# Uniform confidence bands: characterization and optimality*

Joachim Freyberger† Yoshiyasu Rai‡

January 7, 2017

**Abstract**

This paper studies optimal uniform confidence bands for functions $g(x, \beta_0)$, where $\beta_0$ is an unknown parameter vector. While there are many different $1 - \alpha$ confidence bands for the same function, not all $1 - \alpha$ confidence bands are *taut* in the sense that it might be possible to weakly decrease the width of the band for all $x$ and to strictly decrease it for some $x$. We provide a simple characterization of a general class of taut $1 - \alpha$ uniform confidence bands, allowing for both nonlinear and nonparametric functions. Specifically, we show that all taut bands can be obtained from projections on confidence sets for $\beta_0$ and we characterize the class of confidence sets which yield taut bands. Using our simple and constructive characterization of these sets, we then present a computational method for selecting an approximately optimal confidence band for a given objective function, such as minimizing the weighted area. We illustrate the wide applicability of these results in two numerical applications.

**Keywords:** Uniform confidence bands, simultaneous inference, projections, optimality.

*We 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 particpants at UW Madison and the 2016 CEME conference at Duke, and the 2017 ASSA meetings for helpful comments and discussions.

†Department of Economics, University of Wisconsin - Madison. Email: jfreyberger@ssc.wisc.edu.

‡Department of Economics, University of Wisconsin - Madison. Email: yrai@wisc.edu.

# 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.

Generally, a uniform confidence band for a function $g(x,\beta_0)$, where $\beta_0$ 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 one could report 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 and nonparametric functions. Specifically, we show that, under certain restrictions, all taut bands can be obtained from projections on confidence sets for $\beta_0$ and we characterize the class of confidence sets which yield taut bands. Furthermore, we provide a second characterization of taut bands in terms of inversions of suprema of weighted t-statistics.

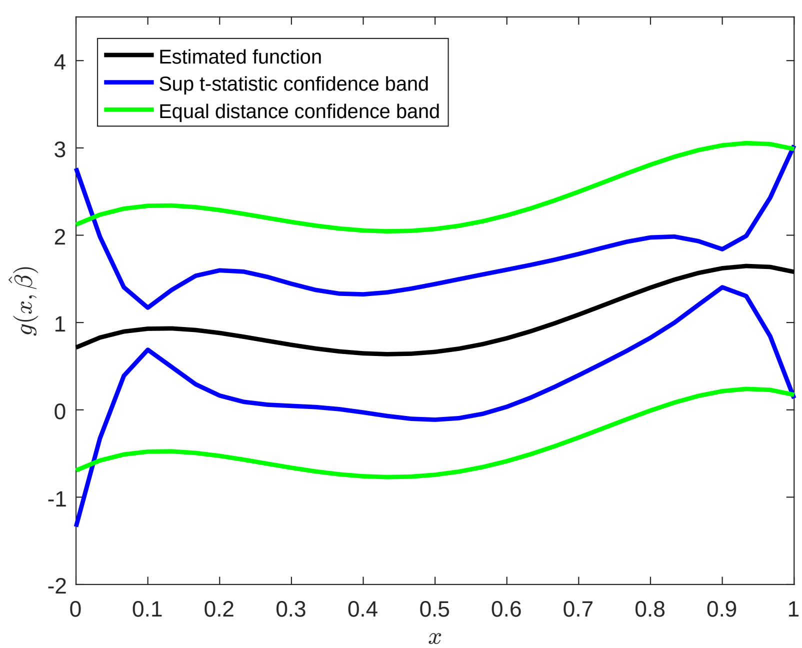
Using our simple and constructive characterization of taut uniform confidence bands, we then present a computational method for selecting approximately optimal bands for different objective functions. Our leading example is the band which minimizes a weighted area. For this example we provide low level 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. The approximation requires calculating Gaussian probabilities over convex polyhedral integration regions and we use recent advances in the computational literature.

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 $\beta_0$, 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 $\beta_0$, and nonparametric functions 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, Levinsohn, and Pakes (1995) and construct confidence bands for price 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. Figure 1 shows the estimated function as well as two different 90% confidence bands. The blue band is based on inversion of a standard sup t-statistic and the green band has equal distance to the estimated function 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 blue band, but not with the green one.

Figure 1: Illustrative example




**Related literature:** The literature on uniform confidence bands goes back to Working and Hotelling (1929) who introduced hyperbolic confidence bands in a simple linear regression model with normal errors. They show 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 refereed to as Scheffé bands due to his seminal work on multiple hypothesis testing (Scheffé, 1953). 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 show 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) discuss 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) construct 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, 2015), 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 paper 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 have 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) considers optimality of bands which satisfy a bound on an expected coverage measure instead of having a certain coverage probability. Naiman (1987) characterizes 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. 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.

Finally, notice that many of these characterization and optimality results are obtained in a regression setting with normally distributed 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 for example 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 $\hat{\beta} \sim N(\beta_0, \Sigma)$. Band 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

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

such that the set $\{\beta : g_l(x, \hat{\beta}) \leq g(x, \beta) \leq g_u(x, \hat{\beta}) \text{ for all } x \in \mathcal{X}\}$ 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 $\alpha$ 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 $x \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ and $|p_k(x)| < \infty$ for all $x \in \mathcal{X}$ and $k \in \{1, \ldots, K\}$.

**Assumption 2.** $\hat{\beta} \sim N(\beta_0, \Sigma)$ and $\Sigma$ 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 $\beta_0$ 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 $\Sigma$ 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.

[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 below.

**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*

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

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

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

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

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

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

The first part of the lemma shows that a confidence band obtained by a projection on a confidence set for $\beta_0$ 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 $x \in \mathcal{X}$ $g_l(x, \hat{\beta}) = \inf_{\beta \in CI(\hat{\beta})} p(x)'\beta \leq p(x)'\beta_0$ and similarly $p(x)'\beta_0 \leq g_u(x, \hat{\beta})$. 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 yield taut bands in $\mathcal{C}$. The next theorem provides a simple characterization of the confidence sets for $\beta_0$, 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*

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

   *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* $w_l(x), w_u(x) \geq 0$ *such that*

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

*and*

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

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

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

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 $g_u^*(x,\hat\beta) \leq g_u(x,\hat\beta)$, and $g_l^*(x,\hat\beta) \geq g_l(x,\hat\beta)$ 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 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

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

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

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

yields nonconservative confidence bands. In the appendix 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

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

where $c_{2,1-\alpha}$ is the $1-\alpha$ critical value of the $\chi^2_2$ 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

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

Figure 2 illustrates the situation. It shows the ellipse $CI(\hat{\beta})$ as well as the lines of $\beta$ 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 $p(x) = (0, 1)'$, $p(x) = (1, 1)'$, and $p(x) = (1, -1)'$. 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, $g_l(\bar{x}, \hat{\beta}) > p(\bar{x})'\bar{\beta}_1$ 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 $\hat{\beta}_2$ and $\hat{\beta}_3$, such a line does not exist and hence, both of these vectors are in $CI_p(\hat{\beta})$. Therefore, the red area in Figure 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 $\{\beta \in \mathbb{R}^K : c_l(x) \leq p(x)'(\beta - \hat{\beta}) \leq c_u(x) \text{ for all } x \in \mathcal{X}\}$. 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.

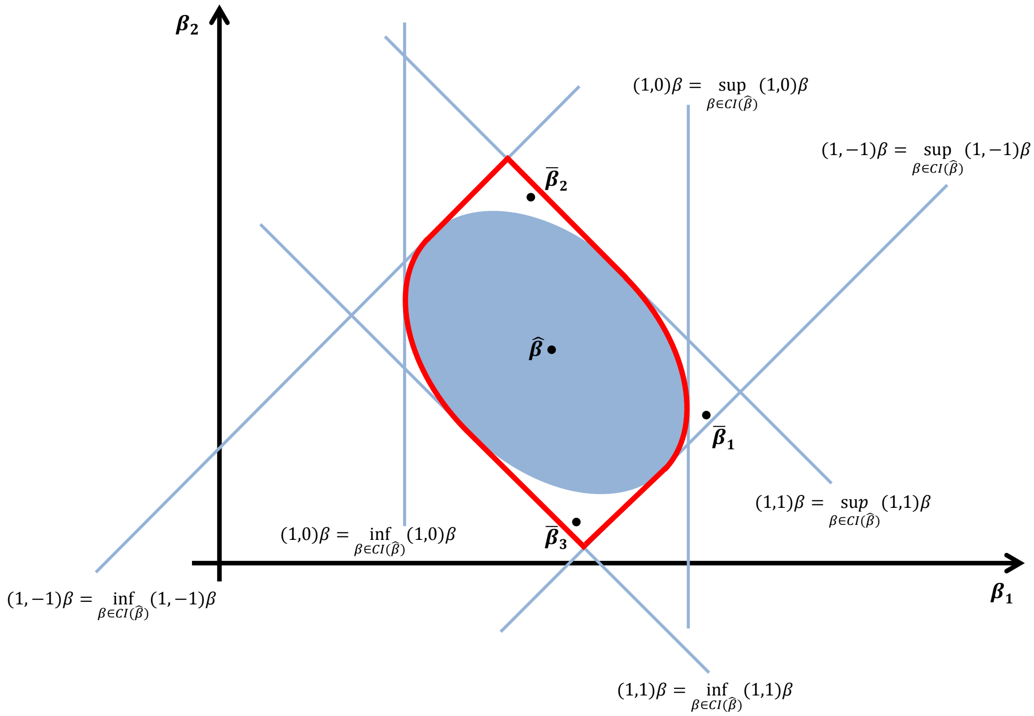
As shown above, for any taut band in $\mathcal{C}$, there exists a nonrandom set $S \subset \mathbb{R}^K$ such that

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

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

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




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* $\Omega$ *be a positive definite symmetric matrix and define* $\nu(x) = \sqrt{p(x)'\Omega p(x)}$*. Then* $\nu(x) > 0$ *for all* $x$*. Let* $w(x) = \frac{\sigma(x)}{\nu(x)}$*. Then there exists a constant* $c < \infty$ *such that*

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

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

To prove this result, we show 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 a set that satisfies the properties in Theorem 1. The next result states that equal distance 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*

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

*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 equal distance 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 $P(2|Z| \leq c) = 1 - \alpha$. The equal distance band is $x\hat{\beta} \pm c$. It then also holds by construction that

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

and hence another $1 - \alpha$ confidence band is $x\hat{\beta} \pm \frac{c}{2}|x|$, which reduces the widths of the equal distance 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:

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

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 following discussion 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

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

or, by Assumption 2,

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

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

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

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

**Assumption 4.** $h(\cdot, \cdot)$ is a nonrandom functional and $h(c_l^1(\cdot), c_u^1(\cdot)) \geq h(c_l^2(\cdot), c_u^2(\cdot))$ 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

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

The minimum area criterion of the implied confidence set for $\beta_0$, 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

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

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

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

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

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

by

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

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

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

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.

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

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

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

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 $[\bar{g}_l^J(x,\hat{\beta}),\, \bar{g}_u^J(x,\hat{\beta})]$ is a taut band in $\mathcal{C}$ by Theorem 1. The next theorem states conditions under which it is approximately optimal.

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

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

*and*

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

*Then*

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

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 Section 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. While we only have to solve a finite dimensional minimization problem to obtain an approximately optimal confidence band, calculating the constraint is computationally challenging. We discuss this issue in Section 4.

# 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 $\Sigma$. The goal is to use the asymptotic distribution and the previous results to obtain optimal and asymptotically valid uniform confidence bands for

[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 Sections A.1 and A.2 for related discussions.

$p(x)'\beta_0$. 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 $\{\beta : g_l(x,\hat{\beta},\hat{\Sigma}) \leq p(x)'\beta \leq g_u(x,\hat{\beta},\hat{\Sigma}) \text{ for all } x \in \mathcal{X}\}$ 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

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

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

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

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 $\beta_0$ of the form

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

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

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

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 $\Sigma$. To better understand it write

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

Hence, as long as $\hat{\Sigma}$ is a consistent estimator of $\Sigma$, any rule of choosing a band that depends continuously on $\Sigma$ 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

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

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.

[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.

[4] In Section 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.

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

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

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 $\Sigma$. For example, in Section A.2 in the appendix 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 $\Sigma$.

However, notice 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 $\Sigma$. As a simple 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 $\Sigma$ is a diagonal matrix. Then the constraint can be written as $P(c_l(x_1) \leq \sigma_1 Z_1 \leq c_u(x_1),\ c_l(x_2) \leq \sigma_2 Z_2 \leq c_u(x_2)) = 1-\alpha$. Now suppose that $h(c_l(x), c_u(x)) = \min\{c_u(x_1), c_u(x_2)\}$. 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

$$(\bar{c}_u(x_1), \bar{c}_u(x_2)) = (c_{1-\alpha}\sigma_1, \infty) \text{ if } \sigma_1 \leq \sigma_2 \text{ and } (\bar{c}_u(x_1), \bar{c}_u(x_2)) = (\infty, c_{1-\alpha}\sigma_2) \text{ if } \sigma_1 > \sigma_2,$$

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 function of $x$ and $\beta$ 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 $\beta$. As defined below, we only consider confidence bands based on this linear approximation of $g(x,\beta)$ .

**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 $\{\beta : g_l(x,\hat{\beta},\hat{\Sigma}) \leq \nabla_\beta g(x,\hat{\beta})'\beta \leq g_u(x,\hat{\beta},\hat{\Sigma}) \text{ for all } x \in \mathcal{X}\}$ 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 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

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

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

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

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

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

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

*Then as $n\to\infty$*

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

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 $\|\nabla_\beta g(x,\beta_1) - \nabla_\beta g(x,\beta_2)\| \leq C\|\beta_1 - \beta_2\|$ 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 $\{\beta : g_l(x, \hat{\beta}_{K_n}, \hat{\Sigma}) \leq p_{K_n}(x)'\beta \leq g_u(x, \hat{\beta}_{K_n}, \hat{\Sigma}) \ \forall x \in \mathcal{X}\}$ and 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

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

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

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

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)* $\sup_{S \in \mathcal{S}} \left| P\big(\sqrt{n}\Sigma^{-1/2}(\hat{\beta}_{K_n} - E(\hat{\beta}_{K_n})) \in S\big) - P(Z \in S) \right| \to 0$ *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. See for example Chen (2007) for sufficient conditions.

# 4 Numerical examples

In this section we demonstrate the results using 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 elasticities implied by estimated parameters of a structural model of demand. In both applications 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 2.3, we obtain an approximation of the optimal band by first solving

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

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 a projecting on

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

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

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

We approximate the probability using an expectation propagation algorithm recently proposed by Cunningham et al. (2011), which provides fast and accurate calculations of Gaussian probabilities over convex polyhedral regions.[5] Relative to simulating the probability,

[5] See also Mandt et al. (2016) for an application of this algorithm to a correlated probit model.

an advantage of this algorithm is that the approximation is a smooth function of the vectors $(c_l(x_1), \ldots, c_l(x_J))$ and $(c_u(x_1), \ldots, c_u(x_J))$. We further improve the accuracy of the approximation by splitting up the integration region into subregions of the form

$$\{z \in \mathbb{R}^K : c_l(x_j) \leq p(x_j)'\Sigma^{1/2} z \leq c_u(x_j) \text{ for all } j = 1, \ldots, J,\ z_1 \geq 0,\ z_2 \leq 0,\ \ldots,\ z_K \leq 0\}.$$

In our experience, the algorithm is fast, accurate, and stable. Finally, once we find an optimal solution $(\bar{c}_l(x_1), \ldots, \bar{c}_l(x_J))$ and $(\bar{c}_u(x_1), \ldots, \bar{c}_u(x_J))$ we eliminate the approximation error from the algorithm to guarantee the correct coverage probability. To do so, we pick the constant $b$ such that

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

and the left hand side can be calculated arbitrarily accurately using simulations.[6] Notice that in this way the resulting 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 the Scheffé band.

## 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 200 to estimate $\beta_0$ by OLS and $\Sigma$ using heteroskedasticity robust standard errors.

Figure 3 shows the estimated function and three 90% uniform confidence bands for a representative data set, 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 equal distance to the estimated function for all $x$), and the confidence band which minimizes the total area. To solve for the optimal bands we use the grid $x_j = \frac{j}{30}$ for $j = 0, \ldots, 30$. The

[6] As opposed to this one time calculation, using simulations to approximate the integral in the optimization has the disadvantages that the constraint is nonsmooth in the parameters and that it is computationally expensive with a large number of draws. Consequently, the results are numerically unstable.

Figure 3: Regression - optimal unweighted area band

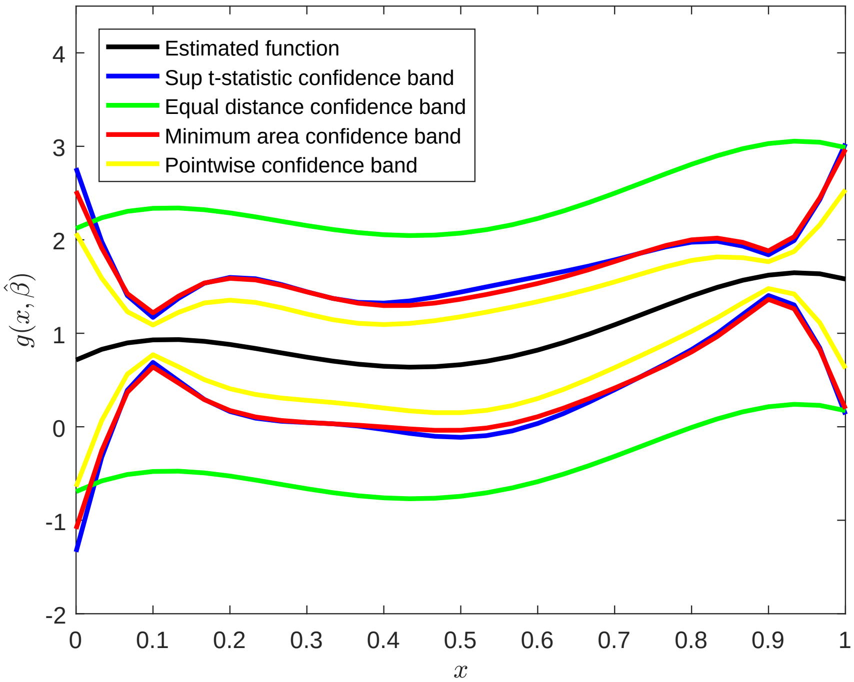



Figure 4: Regression - optimal weighted area band

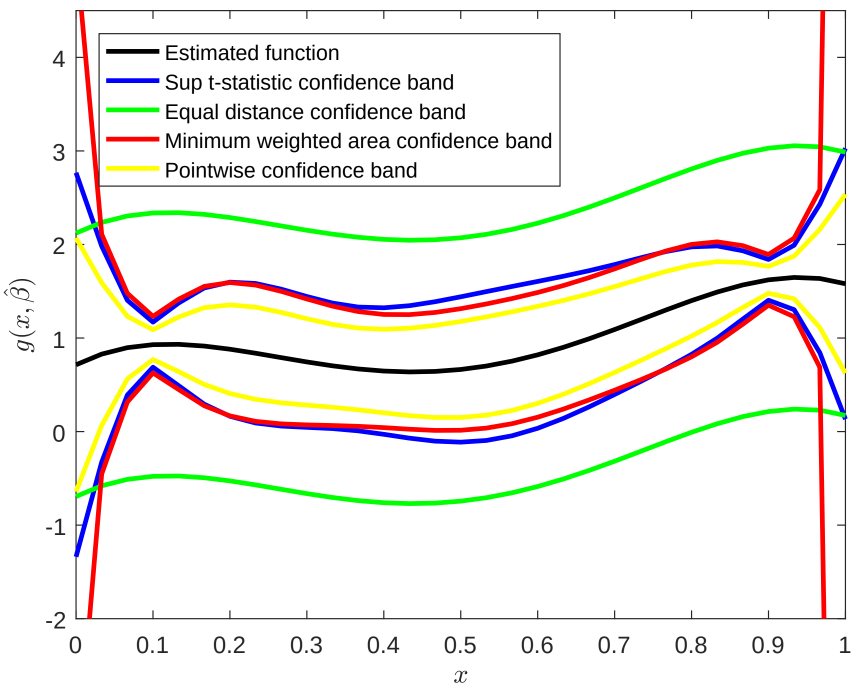


figure also shows pointwise confidence bands as a reference because no symmetric uniform confidence band can be inside the pointwise band for any $x$. Among these bands, the equal distance band is least preferred for most objective functions, unless one places extremely high priority on small values of $x$. The minimum area band looks similar to the unweighted sup t-statistic band for all $x$, but is narrower for $x$ around 0.5. The total area of this band is around 3% smaller compared to the unweighted sup t-statistic band.

Figure 4 shows the same bands, except that the red band now corresponds to the one that minimizes a weighted area where the weight function is the density of $X$. This band is narrower than the unweighted sup t-statistic band for all $x \in [0.2, 0.75]$, which covers around 75% of the data. The cost is that the band is much wider in areas with little data.

### 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 heterogenous 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

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

where $F_\eta$ is the log-normal distribution, $\xi_{jt}$ 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 $\beta_0$ can be estimated from observed market shares, prices, product characteristics, and instruments using a GMM approach.

For a fixed market $t$ let

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

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 $\beta_0$ 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 to 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 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).

Figure 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. 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 equal distance band is narrower than the sup t-statistic band for the majority of the observed prices. Moreover, the minimum area band is narrower than the sup t-statistic band for more than 90% of prices.

Figure 6 shows the same bands, except that the red 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 the pointwise band around the median price of \$7.8.

# 5 Conclusion

In this paper we provide a method for constructing optimal confidence bands for a general class of functions, which includes nonlinear and nonparametric functions. To obtain these results we show that any taut $1-\alpha$ confidence band can be obtained using a projection or inversion of a weighted t-statistic. We then characterize the class of confidence sets for $\beta_0$ 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, 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

Figure 5: Elasticities - optimal unweighted area band

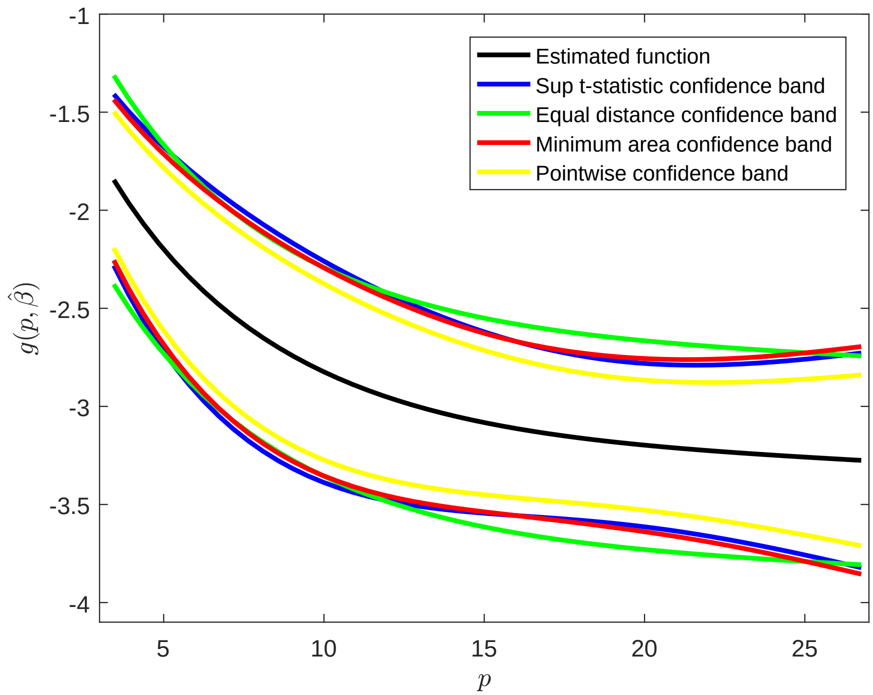
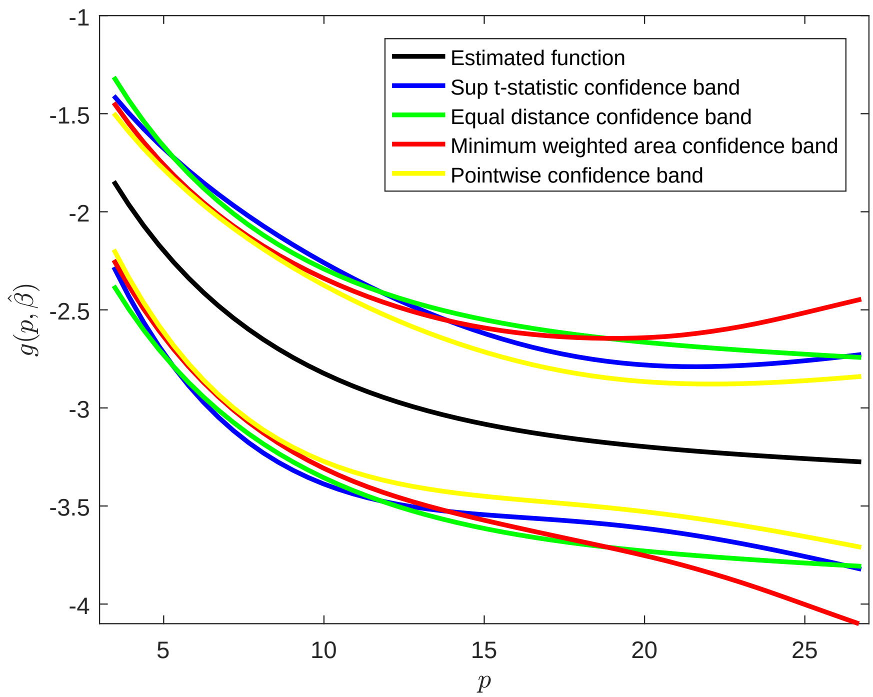



Figure 6: Elasticities - optimal weighted area band



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 $h(c_l(\cdot), c_u(\cdot)) = \max_{x\in\mathcal{X}} P(c_l(x) \leq p(x)'\Sigma^{1/2}Z \leq c_u(x))$. 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 band with equal distance to the estimated function 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)).

# 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:

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

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

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

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 asymptotically valid. As in Section 2.4, let $\{\bar{c}_l(x_j), \bar{c}_u(x_j)\}_{j=1}^J$ be an optimal solution to

$$\begin{aligned} \min_{c_l(x_j),\, c_u(x_j),\, j=1,\ldots,J} \quad & \sum_{j=1}^{J-1} (c_u(x_j) - c_l(x_j))\, w_X(x_j)(x_{j+1} - x_j) \\ \text{s.t.} \quad & P(c_u(x_j) \leq p(x_j)'\Sigma^{1/2}Z \leq c_l(x_j) \text{ for all } j = 1, \ldots, J) = 1 - \alpha. \end{aligned}$$

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

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

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

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

and

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

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 $\int_{\mathcal{X}} |\tilde{p}(x)'\gamma| w_X(x)dx \geq \varepsilon$.

It is easy to show that a simply sufficient condition for $\int_{\mathcal{X}} |\tilde{p}(x)'\gamma| w_X(x)dx \geq \varepsilon$ 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 $p(x) = (b_1 x, \ldots, b_K x)'$ for constants $b_1, \ldots, b_K$. Hence, $\int_{\mathcal{X}} p(x)p(x)'w_X(x)dx$ has rank 1, but with $B = (b_1, \ldots, b_K)'$ 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 $\sum_{j=1}^{J-1} |\tilde{p}(x_j)'\gamma| w_X(x_j)(x_{j+1} - x_j) \geq \varepsilon$, 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 $\tilde{p}(x_1), \ldots, \tilde{p}(x_J)$ 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 C.2.

**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

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

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

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

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 C.2.

**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*

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

# B Useful lemmas

In the lemmas below we use the following notation. For $S \subset \mathbb{R}^K$, let $conv(\cdot)$ denote the convex hull of $S$. For $S_1 \subset \mathbb{R}^K$ and $S_2 \subset \mathbb{R}^K$ and $\lambda \in (0,1)$ let $\lambda S_1 \oplus (1-\lambda)S_2 = \{s \in R^K : s = \lambda s_1 + (1-\lambda)s_2 \text{ for some } s_1 \in S_1 \text{ and } s_2 \in S_2\}$.

**Lemma A3.** *Let $CI_1 = \{\beta : \beta - \hat{\beta} \in S_1\}$ and $CI_2 = \{\beta : \beta - \hat{\beta} \in S_2\}$, where $S_1$ is closed, $S_2$ is convex, and both sets are nonrandom. Suppose $CI_1 \subseteq CI_2$ and that*

$$P(\beta_0 \in CI_1) = P(\beta_0 \in CI_2) = 1 - \alpha$$

*for some $\alpha \in (0,1)$. Then $CI_1 = CI_2$.*

*Proof.* Suppose there is $\bar{s} \in S_2$, but $\bar{s} \notin S_1$. Since $S_1$ is closed, it follows that there exits $\delta > 0$ such that $\inf_{s \in S_1} \|\bar{s} - s\| \geq \delta$. Let $S_3 = conv(S_1 \cup \bar{s}) \subseteq S_2$. Since $S_1$ has positive Lebesgue measure, also $S_3$ has positive Lebesgue measure, and since it is convex, it contains a compact ball $B$ of positive measure (see for example Corollary 2.4.9 and Proposition 4.10.11 of Bogachev (1998)). Now for any $\gamma > 0$ define $S_\gamma = \{s : (1-\alpha)\bar{s} + \gamma s_B \text{ for some } s_B \in B\}$. Notice that $S_\gamma \subset S_3$, but for $\gamma$ small enough (but positive), $S_\gamma \cap S_1 = \emptyset$ because $\inf_{s \in S_1} \|\bar{s} - s\| \geq \delta$. It follows that $S_3 \backslash S_1$ has positive Lebesgue measure as well and therefore

$$1 - \alpha = P(\beta_0 - \hat{\beta} \in S_2) \geq P(\beta_0 - \hat{\beta} \in S_3) > P(\beta - \hat{\beta} \in S_1) = 1 - \alpha,$$

which is a contradiction. □

**Lemma A4.** *Let $Z \sim N(0, I_{K_n \times K_n})$, $\Sigma \in \mathbb{R}^{K_n \times K_n}$ be positive definite, $p(x) \in \mathbb{R}^{K_n}$, and $\sigma(x) = \sqrt{p(x)'\Sigma p(x)} \in (0,\infty)$ for all $x \in \mathcal{X}$. Let $c_u(x)$ and $c_l(x)$ be functions such that*

$$P\left(c_l(x) \leq \frac{p(x)'\Sigma^{1/2}Z}{\sigma(x)} \leq c_u(x)\,\forall x \in \mathcal{X}\right) = 1 - \alpha,$$

*where $\alpha \in (0, 1/2)$. Let $\varepsilon_n \to 0$ such that $K_n\varepsilon_n \to 0$. Then*

$$P\left(c_l(x) - \varepsilon_n \leq \frac{p(x)'\Sigma^{1/2}Z}{\sigma(x)} \leq c_u(x) + \varepsilon_n\,\forall x \in \mathcal{X}\right) \to 1 - \alpha.$$

*Proof.* Note that $c_l(x) < -\Phi^{-1}(1-\alpha)$ and $c_u(x) > \Phi^{-1}(1-\alpha)$, where $\Phi$ denotes the standard norm cdf. Let

$$\gamma_n^1 = \frac{\Phi^{-1}(1-\alpha) + |\varepsilon_n|}{\Phi^{-1}(1-\alpha)} \qquad \text{and} \qquad \gamma_n^2 = \frac{\Phi^{-1}(1-\alpha) - |\varepsilon_n|}{\Phi^{-1}(1-\alpha)}.$$

Since $\Phi^{-1}(1-\alpha) > 0$, $\gamma_n^1$ and $\gamma_n^2$ are well defined and positive for $\varepsilon_n$ close enough to 0. Define

$$S_1 = \left\{ z \in \mathbb{R}^{K_n} : c_l(x) \leq \frac{p(x)'\Sigma^{1/2}z}{\sigma(x)} \leq c_u(x) \, \forall x \in \mathcal{X} \right\},$$

$$S_{2,n} = \left\{ z \in \mathbb{R}^{K_n} : c_l(x) - \varepsilon_n \leq \frac{p(x)'\Sigma^{1/2}z}{\sigma(x)} \leq c_u(x) + \varepsilon_n \, \forall x \in \mathcal{X} \right\}$$

and

$$S_1^\gamma = \left\{ z \in \mathbb{R}^{K_n} : \gamma c_l(x) \leq \frac{p(x)'\Sigma^{1/2}z}{\sigma(x)} \leq \gamma c_u(x) \, \forall x \in \mathcal{X} \right\}$$

for $\gamma > 0$. By definition,

$$P(Z \in S_1^{\gamma_n^2}) \leq P(Z \in S_{2,n}) \leq P(Z \in S_1^{\gamma_n^1}) \tag{2}$$

holds. By a change of variables,

$$\begin{aligned} P\left(Z \in S_1^\gamma\right) &= \int \mathbf{1}\left(z \in S_1^\gamma\right) \phi(z) dz \\ &= \gamma^{K_n} \int \mathbf{1}\left(\gamma z \in S_1^\gamma\right) \phi(\gamma z) dz \\ &= \gamma^{K_n} \int \mathbf{1}\left(z \in S_1\right) \phi(\gamma z) dz. \end{aligned}$$

Since $\phi(\gamma z) < \phi(z)$ for $\gamma > 1$ and $\phi(\gamma z) > \phi(z)$ for $\gamma < 1$,

$$0 \leq P\left(Z \in S_1^{\gamma_n^1}\right) - P\left(Z \in S_1\right) \leq P\left(Z \in S_1\right)\left((\gamma_n^1)^{K_n} - 1\right)$$

and

$$0 \geq P\left(Z \in S_1^{\gamma_n^2}\right) - P\left(Z \in S_1\right) \geq P\left(Z \in S_1\right)\left((\gamma_n^2)^{K_n} - 1\right).$$

Note that

$$(\gamma_n^1)^{K_n} = \left(1 + \frac{|\varepsilon_n|}{\Phi^{-1}(1-\alpha)}\right)^{K_n} = \left(1 + \frac{K_n|\epsilon_n|}{K_n\Phi^{-1}(1-\alpha)}\right)^{K_n} = \exp\left(\frac{K_n|\varepsilon_n|}{\Phi^{-1}(1-\alpha)}\right) + o(1).$$

Thus, if $K_n|\varepsilon_n| \to 0$, then

$$P(Z \in S_1^{\gamma_n^1}) \to P(Z \in S_1) \quad \text{and} \quad P(Z \in S_1^{\gamma_n^2}) \to P(Z \in S_1).$$

Together with (2) we can conclude $P(Z \in S_{2,n}) \to 1 - \alpha$. $\square$

**Lemma A5.** *Let $\lambda \in (0,1)$. Let $\nu$ denote the Lebesgue measure on $\mathbb{R}^K$, let $\Phi$ denote the $K$-dimensional standard normal measure, and let $\phi$ denote the standard normal pdf. Let $S_1$ and $S_2$ be bounded sets on $\mathbb{R}^K$. Suppose there is a set $S_3 \subseteq \lambda S_1 \oplus (1-\lambda)S_2$ with $\nu(S_3) > 0$ such that for all $s_3 \in S_3$ and some $\delta > 0$*

$$\inf_{(s_1,s_2)\in S_1\times S_2:\lambda s_1+(1-\lambda)s_2=s_3} \|s_1 - s_2\|^2 > \delta.$$

*Then*

$$\Phi(\lambda S_1 \oplus (1-\lambda)S_2) \geq \Phi(S_1)^\lambda \Phi(S_2)^{1-\lambda} + \inf_{s\in S_1\cup S_2} \phi(s)\left(\exp\left(\frac{\lambda(1-\lambda)\delta}{2}\right) - 1\right)\nu(S_3).$$

*Proof.* From simple algebra it follows that

$$\ln\phi(\lambda s_1 + (1-\lambda)s_2) - (\lambda\ln\phi(s_1) + (1-\lambda)\ln\phi(s_2)) = \frac{\lambda(1-\lambda)}{2}\|s_1 - s_2\|^2$$

and thus

$$\phi(\lambda s_1 + (1-\lambda)s_2) = \phi(s_1)^\lambda\phi(s_2)^{1-\lambda}\exp\left(\frac{\lambda(1-\lambda)}{2}\|s_1 - s_2\|^2\right).$$

Now define

$$h(x) = \sup_{y\in\mathbb{R}^K} \phi\left(\frac{x-y}{\lambda}\right)^\lambda \mathbf{1}\left(\frac{x-y}{\lambda} \in S_1\right)\phi\left(\frac{y}{1-\lambda}\right)^{1-\lambda}\mathbf{1}\left(\frac{y}{1-\lambda} \in S_2\right).$$

Notice that

$$x = \lambda\frac{x-y}{\lambda} + (1-\lambda)\frac{y}{1-\lambda}$$

and therefore, log-concavity of $\phi(\cdot)$ implies that for all $y \in \mathbb{R}^K$

$$\phi(x) \geq \phi\left(\frac{x-y}{\lambda}\right)^\lambda \phi\left(\frac{y}{1-\lambda}\right)^{1-\lambda}.$$

Moreover, $\mathbf{1}(x \in S_1 \oplus (1-\lambda)S_2) \geq \mathbf{1}\left(\frac{x-y}{\lambda} \in S_1\right)\mathbf{1}\left(\frac{y}{1-\lambda} \in S_2\right)$ and thus for all $x \in \mathbb{R}^K$

$$\phi(x)\mathbf{1}(x \in S_1 \oplus (1-\lambda)S_2) \geq h(x).$$

Similarly, if $x \in S_3$, then for any $y \in \mathbb{R}^K$

$$\phi(x) \geq \phi\left(\frac{x-y}{\lambda}\right)^\lambda \phi\left(\frac{y}{1-\lambda}\right)^{1-\lambda}\exp\left(\frac{\lambda(1-\lambda)}{2}\left\|\frac{x-y}{\lambda} - \frac{y}{1-\lambda}\right\|^2\right).$$

If $\frac{x-y}{\lambda} \in S_1$ and $\frac{y}{1-\lambda} \in S_2$, then $\left\|\frac{x-y}{\lambda} - \frac{y}{1-\lambda}\right\|^2 \geq \delta$ and $x \in \lambda S_1 \oplus (1-\lambda)S_2$. Therefore, for all $x \in S_3$

$$\phi(x) \geq h(x)\exp\left(\frac{\lambda(1-\lambda)\delta}{2}\right)$$

Moreover, take any $x \in \lambda S_1 \oplus (1-\lambda)S_2$. Then there exist $s_1 \in S_1$ and $s_2 \in S_2$ such that $x = \lambda s_1 + (1-\lambda)s_2$. Let $y = (1-\lambda)s_2$. Then $s_1 = \frac{x-y}{\lambda} \in S_1$. Thus

$$h(x) \geq \phi(s_1)^{\lambda} \mathbf{1}(s_1 \in S_1) \phi(s_2)^{1-\lambda} \mathbf{1}(s_2 \in S_2) \geq \inf_{s \in S_1 \cup S_2} \phi(s) > 0.$$

We now get

$$\begin{aligned}
\phi(x)\mathbf{1}(x \in \lambda S_1 \oplus (1-\lambda)S_2) &= \phi(x)\mathbf{1}(x \in \lambda S_1 \oplus (1-\lambda)S_2)\mathbf{1}(x \notin S_3) + \phi(x)\mathbf{1}(x \in S_3) \\
&\geq h(x)\mathbf{1}(x \notin S_3) + h(x)\exp\left(\frac{\lambda(1-\lambda)\delta}{2}\right)\mathbf{1}(x \in S_3) \\
&= h(x) + h(x)\left(\exp\left(\frac{\lambda(1-\lambda)\delta}{2}\right) - 1\right)\mathbf{1}(x \in S_3) \\
&\geq h(x) + \inf_{s \in S_1 \cup S_2} \phi(s)\left(\exp\left(\frac{\lambda(1-\lambda)\delta}{2}\right) - 1\right)\mathbf{1}(x \in S_3).
\end{aligned}$$

By Theorem 1.8.2 in Bogachev (1998),

$$\int h(x)dx \geq \Phi(S_1)^{\lambda}\Phi(S_2)^{1-\lambda},$$

Taking the integral of the inequality above yields

$$\Phi(\lambda S_1 \oplus (1-\lambda)S_2) \geq \Phi(S_1)^{\lambda}\Phi(S_2)^{1-\lambda} + \inf_{s \in S_1 \cup S_2} \phi(s)\left(\exp\left(\frac{\lambda(1-\lambda)\delta}{2}\right) - 1\right)\nu(S_3).$$

□

**Lemma A6.** *Let $\lambda \in (0,1)$. Let $S_1$ and $S_2$ be convex and compact subsets of $\mathbb{R}^K$ such that for some $C > 0$, $\underline{S} \equiv \{s : \|s\| \leq C\} \subset S_1$ and $\underline{S} \subset S_2$. Let $c_1(x) = \sup_{\gamma \in S_1} p(x)'\gamma$ and $c_2(x) = \sup_{\gamma \in S_2} p(x)'\gamma$, where $\sup_{x \in \mathcal{X}} \|p(x)\| < \infty$. If $\sup_{x \in \mathcal{X}} |c_1(x) - c_2(x)| > \varepsilon$ for some $\varepsilon > 0$, then there exists a set $S_3 \subseteq \lambda S_1 \oplus (1-\lambda)S_2$ and a $\delta > 0$ such that for all $s_3 \in S_3$*

$$\inf_{(s_1,s_2) \in S_1 \times S_2 : \lambda s_1 + (1-\lambda)s_2 = s_3} \|s_1 - s_2\| > \delta$$

*and $\nu(S_3) > \delta$, where $\delta$ only depends on $\lambda$, $\varepsilon$, $\sup_{x \in \mathcal{X}} \|p(x)\|$, $C$, and $\sup_{\gamma \in S_1 \cup S_2} \|\gamma\|$.*

*Proof.* Without loss of generality assume that for some $\bar{x} \in \mathcal{X}$ it holds that $c_1(\bar{x}) - c_2(\bar{x}) > \varepsilon$. Also notice that $c_1(\bar{x}) = p(\bar{x})'\gamma_1$ for $\gamma_1 \in S_1$ and $c_2(\bar{x}) = p(\bar{x})'\gamma_2$ for $\gamma_2 \in S_2$. Let $c_3 = \lambda c_1(\bar{x}) + (1-\lambda)c_2(\bar{x})$. Let $c$ be the midpoint of $(c_2(\bar{x}), c_3)$, which is

$$c = c_2(\bar{x}) + \frac{1}{2}\lambda(c_1(\bar{x}) - c_2(\bar{x})),$$

and define

$$H_1 = \{\gamma \in \mathbb{R}^K : p(\bar{x})'\gamma \geq c\} \qquad \text{and} \qquad H_2 = \{\gamma \in \mathbb{R}^K : p(\bar{x})'\gamma \leq c_2(\bar{x})\}.$$

Now define $S_3 = \lambda S_1 \oplus (1-\lambda)S_2 \cap H_1$. Notice that $\lambda\gamma_1 + (1-\lambda)\gamma_2 \in H_1$ and thus, $S_3$ is not empty. We will now prove that $S_3$ satisfies the properties in the lemma.

Let $A = conv(\underline{S} \cup \lambda\gamma_1 + (1-\lambda)\gamma_2)$. Since $\underline{S} \subseteq \lambda S_1 \oplus (1-\lambda)S_2$, $\lambda\gamma_1 + (1-\lambda)\gamma_2 \in \lambda S_1 \oplus (1-\lambda)S_2$, and $\lambda S_1 \oplus (1-\lambda)S_2$ is convex, it follows that $A \subseteq \lambda S_1 \oplus (1-\lambda)S_2$.

Let $\underline{c} = \inf_{\gamma\in\underline{S}} p(\bar{x})'\gamma$. Then for $\bar{\lambda} = \frac{c-\underline{c}}{c_3-\underline{c}}$ and $\gamma \in (\bar{\lambda}(\lambda\gamma_1 + (1-\lambda)\gamma_2) \oplus (1-\bar{\lambda})\underline{S})$ we have $p(\bar{x})'\gamma \geq c$ and thus,

$$(\bar{\lambda}(\lambda\gamma_1 + (1-\lambda)\gamma_2) \oplus (1-\bar{\lambda})\underline{S}) \subseteq H_1.$$

Moreover,

$$(\bar{\lambda}(\lambda\gamma_1 + (1-\lambda)\gamma_2) \oplus (1-\bar{\lambda})\underline{S}) \subseteq A \subseteq \lambda S_1 \oplus (1-\lambda)S_2.$$

Together this means that

$$(\bar{\lambda}(\lambda\gamma_1 + (1-\lambda)\gamma_2) \oplus (1-\bar{\lambda})\underline{S}) \subseteq S_3.$$

The set $(\bar{\lambda}(\lambda\gamma_1 + (1-\lambda)\gamma_2) \oplus (1-\bar{\lambda})\underline{S})$ is a ball with center at $\bar{\lambda}(\lambda\gamma_1 + (1-\lambda)\gamma_2)$ and radius

$$(1-\bar{\lambda})C = C\frac{c_3 - c}{c_3 - \underline{c}} \geq \frac{C}{2}\frac{\lambda\varepsilon}{c_3 - \underline{c}} > 0.$$

Let $\bar{c} = \sup_{\gamma\in S_1\cup S_2} p(\bar{x})'\gamma$. Then the radius of the ball, which is contained in $S_3$, is at least $\frac{C}{2}\frac{\lambda\varepsilon}{\bar{c}-\underline{c}} > 0$ and therefore, $\nu(S_3) > 0$.

Next, define $D = (S_1 \cap H_1)$ which is nonempty. Also notice that $S_2 \subseteq H_2$. Therefore,

$$\inf_{(s_1,s_2)\in D\times S_2} \|s_1 - s_2\| \;\; \geq \;\; \inf_{(s_1,s_2)\in H_1\times H_2} \|s_1 - s_2\| \geq \frac{c - c_2(\bar{x})}{\|p(\bar{x})\|} \geq \frac{\lambda\varepsilon}{2\sup_{x\in\mathcal{X}}\|p(x)\|}$$

Now take $s_3 \in S_3$. Then $p(x)'s_3 \geq c$. Write $s_3 = \lambda s_1 + (1-\lambda)s_2$ for $s_1 \in S_1$ and $s_2 \in S_2$. Suppose that $s_1 \notin D$. Then $s_1 \notin H_1$, which implies that $p(\bar{x})'s_1 < c$. But since it also holds that $s_2 \in S_2$, we have $p(\bar{x})'s_3 < c$, which would yield the contradiction that $p(x)'s_2 < c$. It follows that for all $s_3 \in S_3$

$$\inf_{(s_1,s_2)\in S_1\times S_2:\lambda s_1+(1-\lambda)s_2=s_3} \|s_1 - s_2\| \geq \inf_{(s_1,s_2)\in D\times S_2} \|s_1 - s_2\| \geq \frac{\lambda\varepsilon}{2\sup_{x\in\mathcal{X}}\|p(x)\|} > 0.$$

The now conclusion follows with $\delta = \min\left\{\frac{\lambda\varepsilon}{2\sup_{x\in\mathcal{X}}\|p(x)\|}, \nu((1-\bar{\lambda})\underline{S})\right\}$. $\square$

# C Proofs

## C.1 Proofs of main results

*Proof of Lemma 1.* (1) Suppose $\beta \in CI(\hat{\beta})$. Then for all $x \in \mathcal{X}$ it holds by definition that $g_l(x, \hat{\beta}) \leq p(x)'\beta$ and $p(x)'\beta \leq g_u(x, \hat{\beta})$. Therefore

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

(2) Let $\mathcal{X}_u = \{x \in \mathcal{X} : w_u(x) > 0\}$ and $\mathcal{X}_l = \{x \in \mathcal{X} : w_l(x) > 0\}$. Then

$$\begin{aligned}
1 - \alpha &= P\left(\sup_{x \in \mathcal{X}} \frac{p(x)'(\beta_0 - \hat{\beta})}{\sigma(x)} w_u(x) \leq c,\ \inf_{x \in \mathcal{X}} \frac{p(x)'(\beta_0 - \hat{\beta})}{\sigma(x)} w_l(x) \geq -c\right) \\
&= P\left(\sup_{x \in \mathcal{X}_u} \frac{p(x)'(\beta_0 - \hat{\beta})}{\sigma(x)} w_u(x) \leq c,\ \inf_{x \in \mathcal{X}_l} \frac{p(x)'(\beta_0 - \hat{\beta})}{\sigma(x)} w_l(x) \geq -c\right) \\
&= P\left(p(x)'\beta_0 \leq g_u(x, \hat{\beta})\ \forall x \in \mathcal{X}_u,\ p(x)'\beta_0 \geq g_l(x, \hat{\beta})\ \forall x \in \mathcal{X}_l\right) \\
&= P\left(g_l(x, \hat{\beta}) \leq p(x)'\beta_0 \leq g_u(x, \hat{\beta}) \text{ for all } x \in \mathcal{X}\right)
\end{aligned}$$

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

*Proof of Theorem 1.* (1): Let $[g_l(x, \hat{\beta}),\ g_u(x, \hat{\beta})] \in \mathcal{C}$ be an arbitrary taut confidence band and define

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

and

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

By Lemma 1, $P(g_l^*(x, \hat{\beta}) \leq g_0(x) \leq g_u^*(x, \hat{\beta})\ \forall x \in \mathcal{X}) \geq 1 - \alpha$. Moreover, $g_u^*(x, \hat{\beta}) \leq g_u(x, \hat{\beta})$, and $g_l^*(x, \hat{\beta}) \geq g_l(x, \hat{\beta})$ for all $x$. Hence, if $[g_l(x, \hat{\beta}),\ g_u(x, \hat{\beta})]$ is taut, it holds that $[g_l(x, \hat{\beta}),\ g_u(x, \hat{\beta})] = [g_l^*(x, \hat{\beta}),\ g_u^*(x, \hat{\beta})]$. Finally, by definition of the class $\mathcal{C}$, there exists a nonrandom set $S \subset \mathbb{R}^K$ such that

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

and similarly

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

It now immediately follows that the taut band $[g_l(x,\hat\beta),\, g_u(x,\hat\beta)] \in \mathcal{C}$ can be obtained by a projection on

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

Now suppose that $[g_l(x,\hat\beta),\, g_u(x,\hat\beta)]$ is a confidence band obtained by a projection on

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

The arguments above imply that there exists a nonrandom set $S$ such that

$$\begin{aligned}[g_l(x,\hat\beta),\, g_u(x,\hat\beta)] &= [p(x)'\hat\beta + \inf_{\gamma\in S} p(x)'\gamma,\, p(x)'\hat\beta + \sup_{\gamma\in S} p(x)'\gamma] \\ &\subseteq [p(x)'\hat\beta + c_l(x),\, p(x)'\hat\beta + c_u(x)]\end{aligned}$$

Since $[g_l(x,\hat\beta),\, g_u(x,\hat\beta)]$ is obtained by a projection on $CI(\hat\beta)$ it also has to hold that

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

and thus

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

It follows that $[g_l(x,\hat\beta),\, g_u(x,\hat\beta)] \in \mathcal{C}$. If $[g_l(x,\hat\beta),\, g_u(x,\hat\beta)]$ was not taut, then there exists $[\tilde g_l(x,\hat\beta),\, \tilde g_u(x,\hat\beta)]$ such that $\widetilde{CI}(\hat\beta) \subseteq CI(\hat\beta)$, where

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

and $P(\beta_0 \in \widetilde{CI}(\hat\beta)) = 1-\alpha$. By Lemma A3, $\widetilde{CI}(\hat\beta) = CI(\hat\beta)$. But since $[g_l(x,\hat\beta),\, g_u(x,\hat\beta)]$ is obtained by a projection on $CI(\hat\beta)$ it holds that $[g_l(x,\hat\beta),\, g_u(x,\hat\beta)] \subseteq [\tilde g_l(x,\hat\beta),\, \tilde g_u(x,\hat\beta)]$ for all $x \in \mathcal{X}$, which is a contradiction.

(2): By the proof of the first part, we can write

$$[g_l(x,\hat\beta),\, g_u(x,\hat\beta)] = \left[p(x)'\hat\beta + \inf_{\gamma\in S} p(x)'\gamma,\, p(x)'\hat\beta + \sup_{\gamma\in S} p(x)'\gamma\right].$$

Let $w_l(x) = -\frac{\sigma(x)}{\inf_{\gamma\in S} p(x)'\gamma}$ if $\inf_{\gamma\in S} p(x)'\gamma > -\infty$ and $w_l(x) = 0$ if $\inf_{\gamma\in S} p(x)'\gamma = -\infty$. Similarly, let $w_u(x) = \frac{\sigma(x)}{\sup_{\gamma\in S} p(x)'\gamma}$ if $\sup_{\gamma\in S} p(x)'\gamma < \infty$ and $w_u(x) = 0$ if $\sup_{\gamma\in S} p(x)'\gamma = \infty$. Then

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

and

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

Moreover, let $\mathcal{X}_u = \{x \in \mathcal{X} : w_u(x) > 0\}$ and $\mathcal{X}_l = \{x \in \mathcal{X} : w_u(x) > 0\}$. Then

$$\begin{aligned}
1-\alpha &= P\left(p(x)'\hat\beta + \inf_{\gamma\in S} p(x)'\gamma \le p(x)'\beta_0 \le p(x)'\hat\beta + \sup_{\gamma\in S} p(x)'\gamma \text{ for all } x \in \mathcal{X}\right) \\
&= P\left(p(x)'\beta_0 \le p(x)'\hat\beta + \inf_{\gamma\in S} p(x)'\gamma \ \forall x \in \mathcal{X}_u, p(x)'\beta_0 \ge p(x)'\hat\beta + \sup_{\gamma\in S} p(x)'\gamma \ \forall x \in \mathcal{X}_l\right) \\
&= P\left(p(x)'\beta_0 \le p(x)'\hat\beta + \frac{\sigma(x)}{w_u(x)} \ \forall x \in \mathcal{X}_u, p(x)'\beta_0 \ge p(x)'\hat\beta - \frac{\sigma(x)}{w_l(x)} \ \forall x \in \mathcal{X}_l\right) \\
&= P\left(\sup_{x\in\mathcal{X}_u} \frac{p(x)'(\beta_0-\hat\beta)}{\sigma(x)} w_u(x) \le 1, \inf_{x\in\mathcal{X}_l} \frac{p(x)'(\beta_0-\hat\beta)}{\sigma(x)} w_l(x) \ge -1\right) \\
&= P\left(\sup_{x\in\mathcal{X}} \frac{p(x)'(\beta_0-\hat\beta)}{\sigma(x)} w_u(x) \le 1, \inf_{x\in\mathcal{X}} \frac{p(x)'(\beta_0-\hat\beta)}{\sigma(x)} w_l(x) \ge -1\right).
\end{aligned}$$

□

*Proof of Corollary 1.* First notice that since $\sigma(x) = \sqrt{p(x)'\Sigma p(x)} > 0$ for all $x \in \mathcal{X}$ and since $\Sigma$ is positive definite, it follows that $p(x)$ is not the zero vector and $\nu(x) = \sqrt{p(x)'\Omega p(x)} > 0$. Next, for any $c_1 > 0$ define $CI_1(\hat\beta, c_1) = \{\beta : (\beta-\hat\beta)\Omega^{-1}(\beta-\hat\beta) \le c_1^2\}$. Then

$$\sup_{\beta\in CI_1(\hat\beta,c_1)} p(x)'\beta = p(x)'\hat\beta + \max_{\gamma:\gamma'\Omega^{-1}\gamma\le c_1^2} p(x)'\gamma.$$

The Lagrangian of the maximization problem is $p(x)'\gamma + \lambda(c_1^2 - \gamma'\Omega^{-1}\gamma)$ with (necessary and sufficient) first order conditions $p(x) = 2\lambda\Omega^{-1}\gamma$ and $\lambda(c_1^2 - \gamma'\Omega^{-1}\gamma) = 0$. Since $p(x) \ne 0$ it follows that $\gamma \ne 0$ and $\lambda \ne 0$ and therefore it is easy to solve for

$$\lambda = \frac{1}{2c_1}\sqrt{p(x)'\Omega p(x)} \qquad \text{and} \qquad \gamma = \frac{c_1 \Omega p(x)}{\sqrt{p(x)'\Omega p(x)}}.$$

Therefore,

$$\sup_{\beta\in CI_1(\hat\beta,c_1)} p(x)'\beta = p(x)'\hat\beta + c_1\sqrt{p(x)'\Omega p(x)}.$$

Analogously,

$$\inf_{\beta\in CI_1(\hat\beta,c_1)} p(x)'\beta = p(x)'\hat\beta - c_1\sqrt{p(x)'\Omega p(x)}.$$

Next define

$$CI(\hat\beta, c_1) = \{\beta : -c_1\sqrt{p(x)'\Omega p(x)} \le p(x)'(\beta-\hat\beta) \le c_1\sqrt{p(x)'\Omega p(x)} \text{ for all } x \in \mathcal{X}\}.$$

Since $CI(\hat{\beta}, c_1)$ is obtained by a projection it has to hold that $CI_1(\hat{\beta}, c_1) \subseteq CI(\hat{\beta}, c_1)$. Next notice that there is a unique $c$ such that $P(\beta_0 \in CI(\hat{\beta}, c)) = 1-\alpha$, because $P(\beta_0 \in CI(\hat{\beta}, c_1))$ is strictly increasing in $c_1$. It follows that

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

where $w(x) = \frac{\sigma(x)}{\nu(x)}$. Finally, since $CI_1(\hat{\beta}, c) \subseteq CI(\hat{\beta}, c)$, we have

$$p(x)'\hat{\beta} + c\sqrt{p(x)'\Omega p(x)} = \sup_{\beta\in CI_1(\hat{\beta},c)} p(x)'\beta \leq \sup_{\beta\in CI(\hat{\beta},c)} p(x)'\beta$$

and by definition of $CI(\hat{\beta}, c)$ it holds that

$$\sup_{\beta\in CI(\hat{\beta},c)} p(x)'\beta \leq p(x)'\hat{\beta} + c\sqrt{p(x)'\Omega p(x)}.$$

Therefore, $\sup_{\beta\in CI(\hat{\beta},c)} p(x)'\beta = p(x)'\hat{\beta} + c\sqrt{p(x)'\Omega p(x)}$ and similarly $\inf_{\beta\in CI(\hat{\beta},c)} p(x)'\beta = p(x)'\hat{\beta} - c\sqrt{p(x)'\Omega p(x)}$. It follows that the confidence band obtained from the sup t-statistic, $[p(x)'\hat{\beta} - c\nu(x),\, p(x)'\hat{\beta} + c\nu(x)]$, coincides with the taut projection confidence band obtained from a projection on $CI(\hat{\beta}, c)$, and by Theorem 1, the confidence band obtained from projecting on $CI(\hat{\beta}, c)$ is taut. $\square$

*Proof of Corollary 2.* Suppose $p_1(x) = 1$ and let

$$CI(\hat{\beta}) = \{\beta : -c \leq p(x)'(\beta - \hat{\beta}) \leq c \text{ for all } x \in \mathcal{X}\}.$$

Then $P(\beta_0 \in CI(\hat{\beta})) = 1 - \alpha$. By Theorem 1 it suffices to show that

$$\sup_{\beta\in CI(\hat{\beta})} p(x)'\beta = p(x)'\hat{\beta} + c \qquad \text{and} \qquad \inf_{\beta\in CI(\hat{\beta})} p(x)'\beta = p(x)'\hat{\beta} - c.$$

By the definition of the supremum,

$$\sup_{\beta\in CI(\hat{\beta})} p(x)'\beta \leq p(x)'\hat{\beta} + c.$$

Now define $\tilde{\beta} = \hat{\beta} + \begin{pmatrix} c & 0 & \dots & 0 \end{pmatrix}'$ and notice that $\tilde{\beta} \in CI(\hat{\beta})$ because $p_1(x) = 1$. Therefore,

$$\sup_{\beta\in CI(\hat{\beta})} p(x)'\beta \geq p(x)'\tilde{\beta} = p(x)'\hat{\beta} + c.$$

Analogously, one can show that $\inf_{\beta\in CI(\hat{\beta})} p(x)'\beta = p(x)'\hat{\beta} - c$. $\square$

*Proof of Theorem 2.* First notice that since $(\bar{c}_l(x),\, \bar{c}_u(x))$ is optimal (and $(\bar{c}_l^J(x),\, \bar{c}_u^J(x))$ is feasible in the original problem)

$$h(\bar{c}_l^J(x),\, \bar{c}_u^J(x)) \geq h(\bar{c}_l(x),\, \bar{c}_u(x)).$$

Next define $\bar{C}I(\hat{\beta}) = \{\beta \in \mathbb{R}^K : \bar{c}_u(x) \leq p(x)'(\beta - \hat{\beta}) \leq \bar{c}_u(x) \text{ for all } x \in \mathcal{X}_J\}$. Moreover, let $\tilde{c}_u(x) = \sup_{\beta \in \bar{C}I(\hat{\beta})} p(x)'\beta$ and $\tilde{c}_l(x) = \inf_{\beta \in \bar{C}I(\hat{\beta})} p(x)'\beta$ for all $x \in \mathcal{X} \backslash \mathcal{X}_J$ and $\tilde{c}_u(x) = \bar{c}_u(x)$ and $\tilde{c}_l(x) = \bar{c}_l(x)$ for all $x \in \mathcal{X}_J$. It follows that

$$\begin{aligned} 1 - \alpha &= P(\bar{c}_l(x) \leq p(x)'\Sigma^{1/2} Z \leq \bar{c}_u(x) \text{ for all } x \in \mathcal{X}) \\ &\leq P(\bar{c}_l(x) \leq p(x)'\Sigma^{1/2} Z \leq \bar{c}_u(x) \text{ for all } x \in \mathcal{X}_J) \\ &= P(\tilde{c}_l(x) \leq p(x)'\Sigma^{1/2} Z \leq \tilde{c}_u(x) \text{ for all } x \in \mathcal{X}_J). \end{aligned}$$

Since $(\bar{c}_l^J(x), \bar{c}_u^J(x))$ is optimal, it follows that

$$h_J(\bar{c}_l^J(x),\, \bar{c}_u^J(x)) \leq h_J(\tilde{c}_l(x),\, \tilde{c}_u(x)) = h_J(\bar{c}_l(x),\, \bar{c}_u(x)).$$

Therefore

$$\begin{aligned} h(\bar{c}_l^J(x),\, \bar{c}_u^J(x)) &= h(\bar{c}_l^J(x),\, \bar{c}_u^J(x)) + h_J(\bar{c}_l^J(x),\, \bar{c}_u^J(x)) - h_J(\bar{c}_l^J(x),\, \bar{c}_u^J(x)) \\ &\leq h(\bar{c}_l^J(x),\, \bar{c}_u^J(x)) - h_J(\bar{c}_l^J(x),\, \bar{c}_u^J(x)) + h_J(\bar{c}_l(x),\, \bar{c}_u(x)) \\ &\to h(\bar{c}_l(x),\, \bar{c}_u(x)). \end{aligned}$$

□

*Proof of Lemma 2.* Define

$$\bar{c}_l(x, \Sigma) = \inf_{\gamma \in \mathbb{R}^K :\, \Sigma^{-1/2}\gamma \in S(\Sigma)} p(x)'\gamma \quad \text{and} \quad \bar{c}_u(x, \Sigma) = \sup_{\gamma \in \mathbb{R}^K :\, \Sigma^{-1/2}\gamma \in S(\Sigma)} p(x)'\gamma.$$

and let $\varepsilon_n \to 0$ such that

$$P\left(\sup_{x \in \mathcal{X}} \frac{|\bar{c}_l(x, \hat{\Sigma}) - \bar{c}_l(x, \Sigma)|}{\sigma(x)} \leq \varepsilon_n\right) \to 1 \quad \text{and} \quad P\left(\sup_{x \in \mathcal{X}} \frac{|\bar{c}_u(x, \hat{\Sigma}) - \bar{c}_u(x, \Sigma)|}{\sigma(x)} \leq \varepsilon_n\right) \to 1.$$

Then

$$\begin{aligned}
CR &\equiv P\left(g_l(x,\hat\beta,\hat\Sigma) \le g(x,\beta_0) \le g_u(x,\hat\beta,\hat\Sigma)\,\forall x \in \mathcal{X}\right)\\
&= P\left(\frac{\bar c_l(x,\hat\Sigma)}{\sigma(x)} \le \frac{\sqrt{n}p(x)'(\beta_0-\hat\beta)}{\sigma(x)} \le \frac{\bar c_u(x,\hat\Sigma)}{\sigma(x)}\,\forall x \in \mathcal{X}\right)\\
&\le P\left(\frac{\bar c_l(x,\Sigma)}{\sigma(x)} - \varepsilon_n \le \frac{\sqrt{n}p(x)'(\beta_0-\hat\beta)}{\sigma(x)} \le \frac{\bar c_u(x,\Sigma)}{\sigma(x)} + \varepsilon_n\,\forall x \in \mathcal{X}\right) + o(1)\\
&= P\left(\frac{\bar c_l(x,\Sigma)}{\sigma(x)} - \varepsilon_n \le \frac{p(x)'\Sigma^{1/2}\sqrt{n}\Sigma^{-1/2}(\beta_0-\hat\beta)}{\sigma(x)} \le \frac{\bar c_u(x,\Sigma)}{\sigma(x)} + \varepsilon_n\,\forall x \in \mathcal{X}\right) + o(1)\\
&= P\left(\frac{\bar c_l(x,\Sigma)}{\sigma(x)} - \varepsilon_n \le \frac{p(x)'\Sigma^{1/2}Z}{\sigma(x)} \le \frac{\bar c_u(x,\Sigma)}{\sigma(x)} + \varepsilon_n\,\forall x \in \mathcal{X}\right) + o(1)\\
&= P\left(\frac{\bar c_l(x,\Sigma)}{\sigma(x)} \le \frac{p(x)'\Sigma^{1/2}Z}{\sigma(x)} \le \frac{\bar c_u(x,\Sigma)}{\sigma(x)}\,\forall x \in \mathcal{X}\right) + o(1)\\
&= 1-\alpha+o(1).
\end{aligned}$$

The fifth line follows from the normal approximation assumption, the sixth line follows from Lemma A4, and the last line follows from the definition of the confidence band.

Analogous arguments yield $CR \ge 1-\alpha+o(1)$ and hence

$$P\left(g_l(x,\hat\beta,\hat\Sigma) \le g(x,\beta_0) \le g_u(x,\hat\beta,\hat\Sigma)\,\forall x \in \mathcal{X}\right) \to 1-\alpha.$$

□

*Proof of Theorem 3.* The result follows from the assumption and the definition of $S^*(\hat\Sigma)$. □

*Proof of Theorem 4.* Let $\varepsilon_n \to 0$ such that

$$P\left(\sup_{x\in\mathcal{X}} \frac{|\bar c_l(x,\hat\beta,\hat\Sigma) - \bar c_l(x,\beta_0,\Sigma)|}{\sigma(x)} \le \varepsilon_n\right) \to 1,$$

$$P\left(\sup_{x\in\mathcal{X}} \frac{|\bar c_u(x,\hat\beta,\hat\Sigma) - \bar c_u(x,\beta_0,\Sigma)|}{\sigma(x)} \le \varepsilon_n\right) \to 1,$$

and

$$P\left(\sup_{x\in\mathcal{X}} \sup_{\beta:\|\beta-\beta_0\|\|\le\|\hat\beta-\beta_0\|} \|\nabla_\beta g(x,\tilde\beta) - \nabla_\beta g(x,\beta_0)\|\|\sqrt{n}(\beta_0-\hat\beta)\|\|\sigma(x)^{-1}| \le \varepsilon_n\right) \to 1.$$

Then

$$\begin{aligned}
CR &\equiv P\left(\bar{g}_l(x,\hat{\beta},\hat{\Sigma}) \le g(x,\beta_0) \le \bar{g}_u(x,\hat{\beta},\hat{\Sigma})\,\forall x \in \mathcal{X}\right) \\
&= P\left(g(x,\hat{\beta}) + \bar{c}_l(x,\hat{\beta},\hat{\Sigma})/\sqrt{n} \le g(x,\beta_0) \le g(x,\hat{\beta}) + \bar{c}_u(x,\hat{\beta},\hat{\Sigma})/\sqrt{n}\,\forall x \in \mathcal{X}\right) \\
&= P\left(\frac{\bar{c}_l(x,\hat{\beta},\hat{\Sigma})}{\sigma(x)} \le \frac{\nabla_\beta g(x,\tilde{\beta})'\sqrt{n}(\beta_0-\hat{\beta})}{\sigma(x)} \le \frac{\bar{c}_u(x,\hat{\beta},\hat{\Sigma})}{\sigma(x)}\,\forall x \in \mathcal{X}\right) \\
&\le P\left(\frac{\bar{c}_l(x,\hat{\beta},\hat{\Sigma})}{\sigma(x)} - \varepsilon_n \le \frac{\nabla_\beta g(x,\beta_0)'\sqrt{n}(\beta_0-\hat{\beta})}{\sigma(x)} \le \frac{\bar{c}_u(x,\hat{\beta},\hat{\Sigma})}{\sigma(x)} + \varepsilon_n\,\forall x \in \mathcal{X}\right) + o(1) \\
&\le P\left(\frac{\bar{c}_l(x,\beta_0,\Sigma)}{\sigma(x)} - 2\varepsilon_n \le \frac{\nabla_\beta g(x,\beta_0)'\sqrt{n}(\beta_0-\hat{\beta})}{\sigma(x)} \le \frac{\bar{c}_u(x,\beta_0,\Sigma)}{\sigma(x)} + 2\varepsilon_n\,\forall x \in \mathcal{X}\right) + o(1) \\
&= P\left(\frac{\bar{c}_l(x,\beta_0,\Sigma)}{\sigma(x)} - 2\varepsilon_n \le \frac{\nabla_\beta g(x,\beta_0)'\Sigma^{1/2}Z}{\sigma(x)} \le \frac{\bar{c}_u(x,\beta_0,\Sigma)}{\sigma(x)} + 2\varepsilon_n\,\forall x \in \mathcal{X}\right) + o(1) \\
&= P\left(\frac{\bar{c}_l(x,\beta_0,\Sigma)}{\sigma(x)} \le \frac{\nabla_\beta g(x,\beta_0)'\Sigma^{1/2}Z}{\sigma(x)} \le \frac{\bar{c}_u(x,\beta_0,\Sigma)}{\sigma(x)}\,\forall x \in \mathcal{X}\right) + o(1) \\
&= 1-\alpha+o(1)
\end{aligned}$$

The second line follows from the mean value theorem, where $\tilde{\beta}$ is an intermediate value and $\|\tilde{\beta}-\beta_0\| \le \|\hat{\beta}-\beta_0\|$. For the third line notice that

$$\left|\frac{\left(\nabla_\beta g(x,\tilde{\beta})' - \nabla_\beta g(x,\beta_0)'\right)\sqrt{n}(\beta_0-\hat{\beta})}{\sigma(x)}\right| \le \|\nabla_\beta g(x,\tilde{\beta}) - \nabla_\beta g(x,\beta_0)\|\|\sqrt{n}(\beta_0-\hat{\beta})\||\sigma(x)^{-1}|$$

and hence

$$P\left(\sup_{x\in\mathcal{X}}\left|\frac{\left(\nabla_\beta g(x,\tilde{\beta})' - \nabla_\beta g(x,\beta_0)'\right)\sqrt{n}(\beta_0-\hat{\beta})}{\sigma(x)}\right| \le \varepsilon_n\right) \to 1.$$

The remaining lines follow similar steps to the ones in the proof of Lemma 2.

Analogous arguments yield $CR \ge 1-\alpha+o(1)$ and hence

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

□

*Proof of Theorem 5.* Let $\varepsilon_n \to 0$ such that $K_n\varepsilon_n \to 0$ and for $n$ large enough

$$\sup_{x\in\mathcal{X}}\frac{|p(x)'E(\hat{\beta}) - g_0(x)|}{\sigma(x)} \le \varepsilon_n$$

for $n$ large enough and

$$P\left(\sup_{x\in\mathcal{X}}\frac{|\bar{c}_l(x,\hat{\Sigma})-\bar{c}_l(x,\Sigma)|}{\sigma(x)}\leq\varepsilon_n\right)\to 1\ \text{ and }\ P\left(\sup_{x\in\mathcal{X}}\frac{|\bar{c}_u(x,\hat{\Sigma})-\bar{c}_u(x,\Sigma)|}{\sigma(x)}\leq\varepsilon_n\right)\to 1.$$

We now get, similar as in the proof of Theorem 4,

$$\begin{aligned}
CR &\equiv P\left(g_l(x,\hat{\beta}_{K_n},\hat{\Sigma})\leq g_0(x)\leq g_l(x,\hat{\beta}_{K_n},\hat{\Sigma})\,\forall x\in\mathcal{X}\right)\\
&= P\left(\frac{\bar{c}_l(x,\hat{\Sigma})}{\sigma(x)}\leq\frac{p(x)'\sqrt{n}(E(\hat{\beta}_{K_n})-\hat{\beta}_{K_n})}{\sigma(x)}+\frac{\sqrt{n}(g_0(x)-p(x)'E(\hat{\beta}_{K_n}))}{\sigma(x)}\leq\frac{\bar{c}_u(x,\hat{\Sigma})}{\sigma(x)}\,\forall x\in\mathcal{X}\right)\\
&\leq P\left(\frac{\bar{c}_l(x,\hat{\Sigma})}{\sigma(x)}-\varepsilon_n\leq\frac{p(x)'\sqrt{n}(E(\hat{\beta}_{K_n})-\hat{\beta}_{K_n})}{\sigma(x)}\leq\frac{\bar{c}_u(x,\hat{\Sigma})}{\sigma(x)}+\varepsilon_n\,\forall x\in\mathcal{X}\right)+o(1)\\
&\leq P\left(\frac{\bar{c}_l(x,\Sigma)}{\sigma(x)}-2\varepsilon_n\leq\frac{p(x)'\sqrt{n}(E(\hat{\beta}_{K_n})-\hat{\beta}_{K_n})}{\sigma(x)}\leq\frac{\bar{c}_u(x,\Sigma)}{\sigma(x)}+2\varepsilon_n\,\forall x\in\mathcal{X}\right)+o(1)\\
&= P\left(\frac{\bar{c}_l(x,\Sigma)}{\sigma(x)}-2\varepsilon_n\leq\frac{p(x)'\Sigma^{1/2}Z}{\sigma(x)}\leq\frac{\bar{c}_u(x,\Sigma)}{\sigma(x)}+2\varepsilon_n\,\forall x\in\mathcal{X}\right)+o(1)\\
&= 1-\alpha+o(1)
\end{aligned}$$

Analogous arguments yield $CR\geq 1-\alpha+o(1)$ and hence

$$P\left(\bar{g}_l(x,\hat{\beta}_{K_n},\hat{\Sigma})\leq g(x,\beta_0)\leq\bar{g}_u(x,\hat{\beta}_{K_n},\hat{\Sigma})\,\forall x\in\mathcal{X}\right)\to 1-\alpha.$$

□

## C.2 Proofs of lemmas from Section A

*Proof of Lemma A1.* Let $C_s$ be a constant and define

$$S_u=\{\gamma\in\mathbb{R}^K:|\gamma_k|\leq C_s\text{ for all }k=1,\ldots,K\},$$

$$S_J=\{\gamma\in\mathbb{R}^K:\bar{c}_l^J(x)\leq p(x)'\Sigma^{1/2}\gamma\leq\bar{c}_u^J(x)\text{ for all }x\in\mathcal{X}_J\}$$

and

$$S=\{\gamma\in\mathbb{R}^K:\bar{c}_l(x)\leq p(x)'\Sigma^{1/2}\gamma\leq\bar{c}_u(x)\text{ for all }x\in\mathcal{X}\}.$$

We first show that if $S\subseteq S_u$ and $S_J\subseteq S_u$ for some $C_s<\infty$, then

$$\left|\int(\bar{c}_u(x)-\bar{c}_l(x))w_X(x)dx-\int(\bar{c}_u^J(x)-\bar{c}_l^J(x))w_X(x)dx\right|\to 0.$$

To do so, let $\varepsilon > 0$. Also notice that, by definition, $\bar{c}_u(x) = \sup_{\gamma \in S} p(x)'\gamma$. Therefore, for all $x, x' \in \mathcal{X}$ it holds that

$$\begin{aligned}
|\bar{c}_u(x) - \bar{c}_u(x')| &= |\sup_{\gamma \in S} p(x)'\gamma - \sup_{\gamma \in S} p(x')'\gamma| \\
&\leq \sup_{\gamma \in S} |p(x)'\gamma - p(x')'\gamma| \\
&\leq \sup_{\gamma \in S} |\nabla p(\tilde{x})'\gamma(x - x')| \\
&\leq \sup_{\gamma \in S_u} \|\gamma\| \sup_{\tilde{x} \in \mathcal{X}} \|\nabla p(\tilde{x})\| |x - x'|,
\end{aligned}$$

where the last line follows from the assumption that $S \subseteq S_u$. Hence,

$$\sup_{x, x' \in \mathcal{X}, x \neq x'} \frac{|\bar{c}_u(x) - \bar{c}_u(x')|}{|x - x'|} \leq \sup_{\gamma \in S_u} \|\gamma\| \sup_{x \in \mathcal{X}} \|\nabla p(x)\|.$$

Moreover,

$$\bar{c}_u(x) \leq \|p(x)\| \sup_{\gamma \in S_u} \|\gamma\|$$

It follows that for all $x, x' \in \mathcal{X}$

$$\begin{aligned}
\frac{|\bar{c}_u(x)w(x) - \bar{c}_u(x')w(x')|}{|x - x'|} &\leq \frac{|(\bar{c}_u(x) - c(x'))w(x)| + |(w(x) - w(x'))\bar{c}_u(x')|}{|x - x'|} \\
&\leq C \sup_{\gamma \in S_u} \|\gamma\| \sup_{x \in \mathcal{X}} \|\nabla p(x)\| + C\|p(x')\| \sup_{\gamma \in S_u} \|\gamma\| \\
&\leq 2C^2 \sup_{\gamma \in S_u} \|\gamma\|
\end{aligned}$$

Next write

$$\begin{aligned}
&\left| \int \bar{c}_u(x) w_X(x) dx - \sum_{j=1}^{J-1} \bar{c}_u(x_j) w_X(x_j)(x_{j+1} - x_j) \right| \\
&= \left| \sum_{j=1}^{J-1} \int_{x_j}^{x_{j+1}} \left( \bar{c}_u(x) w_X(x) - \bar{c}_u(x_j) w_X(x_j) \right) dx \right| \\
&\leq \left| \sum_{j=1}^{J-1} \int_{x_j}^{x_{j+1}} 2C^2 \sup_{\gamma \in S_u} \|\gamma\| |x_{j+1} - x_j| dx \right| \\
&= 2C^2 \sup_{\gamma \in S_u} \|\gamma\| \sum_{j=1}^{J-1} (x_{j+1} - x_j)^2
\end{aligned}$$

It follows that

$$\left| \int \bar{c}_u(x) w_X(x) dx - \sum_{j=1}^{J-1} \bar{c}_u(x_j) w_X(x_j)(x_{j+1} - x_j) \right| \to 0$$

Identical arguments imply that

$$\left| \int \bar{c}_u^J(x) w_X(x) dx - \sum_{j=1}^{J-1} \bar{c}_u^J(x_j) w_X(x_j)(x_{j+1} - x_j) \right| \to 0$$

and we get analogous results for $\bar{c}_l(x)$ and $\bar{c}_l^J(x)$. Theorem 2 now implies that

$$\left| \int (\bar{c}_u(x) - \bar{c}_l(x)) w_X(x) dx - \int (\bar{c}_u^J(x) - \bar{c}_l^J(x)) w_X(x) dx \right| \to 0.$$

Next notice that

$$p(x)'\Sigma^{1/2} Z = \tilde{p}(x)' B' \Sigma^{1/2} Z = \tilde{p}(x)' \tilde{Z},$$

where $\tilde{Z} = B'\Sigma^{1/2} Z \sim N(0, B'\Sigma B)$. It follows that we can write the constraint

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

as

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

where $W \sim N(0, I_{L\times L})$ and $\tilde{\Sigma}^{1/2} = (B'\Sigma B)^{1/2}$. Moreover, the confidence band is a projection of $p(x)'\beta$ on

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

or equivalently a projection of $\tilde{p}(x)'\beta$ on

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

where $\tilde{\beta} = B'\hat{\beta} \sim N(0, B'\Sigma B)$. Since we can always rewrite the problem using the $L$ dimensional vector $\tilde{p}(x)$, it is sufficient to prove that there exists $C_s < \infty$ such that $S \subseteq S_u$ and $S_J \subseteq S_u$ when $B = I_{K\times K}$.

Now notice that from arguments in the proof of Corollary 1, the band resulting from the projection on $\{\beta : (\hat{\beta} - \beta)\Sigma^{-1}(\hat{\beta} - \beta) \leq c_{K,1-\alpha}\}$, where $c_{K,1-\alpha}$ is the $1 - \alpha$ quantile of the $\chi^2_K$ distribution, leads to a conservative band of the form $p(x)'\hat{\beta} \pm c_{K,1-\alpha}\sigma(x)$. The weighted area of this band is

$$2c_{K,1-\alpha} \int \sqrt{p(x)\Sigma p(x)} w_X(x) dx \leq 2c_{K,1-\alpha} \lambda_{max}(\Sigma^{1/2}) \int \|p(x)\| w_X(x) dx < \infty,$$

where $\lambda_{max}(\Sigma^{1/2})$ denotes the largest eigenvalue of $\Sigma^{1/2}$. Define

$$\bar{U} = 2c_{K,1-\alpha} \lambda_{max}(\Sigma^{1/2}) \int \|p(x)\| w_X(x) dx.$$

Let $\gamma \in S$. Then for each $x \in \mathcal{X}$ either $\bar{c}_u(x) \geq |p(x)'\gamma|$ or $\bar{c}_l(x) \leq -|p(x)'\gamma|$. Also notice that $\bar{c}_u(x) \geq 0$ and $\bar{c}_l(x) \leq 0$. It follows from Assumption 8 that for some $\varepsilon > 0$

$$\int (\bar{c}_u(x) - \bar{c}_l(x)) w_X(x) dx \geq \int |p(x)'\gamma| w_X(x) dx \geq \varepsilon \|\gamma\|.$$

Since we also established $\int (\bar{c}_u(x) - \bar{c}_l(x)) w_X(x) d(x) \leq \bar{U}$, it follows that $\|\gamma\| \leq \frac{\bar{U}}{\varepsilon}$ and therefore there exists $C_s$ such that

$$S \subseteq \{\gamma \in \mathbb{R}^K : -C_s \leq \gamma_k \leq C_s \text{ for all } k = 1, \ldots, K\}.$$

Similarly, projecting on the set $\{\beta : (\hat{\beta} - \beta)\Sigma^{-1}(\hat{\beta} - \beta) \leq c_{K,1-\alpha}\}$ for all $p(x_j)$ yields a conservative band and the objective function $h_J$ for this band is

$$2c_{K,1-\alpha} \sum_{j=1}^{J-1} \sqrt{p(x_j)\Sigma p(x_j)} w_X(x_j)(x_{j+1} - x_j).$$

Arguments as in the first part imply that

$$\left| 2c_{K,1-\alpha} \sum_{j=1}^{J-1} \sqrt{p(x_j)\Sigma p(x_j)} w_X(x_j)(x_{j+1} - x_j) - 2c_{K,1-\alpha} \int \sqrt{p(x)\Sigma p(x)} w_X(x) dx \right| \to 0.$$

Therefore, for $J$ large enough

$$\sum_{j=1}^{J-1} (\bar{c}_u^J(x_j) - \bar{c}_l^J(x_j)) w_X(x_j)(x_{j+1} - x_j) \leq 2c_{K,1-\alpha} \sum_{j=1}^{J-1} \sqrt{p(x_j)\Sigma p(x_j)} w_X(x_j)(x_{j+1} - x_j) \leq \bar{U} + 1.$$

Now let $\gamma \in S_J$ and write $\bar{\gamma} = \alpha\gamma$. Then, similar as before there exists $\varepsilon > 0$ such that $\|\gamma\| \leq \frac{\bar{U}+1}{\varepsilon}$ for all $\bar{\gamma} \in S_J$. Thus, there exists a $C_s$ such that

$$S_J \subseteq \{\gamma \in \mathbb{R}^K : -C_s \leq \gamma_k \leq C_s \text{ for all } k = 1, \ldots, K\}.$$

The conclusion now follows from the first part. □

*Proof of Lemma A2.* As in the proof of Lemma A1, we can assume without loss of generality that $B = I_{K \times K}$ because otherwise we can simply work with a transformed problem. Analogously, we can assume without loss of generality that $\Sigma = I_{K \times K}$ (but $\hat{\Sigma} \neq I_{K \times K}$ in general).

The proof proceeds in several steps. First, we show that the bands are based on projections on bounded sets and that the weighted areas corresponding to the optimal solutions are bounded from above and below (in probability). We then show that the optimal bands are symmetric. Next we show that

$$\left|\int \bar{c}_u(x,\hat{\Sigma})w_X(x)dx - \int \bar{c}_u(x,\Sigma)w_X(x)dx\right| \xrightarrow{p} 0.$$

Finally, we show a contradiction if for some $\varepsilon > 0$ and $\delta > 0$ and $n$ large enough

$$P\left(\sup_{x\in\mathcal{X}} |\bar{c}_u(x,\hat{\Sigma}) - \bar{c}_u(x,\Sigma)| > \varepsilon\right) > \delta.$$

For the first step, arguments as in the proof of Lemma A1 imply that there exists a constant $C_s$ such that $S(\Sigma) \subseteq S_u$, where

$$S_u = \{\gamma \in \mathbb{R}^K : |\gamma_k| \leq C_s \text{ for all } k = 1,\ldots,K\},$$

and

$$S(\Sigma) = \{\gamma \in \mathbb{R}^K : \bar{c}_l(x,\Sigma) \leq p(x)'\Sigma^{1/2}\gamma \leq \bar{c}_u(x,\Sigma) \text{ for all } x \in \mathcal{X}\}.$$

Analogous arguments imply that there is a constant $C_s$ such that $S(\hat{\Sigma}) \subseteq S_u$ with probability approaching 1. From the proof Lemma A1 it also follows that $\int \left(\bar{c}_u(x,\Sigma) - \bar{c}_l(x,\Sigma)\right) w_X(x)dx$ is bounded and that $\int(\bar{c}_u(x,\hat{\Sigma})-\bar{c}_l(x,\hat{\Sigma}))w_X(x)dx$ is bounded by a constant with probability approaching 1.

Now notice that $\bar{c}_u(x,\Sigma) \geq \sigma(x)c_{1-\alpha}$ and $\bar{c}_l(x,\Sigma) \leq -\sigma(x)c_{1-\alpha}$, where $c_{1-\alpha}$ denotes the $1-\alpha$ quantile of the standard normal distribution. Therefore,

$$\int \left(\bar{c}_u(x,\Sigma) - \bar{c}_l(x,\Sigma)\right) w_X(x)dx \geq 2c_{1-\alpha}\int \sigma(x)w_X(x)dx > 0$$

and

$$\int \left(\bar{c}_u(x,\hat{\Sigma}) - \bar{c}_l(x,\hat{\Sigma})\right) w_X(x)dx \geq c_{1-\alpha}\int \sigma(x)w_X(x)dx + o_p(1).$$

We now prove that $\bar{c}_l(x,\Sigma) = -\bar{c}_u(x,\Sigma)$. Suppose $\bar{c}_u(x,\Sigma) \neq -\bar{c}_l(x,\Sigma)$ for some $x \in \mathcal{X}$. Let

$$\tilde{c}_l(x,\Sigma) = \bar{c}_l(x,\Sigma) - 1/2(\bar{c}_l(x,\Sigma) + \bar{c}_u(x,\Sigma)) = 1/2\bar{c}_l(x,\Sigma) - 1/2\bar{c}_u(x,\Sigma)$$

and

$$\tilde{c}_u(x,\Sigma) = \bar{c}_u(x,\Sigma) - 1/2(\bar{c}_l(x,\Sigma) + \bar{c}_u(x,\Sigma)) = 1/2\bar{c}_u(x,\Sigma) - 1/2\bar{c}_l(x,\Sigma).$$

Let

$$S_1 = \{z \in \mathbb{R}^K : \bar{c}_l(x,\Sigma) \leq p(x)'\Sigma^{1/2}z \leq \bar{c}_u(x,\Sigma) \text{ for all } x \in \mathcal{X}\}$$

and

$$S_2 = \{z \in \mathbb{R}^K : -\bar{c}_u(x,\Sigma) \leq p(x)'\Sigma^{1/2}z \leq -\bar{c}_l(x,\Sigma) \text{ for all } x \in \mathcal{X}\}.$$

By the symmetry of the normal measure, $P(Z \in S_1) = P(Z \in S_2) = 1 - \alpha$. It follows from Lemmas A5 and A6 that

$$P(\tilde{c}_l(x, \Sigma) \leq p(x)'\Sigma^{1/2}Z \leq \tilde{c}_u(x, \Sigma) \text{ for all } x \in \mathcal{X}) > 1 - \alpha.$$

But

$$\int \left(\tilde{c}_u(x, \Sigma) - \tilde{c}_l(x, \Sigma)\right) w_X(x)dx = \int \left(\bar{c}_u(x, \Sigma) - \bar{c}_l(x, \Sigma)\right) w_X(x)dx,$$

which contradicts that $(\bar{c}_l(x, \Sigma), \bar{c}_u(x, \Sigma))$ is optimal. Therefore, $\bar{c}_u(x, \Sigma) = -\bar{c}_l(x, \Sigma)$.

Next we prove that

$$\left| \int \bar{c}_u(x, \hat{\Sigma})w_X(x)dx - \int \bar{c}_u(x, \Sigma)w_X(x)dx \right| \xrightarrow{p} 0.$$

Let $c_1$ be a constant such that

$$P(-c_1\bar{c}_u(x, \Sigma) \leq p(x)'\hat{\Sigma}^{1/2}Z \leq c_1\bar{c}_u(x, \Sigma) \text{ for all } x \in \mathcal{X} \mid \hat{\Sigma}) = 1 - \alpha.$$

Since

$$|p(x)'(\hat{\Sigma}^{1/2} - \Sigma^{1/2})Z| \leq \|p(x)\|\|\hat{\Sigma}^{1/2} - \Sigma^{1/2}\|\|Z\| = o_p(1),$$

it holds that $|c_1 - 1| = o_p(1)$. Since $\bar{c}_u(x, \hat{\Sigma})$ is optimal and $c_1\bar{c}_u(x, \Sigma)$ is feasible it holds that

$$\begin{aligned}
\int \bar{c}_u(x, \hat{\Sigma})w_X(x)dx &\leq \int \bar{c}_u(x, \Sigma)w_X(x)dx + (c_1 - 1)\int \bar{c}_u(x, \Sigma)w_X(x)dx \\
&\leq \int \bar{c}_u(x, \Sigma)w_X(x)dx + o_p(1).
\end{aligned}$$

Similarly, $|c_2 - 1| = o_p(1)$, where $c_2$ is such that

$$P(-c_2\bar{c}_u(x, \hat{\Sigma}) \leq p(x)'\Sigma^{1/2}Z \leq c_2\bar{c}_u(x, \hat{\Sigma}) \text{ for all } x \in \mathcal{X} \mid \hat{\Sigma}) = 1 - \alpha.$$

Thus,

$$\begin{aligned}
\int \bar{c}_u(x, \Sigma)w_X(x)dx &\leq \int \bar{c}_u(x, \hat{\Sigma})w_X(x)dx + (c_2 - 1)\int \bar{c}_u(x, \hat{\Sigma})w_X(x)dx \\
&\leq \int \bar{c}_u(x, \hat{\Sigma})w_X(x)dx + o_p(1).
\end{aligned}$$

Together

$$\left| \int \bar{c}_u(x, \hat{\Sigma})w_X(x)dx - \int \bar{c}_u(x, \Sigma)w_X(x)dx \right| \xrightarrow{p} 0.$$

Since we assume that $\inf_{x \in \mathcal{X}} \sigma(x) > 0$, it is sufficient to prove that

$$\sup_{x \in \mathcal{X}} |\bar{c}_u(x, \hat{\Sigma}) - \bar{c}_u(x, \Sigma)| = o_p(1).$$

Let $\varepsilon > 0$ and $\delta > 0$ and suppose that

$$P\left(\sup_{x\in\mathcal{X}} |\bar{c}_u(x,\hat{\Sigma}) - \bar{c}_u(x,\Sigma)| > \varepsilon\right) > \delta$$

Define

$$C(\Sigma) = \{\gamma \in \mathbb{R}^K : -\bar{c}_u(x,\Sigma) \leq p(x)'\gamma \leq \bar{c}_u(x,\Sigma) \text{ for all } x \in \mathcal{X}\}.$$

Also let $\lambda \in (0,1)$ and define

$$\bar{c}_u^\lambda(x,\hat{\Sigma}) = \lambda\bar{c}_u(x,\hat{\Sigma}) + (1-\lambda)\bar{c}_u(x,\Sigma)$$

and

$$\hat{C}^\lambda(\hat{\Sigma}) = \{\gamma \in \mathbb{R}^K : -\bar{c}_u^\lambda(x,\hat{\Sigma}) \leq p(x)'\gamma \leq \bar{c}_u^\lambda(x,\hat{\Sigma}) \text{ for all } x \in \mathcal{X}\}.$$

Let $Z \sim N(0, I_{K\times K})$. Since $P(\sup_{x\in\mathcal{X}} |\bar{c}_u(x,\hat{\Sigma}) - \bar{c}_u(x,\Sigma)| > \varepsilon) > \delta$, by Lemmas A5 and A6 there exists a constant $\eta > 0$ such that with probability at least $\delta/2$ and for $n$ large enough

$$P(\hat{\Sigma}^{1/2}Z \in \hat{C}^\lambda(\hat{\Sigma}) \mid \hat{\Sigma}) \geq P(\hat{\Sigma}^{1/2}Z \in C(\hat{\Sigma}) \mid \hat{\Sigma})^\lambda P(\hat{\Sigma}^{1/2}Z \in C(\Sigma) \mid \hat{\Sigma})^{1-\lambda} + \eta$$

Notice that $P(\hat{\Sigma}^{1/2}Z \in C(\hat{\Sigma}) \mid \hat{\Sigma}) = 1-\alpha$. Moreover, since $P(\Sigma^{1/2}Z \in C(\Sigma)) = 1-\alpha$ and

$$|p(x)'(\hat{\Sigma}^{1/2} - \Sigma^{1/2})Z| \leq \|p(x)\|\|\hat{\Sigma}^{1/2} - \Sigma^{1/2}\|\|Z\| = o_p(1),$$

$P(\hat{\Sigma}^{1/2}Z \in C(\Sigma) \mid \hat{\Sigma}) = 1-\alpha+o_p(1)$. It follows that with probability at least $\delta/2$ and for $n$ large enough

$$P(\hat{\Sigma}^{1/2}Z \in \hat{C}^\lambda \mid \hat{\Sigma}) \geq 1-\alpha+\eta+o_p(1).$$

Next let $c = \frac{1-\alpha+\eta/2}{1-\alpha+\eta} \in (0,1)$ and let

$$\tilde{c}^\lambda(x,\hat{\Sigma}) = c(\lambda\bar{c}_u(x,\hat{\Sigma}) + (1-\lambda)\bar{c}_u(x,\Sigma)).$$

Then with probability at least $\delta/2$ and for $n$ large enough

$$\begin{aligned}
&P(-\tilde{c}^\lambda(x,\hat{\Sigma}) \leq p(x)'\hat{\Sigma}^{1/2}Z \leq \tilde{c}^\lambda(x,\hat{\Sigma}) \text{ for all } x \in \mathcal{X} \mid \hat{\Sigma}) \\
&\quad\geq cP(-\bar{c}_u^\lambda(x,\hat{\Sigma}) \leq p(x)'\hat{\Sigma}^{1/2}Z \leq \bar{c}_u^\lambda(x,\hat{\Sigma}) \text{ for all } x \in \mathcal{X} \mid \Sigma) \\
&\quad\geq c(1-\alpha+\eta+o_p(1)) \\
&\quad= 1-\alpha+\eta/2+o_p(1).
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\int \tilde{c}(x,\hat{\Sigma})w_X(x)dx \\
&\quad= c\int \bar{c}_u(x,\hat{\Sigma})w_X(x)dx + c(1-\lambda)\left(\int \bar{c}_u(x,\Sigma)w_X(x)dx - \int \bar{c}_u(x,\hat{\Sigma})w_X(x)dx\right) \\
&\quad= c\int \bar{c}_u(x,\hat{\Sigma})w_X(x)dx + o_p(1).
\end{aligned}$$

It follows that with probability at least $\delta/4$ and for $n$ large enough

$$P(-\tilde{c}^{\lambda}(x,\hat{\Sigma}) \leq p(x)'\hat{\Sigma}^{1/2}Z \leq \tilde{c}^{\lambda}(x,\hat{\Sigma}) \text{ for all } x \in \mathcal{X} \mid \hat{\Sigma}) > 1-\alpha$$

and

$$\int \tilde{c}(x,\hat{\Sigma})w_X(x)dx < \int \bar{c}_u(x,\hat{\Sigma})w_X(x)dx,$$

which would contradict that $\bar{c}_u(x,\hat{\Sigma})$ is optimal.

It therefore has to hold that

$$P\left(\sup_{x\in\mathcal{X}} |\bar{c}_u(x,\hat{\Sigma}) - \bar{c}_u(x,\Sigma)| > \varepsilon\right) \to 0$$

which means that

$$\sup_{x\in\mathcal{X}} |\bar{c}_u(x,\hat{\Sigma}) - \bar{c}_u(x,\Sigma)| = o_p(1).$$

□

# References


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

Bentkus, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference 113*(2), 385–402.

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

Bogachev, V. I. (1998). *Gaussian Measures*. Mathematical surveys and monographs. American Mathematical Society.

Bohrer, R. (1973). Optimality property of scheffé bounds. *The Annals of Statistics 1*(4), 766–772.

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

Chen, X. and T. M. Christensen (2015). Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation. Working paper.

Cunningham, J. P., P. Hennig, and S. Lacoste-Julien (2011). Gaussian probabilities and expectation propagation. Working paper.

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

Gafarov, B., M. Meier, and J. L. Montiel Olea (2016). Projection inference for set-identified SVARs. Working paper.

Gandhi, A. and J.-F. Houde (2016). Measuring substitution patterns in differentiated products industries. Working paper.

Gsteiger, S., F. Bretz, and W. Liu (2011). Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *Journal of Biopharmaceutical Statistics 21*, 708–725.

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

Horowitz, J. L. and S. Lee (2015). Nonparametric estimation and inference under shape restrictions. Working paper.

Kaido, H., F. Molinari, and J. Stoye (2016). Confidence intervals for projections of partially identified parameters. Working paper.

Khorasani, F. and G. A. Milliken (1979). On the exactness of confidence bands about a linear model. *Journal of the American Statistical Association 74*(366), 446–448.

Liu, W. (2010). *Simultaneous inference in regression.* CRC Press.

Liu, W. and P. Ah-Kine (2010). Optimal simultaneous confidence bands in simple linear regression. *Journal of Statistical Planning and Inference 100*, 1225–1235.

Liu, W. and A. J. Hayter (2007). Minimum area confidence set optimality for confidence bands in simple linear regression. *Journal of the American Statistical Association 102*(477), 181–190.

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

Lütkepohl, H., A. Staszewska-Bystrova, and P. Winker (2015). Confidence bands for impulse responses: Bonferroni vs. Wald. *Oxford Bulletin of Economics and Statistics 77*(6), 800–821.

Mandt, S., F. Wenzel, S. Nakajima, J. Cunningham, C. Lippert, and M. Kloft (2016). Sparse estimation in a correlated probit model. Working paper.

Naiman, D. Q. (1983). Comparing Scheffé-type to constant-width confidence bounds in regression. *Journal of the American Statistical Association 78*(384), 906–912.

Naiman, D. Q. (1984a). Average width optimality of simultaneous confidence bounds. *The Annals of Statistics 12*(4), 1199–1214.

Naiman, D. Q. (1984b). Optimal simultaneous confidence bounds. *The Annals of Statistics 12*(2), 702–715.

Naiman, D. Q. (1987). Minimax regret simultaneous confidence bands for multiple regression functions. *Journal of the American Statistical Association 82*(399), 894–901.

Piegorsch, W. W. (1985). Admissible and optimal confidence bands in simple linear regression. *The Annals of Statistics 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). Pointwise and uniform inference for semi-nonparametric gmm models. Working paper.

Working, H. and H. Hotelling (1929). Applications of the theory of error to the interpretation of trends. *Journal of the American Statistical Association 24*(165), 73–85.

Wynn, H. P. and P. Bloomfield (1971). Simultaneous confidence bands in regression analysis. *Journal of the Royal Statistical Society. Series B 33*(2), 202–217.