Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs*

José Luis Montiel Olea — Columbia University

Mikkel Plagborg-Møller — Princeton University

August 8, 2018

Abstract: Simultaneous confidence bands are versatile tools for visualizing estimation uncertainty for parameter vectors, such as impulse response functions. In linear models, it is known that that the sup-t confidence band is narrower than commonly used alternatives, for example Bonferroni and projection bands. We show that the same ranking applies *asymptotically* even in general nonlinear models, such as VARs. Moreover, we provide further justification for the sup-t band by showing that it is the optimal default choice when the researcher does not know the audience's preferences. Complementing existing plug-in and bootstrap implementations, we propose a computationally convenient Bayesian sup-t band with exact finite-sample simultaneous credibility. In an application to SVAR impulse response function estimation, the sup-t band—which has been surprisingly overlooked in this setting—is at least 35% narrower than other off-the-shelf simultaneous bands.

Keywords: Bonferroni adjustment, projection inference, simultaneous Bayesian credibility, simultaneous inference, structural vector autoregression, sup-t confidence band.

JEL codes: C11, C12, C32, C44.

---

*Plagborg-Møller (corresponding author): Julis Romo Rabinowitz Building, Princeton, NJ 08544, `mikkelpm@princeton.edu`. Montiel Olea: 420 W 118th St, New York, NY 10027, `montiel.olea@gmail.com`.
We are grateful for comments from Gary Chamberlain, Aubrey Clark, Max Farrell, Joachim Freyberger, Chris Hansen, Lutz Kilian, Frank Kleibergen, Roger Koenker, Michal Kolesár, Simon Lee, Barbara Rossi, Azeem Shaikh, Neil Shephard, Chris Sims, Jim Stock, Elie Tamer, Harald Uhlig, Mark Watson, five anonymous referees, and seminar participants at several venues.

1 Introduction

Simultaneous inference concerns arise commonly in applied work. Consider for example the classical problem of measuring the response of economic activity to an increase in interest rates. What is the difference between the one-month- and one-year-ahead response of industrial production? Do the short-run effects on industrial production disappear after the first year? These questions require comparing the responses, simultaneously, at different time horizons. Similar issues crop up in event studies in applied microeconomics, or when comparing coefficient estimates across different regression specifications, and so on.

Confidence bands—collections of confidence intervals for each component of a parameter vector—are often used to visualize estimation uncertainty in examples like the ones described above. Unlike the oft-used pointwise confidence band, *simultaneous* confidence bands cover the *entire* parameter vector of interest with (asymptotic) probability at least $1 - \alpha$. For example, a simultaneous confidence band for the responses of industrial production to a monetary shock contains the entire path of responses over time, up to some fixed horizon, with probability at least $1 - \alpha$. This allows for valid confidence statements that compare across horizons, i.e., individual parameters. In contrast to confidence *ellipsoids*, confidence bands are by definition easy to visualize regardless of the dimension of the parameter vector.

While many simultaneous confidence bands have been developed in the literature, there exists little theory to select among these, at least outside the linear regression model. Bands encountered in applied work include Bonferroni, Šidák, projection, and "sup-t" bands. These bands can differ substantially in terms of their width and coverage properties, and unfortunately, existing analyses of confidence bands are not directly applicable to nonlinear or heteroskedastic econometric models. Consequently, in some applications, practitioners have hitherto relied exclusively on simulation evidence to select among different bands. This is the case for impulse response analysis in Vector Autoregressions (VARs).[1]

We fill this gap by providing analytical comparisons of simultaneous confidence bands in an empirically relevant econometric framework encompassing macroeconometric and microeconometric applications. Our goal is to compare bands in terms of performance measures that are relevant to practitioners: the simultaneous coverage probability and the expected width of the confidence bands. Our concrete suggestion for applied researchers is to use the computationally convenient sup-t band as a default procedure to conduct simultaneous inference.[2] Although the sup-t band has a long tradition in nonparametric regression and density estimation, it has been overlooked in other econometric problems—such as impulse response function analysis in VARs and sensitivity analysis in linear regression—in which it performs well. We support our conclusion by establishing three results.

First, we analytically compare the relative widths of popular confidence bands that, asymptotically, fall in a one-parameter class. Bands in the one-parameter class equal a natural point estimator plus/minus a constant c times the vector of pointwise standard errors. In linear models, it is known that the class contains the pointwise, Bonferroni, Šidák, projection, and sup-t bands; and the latter is trivially the narrowest simultaneous confidence band in the class (in every realization of the data). Our contribution is to show that the same ranking of popular bands applies *asymptotically* even in general nonlinear models such as VARs, under weak regularity conditions. To our knowledge, our analysis provides the first analytical ranking of confidence bands in nonlinear models. The gain from using the sup-t band is especially large when the estimators of the individual parameters of interest are highly dependent.[3] Moreover, it is known that the width of the sup-t band increases

[1] See, for example, the simulation studies in Lütkepohl et al. (2015a,b) and Bruder & Wolf (2018).

[2] We have created a full Matlab suite to implement the sup-t band in the generic class of econometric models covered by this paper. The files are available at `https://scholar.princeton.edu/mikkelpm/publications/conf_band`

[3] We also show that the Šidák and Bonferroni bands are narrower than projection-based bands, unless the number of parameters of interest is orders of magnitude larger than the dimension of the underlying model.

very slowly with the number of parameters of interest, so informative simultaneous inference does not require restricting attention *a priori* to a small number of parameters.

Second, we propose a simple Bayesian implementation of the sup-t band with finite-sample credibility $1-\alpha$, which complements existing plug-in and bootstrap implementations. The Bayesian band equals the Cartesian product of the usual component-wise equal-tailed credible intervals, where we "calibrate" the tail probability to achieve the desired *simultaneous* credibility. The procedure is computationally fast, and its only input is a set of posterior draws of the parameters of interest. Our procedure appears to be the first in the literature to achieve finite-sample simultaneous Bayesian credibility. Moreover, we show that the very same algorithm can be used to compute a valid bootstrap sup-t band by replacing the posterior draws of the model parameters by analogous bootstrap draws. We show that the plug-in, bootstrap, and Bayes sup-t bands are all first-order asymptotically equivalent, provided that the bootstrap is consistent and the posterior satisfies a Bernstein-von Mises property.

Third, we show that the optimal band within the one-parameter class, the sup-t band, uniquely minimizes *worst-case regret* among all translation equivariant confidence bands. The regret is the ratio of attained loss to the smallest possible loss under a given loss function. The "worst case" is taken over all possible (homogeneous of degree 1) loss functions that depend on the component-wise lengths of the band. Hence, among the large collection of translation equivariant bands, the sup-t band provides the smallest possible upper bound on regret when the researcher is unsure about the audience's loss functions. To the best of our knowledge, this optimality result has not appeared in the literature before. We establish the decision theoretic result in a linear Gaussian model motivated by our asymptotic analysis.

We illustrate the applicability of our results by computing simultaneous confidence bands for impulse response functions in a Structural Vector Autoregression (SVAR). The relevance of simultaneous inference in the VAR context has been highlighted recently by Jordà (2009), Inoue & Kilian (2013, 2016), and Lütkepohl et al. (2015a,b). While many recent papers have constructed confidence bands for SVAR analysis, the optimality of the sup-t band has been surprisingly overlooked, perhaps due to the nonlinear nature of the problem. Following Gertler & Karadi (2015), we estimate the impulse response function of real output to a monetary policy shock identified by exclusion restrictions or by an external instrument.[4] In this application, the sup-t band is about 35% narrower than the oft-used Bonferroni band. The tightness of the sup-t band matters economically: neither Bonferroni nor projection bands allow us to reject the null hypothesis of monetary policy neutrality; but the sup-t band does. To illustrate the flexibility of our set-up, Appendix A.6 contains an additional application to sensitivity analysis in cross-sectional regression.

Literature. Since the original contribution by Working & Hotelling (1929), the literature on simultaneous confidence bands has grown vastly. Decision-theoretic contributions include Naiman (1984, 1987), Piegorsch (1985a,b), and the comprehensive list of references in Liu (2011). Despite the large body of work and a plethora of simultaneous confidence bands available, there are few concrete recommendations for practitioners outside the linear regression model. We appear to be the first to compare the projection, Šidák, and Bonferroni bands in a general non-linear, heteroskedastic framework relevant to economic applications.

This obvious gap in the study of simultaneous confidence bands has generated renewed interest in the subject. In an insightful recent paper, Freyberger & Rai (2018) propose a novel computational procedure for obtaining an approximately optimal confidence band given a *particular* loss function. Our contribution is complementary: We show that the simple sup-t band has a worst-case regret optimality property when the decision-maker is *unsure* about the appropriate loss function.

There is a growing literature on simultaneous confidence bands for VAR impulse response

[4] Our theoretical results apply to any point identification scheme.

analysis.[5] This literature has hitherto relied on Monte Carlo simulations to compare bands. In Appendix A.4 we argue that previously proposed bands can all be viewed, asymptotically, as variants of bands in the one-parameter class, so we can unambiguously rank these bands in large samples. Our analysis covers projection (also known as "Wald") bands, which are not as obviously related to the other bands. Asymptotically equivalent implementations of the sup-t band for VARs have appeared under the names of "adjusted Wald/Bonferroni band" (Lütkepohl et al., 2015b) and "balanced bootstrap band" (Bruder & Wolf, 2018). To our surprise, it has not been recognized that these bands are (in large samples) particular implementations of the sup-t band and thus can be analytically compared with competing alternatives. By providing a first-order asymptotic ranking of simultaneous bands, we believe that our analytical framework yields additional insights that complement the extensive small-sample simulation studies of Lütkepohl et al. (2015a,b) and Bruder & Wolf (2018).

The time series literature has considered the related issue of constructing simultaneous bands for path forecasts. Because forecast errors are not necessarily Gaussian, our theory—which relies on normality implied by the central limit theorem for the case of parameter estimation—does not apply to the forecasting setting. Wolf & Wunderli (2015) have developed a modified sup-t band for path forecast bands (they do not show that the sup-t band has optimality properties in non-Gaussian settings).

The sup-t band has a long tradition in nonparametric regression and density estimation. Wasserman (2006, chapter 5.7) gives a textbook treatment. Econometric applications include Horowitz & Lee (2012), Chernozhukov et al. (2013), Chernozhukov et al. (2014), and Lee et al. (2017). Unlike these papers, we focus on a finite-dimensional setting that covers different applications, including ours, and we analytically compare several popular bands.

The issue of how to construct optimal simultaneous confidence bands has many parallels

[5] Some authors visualize estimation uncertainty about impulse responses using methods other than simultaneous confidence bands (Sims & Zha, 1999; Inoue & Kilian, 2013, 2016; Baumeister & Hamilton, 2018).

in the multiple hypothesis testing literature (Romano et al., 2010, section 8). Instead of dealing with test power, our analysis focuses directly on the *width* of confidence bands—a key issue for practitioners.[6] Our one-parameter class of confidence bands can be obtained by inverting the class of *single-step* multiple testing procedures, using studentized test statistics. Lehmann & Romano (2005, chapter 9) show that the sup-t band yields an optimal single-step testing procedure under certain equivariance conditions, including permutation equivariance conditions that we do not impose in our analysis. White (2000) and Hansen (2005) construct multiple hypothesis testing procedures that are analogues of the sup-t band. Romano & Wolf (2005, 2007) and List et al. (2016) develop *step-down* multiple testing procedures that improve on the finite-sample power of single-step procedures. Our worst-case regret result does not seem to have a direct parallel in the multiple testing literature.

Outline. Section 2 defines our econometric set-up and the sup-t band. Section 3 analytically compares the widths of several popular confidence bands. Section 4 shows that the sup-t band minimizes worst-case regret. Section 5 contains the empirical application. Section 6 discusses extensions and future research directions. Appendix A, which is available online, contains an application to regression sensitivity analysis, various technical details, and proofs.[7]

Notation. Convergence in probability and distribution are denoted $\xrightarrow{p}$ and $\xrightarrow{d}$. I_p is the $p \times p$ identity matrix, and $\mathbf{0}_p$ is a $p \times 1$ vector of zeros. The p-dimensional normal distribution with mean vector μ and variance matrix Ω is denoted $N_p(\mu, \Omega)$. The $\zeta \in (0,1)$ quantile of the square root of the $\chi^2(p)$ distribution is denoted $\chi_{p,\zeta}$; for $p = 1$, we just write $\chi_\zeta \equiv \chi_{1,\zeta}$. The $\zeta \in (0,1)$ quantile of random variable X is denoted $Q_\zeta(X)$.

[6] We remark that, in terms of (local) power, joint tests based on the sup-t band are neither dominated by nor do they dominate tests based on the Wald ellipsoid.

[7] https://scholar.princeton.edu/mikkelpm/publications/conf_band

2 Econometric set-up and the sup-t band

We start by defining our general econometric framework, simultaneous confidence bands, and the particular band that we recommend for applied use: the sup-t band.

Econometric framework. We seek to construct a simultaneous confidence band for a parameter vector $\theta \in \mathbb{R}^k$. This parameter vector of interest is a possibly nonlinear transformation of the parameter vector $\mu \in \mathbb{R}^p$ of some underlying model. That is, $\theta \equiv h(\mu)$ where $h(\cdot) = (h_1(\cdot), h_2(\cdot), \ldots, h_k(\cdot))'$ is a k-valued function. We assume that the transformation $h(\cdot)$ is continuously differentiable and denote the $k \times p$ Jacobian as $\dot{h}(\cdot) = (\dot{h}_1(\cdot), \ldots, \dot{h}_k(\cdot))'$, where $\dot{h}_j(\mu) \equiv \partial h_j(\mu)/\partial \mu$. We do not restrict the relative magnitudes of the dimensions p and k, except in assuming that they are both finite.

A *simultaneous* $1 - \alpha$ *confidence band* for θ is defined as the Cartesian product

$$\widehat{C} = \widehat{C}_1 \times \widehat{C}_2 \times \ldots \times \widehat{C}_k \tag{1}$$

of data-dependent intervals $\widehat{C}_j \subseteq \mathbb{R}$ that covers the true parameter vector $\theta = (\theta_1, \ldots, \theta_k)'$ with probability at least $1 - \alpha$ asymptotically:

$$\liminf_{n\to\infty} P(\theta \in \widehat{C}) = \liminf_{n\to\infty} P(\theta_j \in \widehat{C}_j \text{ for all } j) \geq 1 - \alpha, \tag{2}$$

where n denotes the sample size. More succinctly, a simultaneous confidence band is a confidence set for θ with *rectangular* structure. The advantage of imposing the rectangular structure is that the confidence band (unlike the Wald ellipsoid) can be easily visualized regardless of the dimensions k and p, allowing readers to flexibly test different hypotheses by visual inspection. We illustrate this well-known property of simultaneous bands in Section 5.

To construct simultaneous confidence bands, we assume the availability of an asymptotically normal estimator $\hat{\mu}$ of the underlying model parameters μ:

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N_p(\mathbf{0}_p, \Omega) \quad \text{as } n \to \infty. \tag{3}$$

We do not restrict the asymptotic correlation structure of $\hat{\mu}$, and we do not assume that the data is i.i.d. We do assume the existence of a consistent estimator $\hat{\Omega}$ of the possibly singular asymptotic variance Ω.[8] As described in Assumption 1 in Appendix A.1, we impose standard weak regularity conditions such that the delta method implies asymptotic normality of the plug-in estimator $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)' \equiv h(\hat{\mu})$ of the parameter vector of interest θ:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N_k(\mathbf{0}_k, \Sigma) \quad \text{as } n \to \infty. \tag{4}$$

Importantly, the asymptotic variance matrix $\Sigma = \dot{h}(\mu)\Omega\dot{h}(\mu)'$ of $\hat{\theta}$ is allowed to be singular, which for example occurs when $k > p$ (Lütkepohl et al., 2015b; Inoue & Kilian, 2016).

One-parameter class. To facilitate comparisons between confidence bands, we introduce a one-parameter class of confidence bands. We argue in Section 3 that the class of one-parameter bands contains many commonly used bands in applied work. First, define the usual pointwise standard error for $\hat{\theta}_j$:

$$\hat{\sigma}_j \equiv n^{-1/2}\sqrt{\dot{h}_j(\hat{\mu})'\hat{\Omega}\dot{h}_j(\hat{\mu})}, \quad j = 1, \ldots, k.$$

The one-parameter class of confidence bands is parameterized by a positive scalar c, called the *critical value*, which governs the width of the confidence band.

[8] Allowing Ω to be singular is important in applications to impulse response function estimation using non-stationary VARs, cf. Section 5.

Definition 1. *For any $c > 0$, define the one-parameter confidence band*

$$\widehat{B}(c) \equiv \left[\hat{\theta}_1 - \hat{\sigma}_1 c, \hat{\theta}_1 + \hat{\sigma}_1 c\right] \times \left[\hat{\theta}_2 - \hat{\sigma}_2 c, \hat{\theta}_2 + \hat{\sigma}_2 c\right] \times \cdots \times \left[\hat{\theta}_k - \hat{\sigma}_k c, \hat{\theta}_k + \hat{\sigma}_k c\right].$$

Any one-parameter band $\widehat{B}(c)$ is the Cartesian product of scaled-up versions of the usual pointwise confidence intervals, where the scaling factor is the same for every element of θ.[9] The parameter c scales up or down the entire confidence band: If $c \leq \tilde{c}$, then $\widehat{B}(c) \subset \widehat{B}(\tilde{c})$ in every realization of the data. The relative widths of any two one-parameter bands is given by the ratio of their critical values c. Our analysis allows for data-dependent choices of c.

Sup-t band. The sup-t band is the member of the one-parameter class with the smallest critical value c that guarantees simultaneous coverage. Any larger value of c results in an unnecessarily wide and conservative band. Lemma 1 in Appendix A.1 shows that, under weak conditions, the asymptotic coverage probability of any one-parameter band $\widehat{B}(c)$ is given by the cumulative distribution function of a maximum of absolute values of correlated normal variables, evaluated at c:

$$P\left(\theta \in \widehat{B}(c)\right) \to P\left(\max_{j=1,\dots,k} \left|\Sigma_{jj}^{-1/2} V_j\right| \leq c\right) \quad \text{as } n \to \infty, \tag{5}$$

where $V = (V_1, \dots, V_k)' \sim N_k(\mathbf{0}_k, \Sigma)$, and Σ_{jj} is the j-th diagonal element of Σ. Define the ζ-quantile of the above random variable, as a function of the variance-covariance matrix Σ:

$$q_\zeta(\Sigma) \equiv Q_\zeta\left(\max_{j=1,\dots,k} \left|\Sigma_{jj}^{-1/2} V_j\right|\right). \tag{6}$$

[9] The band $\widehat{B}(c)$ can also be interpreted as the collection of those parameter vectors θ for which the *largest* component-wise t-statistic does not exceed a critical value c: $\widehat{B}(c) = \{\tilde{\theta} \in \mathbb{R}^k \mid \max_j |\tilde{\theta}_j - \hat{\theta}_j|/\hat{\sigma}_j \leq c\}$. This explains the terminology “sup-t” below.

The sup-t band is obtained by choosing $c = q_{1-\alpha}(\Sigma)$ (the *sup-t critical value*), yielding a simultaneous coverage probability of precisely $1-\alpha$. In practice, we must use an estimate $\hat{\Sigma}$, as described below. Although the sup-t band is well known in the statistics and econometrics literatures, it seems to be less used in applied work, as discussed in Section 1.

Plug-in sup-t implementation. The most straight-forward feasible implementation of the sup-t band plugs in a consistent estimator of the sup-t critical value, using a delta method estimator for the asymptotic variance of $\hat{\theta}$. Algorithm 1 below presents a standard procedure to estimate the plug-in sup-t critical value $q_{1-\alpha}(\Sigma)$. Except for possibly computing derivatives of $h(\cdot)$, the algorithm takes a fraction of a second to run in Matlab.

Algorithm 1 Plug-in sup-t band

1: Compute the Jacobian $\dot{h}(\hat{\mu})$ and delta method variance estimate $\hat{\Sigma} = \dot{h}(\hat{\mu})\hat{\Omega}\dot{h}(\hat{\mu})'$
2: Draw N i.i.d. normal vectors $\hat{V}^{(\ell)} \sim N_k(\mathbf{0}_k, \hat{\Sigma})$, $\ell = 1, \dots, N$
3: Define $\hat{q}_{1-\alpha}$ as the empirical $1-\alpha$ quantile of $\max_j |\hat{\Sigma}_{jj}^{-1/2}\hat{V}_j^{(\ell)}|$ across $\ell = 1, \dots, N$
4: $\widehat{C} = \widehat{B}(\hat{q}_{1-\alpha}) = \times_{j=1}^{k}[\hat{\theta}_j - \hat{\sigma}_j\hat{q}_{1-\alpha}, \hat{\theta}_j + \hat{\sigma}_j\hat{q}_{1-\alpha}]$

Bootstrap/Bayes sup-t implementation. One of the contributions of this paper is to develop an alternative implementation of the sup-t band, which is useful in combination with bootstrap or Bayesian inference strategies. Unlike Algorithm 1, the following bootstrap/Bayes band does not require explicit calculation of standard errors. Instead, the algorithm only requires a valid strategy for generating either bootstrap draws or Bayesian posterior draws of the underlying model parameters. From an asymptotic perspective, Appendix A.2 shows that our suggested bootstrap/Bayes implementation is equivalent with the above plug-in implementation of the sup-t band.

Our algorithm "calibrates" equal-tailed pointwise confidence/credible intervals to achieve a desired simultaneous coverage/credibility level. The Bayesian band appears to be the first

Algorithm 2 Quantile-based bootstrap or Bayes band

1: Let $\hat{P}$ be the bootstrap distribution of $\hat{\mu}$ or the posterior distribution of μ
2: Draw N samples $\hat{\mu}^{(1)}, \dots, \hat{\mu}^{(N)}$ from $\hat{P}$
3: **for** $\ell = 1, \dots, N$ **do**
4: $\quad\hat{\theta}^{(\ell)} = h(\hat{\mu}^{(\ell)})$
5: **end for**
6: Let $\hat{Q}_{j,\zeta}$ denote the empirical ζ quantile of $\hat{\theta}_j^{(1)}, \dots, \hat{\theta}_j^{(N)}$
7: $\hat{\zeta} = \sup\{\zeta \in [\alpha/(2k), \alpha/2] \mid N^{-1} \sum_{\ell=1}^N \mathbb{1}(\hat{\theta}^{(\ell)} \in \times_{j=1}^k [\hat{Q}_{j,\zeta}, \hat{Q}_{j,1-\zeta}]) \geq 1 - \alpha\}$
8: $\widehat{C} = \times_{j=1}^k [\hat{Q}_{j,\hat{\zeta}}, \hat{Q}_{j,1-\hat{\zeta}}]$

generically applicable method for constructing bands with simultaneous credibility.[10] Note that the proposed band is generally asymmetric in finite samples, although we later show that it lies asymptotically in the symmetric one-parameter class from Section 3. Appendix A.2 describes an alternative bootstrap band that is symmetric in finite samples.[11]

The bootstrap band is easy to implement in many applications. Valid bootstrap procedures for $\hat{\mu}$ often exist in smooth i.i.d. models (van der Vaart, 1998, chapter 23) as well as in time series models (Kilian & Lütkepohl, 2017, chapter 12). Unlike the plug-in approach, the bootstrap band does not require computation of the derivatives of $h(\cdot)$. The bootstrap band equals the product of Efron's equal-tailed percentile bootstrap confidence intervals, where the percentile is "calibrated" so that the rectangle $\times_{j=1}^k [\hat{Q}_{j,\zeta}, \hat{Q}_{j,1-\zeta}]$ covers at least a fraction $1 - \alpha$ of the bootstrap draws of $\hat{\theta}$. This can be achieved easily by bisection or other numerical solvers, since the fraction of draws contained in the rectangle is monotonically decreasing in the scalar ζ; moreover, the usual pointwise and Bonferroni bounds imply that

[10] Liu (2011, chapter 2.9) proposes a Bayesian simultaneous confidence band for linear regression. The idea of calibrating the width of a rectangular confidence set to achieve simultaneous confidence/credibility has appeared in different contexts in Kaido et al. (2016) and Gafarov et al. (2016).

[11] We make two comments about the symmetric versions of the sup-t band. First, since the algorithm to achieve symmetry is based on t-statistics, it appears less well motivated from a finite-sample Bayesian perspective, except perhaps when the posterior distribution is exactly Gaussian (as in Liu, 2011, chapter 2.9). Second, the simulations in Appendix A.5 show that for persistent VARs the asymmetric band (based on our Bayesian implementation) can perform better in finite samples than the symmetric plug-in band.

the search can be confined to $\zeta \in [\alpha/(2k), \alpha/2]$ in any finite sample.[12]

The Bayesian band has simultaneous Bayesian credibility equal to $1 - \alpha$, provided we are able to draw from the finite-sample posterior distribution of μ. To see this, let $\hat{P}$ in the algorithm denote the posterior distribution of μ given data $\mathcal{D}$, given some choice of prior. Then, by construction, the set $\widehat{C}$ satisfies $\widehat{P}(\theta \in \widehat{C}) = P(\theta \in \widehat{C} \mid \mathcal{D}) = 1-\alpha$, up to simulation error that vanishes as the number of posterior draws grows large. The simultaneous credible band $\widehat{C}$ is the product of component-wise equal-tailed credible intervals, where the tail probability has been calibrated to yield simultaneous credibility of $1 - \alpha$.

In summary, both the bootstrap and Bayes band are generally applicable and computationally convenient. They do not require explicit calculation of standard errors, and the only numerical optimization necessary is the monotonic, 1-dimensional root finding problem for $\hat{\zeta}$ in Algorithm 2.[13] The bootstrapping of $\hat{\mu}$ or posterior sampling of μ is only performed once and for all (yielding N draws). As a side product, our bootstrap/Bayes algorithm immediately reveals the *pointwise* confidence/credibility level $1 - 2\hat{\zeta}$. Appendix A.2 shows that our implementations of the sup-t band are asymptotically equivalent with the usual plug-in sup-t band under standard regularity conditions like bootstrap consistency and the Bernstein-von Mises property.

3 Comparison of popular confidence bands

In this section we compare the widths of several popular simultaneous confidence bands, e.g., Bonferroni, Šidák, projection, and sup-t. To do this, we show that commonly encountered confidence bands lie asymptotically in a one-parameter class of bands. The sup-t band is

[12] The bootstrap band is not guaranteed to deliver asymptotic refinements relative to the plug-in sup-t band. As the sup-t critical value $q_{1-\alpha}(\Sigma)$ depends on the unknown Σ, it appears difficult to achieve refinements through simple bootstrap procedures.

[13] We use the Matlab routine `fzero` in our applications.

the narrowest simultaneous confidence band within this class by definition, but we are also able to analytically rank the suboptimal bands. These results are well-known for linear models, whereas our contribution is to show that the conclusions extend to our more general *asymptotically linear* framework, which includes nonlinear models such as VARs.

COMMON BANDS. In addition to the sup-t band, many other popular confidence bands in applied work lie in the one-parameter class $\widehat{B}(c)$ defined in Section 2, for particular choices of the critical value c.

The *pointwise band* is the Cartesian product of pointwise confidence intervals, so it corresponds to $c = \chi_{1-\alpha}$.

The *Bonferroni* band is obtained by applying a Bonferroni multiple comparisons adjustment to the pointwise critical value, yielding $c = \chi_{1-\alpha/k}$.[14] Standard arguments (e.g., Alt, 1982) show that the Bonferroni band has asymptotic simultaneous coverage probability greater than $1-\alpha$ if $k > 1$, under the regularity conditions in Appendix A.1.

The *Šidák* band corresponds to that value of c which would yield asymptotic simultaneous coverage of exactly $1-\alpha$ in the special case where the elements $\hat{\theta}_j$ of $\hat{\theta}$ are uncorrelated, yielding $c = \chi_{(1-\alpha)^{1/k}}$. The Šidák (1967) theorem shows that the independent case is "least favorable" for the coverage probability of the multivariate normal distribution. Hence, under the regularity conditions in Appendix A.1, the Šidák band guarantees asymptotic simultaneous coverage of at least $1-\alpha$ regardless of the true correlation structure, but it will typically be conservative.[15]

The *θ-projection* band equals the smallest rectangle in $\mathbb{R}^k$ that contains the usual Wald confidence ellipsoid for θ; this turns out to correspond to a one-parameter band with $c = \chi_{k,1-\alpha}$. If the consistent estimator $\hat{\Sigma} \equiv \dot{h}(\hat{\mu})\hat{\Omega}\dot{h}(\hat{\mu})'$ of Σ is nonsingular, we can define the

[14]Bonferroni adjustments that are not symmetric across the elements $j = 1, \ldots, k$ are also possible, but here we stay within our one-parameter class.

[15]This argument of course relies crucially on asymptotic normality, unlike the Bonferroni adjustment.

Wald ellipse for θ:

$$\widehat{W}_\theta \equiv \left\{\tilde{\theta} \in \mathbb{R}^k \;\middle|\; n(\hat{\theta}-\tilde{\theta})'\hat{\Sigma}^{-1}(\hat{\theta}-\tilde{\theta}) \le \chi^2_{k,1-\alpha}\right\}.$$

Consider the smallest rectangle that contains this ellipse:

$$\widehat{C}_{\theta\text{-proj}} \equiv \bigtimes_{j=1}^{k} \left[\inf\nolimits_{\tilde{\theta}\in\widehat{W}_\theta} \tilde{\theta}_j, \sup\nolimits_{\tilde{\theta}\in\widehat{W}_\theta} \tilde{\theta}_j\right].$$

Since $\widehat{W}_\theta \subset \widehat{C}_{\theta\text{-proj}}$, this θ-projection confidence band automatically has asymptotic simultaneous coverage probability at least $1-\alpha$ under the regularity conditions in Appendix A.1 and provided that Σ is nonsingular. The band is conservative if $k \ge 2$. Straight-forward algebra shows that the θ-projection band is equal to the one-parameter band $\widehat{B}(\chi_{k,1-\alpha})$.[16]

The *μ-projection* band, which equals the smallest rectangle in $\mathbb{R}^k$ that contains all the values of $\theta = h(\mu)$ attained in the usual Wald confidence ellipsoid for μ, is herein shown to be asymptotically equivalent to $c = \chi_{p,1-\alpha}$. If the Wald ellipse for μ is defined as

$$\widehat{W}_\mu \equiv \left\{\tilde{\mu} \in \mathcal{M} \;\middle|\; n(\hat{\mu}-\tilde{\mu})'\hat{\Omega}^{-1}(\hat{\mu}-\tilde{\mu}) \le \chi^2_{p,1-\alpha}\right\},$$

its rectangular projection with respect to the function $h(\cdot)$ is given by[17]

$$\widehat{C}_{\mu\text{-proj}} \equiv \bigtimes_{j=1}^{k} h_j(\widehat{W}_\mu) = \bigtimes_{j=1}^{k} \left[\inf\nolimits_{\tilde{\mu}\in\widehat{W}_\mu} h_j(\tilde{\mu}), \sup\nolimits_{\tilde{\mu}\in\widehat{W}_\mu} h_j(\tilde{\mu})\right].$$

---

[16] Simply maximize/minimize $\tilde{\theta}_j$ subject to the quadratic constraint $n(\hat{\theta}-\tilde{\theta})'\hat{\Sigma}^{-1}(\hat{\theta}-\tilde{\theta}) \le \chi^2_{k,1-\alpha}$, noting that the j-th diagonal element of $\hat{\Sigma}/n$ is $\hat{\sigma}_j^2$. This result does not rely on the specific critical value $\chi^2_{k,1-\alpha}$ in the definition of the Wald ellipse for θ. Hence, the rectangular envelope of the Inoue & Kilian (2016) procedure for constructing confidence bands for VAR impulse response functions also falls within the one-parameter class, although with a non-standard critical value.

[17] The fact that $h_j(\widehat{W}_\mu) = [\inf_{\tilde{\mu}\in\widehat{W}_\mu} h_j(\tilde{\mu}), \sup_{\tilde{\mu}\in\widehat{W}_\mu} h_j(\tilde{\mu})]$ follows from continuity of $h(\cdot)$. The rectangular projection is the smallest rectangle containing the usual projection set $h(\widehat{W}_\mu) = \{h(\tilde{\mu}) \mid \tilde{\mu} \in \widehat{W}_\mu\}$, which in general is not rectangular (and is therefore hard to visualize).

This is a multi-parameter version of the Krinsky & Robb (1986) confidence interval procedure for nonlinearly transformed estimators. The μ-projection band has asymptotic simultaneous coverage probability of at least $1-\alpha$ under the regularity conditions in Appendix A.1, by the standard projection inference argument (e.g., Dufour, 1990). Typically, however, the simultaneous coverage probability will be conservative ("projection bias"). Proposition 2 in Appendix A.1.3 shows the nontrivial result that $\widehat{C}_{\mu\text{-proj}}$ equals the one-parameter band $\widehat{B}(\chi_{p,1-\alpha})$, up to terms of negligible asymptotic order, provided that Ω is nonsingular.

Finally, many bootstrap variants of the bands in the one-parameter class are covered by our theory. If a band has edges that are within asymptotic order $o_p(n^{-1/2})$ of a one-parameter band $\widehat{B}(c)$, then those two bands are equivalent for our purposes: the asymptotic coverage probability is identical, and the ratio of the widths of the component intervals tends to 1 in probability. This follows from the analysis in Appendix A.1.

Illustration of bands. Figure 1 illustrates the various confidence bands in (θ_1, θ_2) space for the case $k = 2$, omitting the μ-projection band.[18] The confidence bands are represented as rectangles due to their Cartesian product structure. The θ-projection band is the smallest rectangle that contains the Wald ellipse for θ, drawn in black. The sup-t band is the smallest band in the one-parameter class that has coverage probability at least $1-\alpha$.

Analytical comparison of bands. We are able to analytically compare the widths of bands in the one-parameter class by comparing their critical values c. We present a detailed analysis in Appendix A.1.4 but summarize the main points here.

By construction, the sup-t band is the narrowest one-parameter band that yields asymptotic simultaneous coverage. All other bands in the one-parameter class either fail to achieve simultaneous coverage (e.g., the pointwise band) or are conservative and unnecessarily wide,

[18] A similar illustration appears in Lütkepohl et al. (2015b, figure S1, supplemental material).

Two-dimensional illustration of Wald ellipse and confidence bands

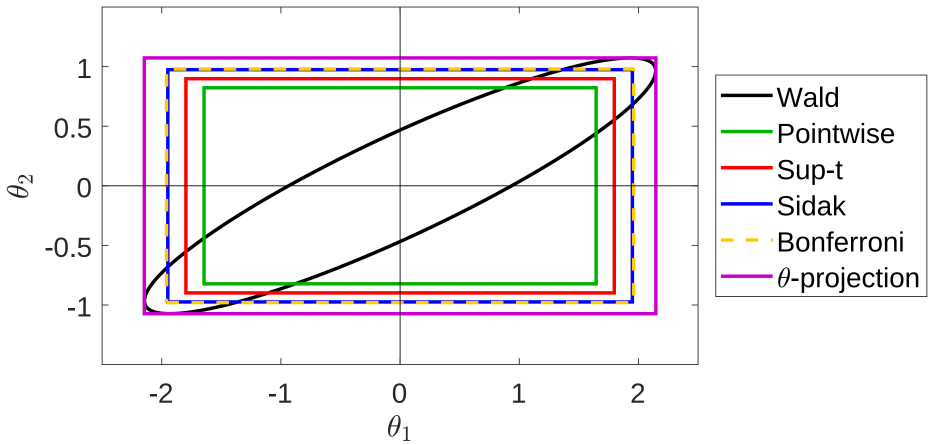

Figure 1: 90% Wald confidence ellipse (black) and rectangular confidence regions (colored rectangles) for the two-dimensional mean $\theta = (\theta_1, \theta_2)'$ of a normally distributed parameter estimator $\hat{\theta}$, given point estimate $\hat{\theta} = (0, 0)'$. The figure assumes the correlation structure $\text{Var}(\hat{\theta}_1) = 1$, $\text{Var}(\hat{\theta}_2) = 0.25$, $\text{Corr}(\hat{\theta}_1, \hat{\theta}_2) = 0.9$.

asymptotically.[19] The sup-t band is especially narrow relative to the other bands when the point estimators $\hat{\theta}_j$ are highly correlated across elements j, as is often the case in applications. If the point estimators $\hat{\theta}_j$ are mutually independent across j, the sup-t band coincides with the Šidák band and virtually coincides with the Bonferroni band in most applications.

Table 1 ranks the critical values c for different confidence bands. The bands in Table 1 are ordered in terms of relative width (narrowest at the top) in the empirically common case $\alpha \leq 0.5$ and $k \leq p$, under the weak assumptions in Appendix A.1. Sup-t is the narrowest band with correct asymptotic coverage, followed by Šidák's band and then Bonferroni. If instead $k \gg p$, then it can happen that μ-projection leads to a narrower band than Šidák or Bonferroni, although the requirements are stringent (cf. Figure 5 in Appendix A.1.5). θ-projection always leads to a wider band than Šidák and Bonferroni for $\alpha \leq 0.5$. The relative widths of the (suboptimal) projection, Bonferroni, and Šidák bands depend only on the

[19] When $k > 1$, the simultaneous coverage probability of the pointwise band equals $1 - \alpha$ if and only if the elements of $\hat{\theta}$ are asymptotically perfectly correlated.

Popular bands in the one-parameter class

Band	Critical value c	Asymptotic coverage
Pointwise	$\chi_{1-\alpha}$	$\leq 1-\alpha$
Sup-t	$q_{1-\alpha}(\Sigma)$	$= 1-\alpha$
Šidák	$\chi_{(1-\alpha)^{1/k}}$	$\geq 1-\alpha$
Bonferroni	$\chi_{1-\alpha/k}$	$\geq 1-\alpha$
θ-projection	$\chi_{k,1-\alpha}$	$\geq 1-\alpha$
μ-projection	$\chi_{p,1-\alpha}$	$\geq 1-\alpha$

Table 1: List of critical values of popular confidence bands in the one-parameter class. The last column describes each band's asymptotic coverage probability.

dimensions p and k and the significance level α, so these bands can be ranked without seeing the data, as we do in Appendix A.1.4. These results should provide analytical guidance to the literature that has used Monte Carlo experiments to evaluate bands for impulse response functions, cf. Appendix A.4.

Lemma 4 in Appendix A.1.5 reviews the important property that the width of the sup-t band increases very slowly with the number k of parameters of interest, specifically at rate $\sqrt{\log k}$. Hence, by using the sup-t band, a researcher can report results for a large number of parameters without sacrificing much power. In contrast, the width of the θ-projection band is highly sensitive to the number of parameters of interest.

4 Decision theoretic justification for the sup-t band

Is the sup-t band in some sense optimal even outside the one-parameter class analyzed in Section 3? We show that that the sup-t band is indeed the unique minimizer of *worst-case regret* among translation equivariant confidence bands. Here "worst case" refers to the choice of loss function. The sup-t band is thus a good default option when the researcher is unsure about the appropriate choice of loss function.

4.1 Gaussian limit experiment and equivariance

Our analysis focuses on a Gaussian limit experiment motivated by the preceding asymptotic analysis. We first describe this Gaussian limit experiment and a translation equivariance concept used in the decision-theoretic analysis.

Finite-sample model. For expositional simplicity, we consider a Gaussian location model with known covariance matrix, where the object of interest is a vector of linear combinations of the unknown mean. For our purposes, it is without loss of generality to assume that the covariance matrix of the data is the identity matrix. Hence, we assume we observe a single multivariate normal draw

$$X \sim N_p(\mu, I_p), \tag{7}$$

where the mean vector $\mu \in \mathbb{R}^p$ is unknown. The distribution of the data under the parameter $\mu \in \mathbb{R}^p$ is denoted P_μ. Let $G = (g_1, \ldots, g_k)' \in \mathbb{R}^{k\times p}$ be a fixed, known matrix, where k may be greater than, equal to, or smaller than p. We assume that each row of G is nonzero, i.e., $g_j \neq \mathbf{0}_p$ for all j. We seek a simultaneous confidence band for the k-dimensional parameter of interest $\theta \equiv G\mu = (g_1'\mu, \ldots, g_k'\mu)'$.[20]

Decision problem: action space, decision rule, loss function. We now specify the decision-maker's action space and loss function. The decision-maker's *action space* is the set $\mathcal{R}$ of closed k-dimensional rectangles (or bands):

$$\mathcal{R} \equiv \left\{ \times_{j=1}^{k} [a_j, b_j] \;\middle|\; a_j, b_j \in \mathbb{R},\, a_j \le b_j,\, j = 1, \ldots, k \right\}.$$

[20]We view the finite-sample Gaussian model as the relevant limit experiment corresponding to the delta method asymptotics of Section 3. Since the finite-sample analysis hinges on properties of $\hat\theta \equiv GX \sim N_k(\theta, G\,\mathrm{Var}(X)G')$, and we do not restrict G, there is no loss of generality in assuming $\mathrm{Var}(X) = I_p$.

A *decision rule* is a map $C\colon \mathbb{R}^p \to \mathcal{R}$ from data to k-dimensional bands. We refer to C as a confidence band and we restrict ourselves to those functions that guarantee a confidence level of at least $1-\alpha$. That is, the set of confidence bands under consideration is

$$\mathcal{C}_{1-\alpha} \equiv \left\{ C\colon \mathbb{R}^p \to \mathcal{R} \;\middle|\; \inf_{\mu \in \mathbb{R}^p} P_\mu(G\mu \in C(X)) \geq 1-\alpha \right\}.$$

In order to define the loss function, we first introduce the class of functions $\mathcal{L}$ that are increasing with respect to the partial order on $\mathbb{R}^k$:

$$\mathcal{L} \equiv \Big\{ L\colon \mathbb{R}^k_+ \to \mathbb{R}_+ \;\Big|\; L(r) \leq L(\tilde{r}) \text{ for all } r, \tilde{r} \in \mathbb{R}^k_+ \text{ s.t. } r \leq \tilde{r} \text{ for all elements,}$$
$$L(r) > 0 \text{ whenever } r > 0 \text{ for all elements} \Big\}.$$

The decision-maker is assumed to have a *loss function* of the form

$$\text{Loss}(R;\mu) \equiv L(b_1 - a_1, \ldots, b_k - a_k) \quad \text{ for } \mu \in \mathbb{R}^p,\ R = \times_{j=1}^k [a_j, b_j] \in \mathcal{R}, \tag{8}$$

which implies that i) the decision maker penalizes bands with large lengths of the component intervals, and ii) the ranking between bands does not depend on μ. In a slight abuse of notation, we write $L(R)$ instead of $\text{Loss}(R;\mu)$ for $L \in \mathcal{L}$ and $R \in \mathcal{R}$.

Translation equivariant bands. For any loss function L in the component-wise length class $\mathcal{L}$, the decision problem of constructing a confidence band in a Gaussian location model is *invariant* under translations of the data X, see Appendix A.3. Motivated by the invariance of the decision problem, we consider the following class of translation equivariant bands:

$$\mathcal{C}_{eq} \equiv \left\{ C\colon \mathbb{R}^p \to \mathcal{R} \;\middle|\; C(x+\lambda) = G\lambda + C(x) \text{ for all } x, \lambda \in \mathbb{R}^p \right\}.$$

Here we have used the Minkowski sum $S + y \equiv \{s + y \mid s \in S\}$, for any set $S \subset \mathbb{R}^p$ and vector $y \in \mathbb{R}^p$. Lemma 6 in Appendix A.3 shows the simple result that[21]

$$\mathcal{C}_{eq} = \{C \colon \mathbb{R}^p \to \mathcal{R} \mid C(x) = Gx + R,\, R \in \mathcal{R}\}\,.$$

Thus, translation equivariant bands are of the form $C(x) = \times_{j=1}^{k} [\hat{\theta}_j - a_j, \hat{\theta}_j + b_j]$ for non-random $a_j, b_j \geq 0$, $j = 1, \ldots, k$. As far as we know, all existing decision theoretic analyses of simultaneous confidence bands impose translation equivariance, cf. the references in Section 1. Bands $\widehat{B}(c)$ in the one-parameter class analyzed in Section 3 are all translation equivariant in the case of linear $h(\cdot)$ and known variance of $\hat{\theta}$. Note that translation equivariance does not require symmetry of the confidence band around $\hat{\theta}$, nor does it require the one-parameter structure.

Risk of translation equivariant bands. The risk of a translation equivariant band equals the realized loss, which depends on neither the data X nor the parameter μ: It is easy to verify that $L(C(X)) = L(R)$ for every realization of the data and that $P_\mu(G\mu \in C(X)) = P(GZ \in R)$, where $Z \sim N_p(\mathbf{0}_p, I_p)$. Moreover, the coverage constraint in the definition of $\mathcal{C}_{1-\alpha}$ reduces the class of bands under consideration to

$$\mathcal{C}_{1-\alpha} \cap \mathcal{C}_{eq} = \Big\{C \colon \mathbb{R}^p \to \mathcal{R} \mid C(x) = Gx + R,\, R \in \mathcal{R}_{1-\alpha}\Big\},$$

where

$$\mathcal{R}_{1-\alpha} \equiv \Big\{R \in \mathcal{R} \;\Big|\; P(GZ \in R) \geq 1 - \alpha,\, Z \sim N_p(\mathbf{0}_p, I_p)\Big\}\,.$$

[21] A similar result appears in Lehmann & Romano (2005, chapter 9.4).

4.2 Sup-t band minimizes worst-case regret

We now show that the sup-t band minimizes *worst-case regret* among translation equivariant confidence bands. Given a loss function, we define regret as the ratio of the actual loss to the smallest possible loss. We consider a decision-maker looking for a confidence band that provides the best guarantee on regret across a range of reasonable loss functions. We restrict the class of loss functions to be homogeneous of degree 1 in lengths.

Smallest possible loss in $\mathcal{R}_{1-\alpha}$. Translation equivariance reduces the decision problem to a finite-dimensional optimization. Given a *particular* loss function $L \in \mathcal{L}$ in component lengths, an optimal equivariant band equals $C(x) = Gx + R^*$, where

$$R^* \in \operatorname*{argmin}_{R \in \mathcal{R}_{1-\alpha}} L(R). \tag{9}$$

This is a $2k$-dimensional optimization problem, with the parameters being the endpoints of the k component intervals of the rectangle R. In general, the solution to the program (9) is not available in closed form.[22] The main practical challenge is that the constraint set $\mathcal{R}_{1-\alpha}$ requires evaluation of the probability that a k-dimensional correlated Gaussian vector GZ lies in a given rectangle. Freyberger & Rai (2018) develop computational tools to numerically solve problems of the form (9) (as well as problems with certain different types of loss functions).

Minimum worst-case regret. In practice, it is difficult to decide on a particular loss function. Researchers reporting confidence bands may recognize that different audience members will care about different features of the parameter vector $\theta = (\theta_1, \dots, \theta_k)'$: For

[22] In certain special cases the solution is simple. For example, under the worst-case length loss function $L(r) = \max_j |r_j|$, the optimal band is the equal-width band of Gafarian (1964).

example, one group of people cares only about θ_1, another group wishes to compare the magnitudes of θ_1 and θ_2, and a third group is interested in the shape of the function $j \mapsto \theta_j$.

In light of this observation, we assume that our decision-maker looks for an equivariant confidence band $C(x) = Gx + R$ that minimizes the *worst-case (relative) regret*

$$\sup_{L \in \mathcal{L}_H} \frac{L(R)}{\inf_{\tilde{R} \in \mathcal{R}_{1-\alpha}} L(\tilde{R})}$$

over R. The worst case is taken over all loss functions in a class $\mathcal{L}_H$, to be defined below. Although we think our definition of worst-case regret is quite reasonable, we were not able to find applications of this exact criterion in the literature.[23] Our definition of regret in terms of a ratio rather than a difference is in line with the relative width comparisons in Section 3.

Loss functions under consideration. We assume that the decision-maker only considers loss functions that are homogeneous of degree 1 in component interval lengths:

$$\mathcal{L}_H \equiv \left\{ L \in \mathcal{L} \;\middle|\; L(\beta r) = \beta L(r) \text{ for all } r \in \mathbb{R}^k_+,\ \beta > 0 \right\}.$$

This class includes the weighted-average loss functions $L(r) = \sum_{j=1}^k w_j r_j$ for $w_j \geq 0$, $\sum_j w_j = 1$ (Hoel, 1951), the maximum-length loss function $L(r) = \max_j |r_j|$ (Gafarian, 1964), and the Euclidean loss $L(r) = \|r\|$ (indeed, any norm on $\mathbb{R}^k$ could be used). The class $\mathcal{L}_H$ excludes the loss function $L(r) = \prod_{j=1}^k r_j$, which measures the volume of a rectangle in $\mathbb{R}^k$. Since the main motivation for using confidence bands is to visualize k-dimensional uncertainty in a two-dimensional figure, the volume loss function is not of primary concern.

[23] Notice that this notion of worst-case regret is different from the common notion of minimax regret in decision theory, where the worst case is taken over possible values for the unknown parameter, here denoted μ (Berger, 1985, chapter 5). These two notions can be reconciled if we think of the audience's preferences L as an unknown parameter that implicitly enters the overall loss function. Our criterion is also different from that of Naiman (1987), who considers the worst case over possible covariate values at which a linear regression line is to be evaluated.

Main result. We are now ready to state the result on worst-case regret. In the present Gaussian limit experiment, we define the sup-t band as

$$C_{\text{sup}}(x) \equiv Gx + R_{\text{sup}}, \quad R_{\text{sup}} \equiv \bigtimes_{j=1}^{k} \left[-\|g_j\| q_{1-\alpha}(GG'), \|g_j\| q_{1-\alpha}(GG')\right],$$

where g_j' is the j-th row of $G = (g_1, \dots, g_k)'$, $\|\cdot\|$ is the Euclidean norm, and $q_{1-\alpha}(\cdot)$ is defined in equation (6).

Proposition 1. *For any $R \in \mathcal{R}_{1-\alpha}$ such that $R \neq R_{\text{sup}}$,*

$$\sup_{L \in \mathcal{L}_H} \frac{L(R)}{\inf_{\tilde{R} \in \mathcal{R}_{1-\alpha}} L(\tilde{R})} > \sup_{L \in \mathcal{L}_H} \frac{L(R_{\text{sup}})}{\inf_{\tilde{R} \in \mathcal{R}_{1-\alpha}} L(\tilde{R})} = \frac{q_{1-\alpha}(GG')}{\chi_{1-\alpha}}.$$

Proof. See Appendix A.7.7. □

Thus, the computationally convenient sup-t band is the translation equivariant band which provides the best possible guarantee on regret across a range of reasonable loss functions. The sup-t band will generally not be optimal—i.e., solve the optimization problem (9)—for a *particular* loss function L.[24] However, Proposition 1 gives a sense in which, among the many possible equivariant confidence bands, the sup-t band is a particularly good default choice in applications where there is no single appropriate loss function. In particular, any equivariant simultaneous confidence band other than the sup-t band yields strictly larger regret than the sup-t band for *some* loss function $L \in \mathcal{L}_H$.

[24] The method of Piegorsch (1985b) may be used to derive a particular loss function with respect to which the sup-t band is optimal. It is interesting that the sup-t band is very close to the numerically computed optimal bands in the empirical applications of Freyberger & Rai (2018, section 4).

5 Application: Impulse response functions

In this section we apply our theory on the relative performance of popular simultaneous confidence bands. As discussed above, our framework applies to many econometric settings where the researcher seeks to visualize the joint uncertainty across several parameters. Our main application below shows that the sup-t band improves on popular approaches to constructing simultaneous confidence bands for impulse response functions in VARs. To illustrate the flexibility of our set-up, the additional application in Appendix A.6 uses a simultaneous confidence band to visualize joint uncertainty in a sensitivity analysis of a linear regression coefficient estimated using different sets of control variables.

Here we estimate the effects of monetary policy shocks on output using recursive and instrumental variable identification strategies for VARs, as in Gertler & Karadi (2015).[25] The impulse reponse analysis in their paper suggests that a monetary shock that increases the one-year government bond yield by 20 basis points on impact causes industrial production to gradually decline, the maximum decline of 0.4% occurring about 2 years after the shock. Their reported confidence bands, however, do not account for multiple comparisons.

The VAR literature has developed several procedures for conducting simultaneous inference on impulse response functions. In our view, theoretical comparisons and recommendations for practice are still lacking (see Appendix A.4 for a literature review). Our nonlinear framework in Section 3 allows us to analytically compare the various confidence bands and provides theoretical support for the sup-t band. To our knowledge, other comprehensive comparisons of confidence bands for impulse response functions rely exclusively on Monte Carlo experiments for particular data generating processes (Lütkepohl et al., 2015a,b). Also, as far as we know, the sup-t band has not previously been exploited to conduct inference on

[25] Their paper provides a very detailed analysis of the channels of transmission of monetary policy, but we focus on the effect on output.

impulse response functions.[26]

Our empirical analysis shows that the economic conclusions of Gertler & Karadi (2015) remain valid when adjusting for multiple comparisons using the sup-t band, but not using the wider Bonferroni or projection bands. The suboptimal—but commonly used—Bonferroni or projection bands cannot rule out the possibility that the effect of a monetary shock on output is zero at all horizons. In contrast, with the sup-t band, the effect of surprise monetary policy tightenings on output is significantly negative 1–3 years after the shock. We now present the details of the VAR specification and the empirical results. See Appendix A.4 for a review of the standard structural VAR model.

Specification. We use the data and specification in Gertler & Karadi (2015) to perform inference on the impulse response function of industrial production (IP) to a contractionary monetary policy shock. The VAR baseline specification uses monthly U.S. data on industrial production, the consumer price index, the 1-year government bond yield, and the Excess Bond Premium (EBP) of Gilchrist & Zakrajšek (2012), from July 1979 to June 2012.[27] The VAR is estimated in levels and has 12 lags. The number of model parameters is then $p = 206$ (+4 for the external instrument specification mentioned below). We consider impulse responses out to 36 months, i.e., $k = 37$ parameters of interest.[28]

Gertler & Karadi (2015) consider two identification schemes. First, as a baseline, they use a *recursive* identification which imposes exclusion restrictions on the impact responses: Industrial production and the consumer price index do not respond to monetary policy shocks on impact, and the bond yield does not respond to EBP shocks on impact. Their preferred specification, however, identifies the monetary shock using an *external instrument*

[26] However, we argue in Appendix A.2 that the "adjusted Bonferroni/Wald" methods of Lütkepohl et al. (2015a,b) are closely related to a bootstrap implementation of the sup-t band.

[27] The data is available online: `https://www.aeaweb.org/articles?id=10.1257/mac.20130329`

[28] Appendix A.4 gives further details of the bootstrap, posterior sampling, and other technical aspects.

given by changes in federal funds futures prices in small time windows around scheduled Federal Open Market Committee announcements.[29]

Results. Figure 2 shows that the sup-t band for the impulse response function of output is about 35% narrower than the Šidák and Bonferroni bands in the external instrument specification. The thick line shows the point estimate (the same as in Gertler & Karadi, 2015), while the thin lines show various 68% simultaneous confidence bands, as well as the 68% pointwise band (the narrowest).[30] While the Bonferroni band contains zero at all horizons, the plug-in sup-t band does not contain zero at any horizon from 13–36 months. Hence, the sup-t band allows us to reject (at level $\alpha = 32\%$) the null hypothesis that the monetary shock has no effect on output, while the Bonferroni band does not allow such rejection. In fact, the sup-t band allows the researcher to reject the null that the impulse response function is nonnegative at *some* horizon from 13–36 months. Moreover, Figure 2 demonstrates the general utility of simultaneous confidence bands: The *pointwise* significantly positive response at the 2-month horizon is not statistically significant when we adjust for multiple comparisons across horizons, even using the sup-t band.

Figure 3 compares the plug-in, bootstrap, and Bayes implementations of the sup-t band for the case of recursive identification.[31] The three bands are similar at short horizons but diverge somewhat at longer horizons. Even adjusting for multiple comparisons, the output response at horizons 1–3 months is significantly positive (the impact response is zero by

[29]We use the three-month-ahead futures series "ff4" preferred by Gertler & Karadi (2015). The data for this series starts in January 1990.

[30]The figure shows that the Šidák and Bonferroni bands virtually coincide, although the latter is slightly wider, which conforms to the theory in Section 3. Appendix A.4 reports that the θ-projection and (especially) μ-projection bands are much wider than the other bands, as theory predicts.

[31]We use a homoskedastic recursive residual bootstrap and maximally diffuse normal-inverse-Wishart prior. See Appendix A.4 for details of the bootstrap/Bayes implementations and for plots of pointwise, Šidák, Bonferroni, and projection bands for the recursive specification. We do not report bootstrap/Bayes bands for the external instrument specification, as no off-the-shelf Bayesian implementation currently exists (Caldara & Herbst, 2016, assume the instrument is white noise, which is counterfactual in our application).

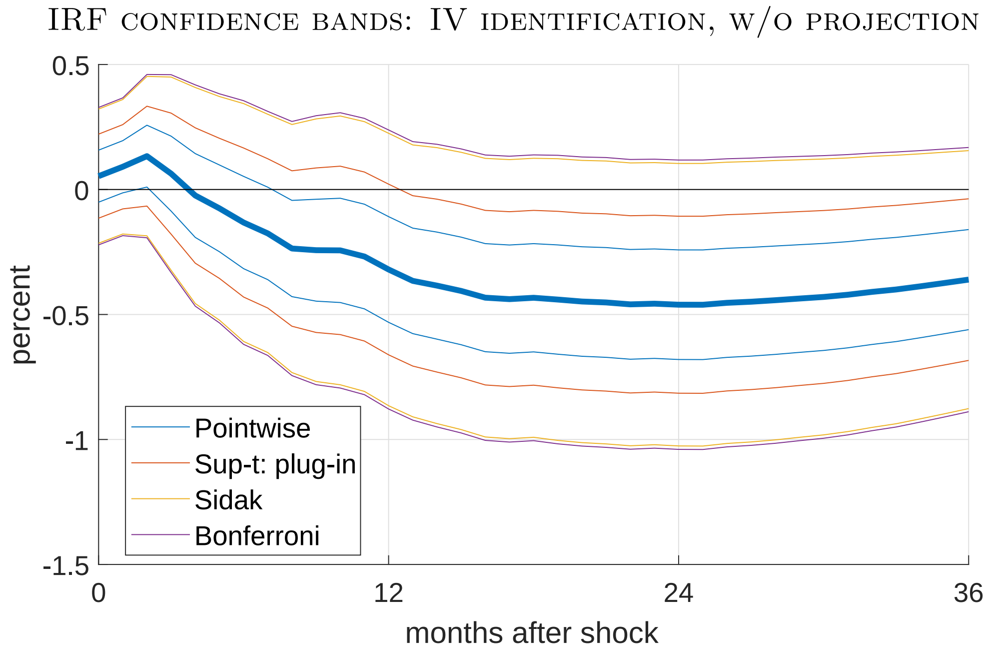

Figure 2: 68% confidence bands for impulse response function of industrial production to a 1 standard deviation contractionary monetary policy shock, external instrument identification, without projection bands. Thick line: point estimate. Thin lines: confidence bands (legend is ordered with respect to width). Šidák and Bonferroni bands virtually coincide.

assumption). Gertler & Karadi (2015) conclude that the difference between the recursively identified results and the external instrument results call into question the recursive exclusion restrictions.

While our results are consistent with some of the simulation evidence provided by Lütkepohl et al. (2015a,b), the analytical perspective in this paper yields additional insights. First, we showed that the narrowness of the Bonferroni approach relative to projection approaches is not accidental. Second, the theoretical viewpoint suggested an improved band, the sup-t band, which had not been exploited in a VAR context, despite being well known in nonparametric econometrics. Finally, we proposed a Bayesian version of the sup-t band, which may be particularly attractive due to the prevalence of Bayesian procedures in VAR studies.

Simulation study. Appendix A.5 presents a modest simulation study of the small-sample coverage probability and average width of the various simultaneous confidence bands for VAR

IRF confidence bands: recursive identification, sup-t implementations

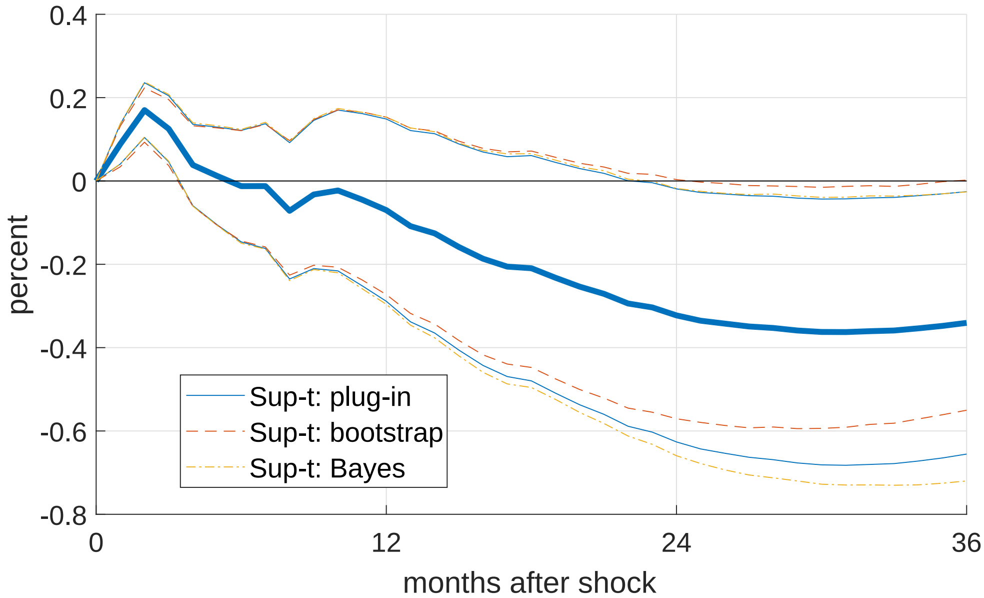

Figure 3: 68% confidence bands for the impulse response function of industrial production to a 1 standard deviation contractionary monetary policy shock, recursive identification, different sup-t implementations. Thick line: point estimate. Thin lines: confidence bands.

impulse response functions. We consider bivariate VARs identified recursively or using external instruments. We find that all versions of the sup-t band yield satisfactory simultaneous coverage probability when the data is moderately persistent. For highly persistent data, only the Bayesian sup-t band performs adequately. The sup-t bands tend to be 20–25% narrower than the Bonferroni and Šidák bands for simultaneous confidence level $1 - \alpha = 68\%$, and 10–20% narrower for $1 - \alpha = 90\%$. Projection bands tend to be unnecessarily wide.

6 Discussion and extensions

Our analysis provides analytical comparisons between popular simultaneous confidence bands in a generally applicable nonlinear framework. The sup-t band emerges as a good default choice, for three reasons. First, it dominates Šidák, Bonferroni, and projection strategies asymptotically, and we have demonstrated that the gain can be large in applications. Second, it uniquely minimizes worst-case regret among translation equivariant confidence bands when the researcher is unsure about the appropriate choice of loss function, for example because different audience members have different parameters of interest. Third, the sup-t band is usually quickly computable in its plug-in, bootstrap, or Bayesian implementations. The Bayesian band is attractive for finite-sample Bayesian inference, but all implementations are asymptotically equivalent from a frequentist perspective under weak conditions.

Extension: generalized error rate control. In some applications it is desirable to replace the simultaneous coverage requirement (2) with the requirement that the band $\widehat{C}$ should with probability $1-\alpha$ cover at least $k-m$ of the parameters, asymptotically:

$$\liminf_{n\to\infty} P\left(\sum_{j=1}^{k} \mathbb{1}(\theta_j \notin \widehat{C}_j) \leq m\right) \geq 1-\alpha,$$

which reduces to condition (2) for $m = 0$. The above condition is referred to as controlling the *generalized familywise error rate* in the multiple testing literature (Romano et al., 2010, section 8.2; Wolf & Wunderli, 2015). For example, in some applications of Bayesian impulse response function analysis, it may be more useful to be 95% sure that the band contains the true responses for at least 20 out of 25 horizons, rather than being 68% sure that the band covers the true responses for *all* horizons.[32]

It is straight-forward to modify the sup-t implementations in this paper to impose the generalized error rate coverage constraint for a given m. Simply replace all instances of the "max" operator with the $(m+1)$-th order statistic. Similarly, for the quantile-based

[32] A distinct but related idea is to control the weighted average coverage probability across different horizons, see Wasserman (2006, sec. 5.8). We leave this possibility to future research.

bootstrap and Bayes bands, we replace the definition of $\hat{\zeta}$ in Algorithm 2 with

$$\hat{\zeta} = \sup\left\{\zeta \in \left[\frac{\alpha}{2k}, \frac{\alpha}{2}\right] \;\middle|\; \frac{1}{N}\sum_{\ell=1}^{N} \mathbb{1}\left(\sum_{j=1}^{k} \mathbb{1}(\hat{\theta}_j^{(\ell)} \notin [\hat{Q}_{j,\zeta}, \hat{Q}_{j,1-\zeta}]) \leq m\right) \geq 1-\alpha\right\}.$$

Future research directions. Our analysis focused on the case of point identification. More research is needed on partially identified applications where confidence bands are of interest, such as VAR analysis under sign restrictions. The results of this paper remain useful for inference on point identified features of the identified set.[33] Moreover, the Bayesian sup-t band may be used for subjective Bayesian analysis even under partial identification.[34]

It would be useful to further investigate the small-sample properties of the sup-t band. Intuitively, the accuracy of the sup-t band will depend on (i) how well $n^{-1}\hat{\Sigma}$ estimates the variance of $\hat{\theta}$, and (ii) how close the distribution of $\max_j \hat{\Sigma}_{jj}^{-1/2}\sqrt{n}|\hat{\theta}_j - \theta_j|$ is to the distribution of a maximum of absolute values of correlated normal variables. When $k = \dim(\theta)$ is large, it may be possible to improve on the plug-in estimator of $\hat{\Sigma}$ using shrinkage. Furthermore, asymptotics as $k \to \infty$ for the maximum t-statistic—as used in the Gaussian process literature—may yield higher-order refinements of the fixed-k limit theory.

We assumed smoothness of the transformation $h(\cdot)$ mapping underlying model parameters into parameters of interest. In highly nonlinear problems, it is possible that the delta method linearization of $h(\cdot)$ used in this paper is an unreliable guide to finite-sample performance. In such cases, alternative asymptotic sequences that do not imply asymptotic linearity may yield more useful results (e.g., Andrews & Mikusheva, 2016). In applications where continuous differentiability of $h(\cdot)$ fails entirely at economically plausible parameter values, our limit theory must be appropriately modified (e.g., Kitagawa et al., 2017).

[33] E.g., the largest VAR impulse responses in the identified set at certain horizons (Giacomini & Kitagawa, 2015; Gafarov et al., 2016), as long as the continuous differentiability assumption is satisfied.

[34] Baumeister & Hamilton (2018) make this argument for the pointwise credible band.

References

Alt, F. B. (1982). Bonferroni Inequalities and Intervals. In S. Kotz & N. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, volume 1 (pp. 294–300). Wiley.

Andrews, I. & Mikusheva, A. (2016). A Geometric Approach to Nonlinear Econometric Models. *Econometrica*, *84*(3), 1249–1264.

Baumeister, C. & Hamilton, J. D. (2018). Inference in Structural Vector Autoregressions When the Identifying Assumptions are Not Fully Believed: Re-evaluating the Role of Monetary Policy in Economic Fluctuations. *Journal of Monetary Economics*. Forthcoming.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer.

Bruder, S. & Wolf, M. (2018). Balanced Bootstrap Joint Confidence Bands for Structural Impulse Response Functions. *Journal of Time Series Analysis*. Forthcoming.

Caldara, D. & Herbst, E. (2016). Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs. Finance and Economics Discussion Series 2016-049, Board of Governors of the Federal Reserve System.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics*, *42*(5), 1787–1818.

Chernozhukov, V., Fernández-Val, I., & Melly, B. (2013). Inference on Counterfactual Distributions. *Econometrica*, *81*(6), 2205–2268.

Dufour, J.-M. (1990). Exact Tests and Confidence sets in Linear Regressions with Autocorrelated Errors. *Econometrica*, *58*(2), 475–494.

Freyberger, J. & Rai, Y. (2018). Uniform confidence bands: characterization and optimality. *Journal of Econometrics*, *204*(1), 119–130.

Gafarian, A. V. (1964). Confidence Bands in Straight Line Regression. *Journal of the American Statistical Association*, *59*(305), 182–213.

Gafarov, B., Meier, M., & Montiel Olea, J. L. (2016). Projection Inference for Set-Identified SVARs. Manuscript, Columbia University.

Gertler, M. & Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, *7*(1), 44–76.

Giacomini, R. & Kitagawa, T. (2015). Robust inference about partially identified SVARs. Manuscript, University College London.

Gilchrist, S. & Zakrajšek, E. (2012). Credit Spreads and Business Cycle Fluctuations. *American Economic Review*, *102*(4), 1692–1720.

Hansen, P. R. (2005). A Test for Superior Predictive Ability. *Journal of Business & Economic Statistics*, *23*(4), 365–380.

Hoel, P. G. (1951). Confidence Regions for Linear Regression. In Neyman, J. (Ed.), *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*, (pp. 75–81). University of California Press.

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

Inoue, A. & Kilian, L. (2013). Inference on impulse response functions in structural VAR models. *Journal of Econometrics*, *177*(1), 1–13.

Inoue, A. & Kilian, L. (2016). Joint confidence sets for structural impulse responses. *Journal of Econometrics*, *192*(2), 421–432.

Jordà, Ò. (2009). Simultaneous Confidence Regions for Impulse Responses. *Review of Economics and Statistics*, *91*(3), 629–647.

Kaido, H., Molinari, F., & Stoye, J. (2016). Confidence Intervals for Projections of Partially Identified Parameters. Cemmap working paper CWP02/16.

Kilian, L. & Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Kitagawa, T., Montiel Olea, J. L., & Payne, J. (2017). Posterior Distribution of Nondifferentiable Functions. Manuscript, Columbia University.

Krinsky, I. & Robb, A. L. (1986). On Approximating the Statistical Properties of Elasticities. *Review of Economics and Statistics*, *68*(4), 715–719.

Lee, S., Okui, R., & Whang, Y.-J. (2017). Doubly Robust Uniform Confidence Band for the Conditional Average Treatment Effect Function. *Journal of Applied Econometrics*, *32*(7), 1207–1225.

Lehmann, E. L. & Romano, J. P. (2005). *Testing Statistical Hypotheses* (3rd ed.). Springer Texts in Statistics. Springer.

List, J. A., Shaikh, A. M., & Xu, Y. (2016). Multiple Hypothesis Testing in Experimental Economics. Manuscript, University of Chicago.

Liu, W. (2011). *Simultaneous Inference in Regression*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.

Lütkepohl, H., Staszewska-Bystrova, A., & Winker, P. (2015a). Comparison of methods for constructing joint confidence bands for impulse response functions. *International Journal of Forecasting*, *31*(3), 782–798.

Lütkepohl, H., Staszewska-Bystrova, A., & Winker, P. (2015b). Confidence Bands for Impulse Responses: Bonferroni vs. Wald. *Oxford Bulletin of Economics and Statistics*, *77*(6), 800–821.

Naiman, D. Q. (1984). Average Width Optimality of Simultaneous Confidence Bounds. *Annals of Statistics*, *12*(4), 1199–1214.

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. (1985a). Admissible and Optimal Confidence Bands in Simple Linear Regression. *Annals of Statistics*, *13*(2), 801–810.

Piegorsch, W. W. (1985b). Average-Width Optimality for Confidence Bands in Simple Linear Regression. *Journal of the American Statistical Association*, *80*(391), 692–697.

Romano, J. P., Shaikh, A. M., & Wolf, M. (2010). Hypothesis Testing in Econometrics. *Annual Review of Economics*, *2*(1), 75–104.

Romano, J. P. & Wolf, M. (2005). Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica*, *73*(4), 1237–1282.

Romano, J. P. & Wolf, M. (2007). Control of generalized error rates in multiple testing. *Annals of Statistics*, *35*(4), 1378–1408.

Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, *62*(318), 626–633.

Sims, C. A. & Zha, T. (1999). Error Bands for Impulse Responses. *Econometrica*, *67*(5), 1113–1155.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer Texts in Statistics. Springer.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, *68*(5), 1097–1126.

Wolf, M. & Wunderli, D. (2015). Bootstrap Joint Prediction Regions. *Journal of Time Series Analysis*, *36*(3), 352–376.

Working, H. & Hotelling, H. (1929). Application of the Theory of Error to the Interpretation of Trends. *Journal of the American Statistical Association*, *24*(Supplement), 73–85.