9469| 0.9506|
| One-sided lower               |       |    |     |     |     |     |     |
| Confidence sets               |       |    |     |     |     |     |     |
| 0.5                           |       |    | 0.9551| 0.9795| 0.9989| 1   | 1 |
| 0.1                           |       |    | 0.9325| 0.9494| 0.9462| 0.9602| 0.9677|
| 0.01                          |       |    | 0.9492| 0.9472| 0.9476| 0.9517| 0.9535|

Note: The design points are evenly spread in the square space $[-1, 1]^2$ by $q$. Hence the sample size is $n = 3^2$ for $q = 0.5, n = 21^2$ for $q = 0.1$ and $n = 201^2$ for $q = 0.01$. The coverage of the upper and lower one-sided confidence sets are calculated with 10,000 simulations.
When \( \beta = (\lambda, 0, 0, 0, 0) \), the level set is the covariate space \( K \) and hence we again check whether each confidence set covers \( K \). The results are given in Table 1. It is shown that for all three designs, the coverage of the confidence sets are reasonably close to the nominal 95\%. We didn’t consider a sample size smaller than \( 3^5 = 243 \) here because parameter estimation based on fewer than 3 data points per covariate is not ideal, and the performance of our confidence set is almost entirely based on the accuracy of the parameter estimation. However, we have illustrated the performance of the confidence sets with smaller sample sizes and fewer covariates, for example, in Table A1 we use \( n = 5^2 = 25 \) (\( q = 0.5 \)) and \( n = 11^2 = 121 \) (\( q = 0.2 \)).

For the logistic regression, we generated binary data from \( \text{Bernoulli}(p) \) where \( p \) is given by \( p = \logit^{-1}(\hat{x}\beta) \) and \( \hat{x}\beta = \beta_0 + \beta_1x_1 + \beta_2x_2 \). The upper and lower 1-sided 95\% confidence sets for each level set with level \( L_0 = \{0.1, 0.3, 0.5, 0.7, 0.9\} \) are constructed, and the coverage is assessed at \( \beta = (\logit(\lambda), 0, 0) \) and \( \beta = (\logit(\lambda), 0, 0) \) with \( \lambda = \logit(L_0) \) respectively. The simulation results are given in Table 2. It is shown that the coverage depends heavily on the convergence of the maximum likelihood estimate to multivariate normality, that is, it is closer to the nominal 95\% when the sample size increases (as expected). If the sample sizes of successes and failures are both large, then the upper and lower one-sided confidence sets for the level set should be approximately 95\%. Note that logistic regression usually requires a large sample size especially when the outcome number of success/failure is very small (eg \( L_0 = 0.1 \)); hence the performance of the confidence set is worse at the two ends and better when \( L_0 = 0.5 \). Similar results can be found in Table A2 in Appendix A.2 for logistic regression with five covariates. However, as we pointed out in Section 2, the two-sided confidence sets are not exact but conservative in the sense that the coverage is at least 95\%; this is confirmed by the simulation results given in Table A3 in Appendix A.2.

In Tables 1 and 2, the design points are evenly chosen from the covariate space. However, this is rarely the case in practice. As the coverage depends on the accuracy of parameter estimate, it is understandable that the coverage will be worse when the design data points are not evenly chosen. In order to demonstrate the coverage of our confidence set, we conducted further simulations where 25 data points are randomly chosen from \([-1, 1]^2\) for two covariate linear and logistic regression models. The results are given in Table A4 in Appendix A.2.

4 | EXAMPLES

In this section, two examples are used to illustrate the confidence sets given in (1)–(3) in linear regression and logistic regression settings. From Section 2, simultaneous confidence bands in (4)–(6) need to be constructed first in order to construct the confidence sets for \( G \) in (1)–(3). The two most popular forms of \( m(x) \) are \( m(x) = \sqrt{x^T(X^TX)^{-1}x} \) and \( m(x) = 1 \), corresponding to the hyperbolic and constant-width confidence bands, respectively. As the hyperbolic band is often better than the constant-width band in the sense that it has smaller average band width and smaller minimum area where the terms average width and minimum area are defined in Liu and Hayter,\(^{25}\) hyperbolic confidence bands are used in both examples in this section. The R code for all the computation in this section is available from the authors.

4.1 | Wisconsin–Madison sheep dive trial

In the Wisconsin–Madison sheep dive trial, 1108 dives were performed and recorded. Following Li et al,\(^1\) logistic regression is used to model the relationship between the mortality rate \( p \) and the two covariates \( x_1 \), the log base 10 exposure depth, and \( x_2 \), the log base 10 exposure duration: \( \logit(p) = \beta_0 + \beta_1x_1 + \beta_2x_2 \); here \( \logit(p) = \log(p/(1-p)) \) which is monotone increasing in \( p \in (0, 1) \). Based on the recorded data, the MLE \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_1)^T \) is calculated to be \((-19.253, 14.196, 3.758)^T\) and the approximate covariance matrix of \( \hat{\beta} \) is

\[
\hat{I}^{-1} = \begin{pmatrix}
5.0004779 & -5.5146133 & -0.8322975 \\
-5.5146133 & 7.1346280 & 0.7606109 \\
-0.8322975 & 0.7606109 & 0.1648114
\end{pmatrix}.
\]

Hence, \( \hat{\beta} \) has approximate normal distribution \( N_3(\beta, \hat{I}^{-1}) \).

A major goal of this study as described in Li et al\(^1\) was to determine the ranges of dive depth and duration that correspond to relatively low mortality rates. This knowledge will be invaluable for safety in human dives. Set \( p = .05 \) as a
threshold of low mortality rate. From the recorded 1108 dives, the minimum value of \( x_1 \) is \( \min(x_1) = 0.314 \) the maximum value of \( x_1 \) is \( \max(x_1) = 0.714 \), \( \min(x_2) = 1.301 \) and \( \max(x_2) = 3.158 \). Hence, it is important to identify the set

\[
\{ x \in K : \beta_0 + \beta_1 x_1 + \beta_2 x_2 \leq \logit(0.05) \} = \{ x \in K : -x^T \beta \geq -\logit(0.05) \},
\]

where \( K = \{ x = (x_1, x_2)^T : 0.314 \leq x_1 \leq 0.714, 1.301 \leq x_2 \leq 3.158 \} \). The set above is just the \( \lambda \)-level set of the regression function \(-x^T \beta\) with \( \lambda = -\logit(0.05) \). So the method of Section 2 can be used to construct the confidence sets in (1-3) for this \( G_1 \).

From Section 2, simultaneous confidence bands for \(-x^T \beta\) over \( x \in K \) need to be constructed first in order to construct the confidence sets. Note, however, only approximate \( 1 - \alpha \) confidence bands of the forms in (4-6), with \( \gamma = 1, \nu = \infty \) and \((X^TX)^{-1}\) replaced with \( \hat{I}^{-1} \), can be constructed by using the approximate normal distribution \( N(\hat{\beta}, \hat{I}^{-1}) \) of \( \hat{\beta} \). Hence the confidence sets for \( G_1 \) are also of approximate \( 1 - \alpha \) level. For \( 1 - \alpha = 95\% \) and \( K \) given above, the critical values \( c_{1,0.05}^1 \) and \( c_{2,0.05}^1 \) are computed to be 2.483 and 2.728, respectively, by using the method of Liu et al \( ^{30} \) (see also Liu \( ^{26} \) Section 3.2 or use the R code provided). Note that only the parameter estimates and the corresponding variance matrix rather than the full data set is needed in computing our confidence sets.

Figure 1a plots the 1-sided upper simultaneous confidence band for \(-x^T \beta\) and the horizontal plane at height \( \lambda = -\logit(0.05) \) over the rectangular region \( K \).

Figure 1b plots the 1-sided upper confidence set \( \hat{G}_{2,1u} \), with the region \( K \) given by the rectangle in solid line and \(-x^T \beta = -\logit(0.05) \) given by the dashed line. Note that the curvilinear-boundary of \( \hat{G}_{2,1u} \) is given by the projection, to the \( x \)-plane, of the intersection between the horizontal plane at height \(-\logit(0.05)\) and the 1-sided upper simultaneous confidence band over the region \( x \in K \) in Figure 1(a). The upper confidence set \( \hat{G}_{2,1u} \) tells us that, with 95% confidence level, within \( K \) only those dives with \( x \) in \( \hat{G}_{2,1u} \) may have mortality rate smaller than or equal to 0.05. Hence a dive with \( x \in K \setminus \hat{G}_{2,1u} \) should be considered too dangerous in terms of mortality rate.

Similarly, Figure 1c plots the 1-sided lower confidence set \( \hat{G}_{2,1l} \) in the \( x \)-plane. Note that the curvilinear-boundary of \( \hat{G}_{2,1l} \) is given by the projection, to the \( x \)-plane, of the intersection between the horizontal plane at height \(-\logit(0.05)\) and the 1-sided lower simultaneous confidence band for \(-x^T \beta\) over the region \( K \). The lower confidence set \( \hat{G}_{2,1l} \) tells us that, with 95% confidence level, dives with \( x \in \hat{G}_{2,1l} \) have mortality rate smaller than or equal to 0.05. Hence these dives may be considered “safe.”

Figure 1d plots the two-sided confidence set \( \hat{G}_{2,2l} \) in the \( x \)-plane. Note that the curvilinear-boundaries of \( \hat{G}_{2,2l} = \hat{G}_{2,1l} \cup \hat{G}_{2,1u} \) are given by the projection, to the \( x \)-plane, of the intersection between the horizontal plane at height \(-\logit(0.05)\) and the two-sided simultaneous confidence band for \(-x^T \beta\) over the region \( K \). The two-sided confidence set \( \hat{G}_{2,2l} \) tells us that, with 95% confidence level, dives with \( x \in \hat{G}_{2,2l} \) are considered as dangerous, dives with \( x \in \hat{G}_{2,2l} \) are considered as safe, and dives with \( x \in \hat{G}_{2,2u} \) are possibly dangerous, in terms of mortality.

If one feels mortality rate 0.05 is too high, one may want to try 0.01, for example, and construct the corresponding confidence set \( \hat{G}_{2,1l} \). Indeed one can construct confidence sets \( \hat{G}_{2,1l}, \hat{G}_{2,1l}, \ldots \) for any sequence of \( \lambda_1, \lambda_2, \ldots \). Theorem 2 guarantees that the simultaneous confidence level of this sequence of lower confidence sets is still approximately \( 1 - \alpha = 95\% \). R codes used to produce the plots are provided in the supplementary materials.

### 4.2 | Infants’ blood pressure example

In Example 1.1 of Liu, \( ^{26} \) a linear regression model of systolic blood pressure (\( Y \)) on the two covariates birth weight in oz (\( x_1 \)) and age in days (\( x_2 \)) of an infant is considered

\[
Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.
\]

The measurements on \( n = 17 \) infants in Liu \( ^{26} \) (table 1.1) are provided in Appendix A.3. Based on these measurements, the linear regression model provides a good fit with \( R^2 = 95\% \). The observed values of \( x_1 \) range from 92 to 149, and the observed values of \( x_2 \) range from 2 to 5; hence, we set \( K = \{ x = (x_1, x_2)^T : 92 \leq x_1 \leq 149.2, 2 \leq x_2 \leq 5 \} \). It is of interest to identify infants, in terms of \( x = (x_1, x_2)^T \in K \), that have mean systolic blood pressure larger than 99, assuming systolic blood pressure larger than 99 is deemed to be too high. Therefore, the level set \( G = G(99) = \{ x \in K : \beta_0 + \beta_1 x_1 + \beta_2 x_2 \geq 99 \} \) is of interest. This example is just to illustrate how the proposed method can be applied to linear regression models. Note
that in practice, small sample size may result in inaccurate estimates of parameters and hence the resultant estimated level set and its confidence set.

In this example, with \( p = 2, n = 17, 1 - \alpha = 95\% \) and the given design matrix \( X \), the fitted model is \( \hat{y} = 47.58 + 0.18x_1 + 5.31x_2 \); \( c_1^2 \) is computed to be 3.11 and \( c_2^2 \) is computed to be 2.77.

Figure 2a plots the 1-sided upper confidence set \( \hat{G}_{1u} \) in the \( x \)-plane, with the region \( K \) given by the rectangle in solid line. Note that the curvilinear-boundary of \( \hat{G}_{1u} \) is given by the projection, to the \( x \)-plane, of the intersection between the horizontal plane at height \( \lambda = 99 \) and the one-sided upper simultaneous confidence band over the region \( x \in K \). The upper confidence set \( \hat{G}_{1u} \) tells us that, with 95% confidence level, only those infants having \( x \in \hat{G}_{1u} \) may have mean systolic blood pressure larger than or equal to 99. Hence, \( x \in \hat{G}_{1u} \) could be used as a screening criterion for further medical check due to concerns over too high systolic blood pressure.
Similarly, Figure 2b plots the one-sided lower confidence set \( \hat{G}_{1l} \) in the \( x \)-plane. Note that the curvilinear-boundary of \( \hat{G}_{1l} \) is given by the projection, to the \( x \)-plane, of the intersection between the horizontal plane at height \( \lambda = 99 \) and the 1-sided lower simultaneous confidence band over the region \( K \). The lower confidence set \( \hat{G}_{1l} \) tells us that, with 95% confidence level, infants having \( x \in \hat{G}_{1u} \) do have mean systolic blood pressure larger than or equal to 99. Hence these infants should have further medical check due to concerns over excessive high systolic blood pressure.

Figure 2c plots the two-sided confidence set \([\hat{G}_{2l}, \hat{G}_{2u}]\) in the \( x \)-plane. Note that the curvilinear-boundaries of \([\hat{G}_{2l}, \hat{G}_{2u}]\) are given by the projection, to the \( x \)-plane, of the intersection between the horizontal plane at height \( \lambda = 99 \) and the two-sided confidence band over the region \( K \). The two-sided confidence set tells us that, with 95% confidence level, infants having \( x \in K \setminus \hat{G}_{2u} \) are not of concern, infants having \( x \in \hat{G}_{2l} \) are of concern, and infants having \( x \in \hat{G}_{2u} \) are possibly of concern, in terms of excessive high mean systolic blood pressure.

Figure 2d plots \( \hat{G}_{1u}, \hat{G}_{1l} \) and \([\hat{G}_{2l}, \hat{G}_{2u}]\) in the same picture for the purpose of comparison. It is clear from the figure that \( \hat{G}_{2l} \subseteq \hat{G}_{1l} \subseteq \hat{G}_{1u} \subseteq \hat{G}_{2u} \) as pointed out in Section 2.
Note that when $\lambda$ is large, 100 say, the horizontal plane at height $\lambda$ and the one-sided lower simultaneous confidence band do not intersect over the region $K$. In this case, the one-sided lower confidence set $\hat{G}_1$ is an empty set. Similar observations hold for other confidence sets.

5 | CONCLUSION AND DISCUSSION

In this article, the construction of confidence sets for the level set of linear regression is considered. Upper, lower and two-sided confidence sets of level $1 - \alpha$ are constructed for the normal-error linear regression. It is shown that these confidence sets are constructed from the corresponding $1 - \alpha$ level simultaneous confidence bands. Hence, these confidence sets and simultaneous confidence bands are closely related.

It is noteworthy that the sample size $n$ only needs to satisfy $n = p - 1 \geq 1$, i.e. $n \geq p + 2$, so that the regression coefficients $\beta$ and the error variance $\sigma^2$ can be estimated. So long as $n \geq p + 2$, the theorem in Section 2 holds. A larger sample size $n$ will make the confidence sets closer to the level set, which is similar to the usual confidence sets for the mean of a normally distributed population. Hence, the method for linear regression provided in this article is much simpler than that for nonparametric regression and density level sets (cf. Mammen and Polonik, 18 Chen et al, 16 Qiao and Polonik 17).

In Theorem 1 in Section 2, the minimum coverage probability over the whole parameter space $\beta \in \mathbb{R}^{p+1}$ and $\sigma > 0$ is sought since no assumption is made about any prior information on $\beta$ or $\sigma > 0$. If it is known a priori that $\beta$ and $\sigma$ are in a restricted space, then the usual estimators $\hat{\beta}$ and $\hat{\sigma}$ should be replaced by the maximum likelihood estimators over the restricted space, and the minimum coverage probability should also be over this restricted space. This situation becomes more complicated and is beyond the scope of this article.

The construction method is then extended to other parametric regression models where the mean response depends on a linear predictor through a monotonic link function. Examples are generalized linear models, linear mixed models and generalized linear mixed models. The illustrative example in Section 4.1 involves a generalized linear model. Therefore the method proposed in this article is widely applicable. Note that among the nonlinear regression models mentioned only logistic regression is considered as in the motivating example and explored in this example. However, it would be reasonable to expect that Poisson regression would have similar, or better performance since convergence of the parameter estimates to multivariate normality should be faster than for logistic regression.

So far, we only consider linear regression or generalized linear models without interaction terms between the covariates. For (generalized) linear models with interaction terms, the construction methods of our confidence sets can be applied readily in theory; however, the computation of the critical constants $c_1^L$ and $c_2^L$ of the simultaneous confidence bands will be more challenging since we can no longer use linearly constrained quadratic programming. We will investigate this case in our future work. It is clear that confidence sets for level sets are based on the construction of simultaneous confidence bands. There seems to be no research on the impact of model misspecification on the construction of simultaneous confidence sets. This warrants further research.

We are unable to establish thus far whether the two-sided confidence set $[\hat{G}_{1,2l}, \hat{G}_{1,2u}]$ is of confidence level $1 - \alpha$ exactly. Construction of a two-sided confidence set of exact confidence level $1 - \alpha$ is clearly of interest and warrants further research. We are actively researching on this.

There is a practical question “How can we best communicate the confidence sets to clinical researchers?” While it is said “a picture is worth a thousand words,” how to visualize high ($n > 3$)-dimensional data have always be an issue in reality. The common approach seems to use multiple plots, each varying two covariates while fixing the other $n - 2$ covariates. On the other hand, the most important clinical consideration might be whether a specific patient’s covariate values lie within a confidence set, which could be achieved via an interactive app such as those produced by Shiny in R. We welcome suggestions from the statistics community on this issue.

ACKNOWLEDGMENT

We thank Professor Jialiang Li of National University of Singapore for his help in preparing this article. We also thank the anonymous reviewers and the associate editor for their constructive comments which have greatly improved this article.

DATA AVAILABILITY STATEMENT

The full data set used in the Infant Blood Pressure example is available in Appendix A.3; the parameter estimates and the corresponding variance matrix (which are the only information needed in constructing the confidence sets) in the Sheep DCS example is available within the R code provided as supplementary material.
WAN et al.

ORCID

Fang Wan  https://orcid.org/0000-0001-5107-9757
Frank Bretz  https://orcid.org/0000-0002-2008-8340

REFERENCES

1. Li J, Nordheim E, Zhang C, Lehner C. Estimation and confidence regions for multi-dimensional effective dose. Biomet J. 2008;50(1):110-122.
2. Tompsett D, Biedermann S, Liu W. Simultaneous confidence sets for several effective doses. Biomet J. 2018;60:703-720.
3. Scott C, Davenport M. Regression level set estimation via cost-sensitive classification. IEEE Trans Signal Process. 2007;55:2752-2757.
4. Dau H, Laloiś T, Servien R. Exact asymptotic limit for kernel estimation of regression level sets. Stat Probab Lett. 2020;161:108721.
5. Wang R, Lagakos S, Ware J, Hunter D, Drazen J. Statistics in medicine 97 reporting of subgroup analyses in clinical trials. New Engl J Med. 2007;357(21):2189-2194.
6. Herrera F, Carmona CJ, González P, Del Jesus MJ. An overview on subgroup discovery: foundations and applications. Knowl Inform Syst. 2011;29(3):495-525.
7. Ting N, Cappelleri J, Ho S, Chen DG. Design and Analysis of Subgroups with Biopharmaceutical Applications. Cham: Springer Nature Switzerland AG; 2020.
8. Cavalier L. Nonparametric estimation of regression level sets. Statistics. 1997;29(2):131-160.
9. Polonik W, Wang Z. Estimation of regression contour clusters: an application of the excess mass approach to regression. J Multivariate Anal. 2005;94:227-249.
10. Willett R, Nowak R. Minimax optimal level-set estimation. IEEE Trans. Image Process. 2007;16(12):2965-2979.
11. Reeve H, Cannings T, Samworth R. Optimal subgroup selection. New York: Cornell University; 2021. https://arxiv.org/abs/2109.01077, Accessed January 1, 2023.
12. Hartigan J. Estimation of a convex density contour in two dimensions. J Amer Statist Assoc. 1987;82(397):267-270.
13. Tsybakov A. On nonparametric estimation of density level sets. Ann Statist. 1997;25(3):948-969.
14. Cadre B. Kernel estimation of density level sets. J Multivariate Anal. 2006;97(4):999-1023.
15. Mason D, Polonik W. Asymptotic normality of plug-in level set estimates. Ann Appl Probab. 2009;19(3):1108-1142.
16. Chen YC, Genovese C, Wasserman L. Density level sets: asymptotics, inference, and visualization. J Amer Statist Assoc. 2017;112(520):1684-1696.
17. Qiao W, Polonik W. Nonparametric confidence regions for level sets: statistical properties and geometry. Electron J Stat. 2019;13:985-1030.
18. Mammen E, Polonik W. Confidence regions for level sets. J Multivar Anal. 2013;122:202-214.
19. Wynn H, Bloomfield P. Simultaneous confidence bands in regression analysis. JRSS (B). 1971;33:202-217.
20. Naiman D. Optimal simultaneous confidence bounds. Ann Statist. 1984;12:702-715.
21. Naiman D. Conservative confidence bands in curve estimation: linear regression. Ann Statist. 1986;14:896-906.
22. Piegorsch W. Admissible and optimal confidence bands in simple linear regression. Ann Statist. 1985a;13:801-810.
23. Piegorsch W. Average-width optimality for confidence bands in simple linear regression. J Amer Statist Assoc. 1985b;80:692-697.
24. Sun J, Loader C. Simultaneous confidence bands for linear regression and smoothing. Ann Statist. 1994;22:1328-1346.
25. Liu W, Hayter A. Minimum area confidence set optimality for confidence bands in simple linear regression. J Amer Statist Assoc. 2007;102:181-190.
26. Liu W. Simultaneous Inference in Regression. Boca Raton, FL: CRC Press; 2010.
27. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biomet J. 2008;50(3):346-363.
28. Wan F, Liu W, Han Y, Bretz F. An exact confidence set for a maximum point of a univariate polynomial function in a given interval. Technometrics. 2015;57(4):559-565.
29. Wan F, Liu W, Bretz F, Han Y. Confidence sets for optimal factor levels of a response surface. Biometrics. 2016;72(4):1285-1293.
30. Liu W, Jamshidian M, Zhang Y, Donnelly J. Simulation-based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals. J Comput Graph Statist. 2005;14:459-484.
31. Liu W, Wynn H, Hayter A. Statistical inferences for linear regression models when the covariates have functional relationships: polynomial regression. J Statist Comput Simul. 2008;78:315-324.
32. McCulloch C, Searle S. Generalized, Linear, and Mixed Models. New York: Wiley; 2001.
33. Faraway J. Extending the Linear Model with R. 2nd ed. Boca Raton, FL: CRC Press; 2016.

How to cite this article: Wan F, Liu W, Bretz F. Confidence sets for a level set in linear regression. Statistics in Medicine. 2024;43(6):1103-1118. doi: 10.1002/sim.9996
APPENDIX

A.1 Proofs of the two theorems

In this appendix proofs of the two theorems in Section 2 are sketched.

Firstly, consider Theorem 1. For proving the statement in (7), we have

\[
\{ G_1 \subseteq \hat{G}_{1,1u} \} = \{ \forall x \in G_1 : x \in \hat{G}_{1,1u} \}
\]

\[
= \{ \forall x \in G_1 : \tilde{x}^T \hat{\beta} + c_i \hat{\delta} m(x) \geq \lambda \}
\]

\[
= \{ \forall x \in G_1 : \tilde{x}^T (\hat{\beta} - \beta) + c_i \hat{\delta} m(x) \geq \lambda - \tilde{x}^T \beta \}
\]

\[
\supseteq \{ \forall x \in G_1 : \tilde{x}^T (\hat{\beta} - \beta) + c_i \hat{\delta} m(x) \geq 0 \}
\]

where the second equation follows directly from the definition of \( \hat{G}_{1,1u} \), the first \( \supseteq \) follows directly from the definition of \( G_1 \), and the second \( \supseteq \) follows directly from the fact that \( G_1 \subseteq K \). It follows therefore

\[
P \{ G_1 \subseteq \hat{G}_{1,1u} \} \geq P \{ \forall x \in K : \tilde{x}^T (\hat{\beta} - \beta) + c_i \hat{\delta} m(x) \geq 0 \} = 1 - \alpha. \tag{A1}
\]

where the last equality is directly due to the fact that \( \tilde{x}^T \hat{\beta} + c_i \hat{\delta} m(x) \) is an upper simultaneous confidence band for \( \tilde{x}^T \beta \) over \( x \in K \) of exact \( 1 - \alpha \) level, as given in (1).

Next we show that the minimum probability over \( \beta \in \mathbb{R}^{p+1} \) and \( \sigma > 0 \) in statement (7) is \( 1 - \alpha \), attained at \( \beta = (\lambda, 0, \ldots, 0)^T \). At \( \beta = (\lambda, 0, \ldots, 0)^T \), we have \( G_1 = K \) and \( \lambda = \tilde{x}^T \beta \), and so

\[
\{ G_1 \subseteq \hat{G}_{1,1u} \} = \{ \forall x \in K : x \in \hat{G}_{1,1u} \}
\]

\[
= \{ \forall x \in K : \tilde{x}^T \hat{\beta} + c_i \hat{\delta} m(x) \geq \lambda \}
\]

\[
= \{ \forall x \in K : \tilde{x}^T (\hat{\beta} - \beta) + c_i \hat{\delta} m(x) \geq \lambda - \tilde{x}^T \beta \}
\]

\[
= \{ \forall x \in K : \tilde{x}^T (\hat{\beta} - \beta) + c_i \hat{\delta} m(x) \geq 0 \}
\]

which gives

\[
P \{ G_1 \subseteq \hat{G}_{1,1u} \} = P \{ \forall x \in K : \tilde{x}^T (\hat{\beta} - \beta) + c_i \hat{\delta} m(x) \geq 0 \} = 1 - \alpha. \tag{A2}
\]

The combination of (A1) and (A2) proves the statement in (7).

Now we prove the statement in (8). For a given set \( A \subseteq K \), let \( A^c \) denote the complement set within \( K \), i.e. \( A^c = K \setminus A \).

We have

\[
\{ \hat{G}_{1,1l} \subseteq G_1 \} = \{ G_1^c \subseteq \hat{G}_{1,1l}^c \}
\]

\[
= \{ \forall x \in G_1^c : x \in \hat{G}_{1,1l}^c \}
\]

\[
= \{ \forall x \in G_1^c : \tilde{x}^T \hat{\beta} - c_i \hat{\delta} m(x) < \lambda \}
\]

\[
= \{ \forall x \in G_1^c : \tilde{x}^T (\hat{\beta} - \beta) - c_i \hat{\delta} m(x) < \lambda - \tilde{x}^T \beta \}
\]

\[
\supseteq \{ \forall x \in G_1^c : \tilde{x}^T (\hat{\beta} - \beta) - c_i \hat{\delta} m(x) \leq 0 \}
\]

\[
\supseteq \{ \forall x \in K : \tilde{x}^T (\hat{\beta} - \beta) - c_i \hat{\delta} m(x) \leq 0 \},
\]

where the third equation follows directly from the definition of \( \hat{G}_{1,1l} \) (or \( \hat{G}_{1,1l}^c \)), the first \( \supseteq \) follows directly from the definition of \( G_1 \) (or \( G_1^c \)), and the second \( \supseteq \) follows directly from the fact that \( G_1^c \subseteq K \). It follows therefore

\[
P \{ \hat{G}_{1,1l} \subseteq G_1 \} \geq P \{ \forall x \in K : \tilde{x}^T (\hat{\beta} - \beta) - c_i \hat{\delta} m(x) \leq 0 \} = 1 - \alpha. \tag{A3}
\]
where the last equality is directly due to the fact that $\hat{x}^T \hat{\beta} - c^*_T \hat{\sigma}(x)$ is a lower simultaneous confidence band for $\hat{x}^T \beta$ over $x \in K$ of exact $1 - \alpha$ level, as given in (2).

Next we show that the minimum probability over $\beta \in \mathbb{R}^{p+1}$ and $\sigma > 0$ in statement (8) is $1 - \alpha$, attained at $\beta = (\lambda^-, 0, \ldots, 0)^T$, where $\lambda^-$ denotes a number that is infinitesimally smaller than $\lambda$. At $\beta = (\lambda^-, 0, \ldots, 0)^T$, we have $\hat{G}_a = K$ and so

$$\{\hat{G}_{a,1} \subseteq G_1\} \{G_a \subseteq \hat{G}_{a,1}^c\} = \{\forall x \in G_a : x \in \hat{G}_{a,1}^c\} = \{\forall x \in K : x \in \hat{G}_{a,1}^c\} = \{\forall x \in K : \hat{x}^T \hat{\beta} - c^*_T \hat{\sigma}(x) < \lambda\} = \{\forall x \in K : \hat{x}^T (\hat{\beta} - \beta) - c^*_T \hat{\sigma}(x) < \lambda - \hat{x}^T \beta\} = \{\forall x \in K : \hat{x}^T (\hat{\beta} - \beta) - c^*_T \hat{\sigma}(x) < 0\}$$

which gives

$$P\{\hat{G}_{a,1} \subseteq G_1\} = P\{\forall x \in K : \hat{x}^T (\hat{\beta} - \beta) - c^*_T \hat{\sigma}(x) < 0\} = 1 - \alpha. \quad (A4)$$

The combination of (A3) and (A4) proves the statement in (8).

The statement (9) can be proved by combining the arguments that establish (A1) and (A3) above to establish that

$$\{\hat{G}_{a,2} \subseteq G_2 \subseteq \hat{G}_{a,2,1}\} \supseteq \{\forall x \in K : -c^*_T \hat{\sigma}(x) \leq \hat{x}^T (\hat{\beta} - \beta) < c^*_T \hat{\sigma}(x)\};$$

details are omitted here to save space. Unfortunately, a least favorable configuration of $\beta$ that achieves the coverage probability $1 - \alpha$ cannot be identified in this case, and so $1 - \alpha$ is only a lower bound on the confidence level.

Now consider Theorem 2. For proving the statement in (11), we have

$$\{G_a \subseteq \hat{G}_{a,1} \forall \lambda \in \mathbb{R}^1\} = \cap_{\beta \in \mathbb{R}^{p+1}} \{G_a \subseteq \hat{G}_{a,1u}\} \supseteq \cap_{\beta \in \mathbb{R}^{p+1}} \{\forall x \in K : \hat{x}^T (\hat{\beta} - \beta) + c^*_T \hat{\sigma}(x) \geq 0\} \quad (A5)$$

where the “$\supseteq$” in (A5) follows directly from the proof of the statement in (7) above, and the second “$=$” follows directly since each set in (A5) has nothing to do with $\lambda$. It follows therefore

$$P\{G_a \subseteq \hat{G}_{a,1u} \forall \lambda \in \mathbb{R}^1\} \geq P\{\forall x \in K : \hat{x}^T (\hat{\beta} - \beta) + c^*_T \hat{\sigma}(x) \geq 0\} = 1 - \alpha. \quad (A6)$$

On the other hand, it is clear that $\{G_a \subseteq \hat{G}_{a,1u} \forall \lambda \in \mathbb{R}^1\} \subseteq \{G_a \subseteq \hat{G}_{a,1u}\}$ and so

$$\inf_{\beta \in \mathbb{R}^{p+1}, \sigma > 0} P\{G_a \subseteq \hat{G}_{a,1u} \forall \lambda \in \mathbb{R}^1\} \leq \inf_{\beta \in \mathbb{R}^{p+1}, \sigma > 0} P\{G_a \subseteq \hat{G}_{a,1u}\} = 1 - \alpha. \quad (A7)$$

The combination of (A6) and (A7) clearly gives the statement in (11).

The statements in (12)–(13) of Theorem 2 can be proved in the similar way, and so the details are omitted to save space.

**A.2 Additional simulation studies**

In this section, some additional simulation results mentioned in Section 3 are included (Tables A1-A5).
### TABLE A1  Coverage of 95% confidence sets for the linear model $Y = \beta x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ with $\epsilon \sim N(0, 1)$.

| Coverage of 95% confidence sets | Level $\lambda$ | $-0.9$ | $-0.5$ | $0$ | $0.5$ | $0.9$ |
|---------------------------------|----------------|--------|--------|-----|------|------|
| One-sided upper confidence sets | 0.5            | 0.9454 | 0.9490 | 0.9448 | 0.9472 | 0.9458 |
|                                 | 0.2            | 0.9506 | 0.9486 | 0.9435 | 0.9505 | 0.9554 |
| One-sided lower confidence sets  | 0.5            | 0.9482 | 0.9529 | 0.9458 | 0.9477 | 0.9482 |
|                                 | 0.2            | 0.9533 | 0.9476 | 0.9437 | 0.9443 | 0.9485 |
| Two-sided confidence sets 1      | 0.5            | 0.9680 | 0.9717 | 0.9706 | 0.9721 | 0.9699 |
|                                 | 0.2            | 0.9733 | 0.9720 | 0.9705 | 0.9739 | 0.9787 |
| Two-sided confidence sets 2      | 0.5            | 0.9483 | 0.9529 | 0.9458 | 0.9477 | 0.9482 |
|                                 | 0.2            | 0.9733 | 0.9702 | 0.9673 | 0.9709 | 0.9749 |

Note: The design points are evenly spread in the square space $[-1, 1]^2$ by $q$. Hence, the sample size used here is $5^2$ for $q = 0.5$ and $11^2$ for $q = 0.2$. Each with 10,000 simulation. The coverage of the upper confidence sets is computed with the parameter $\beta = (\lambda, 0, 0)$ and the coverage of the lower confidence sets are computed with $\beta = (\lambda - 0, 0)$.

### TABLE A2  Coverage of 95% confidence sets for the logistic model $Y = Bern(p)$, $logit(p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_5 x_5$.

| Coverage of 95% confidence sets | Level $L_0$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---------------------------------|-------------|-----|-----|-----|-----|-----|
| One-sided upper confidence sets | 0.5         | 0.9624 | 0.9519 | 0.9515 | 0.9438 | 0.9427 |
|                                 | 0.2         | 0.9499 | 0.9485 | 0.9437 | 0.9508 | 0.9514 |
|                                 | 0.1         | 0.9495 | 0.9554 | 0.9553 | 0.9543 | 0.9470 |
| One-sided lower confidence sets  | 0.5         | 0.9432 | 0.9525 | 0.9513 | 0.9531 | 0.9663 |
|                                 | 0.2         | 0.9514 | 0.9508 | 0.9494 | 0.9485 | 0.9499 |
|                                 | 0.1         | 0.9470 | 0.9543 | 0.9536 | 0.9554 | 0.9495 |

Note: The design points are evenly spread in the square space $[-1, 1]^2$ by $q$. The sample size used here is $n = 5^5$ for $q = 0.5$, $n = 11^5$ for $q = 0.2$ and $n = 21^5$ for $q = 0.1$. The coverage of the upper confidence sets is computed with the parameter $\beta = (\lambda, 0, 0, 0, 0, 0)$ and the coverage of the lower confidence sets are computed with $\beta = (\lambda - 0, 0, 0, 0, 0)$ where $\lambda = logit(L_0)$. Each with 10,000 simulations.

### TABLE A3  Coverage of 95% confidence sets for the $\lambda$-level set of the logistic model $Y = Bern(p)$, $logit(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

| Coverage of 95% two-sided confidence sets | Level $L_0$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|------------------------------------------|-------------|-----|-----|-----|-----|-----|
| $\beta = (\lambda, 0, 0)$               | 0.1         | 0.9855 | 0.9784 | 0.9768 | 0.9731 | 0.9617 |
|                                         | 0.01        | 0.9767 | 0.9757 | 0.9761 | 0.9723 | 0.9744 |
| $\beta = (\lambda - 0, 0)$              | 0.1         | 0.9617 | 0.9703 | 0.9733 | 0.9811 | 0.9868 |
|                                         | 0.01        | 0.9721 | 0.9705 | 0.9723 | 0.9743 | 0.9746 |

Note: The design points are evenly spread in the square space $[-1, 1]^2$ by $q$. The sample size used here is $n = 21^2$ for $q = 0.1$ and $n = 201^2$ for $q = 0.01$. The coverage of the two-sided confidence sets is computed with $\beta = (\lambda, 0, 0)$ and $\beta = (\lambda - 0, 0)$ where $\lambda = logit(L_0)$. Each with 10,000 simulations.
**TABLE A4** Coverage of 95% confidence sets for the level set in linear and logistic regressions.

| Coverage of 95% confidence sets | Linear regression | Logistic regression |
|---------------------------------|-------------------|---------------------|
| **Y = \( \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \)** | **Level \( \lambda \)** | **Level \( L_0 \) (where \( \lambda = \text{logit}(L_0) \))** |
| **Level \( \lambda \)** | -0.9 | -0.5 | 0 | 0.5 | 0.9 |
| One-sided upper confidence sets | 0.9480 | 0.9480 | 0.9470 | 0.9494 | 0.9469 |
| One-sided lower confidence sets | 0.9508 | 0.9539 | 0.9503 | 0.9499 | 0.9503 |
| **Logistic regression** | **Level \( L_0 \)** | **One-sided upper confidence sets** | **One-sided lower confidence sets** |
| **Logistic regression** | **Level \( L_0 \) (where \( \lambda = \text{logit}(L_0) \))** | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| One-sided upper confidence sets | 1 | 1 | 0.9983 | 0.9827 | 0.9619 |
| One-sided lower confidence sets | 0.9579 | 0.9816 | 0.9989 | 1 | 1 |

**Note:** The covariate region is assumed to be \( 92 \leq x_1 \leq 149 \) and \( 2 \leq x_2 \leq 5 \), the same as in the infant blood pressure example in Section 4.2, and there are 100 design points evenly spread on the covariate region. The regression model is \( y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \), where \( x = (x_1, x_2) \). The estimated parameter in Section 4.2 is \( \beta = (47.58, 0.18, 5.31) \), hence we assess the coverage at 7 different linear combinations of \( \beta \) and \( \beta_0 = (90, 0, 0) \). Note that with the same simulation setting, the coverage of the 1-sided upper confidence set at \((90, 0, 0)\) is 0.9507 and the coverage of the 1-sided lower confidence set at \((90^-, 0, 0)\) is 0.9500. Each with 10,000 simulations.

**A.3 Data in Example 2**

Data for Infant’s blood pressure example is given in Table A6.
**TABLE A6** Infant’s blood pressure data.

| Birth weight in oz | Age in days | Systolic blood pressure |
|--------------------|-------------|-------------------------|
| 125                | 3           | 86                      |
| 101                | 4           | 87                      |
| 104                | 4           | 87                      |
| 143                | 5           | 100                     |
| 92                 | 5           | 89                      |
| 119                | 3           | 86                      |
| 100                | 4           | 89                      |
| 149                | 3           | 89                      |
| 133                | 2           | 83                      |
| 120                | 4           | 92                      |
| 118                | 4           | 88                      |
| 94                 | 3           | 79                      |
| 131                | 5           | 98                      |
| 93                 | 4           | 85                      |
| 94                 | 4           | 87                      |
| 121                | 5           | 97                      |
| 96                 | 4           | 87                      |