Journal of Econometrics 204 (2018) 119–130

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Uniform confidence bands: Characterization and optimality[☆]

Joachim Freyberger *, Yoshiyasu Rai

Department of Economics, University of Wisconsin, Madison, United States

ARTICLE INFO

Article history:
Received 11 August 2017
Received in revised form 11 August 2017
Accepted 26 January 2018
Available online 5 February 2018

JEL classification:
C13

Keywords:
Uniform confidence bands
Simultaneous inference
Projections
Optimality

ABSTRACT

This paper studies optimal uniform confidence bands for functions $g(x, \beta_0)$, where β_0 is an unknown parameter vector. We provide a simple characterization of a general class of taut $1-\alpha$ uniform confidence bands, allowing for both nonlinear functions and nonparametrically estimated functions. Specifically, we show that all taut bands can be obtained from projections on confidence sets for β_0 and we characterize the class of sets which yield taut bands. Using these results, we then present a computational method for selecting an approximately optimal confidence band for a given objective function. We illustrate the applicability of these results in numerical applications.

1. Introduction

Uniform confidence bands for functions are useful to summarize statistical uncertainty in both parametric and nonparametric models. They allow the reader to easily assess statistical accuracy and perform various hypothesis tests about the function without access to the data. While there are many different $1-\alpha$ confidence bands for the same function, so far there is little guidance in the literature on which one to choose in practice.

A uniform confidence band for a function $g(x, \beta_0)$, where β_0 is an unknown parameter vector, consists of upper and lower bound functions $\hat{g}_u(x)$ and $\hat{g}_l(x)$, such that $g(x, \beta_0)$ is contained in $[\hat{g}_l(x), \hat{g}_u(x)]$ for all x with probability $1-\alpha$. Many different $1-\alpha$ confidence bands could be reported in a given application. The choice is important because not all $1-\alpha$ confidence bands are *taut* in the sense that it might be possible to weakly decrease the width of the interval for all x and to strictly decrease it for some x while keeping the same coverage probability (see Section 2 for a formal definition). Moreover, even two taut confidence bands for the same function can have very different shapes and properties. In this paper, we provide a simple characterization of a general class of taut $1-\alpha$ confidence bands, allowing for both nonlinear functions and nonparametrically estimated functions. Specifically, we show that, under certain restrictions, all taut bands can be obtained from projections on confidence sets for β_0 and we characterize the class of confidence sets which yield taut bands. We provide a second characterization of taut bands in terms of inversions of suprema of weighted t-statistics.

Using our simple and constructive characterization of taut uniform confidence bands, we then present a computational method for selecting approximately optimal bands for different objective functions. Our leading example is the band which minimizes a weighted area. For this example we provide low level conditions for the selected band to be approximately optimal and asymptotically valid. The general results in the paper also apply to a variety of other objective functions, such as minimizing average marginal coverage probabilities.

As a starting point we consider confidence bands for functions of the form $g(x, \beta_0) = p(x)'\beta_0$. We also assume that we have an estimator $\hat{\beta}$ of β_0, where $\hat{\beta} \sim N(\beta_0, \Sigma)$. Due to the normality assumption, the first set of results are exact finite sample results. We then discuss extensions of these results to asymptotic approximations using $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma)$, nonlinear functions satisfying $g(x, \beta) \approx g(x, \beta_0) + \nabla_\beta g(x, \beta_0)'(\beta - \beta_0)$ in a neighborhood of β_0, and nonparametric estimators using a finite dimensional approximation $g(x) \approx p_K(x)'\beta_K$.

We illustrate the wide applicability of these results in two numerical applications. First, we consider a regression model with heteroskedasticity and simulated data. Second, we use data from Berry et al. (1995) and construct confidence bands for price

☆ We thank a co-editor, an associate editor, and a referee for valuable comments which have helped to improve the paper. We also thank Jean-Marie Dufour, Amit Gandhi, Bruce Hansen, Matt Masten, Pepe Montiel Olea, Jack Porter, Andres Santos, Xiaoxia Shi as well as seminar and conference participants at UW Madison, the 2016 CEME conference at Duke, and the 2017 ASSA meetings for helpful comments and discussions.

* Corresponding author.
E-mail addresses: jfreyberger@ssc.wisc.edu (J. Freyberger), yrai@wisc.edu (Y. Rai).

https://doi.org/10.1016/j.jeconom.2018.01.006
0304-4076/

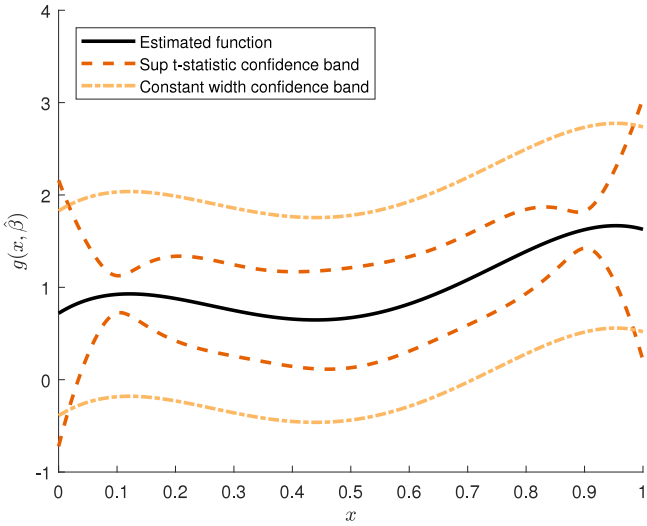

Fig. 1. Illustrative example.

elasticities implied by the estimated parameters of a structural model of demand.

Illustrative example: The following example illustrates that the choice of the uniform confidence band can be important. In this example $g(x, \beta) = \beta_1 + x\beta_2 + x^2\beta_3 + x^3\beta_4 + x^4\beta_5$ and $g(x, \beta_0) = E[Y \mid X = x]$. Section 4.1 explains the DGP, the estimator, and the confidence bands in detail. Fig. 1 shows the estimated function as well as two different 90% confidence bands. The dashed band is based on inversion of a standard sup t-statistic and the dotted–dashed band has a constant width for all x. Both of these bands are taut, have the same coverage probability, and are of the form $g(x, \hat{\beta}) \pm c(x)$. However, depending on how much importance a researcher places on different values of x, one might have clear preferences for one over the other. Moreover, hypothesis tests can lead to different outcomes depending on the band reported. For example, the null hypothesis that $g(x, \beta_0)$ is constant can be rejected with the sup t-statistic band, but not with the constant width band.

Related literature: The literature on uniform confidence bands goes back to Working and Hotelling (1929) who introduced hyperbolic bands in a simple linear regression model with normal errors. They showed that such a band can be obtained from a projection on an ellipse shaped confidence region, although this construction usually leads to conservative bands. These bands are often referred to as Scheffé bands due to his seminal work on multiple hypothesis testing (Scheffé, 1953). The width of the band based on the sup t-statistic is suitably smaller than but proportional to that of the Scheffé band and is thus also sometimes referred to as the Scheffé band. A variety of other bands, such as two or three segment bands or constant width bands, have been proposed in the literature (see Liu (2010) for an excellent overview). The first definition of taut bands we are aware of has been provided by Wynn and Bloomfield (1971), in a less general framework, who also showed in a linear regression with homoskedastic errors that all taut bands can be obtained by a projection (see also Khorasani and Milliken (1979) and Naiman (1984a) for characterization results in linear models). Our characterization of projection bands is more constructive which allows us, among others, to select (approximately) optimal bands using this result. Moreover, we start with more primitive assumptions, relate projection bands to bands obtained by t-statistic inversion, and extend the characterization to more general settings including nonlinear models. Gsteiger et al. (2011) discussed a confidence band for nonlinear regression models. Our results provide a formal justification for this band as well as various other ones. More recently, Belloni et al. (2015) constructed hyperbolic confidence bands for nonparametric regression functions using a series estimator and relying on high dimensional normal approximations. Other recent work on uniform confidence bands in nonparametric models includes Horowitz and Lee (2012, 2017), Chen and Christensen (2015), and Tao (2016). Projection based confidence regions have also been used in various other settings, such as Dufour (1990), Lütkepohl et al. (2015), Gafarov et al. (2016), and Kaido et al. (2016), but these papers do not consider our characterization and optimality results.

Related work on optimality properties of uniform confidence bands includes Bohrer (1973) who proved that Scheffé bands have the smallest average width with respect to the Lebesgue measure when the support is an ellipse (hence there cannot be an intercept in a linear model). Other papers extended this result to show that for certain confidence bands in a regression framework, there exist weight functions such that the bands minimize a weighted area (see Naiman (1984a) and Piegorsch (1985)). Our results imply the reverse, namely that one can find the optimal band for a given weight function. Naiman (1984b) considered optimality of bands which satisfy a bound on an expected coverage measure instead of having a certain coverage probability. Naiman (1987) characterized certain minimax regret bands. More recent papers such as Liu and Hayter (2007), Liu et al. (2009), and Liu and Ah-Kine (2010) are concerned with confidence bands which have optimality properties in terms of the implied confidence set for the parameter. Recently, Montiel Olea and Plagborg-Møller (2017) showed that a band obtained from inverting a sup t-statistic minimizes worst-case regret among a class of confidence bands. Relatedly, as we discuss further in Section 5, this band is also optimal in a minimax sense, but other bands are optimal with different criterion functions. To the best of our knowledge, we are the first to provide a constructive method to obtain optimal uniform confidence bands in a general class of models and general criterion functions.

Finally, note that many of these characterization and optimality results are obtained in a regression setting with normal and homoskedastic errors and rely on certain algebraic features of that model. Thus, not all of these results carry over to a general setting. For example, in a simple linear regression model, the width of the hyperbolic band is minimized at the mean value of the regressor, upper bound functions of taut bands are convex, and the constant width band is taut. These features do not hold more generally and hence optimality considerations can be more important in other settings (see e.g. Corollary 2).

2. Finite sample results

In this section we consider uniform confidence bands for a function $g(x, \beta_0) = p(x)'\beta_0$, where $p(x) \in \mathbb{R}^K$ is a vector of transformations of a vector $x \in \mathbb{R}^{d_x}$, and we have an estimator $\hat{\beta} \sim N(\beta_0, \Sigma)$. Bands are defined over a set $\mathcal{X} \subseteq \mathbb{R}^{d_x}$. The next section extends these results to asymptotic approximations, nonlinear functions, and nonparametric estimators.

2.1. Definitions and assumptions

In our setting the only information in the data about $g(x, \beta_0)$ is the estimator $\hat{\beta} \sim N(\beta_0, \Sigma)$. Thus, we restrict ourselves to confidence bands of the form $[g_l(x, \hat{\beta}), g_u(x, \hat{\beta})]$. Furthermore, we impose a regularity condition on the bands and only consider bands in the class $\mathcal{C}$ described in the following definition, where $\alpha \in (0, 1)$.

Definition 1. Let $\mathcal{C}$ be the class of confidence bands of the form $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})]$ with

$$P\left(g_l(x,\hat{\beta}) \le p(x)'\beta_0 \le g_u(x,\hat{\beta}) \text{ for all } x \in \mathcal{X}\right) = 1-\alpha$$

such that the set $\{\beta : g_l(x,\hat{\beta}) \le g(x,\beta) \le g_u(x,\hat{\beta}) \text{ for all } x \in \mathcal{X}\}$ can be written as $\{\beta : \hat{\beta} - \beta \in S\}$ where S is nonrandom.

All commonly used confidence bands, such as Scheffé bands, two or three segment bands, and constant width bands are in $\mathcal{C}$. The restriction implies that the shape of S is invariant to the realization of $\hat{\beta}$ and we also rule out, for example, that one randomly selects one of several valid confidence bands. Without such restrictions, one valid confidence band is $(-\infty,\infty)$ with probability $1-\alpha$ and $\emptyset$ with probability α for all $x \in \mathcal{X}$, and thus, optimal confidence bands would be much harder to characterize.

Interestingly, it might be possible to decrease the width of a given band in $\mathcal{C}$ without changing the coverage probability. In particular, a band in $\mathcal{C}$ might not be *taut*, which is formally defined as follows.[1]

Definition 2. Let $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})] \in \mathcal{C}$. The confidence band is called *taut* if there is no $[g_l^*(x,\hat{\beta}), g_u^*(x,\hat{\beta})] \in \mathcal{C}$ such that $[g_l^*(x,\hat{\beta}), g_u^*(x,\hat{\beta})] \subseteq [g_l(x,\hat{\beta}), g_u(x,\hat{\beta})]$ for all $x \in \mathcal{X}$ and all $\hat{\beta} \in \mathbb{R}^K$ and $[g_l^*(x,\hat{\beta}), g_u^*(x,\hat{\beta})] \subsetneq [g_l(x,\hat{\beta}), g_u(x,\hat{\beta})]$ and some $x \in \mathcal{X}$ and some $\hat{\beta} \in \mathbb{R}^K$. Bands that are not taut are called *slack*.

We provide a simple example of a slack band in $\mathcal{C}$ at the end of Section 2.2. For the remainder of this section we also impose the following three assumptions.

Assumption 1. $g(x,\beta) = p(x)'\beta$, where $p(x) \in \mathbb{R}^K$ is a vector of transformations of $x \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$.

Assumption 2. $\hat{\beta} \sim N(\beta_0, \Sigma)$ and Σ is positive definite and known.

Assumption 3. $\sigma(x) \equiv \sqrt{p(x)'\Sigma p(x)} > 0$ for all $x \in \mathcal{X}$.

For the results in this section the normality assumption is not critical and we could instead assume that $\hat{\beta} \sim F$ with $E(\hat{\beta}) = \beta_0$. However, without normality the distribution of $\hat{\beta} - \beta_0$ might depend on β_0 and thus the invariance assumption in Definition 1 might not be reasonable. Moreover, the extensions in Section 3 rely on asymptotic normality. In that section we also allow for an estimated covariance matrix. Finally, notice that since Σ is assumed to be positive definite, Assumption 3 is equivalent to $p(x)$ not being the zero vector for any $x \in \mathcal{X}$. This assumption could easily be relaxed and is mainly used to ensure that the standard sup t-statistic is well defined.

2.2. Characterization of taut confidence bands

We now characterize taut confidence bands in $\mathcal{C}$ in two ways, namely by a projection method and by inversion of a weighted t-statistic. The following lemma describes both methods and shows validity of the bands obtained in these ways.

Lemma 1. *Suppose that Assumptions 1–3 hold.*

1. *Let $CI(\hat{\beta}) \subset \mathbb{R}^K$ be such that $P(\beta_0 \in CI(\hat{\beta})) = 1-\alpha$. For all $x \in \mathcal{X}$ let*

$$g_l(x,\hat{\beta}) = \inf_{\beta \in CI(\hat{\beta})} p(x)'\beta \quad \text{and} \quad g_u(x,\hat{\beta}) = \sup_{\beta \in CI(\hat{\beta})} p(x)'\beta.$$

Then

$$P\left(g_l(x,\hat{\beta}) \le p(x)'\beta_0 \le g_u(x,\hat{\beta}) \text{ for all } x \in \mathcal{X}\right) \ge 1-\alpha.$$

2. *Let $w_l(x), w_u(x) \ge 0$ be known functions. Suppose there is a constant $c > 0$ such that*

$$P\left(\sup_{x\in\mathcal{X}} \frac{p(x)'(\beta_0-\hat{\beta})}{\sigma(x)} w_u(x) \le c,\ \inf_{x\in\mathcal{X}} \frac{p(x)'(\beta_0-\hat{\beta})}{\sigma(x)} w_l(x) \ge -c\right) = 1-\alpha.$$

Then $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})] \in \mathcal{C}$, where

$$g_l(x,\hat{\beta}) = \begin{cases} p(x)'\hat{\beta} - \dfrac{c\sigma(x)}{w_l(x)} & \text{if } w_l(x) > 0 \\ -\infty & \text{if } w_l(x) = 0 \end{cases}$$

$$g_u(x,\hat{\beta}) = \begin{cases} p(x)'\hat{\beta} + \dfrac{c\sigma(x)}{w_u(x)} & \text{if } w_u(x) > 0 \\ \infty & \text{if } w_u(x) = 0 \end{cases}$$

The first part of the lemma shows that a confidence band obtained by a projection on a confidence set for β_0 yields coverage at least $1-\alpha$, but it might be conservative. This well known result follows from the simple observation that if $\beta_0 \in CI(\hat{\beta})$, then for all $x \in \mathcal{X}$ $g_l(x,\hat{\beta}) = \inf_{\beta\in CI(\hat{\beta})} p(x)'\beta \le p(x)'\beta_0$ and similarly $p(x)'\beta_0 \le g_u(x,\hat{\beta})$. The second part describes confidence bands obtained from inversions of weighted sup and inf t-statistics, which are by construction in $\mathcal{C}$, but might not be taut. Therefore, neither the projection method nor t-statistic inversion necessarily yields taut bands in $\mathcal{C}$. The next theorem provides a simple characterization of the confidence sets for β_0, which lead to taut bands in $\mathcal{C}$ with the projection method. It also shows that all taut bands in $\mathcal{C}$ can be obtained by either the projection method or by inversion of a weighted t-statistic.

Theorem 1. *Suppose that Assumptions 1–3 hold.*

1. *A band $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})]$ is in $\mathcal{C}$ and taut if and only if $g_u(x,\hat{\beta}) = \sup_{\beta\in CI(\hat{\beta})} p(x)'\beta$ and $g_l(x,\hat{\beta}) = \inf_{\beta\in CI(\hat{\beta})} p(x)'\beta$, where*

$$CI(\hat{\beta}) = \{\beta \in \mathbb{R}^K : c_l(x) \le p(x)'(\beta-\hat{\beta}) \le c_u(x) \text{ for all } x \in \mathcal{X}\}$$

for some $c_l(x)$ and $c_u(x)$ and $P(\beta_0 \in CI(\hat{\beta})) = 1-\alpha$.

2. *Let $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})] \in \mathcal{C}$ be a taut confidence band. Then there exist weight functions $w_l(x), w_u(x) \ge 0$ such that*

$$P\left(\sup_{x\in\mathcal{X}} \frac{p(x)'(\beta_0-\hat{\beta})}{\sigma(x)} w_u(x) \le 1,\ \inf_{x\in\mathcal{X}} \frac{p(x)'(\beta_0-\hat{\beta})}{\sigma(x)} w_l(x) \ge -1\right) = 1-\alpha$$

and

$$g_l(x,\hat{\beta}) = \begin{cases} p(x)'\hat{\beta} - \dfrac{\sigma(x)}{w_l(x)} & \text{if } w_l(x) > 0 \\ -\infty & \text{if } w_l(x) = 0 \end{cases}$$

$$g_u(x,\hat{\beta}) = \begin{cases} p(x)'\hat{\beta} + \dfrac{\sigma(x)}{w_u(x)} & \text{if } w_u(x) > 0 \\ \infty & \text{if } w_u(x) = 0 \end{cases}$$

While the formal proof is in Section S.2 of the supplementary appendix, we now provide the main intuition for this result. First take an arbitrary taut uniform confidence band $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})] \in \mathcal{C}$ and define

$$\widetilde{CI}(\hat{\beta}) = \{\beta : g_l(x,\hat{\beta}) \le p(x)'\beta \le g_u(x,\hat{\beta}) \text{ for all } x \in \mathcal{X}\}.$$

[1] We could have defined the class $\mathcal{C}$ as all confidence bands which have a coverage probability of at least $1-\alpha$. Such a definition does not change the set of taut bands in the class because conservative bands are not taut. It would therefore not affect the characterization in Theorem 1.

Since $P(\beta_0 \in \widetilde{CI}(\hat{\beta})) = 1-\alpha$, Lemma 1 implies that projecting on this set results in another confidence band with coverage at least $1-\alpha$. Denote this confidence band by $[g_l^*(x,\hat{\beta}),\ g_u^*(x,\hat{\beta})]$. By definition of the projection it holds that $g_u^*(x,\hat{\beta}) \le g_u(x,\hat{\beta})$, and $g_l^*(x,\hat{\beta}) \ge g_l(x,\hat{\beta})$ for all $x \in \mathcal{X}$. Hence, if $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})]$ is taut, it has to hold that $[g_l^*(x,\hat{\beta}),\ g_u^*(x,\hat{\beta})] = [g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})]$ for all $x \in \mathcal{X}$. Moreover, by the definition of the class $\mathcal{C}$, there exists a nonrandom set $S \subset \mathbb{R}^K$ such that

$$\begin{aligned} g_l(x,\hat{\beta}) &= \inf_{\beta \in \widetilde{CI}(\hat{\beta})} p(x)'\beta \\ &= p(x)'\hat{\beta} + \inf_{\beta-\hat{\beta} \in \widetilde{CI}(\hat{\beta})-\hat{\beta}} p(x)'(\beta-\hat{\beta}) \\ &= p(x)'\hat{\beta} + \inf_{\gamma \in S} p(x)'\gamma, \end{aligned}$$

and similarly $g_u(x,\hat{\beta}) = p(x)'\hat{\beta} + \sup_{\gamma\in S} p(x)'\gamma$. Therefore, $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})] \in \mathcal{C}$ can be obtained by a projection on

$$\{\beta \in \mathbb{R}^K : \inf_{\gamma\in S} p(x)'\gamma \le p(x)'(\beta-\hat{\beta}) \le \sup_{\gamma\in S} p(x)'\gamma \text{ for all } x \in \mathcal{X}\}.$$

Next, we explain by means of an example why projecting on sets of the form

$$\{\beta \in \mathbb{R}^K : c_l(x) \le p(x)'(\beta-\hat{\beta}) \le c_u(x) \text{ for all } x \in \mathcal{X}\}$$

yields nonconservative confidence bands. In the formal proof we then show that whenever a projection yields a nonconservative band, the band is taut. For the example suppose that $p(x) = (1,x)'$ and $\mathcal{X} = [-1,1]$. Also assume that $\hat{\beta}_1$ and $\hat{\beta}_2$ have the same variance and are negatively correlated. Now consider the confidence band obtained by the projection on

$$CI(\hat{\beta}) = \{\beta \in \mathbb{R}^K : (\beta-\hat{\beta})'\Sigma^{-1}(\beta-\hat{\beta}) \le c_{2,1-\alpha}\}$$

where $c_{2,1-\alpha}$ is the $1-\alpha$ critical value of the χ_2^2 distribution. We will illustrate that the corresponding projection band, denoted by $[g_l(x,\hat{\beta}), g_u(x,\hat{\beta})]$, is conservative.

By definition, this band is conservative if and only if $P(\beta_0 \in CI_p(\hat{\beta})) > 1-\alpha$, where

$$CI_p(\hat{\beta}) = \{\beta \in \mathbb{R}^K : g_l(x,\hat{\beta}) \le p(x)'\beta \le g_u(x,\hat{\beta}) \text{ for all } x \in \mathcal{X}\}.$$

Fig. 2 illustrates the situation. It shows the ellipse $CI(\hat{\beta})$ as well as the lines of β satisfying $p(x)'\beta = \sup_{\beta\in CI(\hat{\beta})} p(x)'\beta$ and $p(x)'\beta = \inf_{\beta\in CI(\hat{\beta})} p(x)'\beta$ for all $x \in \{-1,0,1\}$. Now consider the point $\bar{\beta}_1$, which is not in $CI(\hat{\beta})$. It is also not in $CI_p(\hat{\beta})$ because, as illustrated in the figure, $g_u(\bar{x},\hat{\beta}) < p(\bar{x})'\bar{\beta}_1$ with $\bar{x} = 0$. In particular, there exists a $\bar{x} \in \mathcal{X}$ such that $\bar{\beta}_1$ and $CI(\hat{\beta})$ can be separated with a line $p(\bar{x})'\beta$. For $\bar{\beta}_2$ and $\bar{\beta}_3$, which are also not in $CI(\hat{\beta})$, such a line does not exist and hence, both of these vectors are in $CI_p(\hat{\beta})$. Therefore, the dashed area in Fig. 2 shows the boundary of the set $CI_p(\hat{\beta})$, which is clearly larger than the ellipse $CI(\hat{\beta})$. But since the ellipse covers the true value with probability $1-\alpha$, it follows that $P(\beta_0 \in CI_p(\hat{\beta})) > 1-\alpha$ and that the projection is conservative. These arguments then imply that nonconservative projections are based on confidence sets where each vector not in the set can be separated by a line $p(\bar{x})'\beta$ for some $\bar{x} \in \mathcal{X}$. Equivalently, the closure of the set is of the form $\{\beta \in \mathbb{R}^K : c_l(x) \le p(x)'(\beta-\hat{\beta}) \le c_u(x) \text{ for all } x \in \mathcal{X}\}$. An example here is a set that has the same shape as $CI_p(\hat{\beta})$, but is suitably smaller such that the coverage probability of the band is $1-\alpha$. Finally, notice that if $\mathcal{X} = (-\infty,\infty)$, then projecting on any convex confidence set, including the ellipse, yields a nonconservative band.

As shown above, for any taut band in $\mathcal{C}$, there exists a nonrandom set $S \subset \mathbb{R}^K$ such that

$$g_l(x,\hat{\beta}) = p(x)'\hat{\beta} + \inf_{\gamma\in S} p(x)'\gamma \quad \text{and}$$

$$g_u(x,\hat{\beta}) = p(x)'\hat{\beta} + \sup_{\gamma\in S} p(x)'\gamma.$$

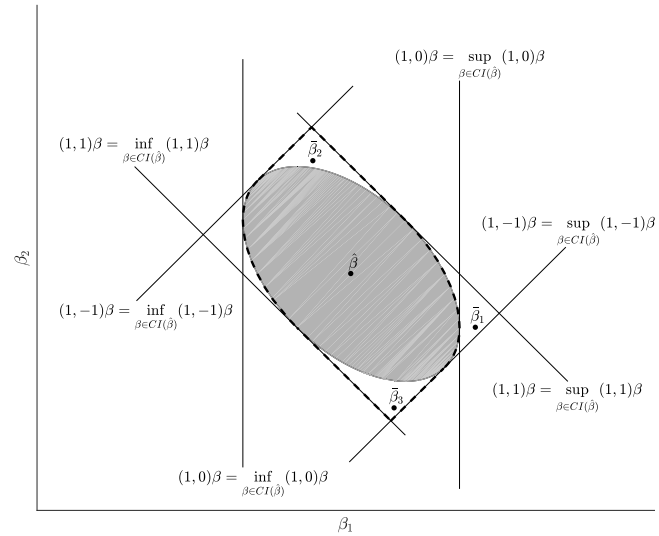

Fig. 2. Illustration of $CI_p(\hat{\beta}) = \{\beta : g_l(x,\hat{\beta}) \le p(x)'\beta \le g_u(x,\hat{\beta})\ \forall x \in \mathcal{X}\}$.

Using this result it is straightforward to construct weight functions such that the band can be obtained by inverting a weighted sup and inf t-statistic.

Theorem 1 implies that any taut confidence band in $\mathcal{C}$ can either be obtained by a projection on a confidence set or by inversion of a weighted sup and inf t-statistic. We now use this result to present an important class of weight functions, which lead to taut and symmetric confidence bands in $\mathcal{C}$. This class, described in the following corollary, is particularly interesting because it includes the case where $w_u(x) = w_l(x) = 1$.

Corollary 1. *Suppose that Assumptions 1–3 hold. Let Ω be a positive definite symmetric matrix and define $\nu(x) = \sqrt{p(x)'\Omega p(x)} > 0$. Let $w(x) = \frac{\sigma(x)}{\nu(x)}$. Then there exists a constant $c < \infty$ such that*

$$P\left(\sup_{x\in\mathcal{X}} \left|\frac{p(x)'(\beta_0-\hat{\beta})}{\sigma(x)} w(x)\right| \le c\right) = 1-\alpha$$

and the confidence band $[p(x)'\hat{\beta} - c\nu(x),\ p(x)'\hat{\beta} + c\nu(x)]$ is a taut confidence band in $\mathcal{C}$.

The arguments of the proof of Corollary 1 imply that $[p(x)'\hat{\beta} - c\nu(x),\ p(x)'\hat{\beta} + c\nu(x)]$ is equivalent to a band obtained from a projection on the set $\{\beta \in \mathbb{R}^K : (\beta-\hat{\beta})\Omega^{-1}(\beta-\hat{\beta}) \le c_1^2\}$, where c_1^2 is chosen such that resulting band has the right coverage probability. The next result states that constant width bands are taut if $p(x)$ includes a constant.

Corollary 2. *Suppose that Assumptions 1–3 hold and that there is a $c < \infty$ such that*

$$P\left(\sup_{x\in\mathcal{X}} \left|p(x)'(\beta_0-\hat{\beta})\right| \le c\right) = 1-\alpha.$$

If $p_1(x) = 1$ for all $x \in \mathcal{X}$, then $[p(x)'\hat{\beta} - c,\ p(x)'\hat{\beta} + c]$ is a taut confidence band in $\mathcal{C}$.

It is also easy to show that if $p(x)$ does not contain a constant, then the constant width band might not be taut. These situations can in particular arise with nonlinear functions discussed in Section 3.2, where the linearized version often does not contain a constant. As a very simple example, suppose that $p(x) = x$, where $\mathcal{X} = [1,2]$ and $\hat{\beta} \sim N(1,1)$. Then $\sup_{x\in\mathcal{X}} |p(x)'(\beta_0-\hat{\beta})| = 2|\beta_0-\hat{\beta}|$. Let $Z \sim N(0,1)$ and let c be such that $P(2|Z| \le c) = 1-\alpha$. The

constant width band is $x\hat{\beta} \pm c$. It then also holds by construction that

$$P\left(\sup_{x\in\mathcal{X}}\left|\frac{x(\beta_0-\hat{\beta})}{\frac{1}{2}|x|}\right| \le c\right) = 1-\alpha$$

and hence another $1-\alpha$ confidence band is $x\hat{\beta} \pm \frac{c}{2}|x|$, which reduces the widths of the constant width band for all $x \neq 2$, and the two bands have the same coverage probability.

2.3. Optimality

Theorem 1 provides a characterization of all taut confidence bands in $\mathcal{C}$ in terms of projections. In this section, we discuss how this result can be used to find a confidence band in this class which has optimality properties. Our leading example is a confidence band, which minimizes a weighted area

$$\int_{\mathcal{X}}\left(g_u(x,\hat{\beta}) - g_l(x,\hat{\beta})\right) w_X(x)dx.$$

A motivation for this objective function is that uniform confidence bands are typically reported to allow a reader to easily assess statistical accuracy and perform various hypothesis tests about the function without access to the data. Since it might be unclear a priori which hypotheses are tested, reporting a small confidence band (in terms of its area) could be desirable. Many other choices of objective functions are possible as well, some of which could relate to optimal testing, such as minimizing a weighted average of marginal coverage rates:

$$\int_{\mathcal{X}} P\left(g_l(x,\hat{\beta}) \le g(x,\beta_0) \le g_u(x,\hat{\beta})\right) w_X(x)dx.$$

We restrict ourselves to the class of taut bands, which is without loss of generality for most reasonable criterion functions. Let $[g_l(x,\hat{\beta}),\ g_u(x,\hat{\beta})]$ be a taut band. Theorem 1 and the discussion after the theorem imply that there exist nonrandom functions $c_l(x)$ and $c_u(x)$ such that $g_l(x,\hat{\beta}) = p(x)'\hat{\beta} + c_l(x)$ and $g_u(x,\hat{\beta}) = p(x)'\hat{\beta} + c_u(x)$. Moreover, it holds that

$$P(c_l(x) \le p(x)'(\beta_0-\hat{\beta}) \le c_u(x) \text{ for all } x\in\mathcal{X}) = 1-\alpha$$

or, by Assumption 2,

$$P(c_l(x) \le p(x)'\Sigma^{1/2}Z \le c_u(x) \text{ for all } x\in\mathcal{X}) = 1-\alpha,$$

where $Z \sim N(0, I_{K\times K})$. Since any taut confidence band is completely determined by $c_l(x)$ and $c_u(x)$, we consider the general optimization problem

$$\min_{c_l(\cdot),\ c_u(\cdot)} h\left(c_l(\cdot),\ c_u(\cdot)\right) \tag{1}$$
$$\text{s.t. } P(c_l(x) \le p(x)'\Sigma^{1/2}Z \le c_u(x) \text{ for all } x\in\mathcal{X}) = 1-\alpha$$

and we assume that $h(\cdot,\cdot)$ satisfies the following assumption.

Assumption 4. $h(\cdot,\cdot)$ is a nonrandom functional and $h(c_l^1(\cdot), c_u^1(\cdot)) \ge h(c_l^2(\cdot), c_u^2(\cdot))$ whenever $[c_l^2(x),\ c_u^2(x)] \subseteq [c_l^1(x),\ c_u^1(x)]$ for all $x\in\mathcal{X}$.

One example satisfying this assumption is the weighted area discussed before where

$$\begin{aligned} &h\left(c_l(\cdot),\ c_u(\cdot)\right) \\ &\quad= \int_{\mathcal{X}} \left(c_u(x) - c_l(x)\right) w_X(x)dx \\ &\quad= \int_{\mathcal{X}} \left(\left(p(x)'\hat{\beta} + c_u(x)\right) - \left(p(x)'\hat{\beta} + c_l(x)\right)\right) w_X(x)dx. \end{aligned}$$

The minimum area criterion of the implied confidence set for β_0, discussed in the introduction, also fits into this framework.

Let $(\bar{c}_l(x), \bar{c}_u(x))$ denote an optimal solution to the minimization problem. Even though $[p(x)'\hat{\beta} + \bar{c}_l(x),\ p(x)'\hat{\beta} + \bar{c}_u(x)]$ is then an optimal confidence band, it is not necessarily taut. We could for example widen a band for one value of x without affecting the weighted area. However, we can always obtain an optimal taut band by projecting on the set

$$\{\beta\in\mathbb{R}^K : \bar{c}_l(x) \le p(x)'(\beta-\hat{\beta}) \le \bar{c}_u(x) \text{ for all } x\in\mathcal{X}\}$$

and if the optimal solution $(\bar{c}_l(x), \bar{c}_u(x))$ corresponds to a taut band, the projection is simply $[p(x)'\hat{\beta} + \bar{c}_l(x),\ p(x)'\hat{\beta} + \bar{c}_u(x)]$. Hence, we can assume without loss of generality that the confidence band $[\bar{g}_l(x,\hat{\beta}),\ \ \bar{g}_u(x,\hat{\beta})]$ corresponding to the optimal solution $(\bar{c}_l(x), \bar{c}_u(x))$ is the projection on

$$\{\beta\in\mathbb{R}^K : \bar{c}_l(x) \le p(x)'(\beta-\hat{\beta}) \le \bar{c}_u(x) \text{ for all } x\in\mathcal{X}\}$$

and can be written as

$$\bar{g}_l(x,\hat{\beta}) = p(x)'\hat{\beta} + \bar{c}_l(x) \quad \text{and} \quad \bar{g}_u(x,\hat{\beta}) = p(x)'\hat{\beta} + \bar{c}_u(x).$$

Then $[\bar{g}_l(x,\hat{\beta}),\ \ \bar{g}_u(x,\hat{\beta})]$ is a taut band in $\mathcal{C}$ with the minimal objective function.

2.4. Approximation

Even in the simple linear model with homoskedastic errors and a bounded regressor, there is no closed form expression for the confidence band which minimizes the total area. Moreover, simply solving the minimization problem numerically is infeasible in practice because $c_l(x)$ and $c_u(x)$ are infinite dimensional. Instead, we now consider confidence bands which are approximately optimal and solve a finite dimensional optimization problem. We first approximate the objective function h by a function h_J and show that replacing h with h_J reduces (1) to a finite dimensional problem. We then provide conditions which guarantee that the band obtained by solving the finite dimensional problem is approximately optimal.

We approximate h by a function h_J using a grid of points $\mathcal{X}_J = \{x_1,\ldots,x_J\} \subseteq \mathcal{X}$. For example, we could approximate

$$h\left(c_l(\cdot), c_u(\cdot)\right) = \int_{\mathcal{X}} \left(c_u(x) - c_l(x)\right) w_X(x)dx,$$

by

$$h_J\left(c_l(\cdot), c_u(\cdot)\right) = \sum_{j=1}^{J-1}\left(c_u(x_j) - c_l(x_j)\right) w_X(x_j)(x_{j+1}-x_j).$$

More generally, we assume that h_J satisfies the following assumption.

Assumption 5. $h_J(\cdot,\cdot)$ satisfies Assumption 4 and $h_J(c_l^1(\cdot), c_u^1(\cdot)) = h_J(c_l^2(\cdot), c_u^2(\cdot))$ whenever $[c_l^1(x),\ c_u^1(x)] = [c_l^2(x),\ c_u^2(x)]$ for all $x\in\mathcal{X}_J$.

Under Assumption 5 it is easy to show that an optimal solution to

$$\min_{c_l(\cdot),\ c_u(\cdot)} h_J(c_l(\cdot), c_u(\cdot))$$
$$\text{s.t. } P(c_l(x) \le p(x)'\Sigma^{1/2}Z \le c_u(x) \text{ for all } x\in\mathcal{X}) = 1-\alpha$$

satisfies $c_l(x) = -\infty$ and $c_u(x) = \infty$ for all $x \notin \mathcal{X}_J$, because $c_l(x)$ and $c_u(x)$ affect the constraint, but not the objective for $x \notin \mathcal{X}_J$. Therefore, we can set $c_l(x) = -\infty$ and $c_u(x) = \infty$ for all $x \notin \mathcal{X}_J$ and minimize the objective function over $\{c_l(x_j), c_u(x_j)\}_{j=1}^J$ only. Hence, we have to solve the following finite dimensional minimization problem:

$$\begin{aligned} \min_{c_l(x_j),\ c_u(x_j),\ j=1,\ldots,J} \quad & h_J(c_l(\cdot), c_u(\cdot)) \\ \text{s.t.} \quad & P(c_l(x) \le p(x)'\Sigma^{1/2}Z \le c_u(x) \text{ for all } x\in\mathcal{X}_J) \\ & = 1-\alpha. \end{aligned}$$

Let $\{c_l^*(x_j),\ c_u^*(x_j)\}_{j=1}^J$ be an optimal solution to this minimization problem. We can now construct an approximately optimal and taut band for all $x \in \mathcal{X}$ by projecting on

$$\{\beta \in \mathbb{R}^K : c_l^*(x) \le p(x)'(\beta - \hat{\beta}) \le c_u^*(x) \text{ for all } x \in \mathcal{X}_J\}$$

and we write the optimal band as $\bar{g}_l^J(x, \hat{\beta}) = p(x)'\hat{\beta} + \bar{c}_l^J(x)$ and $\bar{g}_u^J(x, \hat{\beta}) = p(x)'\hat{\beta} + \bar{c}_u^J(x)$. Then, by Theorem 1, $[\bar{g}_l^J(x, \hat{\beta}),\ \bar{g}_u^J(x, \hat{\beta})]$ is a taut band in $\mathcal{C}$ for any fixed J. The next theorem states conditions under which it is approximately optimal.

Theorem 2. *Suppose Assumptions 4 and 5 hold. Also assume that*

$$\lim_{J\to\infty} \left| h_J\left(\bar{c}_l^J(x),\ \bar{c}_u^J(x)\right) - h\left(\bar{c}_l^J(x),\ \bar{c}_u^J(x)\right) \right| \to 0$$

and

$$\lim_{J\to\infty} \left| h_J\left(\bar{c}_l(x),\ \bar{c}_u(x)\right) - h\left(\bar{c}_l(x),\ \bar{c}_u(x)\right) \right| \to 0.$$

Then

$$\lim_{J\to\infty} \left| h\left(\bar{c}_l^J(x),\ \bar{c}_u^J(x)\right) - h\left(\bar{c}_l(x),\ \bar{c}_u(x)\right) \right| \to 0.$$

The theorem states that the band which solves a finite dimensional optimization problem is approximately optimal as long as h_J approximates h well as $J \to \infty$. In Appendix A.1 we provide low level sufficient conditions for the assumptions of Theorem 2 when the objective is to minimize the weighted area. These conditions then imply that we can obtain an arbitrarily good approximation by picking an arbitrarily fine grid to approximate the integral.[2] In this case we also show that the optimal solution satisfies $\bar{c}_l^J(x) = -\bar{c}_u^J(x)$, which further simplifies the computations.

3. Extensions

In this section we extend the finite sample results to asymptotic approximations, nonlinear functions $g(x, \beta_0)$, and nonparametric functions $g_0(x)$.

3.1. Asymptotic results

We now depart from Assumption 2 and instead assume that $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma)$ and that we have a consistent estimator $\hat{\Sigma}$ of Σ. The goal is to use the asymptotic distribution and the previous results to obtain optimal and asymptotically valid uniform confidence bands for $p(x)'\beta_0$. We restrict the class of confidence bands similar as before, but also add a condition that they are based on the asymptotic distribution.

Definition 3. Let $\hat{\mathcal{C}}$ be the class of confidence bands of the form $[g_l(x, \hat{\beta}, \hat{\Sigma}),\ g_u(x, \hat{\beta}, \hat{\Sigma})]$ such that the set $\{\beta : g_l(x, \hat{\beta}, \hat{\Sigma}) \le p(x)'\beta \le g_u(x, \hat{\beta}, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}\}$ can be written as $\{\beta : \sqrt{n}\hat{\Sigma}^{-1/2}(\beta - \hat{\beta}) \in S(\hat{\Sigma})\}$, where $S(\cdot)$ is a nonrandom function and it does not depend on n, and $P(Z \in S(\hat{\Sigma}) \mid \hat{\Sigma}) = 1 - \alpha$, where $Z \sim N(0, I_{K\times K})$ and $Z \perp\!\!\!\perp \hat{\Sigma}$.

The definition allows for standard ways of obtaining uniform confidence bands in practice. For example, symmetric uniform confidence bands based on a weighted sup t-statistic are constructed by finding a constant $c(\hat{\Sigma})$ such that

$$P\left(\sup_{x\in\mathcal{X}} \left| \frac{p(x)'\hat{\Sigma}^{1/2}Z}{\hat{\sigma}(x)} w(x) \right| \le c(\hat{\Sigma}) \mid \hat{\Sigma}\right) = 1 - \alpha,$$

where $\hat{\sigma}(x) = \sqrt{p(x)'\hat{\Sigma}p(x)}$. The uniform confidence band is then

$$p(x)'\hat{\beta} \pm \frac{c(\hat{\Sigma})}{\sqrt{n}} \frac{\hat{\sigma}(x)}{w(x)}.$$

Therefore, in this case

$$S(\hat{\Sigma}) = \left\{ z \in \mathbb{R}^K : -c(\hat{\Sigma})\frac{\hat{\sigma}(x)}{w(x)} \le p(x)'\hat{\Sigma}^{1/2}z \le c(\hat{\Sigma})\frac{\hat{\sigma}(x)}{w(x)} \text{ for all } x \in \mathcal{X} \right\},$$

which satisfies by construction all properties in Definition 3.

The arguments of the proof of Theorem 1 directly imply that any taut confidence band in $\hat{\mathcal{C}}$ can be obtained by projecting on a confidence set for β_0 of the form

$$\{\beta : c_l(x, \hat{\Sigma}) \le p(x)'\sqrt{n}(\beta - \hat{\beta}) \le c_u(x, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}\},$$

where $P(c_l(x, \hat{\Sigma}) \le p(x)'\hat{\Sigma}^{1/2}Z \le c_u(x, \hat{\Sigma}) \text{ for all } x \in \mathcal{X} \mid \hat{\Sigma}) = 1 - \alpha$ and $c_l(x, \cdot)$ and $c_u(x, \cdot)$ are nonrandom functions and do not depend on n. Thus, for any taut band we have

$$S(\hat{\Sigma}) = \{z \in \mathbb{R}^K : c_l(x, \hat{\Sigma}) \le p(x)'\hat{\Sigma}^{1/2}z \le c_u(x, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}\}.$$

Moreover, using arguments as in the previous section, it follows that

$$g_l(x, \hat{\beta}, \hat{\Sigma}) = p(x)'\hat{\beta} + \frac{1}{\sqrt{n}} \inf_{\gamma\in\mathbb{R}^K:\ \hat{\Sigma}^{-1/2}\gamma\in S(\hat{\Sigma})} p(x)'\gamma$$

and

$$g_u(x, \hat{\beta}, \hat{\Sigma}) = p(x)'\hat{\beta} + \frac{1}{\sqrt{n}} \sup_{\gamma\in\mathbb{R}^K:\ \hat{\Sigma}^{-1/2}\gamma\in S(\hat{\Sigma})} p(x)'\gamma.$$

As explained below, since $\hat{\Sigma}$ is random, the asymptotic coverage probability of such a band might not be $1-\alpha$. We now describe sufficient conditions for asymptotic validity, which is equivalent to $\lim_{n\to\infty} P(\sqrt{n}\hat{\Sigma}^{-1/2}(\beta_0 - \hat{\beta}) \in S(\hat{\Sigma})) = 1 - \alpha$.[3] To state these conditions, let $\mathcal{S}$ denote the class of all convex sets $S \subseteq \mathbb{R}^K$.

Assumption 6. Let $Z \sim N(0, I_{K\times K})$ and let $\Sigma \in \mathbb{R}^{K\times K}$ be positive definite. Suppose that $\sigma(x) = \sqrt{p(x)'\Sigma p(x)} \in (0, \infty)$ for all $x \in \mathcal{X}$,

(i) $\sup_{S\in\mathcal{S}} \left| P(\sqrt{n}\Sigma^{-1/2}(\hat{\beta} - \beta_0) \in S) - P(Z \in S) \right| \to 0$ as $n \to \infty$,

(ii) $\sup_{x\in\mathcal{X}} \frac{|g_l(x,\hat{\beta},\hat{\Sigma}) - g_l(x,\hat{\beta},\Sigma)|}{\sigma(x)} = o_p\left(\frac{1}{\sqrt{n}}\right)$ and
$\sup_{x\in\mathcal{X}} \frac{|g_u(x,\hat{\beta},\hat{\Sigma}) - g_u(x,\hat{\beta},\Sigma)|}{\sigma(x)} = o_p\left(\frac{1}{\sqrt{n}}\right)$.

We get the following result.

Lemma 2. *Suppose $[g_l(x, \hat{\beta}, \hat{\Sigma}),\ g_u(x, \hat{\beta}, \hat{\Sigma})] \in \hat{\mathcal{C}}$ is taut. If Assumption 6 holds, then $\lim_{n\to\infty} P(\sqrt{n}\hat{\Sigma}^{-1/2}(\beta_0 - \hat{\beta}) \in S(\hat{\Sigma})) = 1 - \alpha$.*

The first part of Assumption 6 assumes that $\sqrt{n}\Sigma^{-1/2}(\hat{\beta} - \beta_0)$ converges in distribution to a normal random vector. The requirement that the supremum over all convex sets converges can be verified using primitive sufficient conditions (see Bentkus, 2003). Part (ii) says that the confidence band is continuous in Σ. To better understand it write

$$\sqrt{n}\left| g_l(x, \hat{\beta}, \hat{\Sigma}) - g_l(x, \hat{\beta}, \Sigma) \right| = \left| \inf_{\gamma\in\mathbb{R}^K:\ \hat{\Sigma}^{-1/2}\gamma\in S(\hat{\Sigma})} p(x)'\gamma - \inf_{\gamma\in\mathbb{R}^K:\ \Sigma^{-1/2}\gamma\in S(\Sigma)} p(x)'\gamma \right|.$$

[2] If the optimal solution to the infinite dimensional problem is unique and continuous in $c_l(x)$ and $c_u(x)$, then Theorem 2 also implies that $[\bar{g}_l^J(x, \hat{\beta}),\ \bar{g}_u^J(x, \hat{\beta})]$ converges to $[\bar{g}_l(x, \hat{\beta}),\ \bar{g}_u(x, \hat{\beta})]$ as $J \to \infty$. This is for example the case for the minimum weighted area band. See Appendices A.1 and A.2 for related discussions.

[3] An alternative is to allow the coverage probability to be bigger or equal to $1 - \alpha$ asymptotically. However, the sufficient conditions lead to asymptotically nonconservative bands.

Hence, as long as $\hat{\Sigma}$ is a consistent estimator of Σ, any rule of choosing a band that depends continuously on Σ leads to the right coverage rate asymptotically.[4]

We are particularly interested in confidence bands obtained from solving a minimization problem. Analogously to before, we now solve

$$\min_{c_l(\cdot),\, c_u(\cdot)} h(c_l(\cdot), c_u(\cdot))$$
$$\text{s.t. } P(c_l(x) \le p(x)'\hat{\Sigma}^{1/2}Z \le c_u(x) \text{ for all } x \in \mathcal{X} \mid \hat{\Sigma}) = 1-\alpha.$$

Let $\bar{c}_l(x, \hat{\Sigma})$ and $\bar{c}_u(x, \hat{\Sigma})$ denote the minimizers and let $[\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}), \bar{g}_u(x, \hat{\beta}, \hat{\Sigma})]$ denote the corresponding confidence band. The following theorem now immediately implies that the bands are asymptotically valid under Assumption 6.

Theorem 3. *Suppose* $[\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}), \bar{g}_u(x, \hat{\beta}, \hat{\Sigma})]$ *satisfies Assumption 6. Then*

$$\lim_{n\to\infty} P(\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}) \le p(x)'\beta_0 \le \bar{g}_u(x, \hat{\beta}, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}) = 1-\alpha.$$

A confidence band based on a smooth optimization problem often satisfies Assumption 6, which holds if $[\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}), \bar{g}_u(x, \hat{\beta}, \hat{\Sigma})]$ depends continuously on $\hat{\Sigma}$. For example, in Appendix A.2 we provide primitive sufficient conditions for the confidence band which minimizes a weighted area to be asymptotically valid. Specifically, we show that log-concavity of the normal measure implies that there is a unique confidence band which minimizes the weighted area and thus that the band is continuous in $\hat{\Sigma}$.

However, notice that asymptotic validity might not hold if $[\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}), \bar{g}_u(x, \hat{\beta}, \hat{\Sigma})]$ changes discontinuously with $\hat{\Sigma}$. As an example, suppose that $\mathcal{X} = \{x_1, x_2\}$ and that $p(x_1) = (1, 0)'$ and $p(x_2) = (0, 1)'$. Further suppose that $\hat{\Sigma}$ is a diagonal matrix. Then the constraint can be written as $P(c_l(x_1) \le \hat{\sigma}_1 Z_1 \le c_u(x_1),\ c_l(x_2) \le \hat{\sigma}_2 Z_2 \le c_u(x_2) \mid \hat{\Sigma}) = 1-\alpha$. Now suppose that $h(c_l(\cdot), c_u(\cdot)) = \min\{c_u(x_1), c_u(x_2)\}$. Clearly, for the optimal solution $\bar{c}_l(x_1) = \bar{c}_l(x_2) = -\infty$ and it is easy to show that the optimal solution is

$$(\bar{c}_u(x_1), \bar{c}_u(x_2)) = (c_{1-\alpha}\hat{\sigma}_1, \infty) \text{ if } \hat{\sigma}_1 \le \hat{\sigma}_2 \text{ and}$$
$$(\bar{c}_u(x_1), \bar{c}_u(x_2)) = (\infty, c_{1-\alpha}\hat{\sigma}_2) \text{ if } \hat{\sigma}_1 > \hat{\sigma}_2,$$

where $c_{1-\alpha}$ denotes the $1-\alpha$ quantile of the standard normal distribution. Since the corresponding band changes discontinuously with $\hat{\Sigma}$, one can show that it generally does not have the right coverage probability asymptotically.

3.2. Nonlinear functions

We next apply the previous results to nonlinear functions of x and β by using delta method type arguments. First, we construct a confidence band for a linear approximation of the function. Second, we transform this confidence band to a valid confidence band for the nonlinear function $g(x, \beta_0)$. For the second step to be valid, we assume that

$$\sqrt{n}(g(x, \hat{\beta}) - g(x, \beta_0)) = \sqrt{n}\nabla_\beta g(x, \hat{\beta})'(\hat{\beta} - \beta_0) + o_p(1)$$
$$\xrightarrow{d} N(0, \nabla_\beta g(x, \beta_0)'\Sigma\nabla_\beta g(x, \beta_0)),$$

where $\nabla_\beta g(x, \beta)$ denotes the gradient of $g(x, \beta)$ with respect to β. As defined below, we only consider confidence bands based on this linear approximation of $g(x, \beta)$.

[4] In Appendix A.2 we provide primitive sufficient conditions for Assumption 6 for confidence bands which minimize a weighted area. Moreover, we extend Assumption 6 to nonlinear and nonparametric models in Sections 3.2 and 3.3, respectively.

Definition 4. Let $\hat{\mathcal{C}}_{nl}$ be the class of confidence bands of the form $[g_l(x, \hat{\beta}, \hat{\Sigma}), g_u(x, \hat{\beta}, \hat{\Sigma})]$ such that the set $\{\beta : g_l(x, \hat{\beta}, \hat{\Sigma}) \le \nabla_\beta g(x, \hat{\beta})'\beta \le g_u(x, \hat{\beta}, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}\}$ can be written as $\{\beta : \sqrt{n}\hat{\Sigma}^{-1/2}(\beta - \hat{\beta}) \in S(\hat{\beta}, \hat{\Sigma})\}$, where $S(\cdot, \cdot)$ is a nonrandom function and it does not depend on n, and $P(Z \in S(\hat{\beta}, \hat{\Sigma}) \mid \hat{\beta}, \hat{\Sigma}) = 1-\alpha$, where $Z \sim N(0, I_{K\times K})$ and $Z \perp\!\!\!\perp (\hat{\Sigma}, \hat{\beta})$.

The previous results again imply that all taut confidence bands in $\hat{\mathcal{C}}_{nl}$ can be obtained by a projection and that we can find optimal confidence bands by solving

$$\min_{c_l(\cdot),\, c_u(\cdot)} h(c_l(\cdot), c_u(\cdot))$$
$$\text{s.t. } P(c_l(x) \le \nabla_\beta g(x, \hat{\beta})'\hat{\Sigma}^{1/2}Z \le c_u(x) \text{ for all } x \in \mathcal{X} \mid \hat{\beta}, \hat{\Sigma}) = 1-\alpha.$$

Let $\bar{c}_l(x, \hat{\beta}, \hat{\Sigma})$ and $\bar{c}_u(x, \hat{\beta}, \hat{\Sigma})$ be minimizers such that the projection of $\nabla_\beta g(x, \hat{\beta})'\beta$ on

$$\{\beta \in \mathbb{R}^K : \bar{c}_l(x, \hat{\beta}, \hat{\Sigma}) \le \nabla_\beta g(x, \hat{\beta})'\sqrt{n}(\beta - \hat{\beta}) \le \bar{c}_u(x, \hat{\beta}, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}\}$$

can be written as

$$\left[\nabla_\beta g(x, \hat{\beta})'\hat{\beta} + \frac{1}{\sqrt{n}}\bar{c}_l(x, \hat{\beta}, \hat{\Sigma}),\ \nabla_\beta g(x, \hat{\beta})'\hat{\beta} + \frac{1}{\sqrt{n}}\bar{c}_u(x, \hat{\beta}, \hat{\Sigma})\right].$$

The confidence band for $g(x, \beta_0)$ is now simply

$$\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}) = g(x, \hat{\beta}) + \frac{1}{\sqrt{n}}\bar{c}_l(x, \hat{\beta}, \hat{\Sigma}) \quad \text{and}$$
$$\bar{g}_u(x, \hat{\beta}, \hat{\Sigma}) = g(x, \hat{\beta}) + \frac{1}{\sqrt{n}}\bar{c}_u(x, \hat{\beta}, \hat{\Sigma}).$$

By definition the confidence band is optimal relative to those in $\hat{\mathcal{C}}_{nl}$. The following theorem provides conditions under which the resulting confidence band is also asymptotically valid.

Theorem 4. *Let* $Z \sim N(0, I_{K\times K})$ *and let* $\Sigma \in \mathbb{R}^{K\times K}$ *be positive definite. Suppose that* $\sigma(x) = \sqrt{\nabla_\beta g(x, \beta_0)'\Sigma\nabla_\beta g(x, \beta_0)} \in (0, \infty)$ *for all* $x \in \mathcal{X}$,

(i) $\sup_{S\in\mathcal{S}} \left|P(\sqrt{n}\Sigma^{-1/2}(\hat{\beta} - \beta_0) \in S) - P(Z \in S)\right| \to 0$ *as* $n \to \infty$,

(ii) $\sup_{x\in\mathcal{X}} \frac{|\bar{c}_l(x,\hat{\beta},\hat{\Sigma}) - \bar{c}_l(x,\beta_0,\Sigma)|}{\sigma(x)} = o_p(1)$ *and* $\sup_{x\in\mathcal{X}} \frac{|\bar{c}_u(x,\hat{\beta},\hat{\Sigma}) - \bar{c}_u(x,\beta_0,\Sigma)|}{\sigma(x)} = o_p(1)$,

(iii) $\sup_{x\in\mathcal{X}} \frac{\sup_{\beta:\|\beta-\beta_0\|\le\|\hat{\beta}-\beta_0\|} \|\nabla_\beta g(x,\beta) - \nabla_\beta g(x,\beta_0)\|}{\sigma(x)} = o_p(1)$.

Then as $n \to \infty$

$$P\left(\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}) \le g(x, \beta_0) \le \bar{g}_u(x, \hat{\beta}, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}\right) \to 1-\alpha.$$

Parts (i)–(ii) of the assumptions are analogous to those of Assumption 6. However, notice that here the confidence sets we project on also depend on $\hat{\beta}$ through $\nabla_\beta g(x, \hat{\beta})$. Therefore, the functions $\bar{c}_l(x, \hat{\beta}, \hat{\Sigma})$ and $\bar{c}_u(x, \hat{\beta}, \hat{\Sigma})$ have to be continuous in $\hat{\beta}$ as well to achieve asymptotic validity. Simple sufficient conditions for part (iii) are that $\|\hat{\beta} - \beta_0\| \xrightarrow{p} 0$, $\inf_{x\in\mathcal{X}}\sigma(x) > 0$, and $\|\nabla_\beta g(x, \beta_1) - \nabla_\beta g(x, \beta_2)\| \le C\|\beta_1 - \beta_2\|$ for all $x \in \mathcal{X}$.

3.3. Nonparametric functions

We now generalize the approach to confidence bands for a nonparametric function $g_0(x)$, which we estimate by $p_{K_n}(x)'\hat{\beta}_{K_n}$, where $p_{K_n}(x) \in \mathbb{R}^{K_n}$ and $K_n \to \infty$ as $n \to \infty$. We first use the previous results to construct confidence bands for $p_{K_n}(x)'E(\hat{\beta}_{K_n})$. These bands are then asymptotically valid if the approximation bias $|g_0(x) - p_{K_n}(x)'E(\hat{\beta}_{K_n})|$ converges to 0 fast enough. The bands

are optimal for a given K_n and a given objective, but we do not consider optimal choices of K_n. As before, $\hat{\Sigma}$ denotes the estimated covariance matrix of $\hat{\beta}_{K_n}$. The class of confidence bands is similar as in Section 3.1.

Definition 5. Let $\hat{\mathcal{C}}_{np}$ be the class of confidence bands of the form $[g_l(x, \hat{\beta}_{K_n}, \hat{\Sigma}), g_u(x, \hat{\beta}_{K_n}, \hat{\Sigma})]$ such that the set $\{\beta : g_l(x, \hat{\beta}_{K_n}, \hat{\Sigma}) \leq p_{K_n}(x)'\beta \leq g_u(x, \hat{\beta}_{K_n}, \hat{\Sigma}) \ \forall x \in \mathcal{X}\}$ can be written as $\{\beta : \sqrt{n}\hat{\Sigma}^{-1/2}(\beta - \hat{\beta}_{K_n}) \in S(\hat{\Sigma})\}$, where $S(\cdot)$ is a nonrandom function and it does not depend on n for a given K_n, and $P(Z \in S(\hat{\Sigma}) \mid \hat{\Sigma}) = 1 - \alpha$, where $Z \sim N(0, I_{K_n \times K_n})$ and $Z \perp\!\!\!\perp \hat{\Sigma}$.

Similar as in Section 3.1, symmetric uniform confidence bands based on a weighted sup t-statistic, such as those in Belloni et al. (2015), are in $\hat{\mathcal{C}}_{np}$. The arguments of the proof of Theorem 1 directly imply that any taut confidence band in $\hat{\mathcal{C}}_{np}$ can be obtained by projecting on a confidence set for $E(\hat{\beta}_{K_n})$, which is based on the normal distribution and is conditional on $\hat{\Sigma}$. Therefore, optimal confidence bands can be constructed by solving

$$\min_{c_l(\cdot),\, c_u(\cdot)} h(c_l(\cdot), c_u(\cdot))$$
$$\text{s.t. } P(c_l(x) \leq p_{K_n}(x)'\hat{\Sigma}^{1/2}Z \leq c_u(x) \text{ for all } x \in \mathcal{X} \mid \hat{\Sigma}) = 1 - \alpha.$$

Let $\bar{c}_l(x, \hat{\Sigma})$ and $\bar{c}_u(x, \hat{\Sigma})$ be minimizers such the corresponding confidence band is

$$\bar{g}_l(x, \hat{\beta}_{K_n}, \hat{\Sigma}) = g(x, \hat{\beta}_{K_n}) + \frac{1}{\sqrt{n}}\bar{c}_l(x, \hat{\Sigma}) \quad \text{and}$$
$$\bar{g}_u(x, \hat{\beta}_{K_n}, \hat{\Sigma}) = g(x, \hat{\beta}_{K_n}) + \frac{1}{\sqrt{n}}\bar{c}_u(x, \hat{\Sigma}).$$

The previous arguments imply that this band is optimal relative to all bands in $\hat{\mathcal{C}}_{np}$. The next theorem provides conditions under which the band is asymptotically valid.

Theorem 5. *Let $Z \sim N(0, I_{K_n \times K_n})$ and let $\Sigma \in \mathbb{R}^{K_n \times K_n}$ be positive definite. Suppose that $\sigma(x) = \sqrt{p_{K_n}(x)'\Sigma p_{K_n}(x)} \in (0, \infty)$ for all $x \in \mathcal{X}$,*

(i) $\sup_{S \in \mathcal{S}} \left| P\left(\sqrt{n}\Sigma^{-1/2}(\hat{\beta}_{K_n} - E(\hat{\beta}_{K_n})) \in S\right) - P(Z \in S) \right| \to 0$ *as* $n \to \infty$,

(ii) $\sup_{x \in \mathcal{X}} \frac{|\bar{c}_l(x, \hat{\Sigma}) - \bar{c}_l(x, \Sigma)|}{\sigma(x)} = o_p(1/K_n)$ *and* $\sup_{x \in \mathcal{X}} \frac{|\bar{c}_u(x, \hat{\Sigma}) - \bar{c}_u(x, \Sigma)|}{\sigma(x)} = o_p(1/K_n)$,

(iii) $\sup_{x \in \mathcal{X}} \frac{|p(x)'E(\hat{\beta}_{K_n}) - g_0(x)|}{\sigma(x)} = o(1/K_n)$.

Then as $n \to \infty$ and $K_n \to \infty$

$$P\left(\bar{g}_l(x, \hat{\beta}_{K_n}, \hat{\Sigma}) \leq g_0(\beta) \leq \bar{g}_u(x, \hat{\beta}_{K_n}, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}\right) \to 1 - \alpha.$$

Again, the first two parts of the assumptions are analogous to Assumption 6 and part (iii) is the undersmoothing condition, which typically holds as long as g_0 is sufficiently smooth. See e.g. Chen (2007) for sufficient conditions. An alternative is to adjust the confidence bands to account for a worst-case bias as in Armstrong and Kolesár (2016).

4. Numerical examples

We now demonstrate the results using two numerical applications. First, we consider a regression model with simulated data. Second, we use data from Berry et al. (1995) and construct confidence bands for price elasticities implied by estimated parameters of a structural model of demand. We report confidence bands obtained with t-statistics and different weight functions as well as optimal bands, which minimize a weighted area.

As explained in Section 2.4, we obtain an approximately optimal band by first solving

$$\min_{c_l(x_j),\, c_u(x_j),\, j=1,\ldots,J} \sum_{j=1}^{J-1} \left(c_u(x_j) - c_l(x_j)\right) w_X(x_j)(x_{j+1} - x_j)$$
$$\text{s.t. } P(c_l(x_j) \leq p(x_j)'\hat{\Sigma}^{1/2}Z \leq c_u(x_j)$$
$$\text{for all } j = 1, \ldots, J \mid \hat{\Sigma}) \geq 1 - \alpha.$$

Using an inequality constraint instead of an equality constraint does not change the optimal solution, but it ensures that the feasible region is convex. Let $\{\bar{c}_l(x_j),\ \bar{c}_u(x_j)\}_{j=1}^{J}$ denote the optimal solution. We then obtain the optimal confidence band by projecting on the set $\{\beta \in \mathbb{R}^K : \bar{c}_l(x_j) \leq p(x_j)'(\beta - \hat{\beta}) \leq \bar{c}_u(x_j) \text{ for all } j = 1, \ldots, J\}$.

While we only have to solve a finite dimensional minimization problem, the computationally challenging part is to calculate the left hand side of the constraint

$$P(c_l(x_j) \leq p(x_j)'\hat{\Sigma}^{1/2}Z \leq c_u(x_j) \text{ for all } j = 1, \ldots, J \mid \hat{\Sigma}) \geq 1 - \alpha.$$

Notice that the constraint is equivalent to

$$E\left(\prod_{j=1}^{J} \mathbf{1}(c_l(x_j) \leq p(x_j)'\hat{\Sigma}^{1/2}Z)\mathbf{1}(p(x_j)'\hat{\Sigma}^{1/2}Z \leq c_u(x_j)) \mid \hat{\Sigma}\right)$$
$$\geq 1 - \alpha.$$

We approximate the left hand side of the constraint by

$$\frac{1}{n_s}\sum_{k=1}^{n_s}\prod_{j=1}^{J} K\left(p(x_j)'\hat{\Sigma}^{1/2}Z_k - c_l(x_j)\right) K\left(c_u(x_j) - p(x_j)'\hat{\Sigma}^{1/2}Z_k\right),$$

where $\{Z_k\}_{k=1}^{n_s}$ is an iid sample from the distribution of Z and $K(u)$ is a cdf corresponding to the Epanechnikov kernel function: $K(u) = \frac{3}{4}(\frac{u}{h} + 1) - \frac{1}{4}((\frac{u}{h})^3 + 1)$ for $|\frac{u}{h}| \leq 1$ and $\mathbf{1}(\frac{u}{h} > 1)$ for $|\frac{u}{h}| > 1$. In the simulations below we take $n_s = 50{,}000$ and $h = n_s^{-\frac{1}{2}}$. Notice that as $h \to 0$, $K(u)$ converges to $\mathbf{1}(u \leq 0)$, in which case we would simply use a sample average to approximate an expected value. Our approximation yields almost identical probabilities for any given values of $\{c_l(x_j),\ c_u(x_j)\}_{j=1}^{J}$, but it ensures that the constraint is a smooth function of $\{c_l(x_j),\ c_u(x_j)\}_{j=1}^{J}$. Finally, once we find an optimal solution $\{\bar{c}_l(x_j),\ \bar{c}_u(x_j)\}_{j=1}^{J}$ we can eliminate the approximation error to guarantee the correct coverage probability. To do so, we can pick the constant b such that

$$P(b\bar{c}_l(x_j) \leq p(x_j)'\hat{\Sigma}^{1/2}Z \leq b\bar{c}_u(x_j) \text{ for all } j = 1, \ldots, J \mid \hat{\Sigma}) = 1 - \alpha$$

and the left hand side can be calculated arbitrarily accurately using simulations. In our simulations with a large value of n_s, b is always extremely close to 1. Notice that in this way the uniform confidence band resulting from a projection on

$$\{\beta \in \mathbb{R}^K : b\bar{c}_l(x_j) \leq p(x_j)'(\beta - \hat{\beta}) \leq b\bar{c}_u(x_j) \text{ for all } j = 1, \ldots, J\}$$

is guaranteed to be a taut band in $\mathcal{C}$ and it is approximately optimal. Moreover, the results (in terms of the weighted area for example) can easily be compared to alternative confidence bands, such as bands obtained by weighted sup t-statistic inversion.

4.1. Regression

In this example, let

$$Y = \beta_{0,1} + X\beta_{0,2} + X^2\beta_{0,3} + X^3\beta_{0,4} + X^4\beta_{0,5} + \nu(X)U,$$

where $X \sim Beta(2, 2)$, $U \sim N(0, 1)$, $X \perp\!\!\!\perp U$, and $\nu(X) = 0.3$ if $|X - 0.5| \geq 0.25$ and $\nu(X) = 3$ if $|X - 0.5| < 0.25$. We use a sample size of 500 to estimate β_0 by OLS and Σ using heteroskedasticity robust standard errors. The true regression function is shown in

Table 1
Coverage and average width comparison.

Band	Coverage rate	Total area	Weighted area
Sup t-statistic	0.911	1.003	0.959
Constant width	0.894	2.217	2.217
Minimum area	0.910	0.967	0.919
Minimum weighted area	0.907	1.002	0.891

Fig. 3. We estimate four different 90% uniform confidence bands, namely the band from a sup t-statistic with $w_l(x) = w_u(x) = 1$, the band from a sup t-statistic with $w_l(x) = w_u(x) = \hat{\sigma}(x)$ (which results in a band with the same width for all x), the confidence band which minimizes the total area, and the band which minimizes a weighted area, where the weight function is the density of X. To solve for the optimal bands we use the grid $x_j = \frac{j}{30}$ for $j = 0, \ldots, 30$. Hence $J = 31$. We then use a grid of 301 points to project on the resulting confidence bands. We estimate these bands for 1000 simulated data sets. Table 1 shows the coverage probabilities, the average total area (averaged over the 1000 data sets), and the average weighted area for the four bands. All bands have coverage probabilities close to 90%. The total area of the minimum area band is on average only around 4% smaller than that of the unweighted sup t-statistic band, while the constant width band has a much larger (weighted) area. We solve for the bands using MATLAB's fmincon optimization routine with the interior-point algorithm and the pre-specified tolerance levels. We use closed form expressions of the gradients for both the objective function and the constraint. As starting values, we use a feasible and conservative constant width band (i.e. $c_u(x) = -c_l(x) = 2$). We also use the fact that the optimal band is symmetric and set $c_u(x) = -c_l(x)$, which reduces the number of parameters we optimize over. It takes us on average 271 and 312 seconds to find the band which (approximately) minimizes the total area and the weighted area, respectively.[5]

Figs. 3 and 4 show the true function and the average confidence bands. The figures also show average pointwise confidence bands as a reference because no symmetric uniform confidence band can be inside the pointwise band for any x. The constant width band is least preferred for most objective functions, unless one places extremely high priority on small or large values of x. Fig. 4 illustrates that the band that minimizes the weighted area is narrower than the unweighted sup t-statistic band for all $x \in [0.185, 0.815]$, which covers around 82% of the data. The cost is that the band is wider in areas with little data.

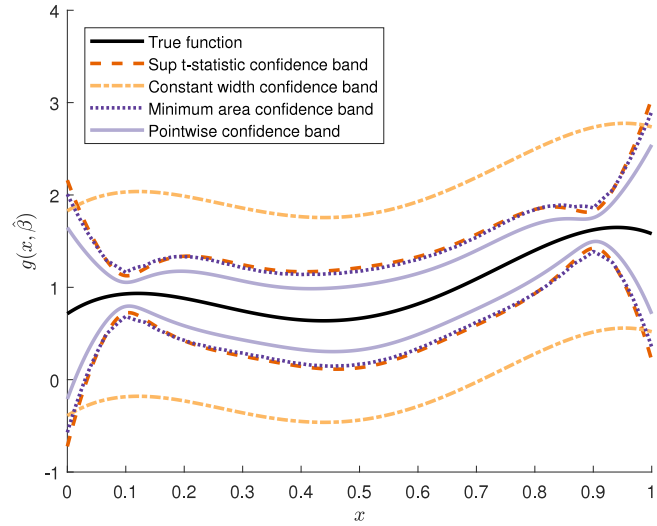

Fig. 3. Regression — optimal unweighted area band.

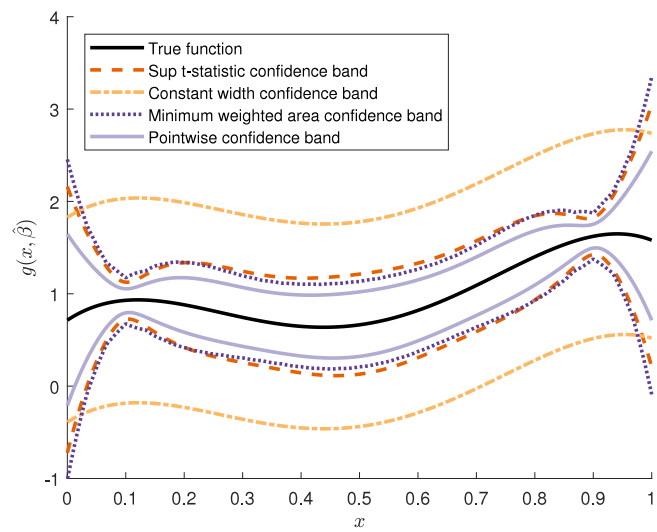

Fig. 4. Regression — optimal weighted area band.

4.2. Demand estimation

In this section we apply the previous results to estimate price elasticities in a simplified version of the well known BLP model proposed by Berry et al. (1995). In this setting each consumer i buys one of J products in market t and chooses the product that maximizes her utility. In our simplified setting, consumers have heterogeneous preferences over prices of product j in market t, denoted by p_{jt}, but homogeneous preferences over other product characteristics w_{jt}. Consequently, the market share of product j in market t, denoted by s_{jt}, can be written as

$$s_{jt} = \int \frac{\exp(w_{jt}'\gamma_0 - p_{jt}\alpha_0 - p_{jt}\sigma_0\eta + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(w_{kt}'\gamma_0 - p_{kt}\alpha_0 - p_{kt}\sigma_0\eta + \xi_{kt})} dF_\eta(\eta)$$
$$j = 1, \ldots, J,$$

where F_η is the log-normal distribution, ξ_{jt} denotes unobserved product characteristics, and the parameter vector is $\beta_0 = (\gamma_0, \alpha_0, \sigma_0)$. Moreover, we assume that there are instruments z_{jt}, such that $E(\xi_{jt}|z_{jt}) = 0$. Berry et al. (1995) show that β_0 can be estimated from observed market shares, prices, product characteristics, and instruments using a GMM approach.

For a fixed market t let us consider $f(p, \beta)$ given in Box I, where $\bar{w}_t$ denotes the median value of w_{jt}, and

$$g(p, \beta) = \frac{p\nabla_p f(p, \beta)}{f(p, \beta)}.$$

Hence, our function of interest $g(p, \beta)$ denotes the price elasticity for a product with the median characteristics added to market t while setting all unobservables to 0.

To estimate β_0 we use a subset of the data from Berry et al. (1995), which contains the retail prices (in \$1000), quantities sold, and several product characteristics of car models marketed between 1971 and 1990 (see their paper for a detailed description of the data). We exclude products with very small market shares and use 2172 model/year observations in total. The vector of product characteristics includes the size of the model defined

[5] To obtain the results in Table 1, we ran the simulations using the resources of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The reported times are based on 100 simulated data sets and we ran these programs using MATLAB R2017a on a desktop computer with an Intel Core i7 processor running at 3.1 GHz. The code is available on one of the authors' website: http://www.ssc.wisc.edu/~jfreyberger/ConfidenceBandsPrograms.

$$f(p,\beta) = \int \frac{\exp(\bar{w}_t'\gamma - p\alpha - p\sigma\eta)}{1 + \exp(\bar{w}_t'\gamma - p\alpha - p\sigma\eta) + \sum_{k=1}^{J} \exp(w_{kt}'\gamma - p_{kt}\alpha - p_{kt}\sigma\eta)} dF_\eta(\eta),$$

Box I.

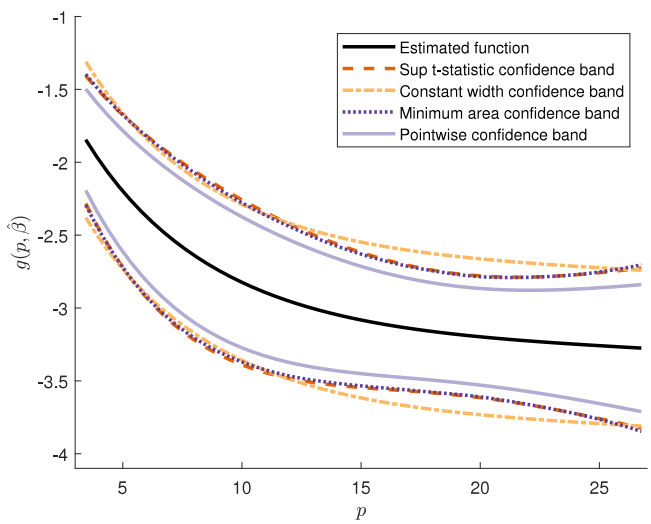

Fig. 5. Elasticities — optimal unweighted area band.

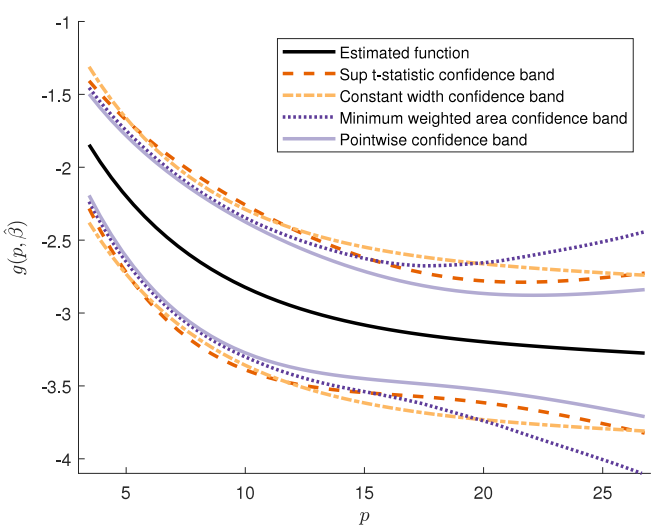

Fig. 6. Elasticities — optimal weighted area band.

as length×width/1000, MPG, horsepower/weight ratio, and an air conditioning dummy. As instruments we use the product characteristics w_{jt} as well as the number of similar products interacted with w_{jt} as recently suggested by Gandhi and Houde (2016).

Fig. 5 shows the estimated price elasticity function and, just as in Section 4.1, three 90% uniform confidence bands, including the band which minimizes the total area, as well as pointwise confidence bands. Here we use a grid of 50 points to calculate the optimal bands, but using 30 or 100 points yields almost identical results. We use $n_s = 50{,}000$ and $h = n_s^{-1/3}$. To calculate the elasticities we use the characteristics from 1971 and in this year around 90% of the prices are below 15 and more than 55% of the prices are between 6 and 12. Hence, the constant width band is narrower than the sup t-statistic band for the majority of the observed prices, and the minimum area band is narrower than the sup t-statistic band for the majority of the observed prices. The total areas of the sup t-statistic band and the constant width band are 1.2% and 9.7% larger than the total area of the optimal band.

Fig. 6 shows the same bands, except that the dotted band now corresponds to the minimum weighted area band, where the weight function is a log-normal approximation of the density of price. This band is narrower than the sup t-statistic band for the vast majority of observed prices and it gets very close to the pointwise band around the median price of \$7.8. The weighted areas of the sup t-statistic band and the constant width band are 15.1% and 13.5% larger than the weighted area of the optimal band.

5. Conclusion

In this paper we provide a method for constructing optimal confidence bands for a general class of functions, which includes nonlinear functions and nonparametrically estimated functions. To obtain these results we show that any taut $1-\alpha$ confidence band in a certain class can be obtained using a projection or inversion of a weighted t-statistic. We then characterize the class of confidence sets for the parameter vector β_0 which lead to nonconservative and taut confidence bands using projections. Our simple characterization allows us to present a computational method for approximating the optimal confidence band for a given objective function. We also provide simple sufficient conditions for asymptotic validity of confidence bands which minimize a weighted area and illustrate the wide applicability of these results in two numerical examples.

In our experience from running a variety of simulations in different settings and with different coverage probabilities, the band from the sup t-statistic with $w_l(x) = w_u(x) = 1$ is usually close to the band which minimizes the total area, and is computationally trivial to obtain. Moreover, since this band is taut and it has the same marginal coverage probability for all $x \in \mathcal{X}$, it is easy to show that it is optimal when the objective is to minimize $h(c_l(\cdot), c_u(\cdot)) = \max_{x\in\mathcal{X}} P(c_l(x) \le p(x)'\Sigma^{1/2}Z \le c_u(x))$. However, we have also seen in Section 4.2 that minimizing a weighted area can lead to a band which is narrower than the (unweighted) sup t-statistic band for the vast majority of x. Similarly, if the constant width band is taut, which follows from the condition in Corollary 2, it is optimal when the objective is to minimize $h(c_l(\cdot), c_u(\cdot)) = \max_{x\in\mathcal{X}} |c_u(x) - c_l(x)|$. However, this band typically performs very poorly in terms of total area (see also Naiman (1983)).

Appendix A. Minimum weighted area

In this section, we provide simple sufficient conditions for the assumptions of Theorems 2 and 3 when the objective is to minimize the weighted area:

$$\min_{c_l(\cdot),\, c_u(\cdot)} \int_{\mathcal{X}} (c_u(x) - c_l(x))\, w_X(x)dx$$
$$\text{s.t. } P(c_l(x) \le p(x)'\Sigma^{1/2}Z \le c_u(x) \text{ for all } x \in \mathcal{X}) = 1-\alpha.$$

As in Section 3.1, let $(\bar{c}_l(x), \bar{c}_u(x))$ denote an optimal solution to the minimization problem such that the corresponding confidence band, which can be obtained by a projection on

$$\{\beta \in \mathbb{R}^K : \bar{c}_l(x) \le p(x)'(\beta - \hat{\beta}) \le \bar{c}_u(x) \text{ for all } x \in \mathcal{X}\},$$

can be written as $\bar{g}_l(x,\hat{\beta}) = p(x)'\hat{\beta} + \bar{c}_l(x)$ and $\bar{g}_u(x,\hat{\beta}) = p(x)'\hat{\beta} + \bar{c}_u(x)$.

A.1. Approximation

We first provide conditions for the confidence band based on the approximate objective function to be approximately optimal. As in Section 2.4, let $\{\bar{c}_l(x_j), \bar{c}_u(x_j)\}_{j=1}^J$ be an optimal solution to

$$\min_{c_l(x_j),\ c_u(x_j),\ j=1,\dots,J} \sum_{j=1}^{J-1} \left(c_u(x_j) - c_l(x_j)\right) w_X(x_j)(x_{j+1} - x_j)$$
$$\text{s.t. } P(c_l(x_j) \le p(x_j)' \Sigma^{1/2} Z \le c_u(x_j) \text{ for all } j = 1, \dots, J) = 1 - \alpha.$$

Let $(\bar{c}_l^J(x), \bar{c}_u^J(x))$ for all $x \notin \mathcal{X}_J$ be such that the confidence band obtained by projecting on

$$\{\beta \in \mathbb{R}^K : \bar{c}_l^J(x) \le p(x)'(\beta - \hat{\beta}) \le \bar{c}_u^J(x) \text{ for all } x \in \mathcal{X}_J\}$$

is $\bar{g}_l^J(x, \hat{\beta}) = p(x)'\hat{\beta} + \bar{c}_l^J(x)$ and $\bar{g}_u^J(x, \hat{\beta}) = p(x)'\hat{\beta} + \bar{c}_u^J(x)$.

The lemma below provides low level conditions for $[\bar{g}_l^J(x, \hat{\beta}), \bar{g}_u^J(x, \hat{\beta})]$ to be approximately optimal when x is a scalar and $\mathcal{X}$ is a bounded interval. The results can be extended to unbounded support or $x \in \mathbb{R}^{d_x}$ using more notation. Let $\bar{x} = \sup \mathcal{X}$ and $\underline{x} = \inf \mathcal{X}$. We impose the following assumptions on the weight function $w_X(x)$ and the vector $p(x)$.

Assumption 7. $\mathcal{X}$ is a bounded interval, $w_X(x)$ and $p(x)$ are differentiable, and there exists a constant C such that

$$\sup_{x\in\mathcal{X}} |w_X(x)| + \sup_{x\in\mathcal{X}} |w_X'(x)| \le C$$

and

$$\sup_{x\in\mathcal{X}} \|p(x)\| + \sup_{x\in\mathcal{X}} \|\nabla p(x)\| \le C.$$

The next assumption imposes a mild support condition.

Assumption 8. There exists a nonrandom matrix $B \in \mathbb{R}^{K\times L}$ and a vector $\tilde{p}(x) \in \mathbb{R}^L$ such that $rank(B) = L$ and $p(x) = B\tilde{p}(x)$ for all $x \in \mathcal{X}$. Furthermore, there exists $\varepsilon > 0$ such that for any $\gamma \in \mathbb{R}^L$ with $\|\gamma\| = 1$ it holds that $\int_{\mathcal{X}} |\tilde{p}(x)'\gamma| w_X(x)dx \ge \varepsilon$.

It is easy to show that a simply sufficient condition for $\int_{\mathcal{X}} |\tilde{p}(x)'\gamma| w_X(x)dx \ge \varepsilon$ for some $\varepsilon > 0$ is that $\sup_{x\in\mathcal{X}} \|\tilde{p}(x)\| < \infty$ and that the matrix $\int_{\mathcal{X}} \tilde{p}(x)\tilde{p}(x)' w_X(x)dx$ has full rank. The matrix B allows $\int_{\mathcal{X}} p(x)p(x)' w_X(x)dx$ to have reduced rank, which can be important in practice. For example, in our empirical application in Section 4.2, $p(x)$ comes from a linearized single index model and is of the form $p(x) = (b_1 x, \dots, b_K x)'$ for constants $b_1, \dots, b_K$. Hence, $\int_{\mathcal{X}} p(x)p(x)' w_X(x)dx$ has rank 1, but with $B = (b_1, \dots, b_K)'$ and $\tilde{p}(x) = x$, Assumption 8 holds as long as $\int_{\mathcal{X}} |x| w_X(x)dx > 0$.

Finally, we impose an assumption on the grid points.

Assumption 9. $x_1 = \underline{x}, x_J = \bar{x}$, and $\sum_{j=1}^{J-1} (x_{j+1} - x_j)^2 \to 0$ as $J \to \infty$. Furthermore, there exists $\varepsilon > 0$ such that for any $\gamma \in \mathbb{R}^L$ with $\|\gamma\| = 1$ and for all J large enough $\sum_{j=1}^{J-1} |\tilde{p}(x_j)'\gamma| w_X(x_j)(x_{j+1} - x_j) \ge \varepsilon$, where $\tilde{p}(x)$ is defined as in Assumption 8.

The first part of the assumption is satisfied with equally spaced grid points for example. Similar as Assumption 8, the second part says that $\tilde{p}(x_1), \dots, \tilde{p}(x_J)$ cannot be in a linear subspace of $\mathbb{R}^L$.

The following lemma shows that under the previous assumptions the band obtained from solving the finite dimensional minimization problem is approximately optimal. The proof is in Section S.2.2 of the supplementary appendix.

Lemma A1. *Suppose Assumptions 1–3 and 7–9 hold. Then*

$$\left| \int_{\mathcal{X}} (\bar{g}_u(x, \hat{\beta}) - \bar{g}_l(x, \hat{\beta})) w_X(x)dx - \int_{\mathcal{X}} (\bar{g}_u^J(x, \hat{\beta}) - \bar{g}_l^J(x, \hat{\beta})) w_X(x)dx \right| \to 0 \text{ as } J \to \infty.$$

A.2. Validity

Let $(\bar{c}_l(x, \hat{\Sigma}), \bar{c}_u(x, \hat{\Sigma}))$ denote an optimal solution to the minimization problem

$$\min_{c_l(\cdot),\ c_u(\cdot)} \int_{\mathcal{X}} (c_u(x) - c_l(x)) w_X(x)dx$$
$$\text{s.t. } P(c_l(x) \le p(x)' \hat{\Sigma}^{1/2} Z \le c_u(x) \text{ for all } x \in \mathcal{X} \mid \hat{\Sigma}) = 1 - \alpha$$

such that the corresponding confidence band, which can be obtained by a projection on

$$\{\beta \in \mathbb{R}^K : \bar{c}_l(x, \hat{\Sigma}) \le p(x)'(\beta - \hat{\beta}) \le \bar{c}_u(x, \hat{\Sigma}) \text{ for all } x \in \mathcal{X}\},$$

can be written as $\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}) = p(x)'\hat{\beta} + \frac{1}{\sqrt{n}}\bar{c}_l(x, \hat{\Sigma})$ and $\bar{g}_u(x, \hat{\beta}) = p(x)'\hat{\beta} + \frac{1}{\sqrt{n}}\bar{c}_u(x, \hat{\Sigma})$.

In this section we provide primitive sufficient conditions for the second part of Assumption 6, namely that

$$\sup_{x\in\mathcal{X}} \frac{|\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}) - \bar{g}_l(x, \hat{\beta}, \Sigma)|}{\sigma(x)} = o_p\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\sup_{x\in\mathcal{X}} \frac{|\bar{g}_u(x, \hat{\beta}, \hat{\Sigma}) - \bar{g}_u(x, \hat{\beta}, \Sigma)|}{\sigma(x)} = o_p\left(\frac{1}{\sqrt{n}}\right).$$

In particular, we obtain the following lemma. The proof is in Section S.2.2 of the supplementary appendix.

Lemma A2. *Suppose Assumptions 1, 3, 7, and 8 hold. Also assume that*

(i) $w_X(x) > 0$ *for all* $x \in (\underline{x}, \bar{x})$ *and* $\inf_{x\in\mathcal{X}} \sigma(x) > 0$,
(ii) $\sup_{x\in X} |\sigma(x) - \hat{\sigma}(x)| = o_p(1)$ *and* $\|\hat{\Sigma}^{1/2} - \Sigma^{1/2}\| = o_p(1)$.

Then

$$\sup_{x\in\mathcal{X}} \frac{|\bar{g}_l(x, \hat{\beta}, \hat{\Sigma}) - \bar{g}_l(x, \hat{\beta}, \Sigma)|}{\sigma(x)} = o_p\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\sup_{x\in\mathcal{X}} \frac{|\bar{g}_u(x, \hat{\beta}, \hat{\Sigma}) - \bar{g}_u(x, \hat{\beta}, \Sigma)|}{\sigma(x)} = o_p\left(\frac{1}{\sqrt{n}}\right).$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2018.01.006.

References

Armstrong, T., Kolesár, M., 2016. Simple and Honest Confidence Intervals in Nonparametric Regression. Working Paper. https://arxiv.org/abs/1606.01200.

Belloni, A., Chernozhukov, V., Chetverikov, D., Kato, K., 2015. Some new asymptotic theory for least squares series: Pointwise and uniform results. J. Econometrics 186 (2), 345–366.

Bentkus, V., 2003. On the dependence of the Berry–Esseen bound on dimension. J. Statist. Plann. Inference 113 (2), 385–402.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.

Bohrer, R., 1973. Optimality property of Scheffé bounds. Ann. Statist. 1 (4), 766–772.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics. In: Handbook of Econometrics, vol. 6, Elsevier, pp. 5550–5623 (Chapter 76).

Chen, X., Christensen, T.M., 2015. Optimal Sup-Norm Rates, Adaptivity and Inference in Nonparametric Instrumental Variables Estimation. Working paper. https://arxiv.org/abs/1508.03365v1.

Dufour, J.-M., 1990. Exact tests and confidence sets in linear regressions with autocorrelated errors. Econometrica 58 (2), 475–494.

Gafarov, B., Meier, M., Montiel Olea, J.L., 2016. Projection inference for set-identified SVARs. Working paper. http://www.joseluismontielolea.com/GMM2_June2016_Final_Submitted.pdf.

Gandhi, A., Houde, J.-F., 2016. Measuring substitution patterns in differentiated products industries. Working paper. https://blogs.cornell.edu/houde/files/2016/01/GH_v5.pdf.

Gsteiger, S., Bretz, F., Liu, W., 2011. Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. J. Biopharm. Statist. 21, 708–725.

Horowitz, J.L., Lee, S., 2012. Uniform confidence bands for functions estimated nonparametrically with instrumental variables. J. Econometrics 168 (2), 175–188.

Horowitz, J.L., Lee, S., 2017. Nonparametric estimation and inference under shape restrictions. J. Econometrics 201 (1), 108–126.

Kaido, H., Molinari, F., Stoye, J., 2016. Confidence intervals for projections of partially identified parameters. Working paper. https://arxiv.org/abs/1601.00934v1.

Khorasani, F., Milliken, G.A., 1979. On the exactness of confidence bands about a linear model. J. Amer. Statist. Assoc. 74 (366), 446–448.

Liu, W., 2010. Simultaneous Inference in Regression. CRC Press.

Liu, W., Ah-Kine, P., 2010. Optimal simultaneous confidence bands in simple linear regression. J. Statist. Plann. Inference 100, 1225–1235.

Liu, W., Hayter, A.J., 2007. Minimum area confidence set optimality for confidence bands in simple linear regression. J. Amer. Statist. Assoc. 102 (477), 181–190.

Liu, W., Hayter, A.J., Piegrosch, W.W., Ah-Kine, P., 2009. Comparison of hyperbolic and constant width simultaneous confidence bands in multiple linear regression under MVCS criterion. J. Multivariate Anal. 100, 1432–1439.

Lütkepohl, Helmut, Staszewska-Bystrova, Anna, Winker, Peter, 2015. Confidence bands for impulse responses: Bonferroni vs. Wald. Oxf. Bull. Econ. Stat. 77 (6), 800–821.

Montiel Olea, J.L., Plagborg-Møller, M., 2017. Simultaneous Confidence Bands: Theoretical Comparisons And Suggestions for Practice. Working Paper. https://scholar.harvard.edu/plagborg/conf_band.

Naiman, D.Q., 1983. Comparing Scheffé-type to constant-width confidence bounds in regression. J. Amer. Statist. Assoc. 78 (384), 906–912.

Naiman, D.Q., 1984a. Average width optimality of simultaneous confidence bounds. Ann. Statist. 12 (4), 1199–1214.

Naiman, D.Q., 1984b. Optimal simultaneous confidence bounds. Ann. Statist. 12 (2), 702–715.

Naiman, D.Q., 1987. Minimax regret simultaneous confidence bands for multiple regression functions. J. Amer. Statist. Assoc. 82 (399), 894–901.

Piegorsch, W.W., 1985. Admissible and optimal confidence bands in simple linear regression. Ann. Statist. 13 (2), 801–810.

Scheffé, H., 1953. A method for judging all contrasts in the analysis of variance. Biometrika 40 (1/2), 87–104.

Tao, J., 2016. Uniform Trinity Tests for Semi/Nonparametric GMM Models. Working paper. https://sites.google.com/a/uw.edu/jing-tao/trinity.pdf?attredirects=0.

Working, H., Hotelling, H., 1929. Applications of the theory of error to the interpretation of trends. J. Amer. Statist. Assoc. 24 (165), 73–85.

Wynn, H.P., Bloomfield, P., 1971. Simultaneous confidence bands in regression analysis. J. R. Stat. Soc. Ser. B Stat. Methodol. 33 (2), 202–217.