Simultaneous Confidence Bands: Theoretical Comparisons and Suggestions for Practice*

José Luis Montiel Olea
Columbia University

Mikkel Plagborg-Møller
Harvard University

April 18, 2017

Abstract: Simultaneous confidence bands are used in applied work to visualize estimation uncertainty for vector-valued parameters. Although many confidence bands have been proposed—e.g., Bonferroni, projection, and sup-t bands—theoretical comparisons and practical recommendations are lacking outside the linear regression model. In a general nonlinear setting, we show that commonly reported confidence bands have the same form, asymptotically: a consistent point estimator for the parameter of interest plus/minus c times the vector of coordinate-wise standard errors. The sup-t band is known to be the narrowest band inside this one-parameter family that achieves simultaneous coverage. We show that, additionally, the sup-t band uniquely minimizes "worst case" regret among *all* translation equivariant bands, where the worst case is taken over possible loss functions in coordinate-wise lengths. Hence, the sup-t band is a good default choice when the researcher does not know the audience's preferences. We propose a simple Bayesian implementation of the sup-t band, which has exact finite-sample simultaneous credibility and is often asymptotically equivalent with standard plug-in or bootstrap implementations. We apply the sup-t band to two settings where it has been overlooked: impulse response function estimation and sensitivity analysis in linear regression. In our applications, the sup-t band is at least 15–35% narrower than other simultaneous bands.

Keywords: Bonferroni adjustment, projection inference, regret, simultaneous Bayesian credibility, simultaneous inference, sup-t confidence band. *JEL codes:* C11, C12, C32, C44.

*Email: montiel.olea@gmail.com and plagborg@fas.harvard.edu. We are grateful for comments from Gary Chamberlain, Max Farrell, Chris Hansen, Lutz Kilian, Roger Koenker, Michal Kolesár, Simon Lee, Azeem Shaikh, Neil Shephard, Chris Sims, Jim Stock, Elie Tamer, Harald Uhlig, and seminar participants at Brown, Chicago, Harvard, Illinois Urbana-Champaign, Rotterdam, Seoul National, and Tilburg universities.

1 Introduction

Simultaneous confidence bands are used in applied work to visualize estimation uncertainty for vector-valued parameters. A confidence band is a rectangular confidence set, i.e., Cartesian product of intervals. This feature makes it straight-forward to plot the band regardless of the dimension of the parameter of interest, in contrast to confidence ellipsoids. Unlike pointwise confidence bands, simultaneous confidence bands cover the *entire* true parameter vector with a pre-specified probability. Hence, simultaneous confidence bands are useful whenever researchers wish to make comparisons across different parameters of interest, such as different points at which a function is evaluated. A confidence band is often more graphically intuitive and versatile than a joint hypothesis test, especially when the researcher does not know the audience's precise hypotheses of interest.

While many simultaneous confidence bands are used in practice, there exists little theory to select among these, at least outside the linear regression model. We review the literature below. Bands commonly 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 decision theoretic analyses of confidence bands are not directly applicable to nonlinear econometric models. Moreover, in some applications, such as confidence band construction for impulse response functions in VAR analysis, practitioners have hitherto relied exclusively on simulation evidence to select among different bands.

We fill this gap by providing analytical comparisons of popular simultaneous confidence bands in an empirically relevant nonlinear econometric framework. 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. We make three contributions to the literature.

First, we analytically compare the relative widths of popular confidence bands under weak assumptions, by showing that many bands 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. This class contains the pointwise, Bonferroni, Šidák, projection, and sup-t bands. The sup-t band is known to be the narrowest simultaneous confidence band (in every realization of the data) within this class. We show that when the estimators of the individual parameters of interest are highly dependent, the sup-t band is substantially narrower than the Bonferroni band. We also show that the Bonferroni band is narrower than projection-based bands, unless the number of parameters of interest is orders

of magnitude larger than the dimension of the underlying model.

Second, 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. We establish the decision theoretic result in a linear Gaussian model motivated by our asymptotic analysis.

Third, we propose a Bayesian implementation of the sup-t band with finite-sample simultaneous credibility $1-\alpha$. The band equals the Cartesian product of component-wise equal-tailed credible intervals, where we "calibrate" the tail probability to achieve the desired *simultaneous* credibility for the band. This procedure is applicable in any setting where we are able to draw from the finite-sample posterior distribution of the underlying model parameters. If the posterior satisfies a Bernstein-von Mises property, we show that the Bayesian band is asymptotically equivalent with standard plug-in and bootstrap implementations of the sup-t band. The plug-in, bootstrap, and Bayes sup-t bands are quickly computable in many econometric models, as they avoid challenging numerical optimization.

We illustrate the general applicability of our methods and theoretical results through two empirical applications. Our main application computes simultaneous confidence bands for impulse response functions in a Vector Autoregression (VAR) identified by exclusion restrictions or by an external instrument.[1] While many recent papers have constructed confidence bands for VAR analysis, there is no analytical guidance for choosing among different bands in this nonlinear setting. Following Gertler & Karadi (2015), we estimate the impulse response function of real output to a monetary policy shock. In this application, the sup-t band, which has been surprisingly overlooked in the VAR literature, is about 35% narrower than the oft-used Bonferroni band. Our second application constructs confidence bands for a linear regression coefficient estimated using different sets of control variables (sensitivity analysis). Here the sup-t band is 15% narrower than Bonferroni.

LITERATURE. Since the original contribution by Working & Hotelling (1929), the literature on simultaneous confidence bands has grown vastly. Decision-theoretic contributions include

[1]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).

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 popular projection, Šidák, and Bonferroni bands in a general non-linear framework relevant to many econometric 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 (2017) 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.

The sup-t band has a long tradition in nonparametric regression and density estimation. Wasserman (2006, chapter 5.7) gives a textbook treatment. Recent papers in econometrics have developed sup-t-type bands with frequentist validity for nonparametric instrumental variable regression (Horowitz & Lee, 2012), counterfactual distributions (Chernozhukov et al., 2013), densities (Chernozhukov et al., 2014), and conditional average treatment effects (Lee et al., 2017). Unlike these papers, we focus on a general but finite-dimensional setting, and we provide analytical comparisons of several popular bands.

The issue of how to construct optimal simultaneous confidence bands has many parallels 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. 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 is optimal within this class under certain equivariance conditions. 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 shows that many popular confidence bands lie in a one-parameter class, allowing analytical comparisons (and ranking) of their relative widths. Section 3 shows that the sup-t band minimizes worst-case regret in a Gaussian limit experiment. Section 4 discusses implementation of the sup-t band using a plug-in approach, a bootstrap

approach, and a novel Bayesian approach. Section 5 applies the theory to construct optimal simultaneous confidence bands for impulse response functions and for regression sensitivity analysis. Section 6 discusses extensions and future research directions. Appendix A contains supplementary results. Proofs are relegated to Appendix B.

NOTATION. Convergence in probability and distribution under the fixed true data generating process 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. If Ω is $p \times p$ and positive semidefinite, $\Omega^{1/2}$ denotes any $p \times p$ matrix such that $\Omega^{1/2}\Omega^{1/2'} = \Omega$. 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 $Q_\zeta(X)$. The Euclidean and maximum norms are denoted $\|\cdot\|$ and $\|\cdot\|_\infty$, respectively. For any set $S \subset \mathbb{R}^p$ and vector $y \in \mathbb{R}^p$, we define $S + y \equiv \{s + y \mid s \in S\}$.

2 Comparison of popular confidence bands

In this section we show that many popular confidence bands lie asymptotically in a one-parameter class of bands, allowing us to analytically compare their relative widths. Bands covered in our analysis include Bonferroni, Šidák, projection, and sup-t bands. The well-known sup-t band is the optimal choice within the one-parameter class. The purpose of this section is to give advice to applied users phrased in terms of a practically important performance measure, namely the relative widths of the confidence bands.

2.1 Model and assumptions

Our framework applies to models with a finite-dimensional parameter vector for which a consistent and asymptotically normal estimator is available. We do not restrict the correlation structure, and we do not assume that the data is i.i.d. We now present our model and assumptions, and we explain what we mean by a simultaneous confidence band.

We seek to construct simultaneous confidence bands for a vector-valued parameter of interest $\theta \in \mathbb{R}^k$, which is given by a possibly nonlinear transformation of an underlying parameter $\mu \in \mathbb{R}^p$. That is, $\theta \equiv h(\mu)$ for a function $h(\cdot)$. We do not restrict the relative magnitudes of the dimensions p and k, except in assuming that they are both finite. A

simultaneous confidence band for θ is a k-fold Cartesian product

$$\widehat{C} = \times_{j=1}^{k} \widehat{C}_j$$

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:

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

where n denotes the sample size. We refer to Cartesian products of intervals as *rectangular* sets. Unlike confidence ellipsoids, rectangular sets are easy to visualize in two dimensions.

Before stating our assumptions, we briefly discuss two examples that motivated our paper.

Example 1 (Impulse response function in a point-identified SVAR)**.** *Consider a finite-order Structural Vector Autoregression (SVAR) in the d time series $y_t = (y_{1,t}, \ldots, y_{d,t})'$ with structural shocks $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{d,t})'$. The parameter of interest is the vector $\theta = (\theta_1, \ldots, \theta_k)'$ that contains the impulse response coefficients at horizons $\ell = 0, 1, \ldots, k-1$ of the first variable $y_{1,t}$ with respect to the first shock $\varepsilon_{1,t}$. The parameter vector μ contains the VAR lag coefficients and innovation variance matrix. If we make assumptions such that θ is point-identified from μ, the map $h(\cdot)$ is well-defined. In most empirical applications we will have $p > k$. A simultaneous confidence band for θ has the interpretation that the* entire *true impulse response function (out to horizon $k-1$) is contained in the band with probability at least $1 - \alpha$, in repeated experiments.*

Example 2 (Sensitivity analysis)**.** *We regress the dependent variable y_t on the covariate x_t and additional controls. We assume that the researcher only cares about the linear regression coefficient on x_t, but is unsure about the right set of controls. Let $w_t^{(1)}, \ldots, w_t^{(k)}$ be k different vectors of potential control variables. Here $\mu = \theta = (\beta_1, \ldots, \beta_k)'$, and β_j is the coefficient on x_t in a population linear projection of y_t on $(1, x_t, w_t^{(j)'})'$. A simultaneous confidence band allows the audience to gauge the statistical significance of differences between coefficients β_j across specifications $j = 1, \ldots, k$.*

We assume that the underlying parameter vector μ can be estimated consistently and asymptotically normally, and that the transformation from μ to θ is smooth. Let the j-th element of the map $h(\cdot)$ be denoted $h_j(\cdot)$, $j = 1, \ldots, k$.

Assumption 1. *The following asymptotic limits are all pointwise as $n \to \infty$, assuming a fixed true data generating process.*

(i) *The true parameter μ lies in the interior of a convex and open parameter space $\mathcal{M} \subset \mathbb{R}^p$.*

(ii) *There exists an estimator $\hat{\mu}$ of μ such that $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N_p(\mathbf{0}_p, \Omega)$. The $p \times p$ matrix Ω is symmetric positive semidefinite (possibly singular).*

(iii) *There exists an estimator $\hat{\Omega}$ of Ω such that $\hat{\Omega} \xrightarrow{p} \Omega$.*

(iv) *The transformation $h\colon \mathcal{M} \to \mathbb{R}^k$ is continuously differentiable on $\mathcal{M}$. Write the Jacobian as $\dot{h}(\cdot) = (\dot{h}_1(\cdot), \ldots, \dot{h}_k(\cdot))' \in \mathbb{R}^{k \times p}$, where $\dot{h}_j(\tilde{\mu}) \equiv \partial h_j(\tilde{\mu})/\partial \tilde{\mu}$ for any $\tilde{\mu} \in \mathcal{M}$.*

(v) *All diagonal elements Σ_{jj} of the $k \times k$ matrix $\Sigma \equiv \dot{h}(\mu)\Omega\dot{h}(\mu)'$ are strictly positive.*

The assumption imposes standard regularity conditions. Observe that we do not restrict the data to be i.i.d. Condition (i) requires μ to lie in the interior of a convex parameter space. Conditions (ii) and (iii) require the existence of a consistent and asymptotically normally estimator $\hat{\mu}$ of μ and a consistent estimator $\hat{\Omega}$ of the asymptotic variance Ω. Note that Ω may be singular, which is important in applications to impulse response function estimation with non-stationary data, cf. Section 5.1. Condition (iv) requires the transformation from underlying model parameters μ to parameters of interest θ to be smooth, as is often the case in applied work. Finally, condition (v) implies that the plug-in estimator $\hat{\theta}_j \equiv h_j(\hat{\mu})$ has non-zero asymptotic variance for each j.

Assumption 1 implies that the plug-in estimator $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)' \equiv h(\hat{\mu})$ for the parameter vector of interest θ is asymptotically normal, with a possibly singular limit. An application of the delta method yields:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N_k(\mathbf{0}_k, \Sigma). \tag{2}$$

The asymptotic variance-covariance matrix Σ of $\hat{\theta}$ is singular in many applications, e.g., if $k > p$. Define the usual pointwise standard error for $\hat{\theta}_j$:

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

Under Assumption 1, a valid *pointwise* $1-\alpha$ confidence interval for parameter θ_j is given by $[\hat{\theta}_j - \hat{\sigma}_j\chi_{1-\alpha}, \hat{\theta}_j + \hat{\sigma}_j\chi_{1-\alpha}]$.[2] However, the Cartesian product of pointwise confidence intervals $j = 1, \ldots, k$ fails to yield simultaneous coverage (1) except in trivial cases, as is well known.

[2]Recall that $\chi^2_{1-\alpha}$ is the $1-\alpha$ quantile of the $\chi^2(1)$ distribution, i.e., the square of the $1-\alpha/2$ quantile of the standard normal distribution.

2.2 One-parameter class

We now introduce a one-parameter class of confidence bands that includes most of the popular choices in applied work, e.g., Bonferroni, Šidák, projection, and sup-t. The class is parameterized by a positive scalar c, which governs the width of the confidence band.

Definition 1. *For any $c > 0$, define the rectangular set*

$$\widehat{B}(c) \equiv \bigtimes_{j=1}^{k} \left[\hat{\theta}_j - \hat{\sigma}_j c, \hat{\theta}_j + \hat{\sigma}_j c\right].$$

$\widehat{B}(c)$ is the Cartesian product of Wald-type confidence intervals for each element of θ. 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 band $\widehat{B}(c)$ can also be interpreted as the collection of those parameter vectors θ for which the *largest* component-wise t-statistic $|\theta_j - \hat{\theta}_j|/\hat{\sigma}_j$ does not exceed a critical value c:

$$\widehat{B}(c) = \left\{ \tilde{\theta} \in \mathbb{R}^k \;\middle|\; \max_{j=1,\ldots,k} \frac{|\tilde{\theta}_j - \hat{\theta}_j|}{\hat{\sigma}_j} \leq c \right\}.$$

The one-parameter class—in addition to containing many commonly encountered confidence bands—has some additional appealing properties. First, each component interval of the band $\widehat{B}(c)$ is centered at the natural point estimate $\hat{\theta}_j$. Second, any band $\widehat{B}(c)$ is *balanced* in the sense of Beran (1988) and Romano & Wolf (2010), i.e., the *pointwise* coverage probabilities $P(\theta_j \in [\hat{\theta}_j - \hat{\sigma}_j c, \hat{\theta}_j + \hat{\sigma}_j c])$ of the component intervals have the same asymptotic limit for all j. Third, the one-parameter class of confidence bands can be obtained by inverting the class of *single-step* multiple hypothesis testing procedures based on studentized test statistics (Romano et al., 2010, section 8).

Now, for the sake of exposition, we state a straightforward result on the coverage probability of any band $\widehat{B}(c)$.

Lemma 1. *Let Assumption 1 hold. Let $\{\hat{a}_j, \hat{b}_j\}_{j=1,\ldots,k}$ be a collection of scalar random variables such that $\hat{a}_j, \hat{b}_j = o_p(n^{-1/2})$ as $n \to \infty$ for $j = 1, \ldots, k$. Then, for any $c > 0$,*

$$P\left(\theta \in \times_{j=1}^{k} \left[\hat{\theta}_j - \hat{\sigma}_j c - \hat{a}_j, \hat{\theta}_j + \hat{\sigma}_j c + \hat{b}_j\right]\right) \to P\left(\max_{j=1,\ldots,k} \left|\Sigma_{jj}^{-1/2} V_j\right| \leq c\right),$$

where $V = (V_1, \ldots, V_k)' \sim N_k(\mathbf{0}_k, \Sigma)$, and Σ_{jj} is the j-th diagonal element of Σ.

Hence, the asymptotic coverage probability of a band in the one-parameter class is determined by the cumulative distribution function of the maximum of the absolute value of k correlated standard normal variables.[3] The quantiles of this maximum will play a central role in the following.

Definition 2. *For any $\zeta \in (0,1)$ and any $k \times k$ symmetric positive semidefinite matrix $\tilde{\Sigma}$ with strictly positive diagonal elements $\tilde{\Sigma}_{jj}$, $j = 1, \ldots, k$, define*

$$q_\zeta(\tilde{\Sigma}) \equiv Q_\zeta \left(\max_{j=1,\ldots,k} \left| \tilde{\Sigma}_{jj}^{-1/2} \tilde{V}_j \right| \right),$$

where $\tilde{V} = (\tilde{V}_1, \ldots \tilde{V}_k)' \sim N_k(\mathbf{0}_k, \tilde{\Sigma})$.

Bands in one-parameter class. Many popular confidence bands in applied work lie in the one-parameter class for a particular choice 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* and *Šidák bands* are obtained by applying multiple comparisons adjustments to the pointwise critical value, yielding $c = \chi_{1-\alpha/k}$ and $c = \chi_{(1-\alpha)^{1/k}}$, respectively. If the asymptotic variance Σ of $\hat{\theta}$ is nonsingular, the *θ-projection* band equals the smallest rectangle in $\mathbb{R}^k$ that contains the usual Wald confidence ellipsoid for θ, which can be seen with a little algebra to correspond to $c = \chi_{k,1-\alpha}$. Finally, the *sup-t band* yields asymptotic simultaneous coverage of exactly $1-\alpha$, which according to Lemma 1 is achieved by $c = q_{1-\alpha}(\Sigma)$ (or a consistent estimate, cf. Section 4). Appendix A.1 provides details on these bands.

Table 1 presents the critical values c for the different confidence bands (the μ-projection band is defined below). The sup-t band achieves asymptotic simultaneous coverage exactly at the pre-specified level. All other bands either fail to achieve simultaneous coverage (the pointwise band) or are conservative and thus unnecessarily wide.

Figure 1 illustrates the various confidence bands in (θ_1, θ_2) space for the case $k = 2$.[4] The various confidence bands are represented as rectangles due to their Cartesian product structure. The θ-projection band is the smallest rectangle that contains the typical Wald ellipse for θ, drawn in black. The sup-t rectangle is the smallest rectangle that has coverage

[3] Analogous expressions are common in the theory of multiple testing (Lehmann & Romano, 2005, chapter 9) and the theory of suprema of Gaussian processes (Giné & Nickl, 2016, chapter 2). The asymptotically negligible random variables $\{\hat{a}_j, \hat{b}_j\}_{j=1,\ldots,k}$ in Lemma 1 allow for analysis of rectangular bands whose edges are all within asymptotic order $o_p(n^{-1/2})$ of a band $\widehat{B}(c)$ in our one-parameter class. This will permit us to consider bands obtained by projection and bootstrap strategies, as explained below.

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

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

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.

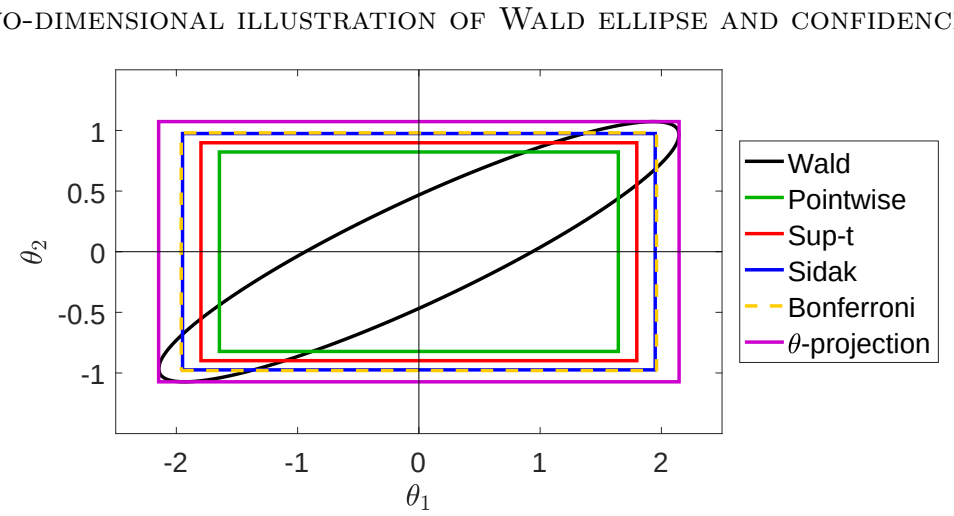

Figure 1: 90% Wald confidence ellipse (black) and rectangular confidence regions (colored rectangles) for the two-dimensional mean $\theta = (\theta_1, \theta_2)'$ of a normally distributed parameter estimator $\hat{\theta}$, given point estimate $\hat{\theta} = (0, 0)'$. The figure assumes the correlation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

probability at least $1-\alpha$ and for which the ratio of its edge lengths equals $\sqrt{\text{Var}(\hat{\theta}_1)/\text{Var}(\hat{\theta}_2)}$ (the requirement imposed by our one-parameter class).

μ-PROJECTION. If the standard errors $\hat{\sigma}_j$ are difficult to compute, it is appealing to obtain a confidence band for θ as the rectangular projection of the Wald confidence set for μ. Assume Ω is positive definite. If the Wald ellipse for μ is defined as

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

its rectangular projection with respect to the function $h(\cdot)$ is given by[5]

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

This is a multi-parameter version of the popular Krinsky & Robb (1986) procedure for constructing confidence intervals for nonlinearly transformed estimators. The μ-projection band has asymptotic simultaneous coverage probability of at least $1-\alpha$ under Assumption 1, by the standard projection inference argument (e.g., Dufour, 1990). Typically, however, the simultaneous coverage probability will be conservative ("projection bias").

We show that the μ-projection band is also contained in our one-parameter class, up to asymptotically negligible terms, provided Assumption 1 holds and Ω is positive definite:

$$\widehat{C}_{\mu\text{-proj}} = \bigtimes_{j=1}^{k} \left[\hat{\theta}_j - \hat{\sigma}_j \chi_{p,1-\alpha} + o_p(n^{-1/2}), \hat{\theta}_j + \hat{\sigma}_j \chi_{p,1-\alpha} + o_p(n^{-1/2}) \right].$$

See Appendix A.1 for a detailed statement.[6]

BOOTSTRAP VERSIONS. Our framework can also accommodate bootstrap versions of the above confidence bands, as discussed in Appendix A.1.

[5] 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 in two dimensions).

[6] A heuristic version of the argument appears in Cox & Ma (1995). The result is well known in the special case of $h(\cdot)$ being a linear map, as it serves as the basis for Scheffé confidence bands in linear regression.

2.3 Comparison of bands

We now analytically compare the relative widths of bands in the one-parameter class. As is well known, the sup-t band is the optimal band within this class. We can compare the (suboptimal) relative widths of the projection, Bonferroni, and Šidák bands without seeing the data. Furthermore, we discuss the conditions under which the sup-t band offers large or small improvements relative to the other bands.

For any two bands in the one-parameter class, the relative ratio of their critical values yields the relative ratio of the lengths of each of their component intervals. Henceforth, we will call this number the *relative width* of the band, which is a well-defined concept within our one-parameter class (outside this class, the relative length of component intervals could vary across components $j = 1, \ldots, k$, and relative lengths could be data-dependent).

1) $c_{\textbf{pointwise}} \leq c_{\textbf{sup-t}} \leq c_{\textbf{Šidák}}$: The sup-t band is optimal within the one-parameter class since it selects the critical value c as the smallest value that guarantees asymptotic simultaneous coverage of $1 - \alpha$.[7] The sup-t critical value depends on the correlation structure Σ of the estimator $\hat{\theta}$, but the pointwise and Šidák critical values constitute its best-case and worst-case values. On the one hand, it is straight-forward to show that the sup-t critical value must weakly exceed the pointwise critical value, with equality only if the elements of $\hat{\theta}$ are asymptotically perfectly correlated. On the other hand, the sup-t critical value is always weakly smaller than both the Šidák critical value and the μ-projection critical value, cf. Appendix A.2 for details. Moreover, if $k \leq p$, the sup-t critical value equals the Šidák critical value if and only if the elements of $\hat{\theta}$ are asymptotically independent. Hence, if $k \leq p$, the pointwise and Šidák bands can be thought of as best-case and worst-case scenarios for the sup-t band, respectively. In applications where the elements of $\hat{\theta}$ are close to uncorrelated, there is little loss in using the simple Šidák band instead of the sup-t band, although the computational cost of the latter band is also small, cf. Section 4.

2) $c_{\textbf{Šidák}}, c_{\textbf{Pointwise}}, c_{\textbf{Bonferroni}}, c_{\theta\textbf{-projection}}$: Our framework allows us to compare the many suboptimal but popular confidence bands. While the sup-t band is easy to compute, cf. Section 4, these other bands are often computationally even simpler, and they are frequently encountered in applied work. Readers who are only interested in the optimality properties of the sup-t band should skip to the next section.

[7] This is well known in the single-step multiple testing literature (Lehmann & Romano, 2005, chapter 9).

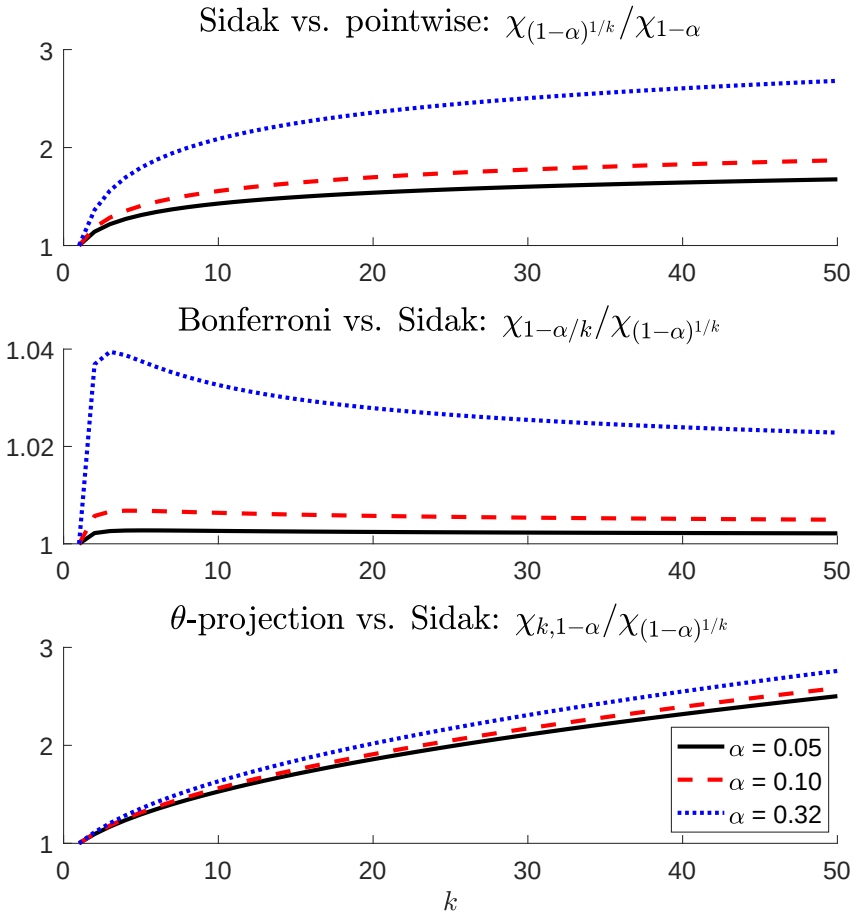

Figure 2: Relative critical values of the pointwise, Šidák, Bonferroni, and θ-projection bands. The dimension $k = \dim(\theta)$ is along the horizontal axis. The three colored curves correspond to the significance levels $\alpha = 0.05$ (black), $\alpha = 0.1$ (red), and $\alpha = 0.32$ (blue).

Except for the sup-t band, the relative widths of all other bands depend only on the significance level α and the dimensions p and k of the model and parameter of interest. From the perspective of first-order asymptotic analysis, no additional information is needed to compare these different bands.[8]

Figure 2 plots the relative widths of the pointwise, Šidák, Bonferroni, and θ-projection confidence bands for different values of the dimension k of θ and different significance levels α. We do not plot the μ-projection critical value $\chi_{p,1-\alpha}$, but it is clear that it exceeds the θ-projection critical value $\chi_{k,1-\alpha}$ if and only if $p > k$.

2.1) $c_{\textbf{Pointwise}} \leq c_{\textbf{Šidák}}$: The first display of the figure shows that, while the relative width of the Šidák and pointwise bands must exceed one, it is below 2 for $k \leq 50$ and $\alpha \leq 0.1$ (hence, this also applies to sup-t vs. pointwise). In fact, Appendix A.2 states the well-known result that the Šidák critical value grows very slowly with k, specifically at rate $\sqrt{\log k}$, so that there is little penalty in terms of width incurred from including additional parameters of interest in θ.[9]

2.2) $c_{\textbf{Šidák}} \leq c_{\textbf{Bonferroni}}, c_{\theta\textbf{-projection}}$: The second display of the figure shows that the Bonferroni critical value always exceeds the Šidák one, but they are within 4% of each other for all common significance levels. Finally, the last display of the figure shows that θ-projection leads to much wider bands than Šidák (and thus sup-t), unless k is very small. Hence, there appears to be no good reason to use θ-projection (with the usual Wald critical value). See Appendix A.2 for analytical results supporting the graphical evidence in Figure 2.

2.3) $c_{\textbf{Šidák}} \leq c_{\mu\textbf{-projection}}$ IN MANY RELEVANT MODELS: The Šidák (and sup-t) bands are narrower than the μ-projection band in most practical cases. While the μ-projection band is always wider than the sup-t band, it can be narrower than the Šidák band if $k \gg p$. However, Appendix A.2 shows that for this to happen at usual significance levels, either the number k of parameter of interest must be in the 1,000s, or the number p of underlying model parameters must be less than 10.

[8] Indeed, researchers can decide on a band before obtaining the relevant data, as long as the model has been specified. The relative widths of the pointwise, Šidák, Bonferroni, and θ-projection bands are the same in any finite sample. However, the comparison with μ-projection is asymptotic.

[9] Of course, the accuracy of the asymptotic normal approximation may deteriorate for large k.

2.4) $c_{\textbf{Bonferroni}} \leq c_{\theta\textbf{-projection}} \leq c_{\mu\textbf{-projection}}$ IF $\alpha < 0.5$ AND $2 \leq k \leq p$: If $\alpha < 0.5$ and $k \geq 2$, then $\chi_{1-\alpha/k} < \chi_{k,1-\alpha}$, i.e., in this case the Bonferroni band is narrower than θ-projection. This result was proven by Alt & Spruill, 1977, although it is seemingly not well known. As a corollary, the Bonferroni band is also narrower than the μ-projection band if $p \geq k$.

SUMMARY. Table 1 shows the various confidence bands ordered in terms of relative width (narrowest at the top) in the empirically common case $\alpha \leq 0.5$ and $k \leq p$. Sup-t is the narrowest band with correct asymptotic coverage, followed by Šidák's band. 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, as shown in Appendix A.2. θ-projection always leads to a wider band than Šidák and Bonferroni for $\alpha \leq 0.5$.

3 Decision theoretic justification for sup-t band

Is the sup-t band in some sense optimal even outside the one-parameter class analyzed in the previous section? 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. Our analysis focuses on a Gaussian limit experiment motivated by the preceding asymptotic analysis.

3.1 Gaussian limit experiment and equivariance

We first describe the Gaussian limit experiment and translation equivariance concept used in the decision-theoretic analysis.

FINITE-SAMPLE MODEL. We specialize our framework to 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, as discussed below. Hence, we assume we observe a single multivariate normal draw

$$X \sim N_p(\mu, I_p), \tag{3}$$

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 $\theta \equiv G\mu = (g_1'\mu, \ldots, g_k'\mu)'$.

We view the finite-sample Gaussian model as the relevant limit experiment corresponding to the delta method asymptotics of Section 2. The draw X plays the same role as the estimator $\hat{\mu}$ in the previous section. Likewise, the matrix G plays the same role as the matrix $\dot{h}(\mu)\Omega^{1/2}$ (as the asymptotic analysis in the previous section relies on a linearization of the function $h(\cdot)$). 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 $\mathrm{Var}(X) = I_p$. Thus, the present finite-sample model allows us to focus on the essentials of the researcher's decision problem.

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):

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

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

$$\mathrm{Loss}(R; \mu) \equiv L(b_1 - a_1, \ldots, b_k - a_k) \quad \text{for } \mu \in \mathbb{R}^p,\, R = \times_{j=1}^k [a_j, b_j] \in \mathcal{R}, \tag{4}$$

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:[10]

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

For ease of exposition, we state a simple characterization of translation equivariance:[11]

Lemma 2. $\mathcal{C}_{eq} = \{ C\colon \mathbb{R}^p \to \mathcal{R} \mid C(x) = Gx + R,\ R \in \mathcal{R} \}$.

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 nonrandom $a_j, b_j \geq 0$, $j = 1, \ldots, k$. 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 2 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

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

where

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

[10] Recall: For $R = \times_{j=1}^k [a_j, b_j]$, $G\lambda + R \equiv \times_{j=1}^k [g_j'\lambda + a_j, g_j'\lambda + b_j] \in \mathcal{R}$, where g_j' is the j-th row of G.

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

3.2 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

$$R^* \in \underset{R \in \mathcal{R}_{1-\alpha}}{\operatorname{argmin}} L(R). \tag{5}$$

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 (5) is not available in closed form.[12] 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 (2017) develop computational tools to numerically solve problems of the form (5) (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 $\theta = (\theta_1, \ldots, \theta_k)'$: For 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.

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

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.[13]

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 $L(r) = \|r\|_\infty$ (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.

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

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

where g_j' is the j-th row of $G = (g_1, \ldots, g_k)'$, and $q_{1-\alpha}(\cdot)$ is given by Definition 1.

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

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

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

[13] 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.

(5)—for a *particular* loss function L.[14] 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.

Remarks.

1. Proposition 1 complements our analysis of the one-parameter class in Section 2. There we showed that the sup-t band is asymptotically preferable to other commonly encountered simultaneous confidence bands. In this section, we have argued that the sup-t band has a particular optimality property also outside the one-parameter class.

2. Proposition 1 implies that any equivariant simultaneous confidence band other than the sup-t band yields strictly larger regret than the sup-t band for *some* loss function $L \in \mathcal{L}_H$. In particular, the proof shows that any simultaneous confidence band other than sup-t must perform relatively poorly against a loss function that returns the length of a single "least favorable" component interval.

3. It is easy to verify that the statement of Proposition 1 continues to hold if we replace the class $\mathcal{L}_H$ of homogeneous-of-degree-1 loss functions with the strictly smaller class of weighted-average loss functions $\{L(r) = \sum_j w_j r_j \mid \sum_j w_j = 1,\ w_j \geq 0,\ j = 1, \ldots, k\}$. We phrase the proposition in terms of the class $\mathcal{L}_H$ because the expression for the worst-case regret of the sup-t band relies intimately on homogeneity of degree 1 of the loss L, as is clear from the proof.

4. Proposition 1 gives an explicit expression for the worst-case regret of the sup-t band; it equals the ratio of the sup-t and pointwise critical values, cf. Section 2.2. It is a consequence of Lemma 3 in Appendix A.2 that the worst-case regret for the sup-t band can be further upper-bounded by the ratio $\chi_{(1-\alpha)^{1/k}}/\chi_{1-\alpha}$ of the Šidák and pointwise critical values, which is independent of G. Hence, the sup-t band delivers guarantees on regret across both loss functions and correlation structures.

4 Implementation of sup-t band

Here we describe how to implement the sup-t band using three computationally simple approaches: a plug-in approach, a bootstrap approach, and a novel Bayesian approach.

[14] Even so, it is interesting that the sup-t band is very close to the numerically computed optimal bands in the empirical applications of Freyberger & Rai (2017, section 4).

These approaches are asymptotically equivalent under regularity conditions. The Bayesian band has finite-sample simultaneous credibility at level $1-\alpha$, provided we are able to draw from the finite-sample posterior distribution of the underlying model parameters μ.

PLUG-IN BAND. The most straight-forward feasible implementation of the sup-t band plugs in a consistent estimator of the sup-t critical value. Algorithm 1 presents a standard procedure to estimate the plug-in sup-t critical value $q_{1-\alpha}(\Sigma)$.

Algorithm 1 Plug-in sup-t band

1: Compute the Jacobian $\dot{h}(\hat{\mu})$ and obtain $\hat{\Sigma} = \dot{h}(\hat{\mu})\hat{\Omega}\dot{h}(\hat{\mu})'$, cf. notation in Assumption 1
2: Draw N i.i.d. normal vectors $\hat{V}^{(\ell)} \sim N_k(\mathbf{0}_k, \hat{\Sigma})$, $\ell = 1, \dots, N$
3: Let $\hat{q}_{1-\alpha}$ be 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 BAND. We also suggest a bootstrap implementation of the sup-t band which is based on a valid strategy for bootstrapping the estimator $\hat{\mu}$. Algorithm 2 defines the procedure for computing the quantile-based bootstrap band $\widehat{C}$. This algorithm also serves as the basis for the Bayesian band defined below.

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: $\widehat{C} = \times_{j=1}^{k}[\hat{Q}_{j,\hat{\zeta}}, \hat{Q}_{j,1-\hat{\zeta}}]$

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 Cartesian 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 the search can be confined to $\zeta \in [\alpha/(2k), \alpha/2]$ in any finite sample.[15]

Appendix A.4 defines an alternative band that bootstraps the sup-t statistic, rather than using quantiles. This band also does not require calculating derivatives of $h(\cdot)$.

BAYESIAN BAND. If we are able to draw from the finite-sample posterior distribution of μ, Algorithm 2 also provides an algorithm for computing a confidence band with simultaneous Bayesian credibility equal to $1-\alpha$. Let $\hat{P}$ in the algorithm denote the posterior distribution of μ given data $\mathcal{D}$. Then by construction the set $\widehat{C}$ satisfies $P(\theta \in \widehat{C} \mid \mathcal{D}) = 1-\alpha$, up to simulation error that vanishes as the number of posterior draws grows large. The simultaneous credible band $\widehat{C}$ is the Cartesian product of component-wise equal-tailed credible intervals, where the tail probability has been calibrated to yield simultaneous credibility of $1-\alpha$.[16]

SUMMARY. The plug-in band requires little computation if the derivatives of $h(\cdot)$ are easy to compute. The Bayesian band is attractive if finite-sample Bayesian inference is desired. The bootstrap band is quite generally applicable and, like the Bayesian band, does not require computation of the derivatives of $h(\cdot)$. The bootstrapping of $\hat{\mu}$ or posterior sampling of μ is only performed once and for all (yielding N draws). The only numerical optimization necessary for the bootstrap and Bayes bands is the monotonic, 1-dimensional root finding problem for $\hat{\zeta}$ in Algorithm 2.[17] As shown in Appendix A.4, the three implementations of the sup-t band are asymptotically equivalent under standard regularity conditions like bootstrap consistency and the Bernstein-von Mises property.

[15] 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. It would be interesting to investigate whether double bootstrap procedures can deliver refinements.

[16] This Bayesian band appears to be the first generically applicable method for constructing bands with simultaneous credibility. 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).

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

5 Applications

In this section we present applications of 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. Here we focus on two concrete applications. Our main application shows that the sup-t band improves on popular approaches to constructing simultaneous confidence bands for impulse response functions in VARs. Our second application 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.

5.1 Impulse response functions in VAR analysis

We estimate the effects of monetary policy shocks on output using recursive and instrumental variable identification strategies for VARs, as in Gertler & Karadi (2015).[18] The impulse reponse function 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.5 for a literature review). Our nonlinear framework in Section 2 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 impulse response functions.[19]

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

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

[19] However, we argue in Appendix A.4 that the "adjusted Bonferroni/Wald" methods of Lütkepohl et al. (2015a,b) are closely related to a bootstrap implementation of the sup-t band.

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.5 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.[20] 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.[21]

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* given by changes in federal funds futures prices in small time windows around scheduled Federal Open Market Committee announcements.[22]

RESULTS. Figure 3 shows that the sup-t band for the impulse response function of output is substantially 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).[23] 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

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

[21] Appendix A.5 gives further details of the bootstrap, posterior sampling, and other technical aspects.

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

[23] The figure shows that the Šidák and Bonferroni bands virtually coincide, although the latter is slightly wider, which conforms to the theory in Section 2. Appendix A.5 reports that the θ-projection and (especially) μ-projection bands are much wider than the other bands, as theory predicts.

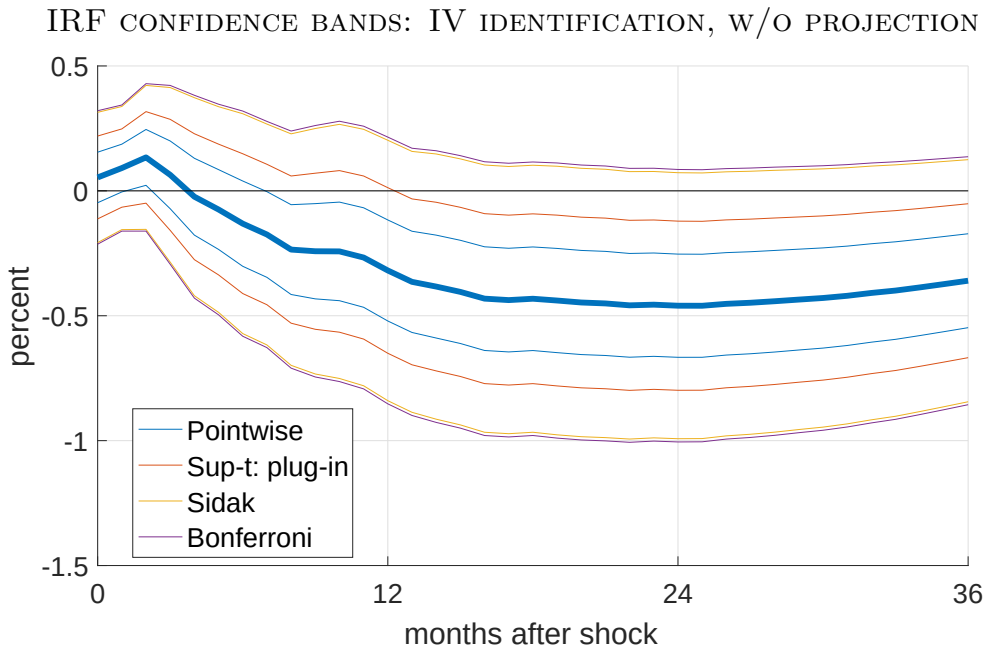

Figure 3: 68% confidence bands for impulse response function of industrial production to a 1 standard deviation contractionary monetary policy shock, external instrument identification, without projection bands. Thick line: point estimate. Thin lines: confidence bands (legend is ordered with respect to width). Šidák and Bonferroni bands virtually coincide.

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 3 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 4 compares the plug-in, bootstrap, and Bayes implementations of the sup-t band for the case of recursive identification.[24] 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 assumption). Gertler & Karadi (2015) conclude that the difference between the recursively identified results and the external instrument results call into question the recursive exclusion

[24]We use a homoskedastic recursive residual bootstrap and maximally diffuse normal-inverse-Wishart prior. See Appendix A.5 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 z_t is white noise, which is counterfactual in our application).

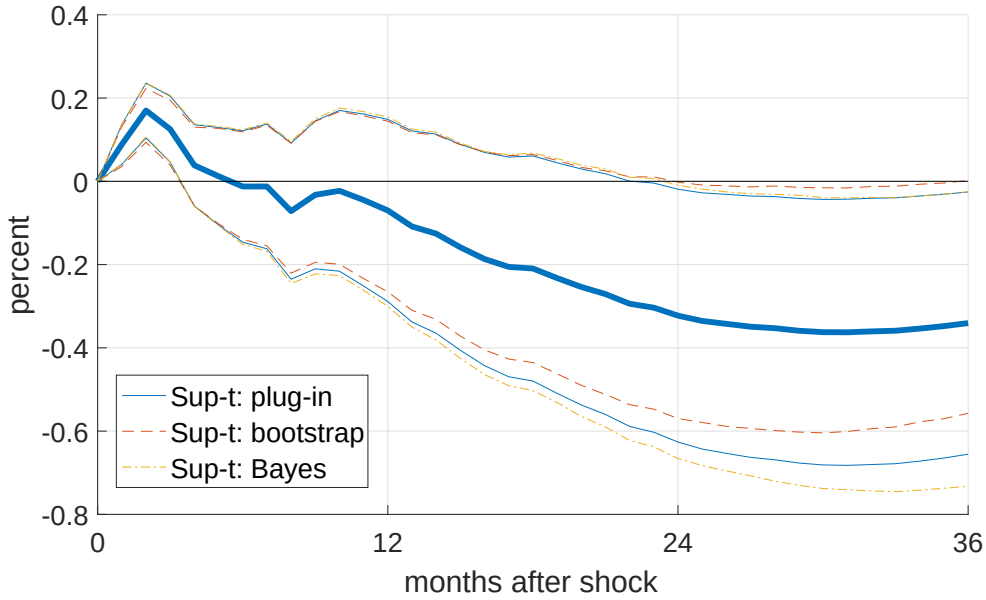

Figure 4: 68% confidence bands for the impulse response function of industrial production to a 1 standard deviation contractionary monetary policy shock, recursive identification, different sup-t implementations. Thick line: point estimate. Thin lines: confidence bands.

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 other areas of 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 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.

5.2 Sensitivity analysis

Here we use a simultaneous confidence band to visualize the joint uncertainty of a linear regression coefficient estimated using different sets of control variables. The simultaneous nature of the band allows comparisons across specifications, in contrast to the common approach of reporting confidence intervals separately for each specification. Our application follows Head et al. (2010) in estimating the effects of gaining political independence on bilateral trade between a former colony and its metropole. We show that the effect on trade 40 years after independence is sensitive to controlling for population and economic development, as economic theory predicts. However, the result is insensitive to controlling for trade agreements and common currency, language, or legal system.

The sup-t band is more attractive for sensitivity analysis than the Bonferroni method, since the point estimators in different specifications will typically be highly correlated. Although we focus here on linear regression, the same method can be applied to sensitivity analysis in many other types of identified economic models. While we have not seen the sup-t band used for visualizing sensitivity analysis before, Berk et al. (2013) and Leeb et al. (2015) analyze the sup-t band as a means for performing valid post-model-selection inference.[25]

Specification. Head et al. (2010) find that bilateral trade between a former colony and its metropole decreases dramatically following independence. They estimate that the full, permanent effect of the separation on trade occurs about 40 years out. Their annual panel data set is based on the International Monetary Fund's Direction of Trade Statistics as well as various data sources for colonial relationships, macroeconomic indicators, trade agreements, etc.[26] We use the sample for the main linear regression specification in Head et al. (2010). The number of dyads (country pairs) is 27,303, while the total number of dyad-year observations is 592,923.

Our parameter of interest is the effect of independence on bilateral trade 40 years after the event. For ease of exposition, we employ the OLS specification of Head et al. (2010) (their results are robust across several other specifications). The regressions include dummies for colonial history and ongoing colonial relationship, as well as dummies for number of years after independence. We focus on the linear projection coefficient on the dummy for 40 years

[25] Several papers have considered the related issue of adjusting critical values to account for "data snooping", e.g., Andrews (1993), Hansen & Timmermann (2012), and Armstrong & Kolesár (2016).

[26] See Head et al. (2010) for details. We thank Keith Head for making code and data available online.

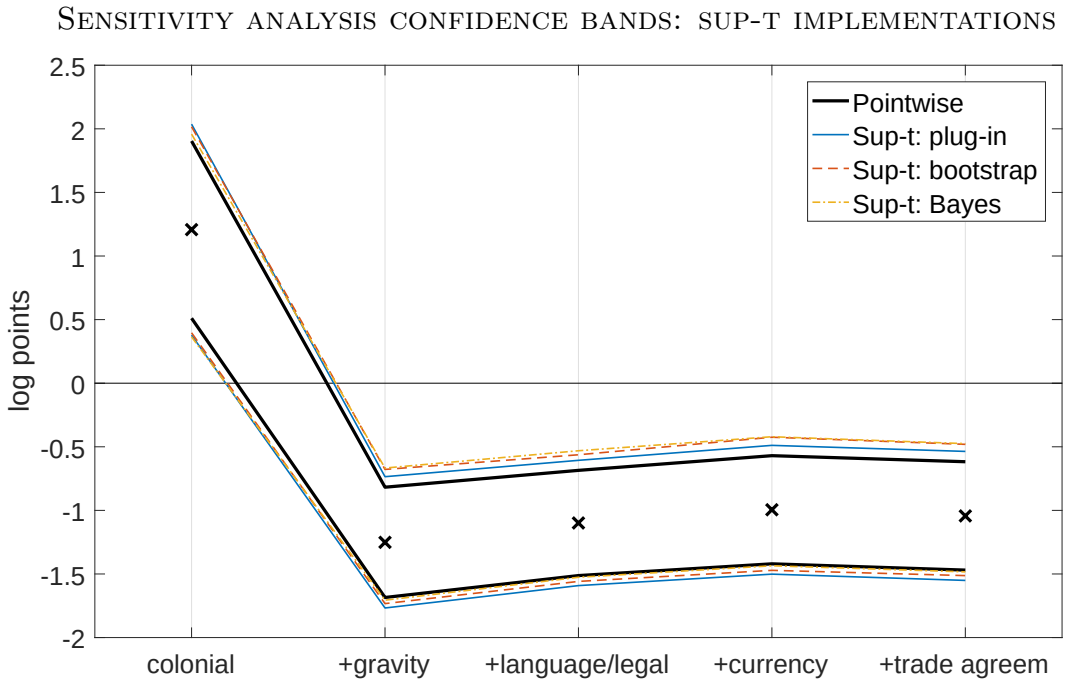

Figure 5: 90% confidence bands for the linear regression coefficient measuring the 40-year effect of independence on log bilateral trade, estimated across five different sets of controls. Crosses: point estimates. Lines: confidence bands. Specifications: "colonial" – colonial history dummy, ongoing colonial relationship dummy, years after independence dummies; "+gravity" – adds log population and log GDP per capita in origin and destination, log distance between origin and destination, shared border dummy; "+language/legal" – further adds dummies for common language and legal system; "+currency" – further adds dummy for common currency; "+trade agreem" – further adds dummies for trade agreements. All specifications include time fixed effects. Cluster variable: dyad.

after independence. We consider five different sets of control variables, as described below.[27] All specifications control for time fixed effects. Standard errors, bootstraps replications, and Bayesian inference is clustered by dyad.[28]

Results. Figure 5 exhibits plug-in, bootstrap, and Bayesian sup-t bands for the 40-year effect of independence estimated across five different sets of control variables. The set of controls expands when moving from left to right; the fifth specification corresponds to the preferred OLS specification of Head et al. (2010). The simultaneous confidence bands allow

[27] As emphasized by Leeb et al. (2015, Remark 2.1(i)), the five estimated coefficients correspond to different linear projections and should not be interpreted as five different estimates of the same parameter in some encompassing model (e.g., the largest model).

[28] We perform Bayesian analysis using the Bayesian bootstrap of Rubin (1981) and Chamberlain & Imbens (2003), a multiplier bootstrap with standard exponential weights. We perform 2,000 multinomial bootstrap and Bayesian bootstrap replications.

the audience to make comparisons across specifications. As expected, the sup-t bands are only about 20% wider than the pointwise band (which does not permit comparison across specifications). In contrast, the Bonferroni band (not shown) is 41% wider than the pointwise band. The three versions of the sup-t band are almost equally wide, but the bootstrap and Bayes bands are shifted slightly upward relative to the plug-in band.

Figure 5 illustrates how the sup-t band can visually communicate which features of the specification are important for the final result, while permitting valid statistical comparisons across specifications. In particular, the bands show that it is crucial to control for post-independence developments in population and GDP per capita, just as economic theory would predict, and as discussed by Head et al. (2010). The first regression specification only uses colonial dummies and years after independence dummies, so it essentially corresponds to an event study comparing colonies that gain independence with colonies that do not. In this specification, the 40-year independence effect on trade is *positive*. The second specification adds traditional "gravity equation" control variables: population and GDP per capita in origin and destination, distance between origin and destination, and a dummy variable for shared border. The plot clearly shows that these particular control variables are driving the highly negative estimated 40-year effect of independence; the point estimate of -1.25 log points corresponds to a 72% reduction in trade. Based on the sup-t bands, the difference between the first and second specifications is statistically significant. However, the remaining three specifications do not yield significantly different results from the second specification.

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. 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 sup-t band is particularly attractive for finite-sample Bayesian inference. If frequentist asymptotic coverage is desired, the plug-in sup-t band is straight-forward to compute if the derivatives of $h(\cdot)$ are simple to evaluate; otherwise, the bootstrap and Bayes implementations only require evaluation of $h(\cdot)$ itself.

EXTENSION: GENERALIZED ERROR RATE CONTROL. In some applications it is desirable to replace the simultaneous coverage requirement (1) 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 (1) 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.

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 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.[29] Moreover, the Bayesian sup-t band may be used for subjective Bayesian analysis even under partial identification.[30]

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.

[29] 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.

[30] Baumeister & Hamilton (2016) make this argument for the pointwise credible band.

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., 2016).

References

Alt, F. & Spruill, C. (1977). A comparison of confidence intervals generated by the Scheffé and Bonferroni methods. *Communications in Statistics – Theory and Methods*, *6*(15), 1503–1510.

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, D. W. K. (1993). Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, *61*(4), 821–856.

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

Armstrong, T. & Kolesár, M. (2016). A Simple Adjustment for Bandwidth Snooping. ArXiv working paper 1412.0267.

Baumeister, C. & Hamilton, J. D. (2016). Optimal Inference about Impulse-Response Functions and Historical Decompositions in Incompletely Identified Structural Vector Autoregressions. Manuscript, University of California San Diego.

Beran, R. (1988). Balanced Simultaneous Confidence Sets. *Journal of the American Statistical Association*, *83*(403), 679–686.

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

Berk, R., Brown, L., Buja, A., Zhang, K., & Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics*, *41*(2), 802–837.

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.

Chamberlain, G. & Imbens, G. W. (2003). Nonparametric Applications of Bayesian Inference. *Journal of Business & Economic Statistics*, *21*(1), 12–18.

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

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

Cox, C. & Ma, G. (1995). Asymptotic Confidence Bands for Generalized Nonlinear Regression Models. *Biometrics*, *51*(1), 142–150.

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

Dunn, O. J. (1958). Estimation of the Means of Dependent Variables. *Annals of Mathematical Statistics*, *29*(4), 1095–1111.

Dunn, O. J. (1959). Confidence Intervals for the Means of Dependent, Normally Distributed Variables. *Journal of the American Statistical Association*, *54*(287), 613–621.

Freyberger, J. & Rai, Y. (2017). Uniform confidence bands: characterization and optimality. Manuscript, University of Wisconsin-Madison.

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.

Giné, E. & Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

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

Hansen, P. R. & Timmermann, A. (2012). Choice of Sample Split in Out-of-Sample Forecast Evaluation. CREATES Research Paper 2012-43.

Head, K., Mayer, T., & Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, *81*(1), 1–14.

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.

Hymans, S. H. (1968). Simultaneous Confidence Intervals in Econometric Forecasting. *Econometrica*, *36*(1), 18–30.

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.

Jordà, Ò., Knüppel, M., & Marcellino, M. (2013). Empirical simultaneous prediction regions for path-forecasts. *International Journal of Forecasting*, *29*(3), 456–468.

Jordà, Ò. & Marcellino, M. (2010). Path forecast evaluation. *Journal of Applied Econometrics*, *25*(4), 635–662.

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. To appear.

Kitagawa, T., Montiel Olea, J. L., & Payne, J. (2016). Posterior Distribution of Nondifferentiable Functions. Manuscript, Columbia University.

Krinsky, I. & Robb, A. L. (1986). On Approximating the Statistical Properties of Elasticities. *The 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*. Forthcoming.

Leeb, H., Pötscher, B. M., & Ewald, K. (2015). On Various Confidence Intervals Post-Model-Selection. *Statistical Science*, *30*(2), 216–227.

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. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag.

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.

Lütkepohl, H., Staszewska-Bystrova, A., & Winker, P. (2016). Calculating Joint Confidence Bands for Impulse Response Functions Using Highest Density Regions. DIW Berlin Discussion Paper 1564.

Mertens, K. & Ravn, M. O. (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review*, *103*(4), 1212–1247.

Montiel Olea, J. L., Stock, J. H., & Watson, M. W. (2016). Uniform Inference in SVARs Identified with External Instruments. Manuscript, Columbia University.

Naiman, D. Q. (1984). Average Width Optimality of Simultaneous Confidence Bounds. *The 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. (1984). *Admissible and Optimal Confidence Bands in Linear Regression*. PhD thesis, Cornell University.

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. *The Annals of Statistics*, *35*(4), 1378–1408.

Romano, J. P. & Wolf, M. (2010). Balanced control of generalized error rates. *The Annals of Statistics*, *38*(1), 598–633.

Rubin, D. B. (1981). The Bayesian Bootstrap. *The Annals of Statistics*, *9*(1), 130–134.

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.

Stock, J. H. & Watson, M. W. (2012). Disentangling the Channels of the 2007–09 Recession. *Brookings Papers on Economic Activity*, (Spring issue), 81–135.

Stock, J. H. & Watson, M. W. (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In J. B. Taylor

& H. Uhlig (Eds.), *Handbook of Macroeconomics*, volume 2A chapter 8, (pp. 415–525). Elsevier.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, *52*(2), 381–419.

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.

A Supplementary results

A.1 Confidence bands in the one-parameter class

Here we provide additional details on some of the bands in the one-parameter class.

BONFERRONI. The Bonferroni band is obtained by applying a Bonferroni multiple comparisons adjustment to the pointwise critical value. 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. Standard arguments (e.g., Alt, 1982) show that the Bonferroni band has asymptotic simultaneous coverage probability strictly greater than $1 - \alpha$ if $k > 1$, under Assumption 1.

ŠIDÁK. The Šidák band can be seen as 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. The famous Šidák (1967) theorem shows that the independent case is "least favorable" for the coverage probability of the multivariate normal distribution. Hence, under Assumption 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.[31]

θ-PROJECTION. A well-known strategy for building a confidence band for θ is to find the smallest rectangle that contains the Wald ellipse for θ. If the consistent estimator $\hat{\Sigma} = \dot{h}(\hat{\mu})\hat{\Omega}\dot{h}(\hat{\mu})'$ of Σ is nonsingular, we can define the Wald ellipse for θ:

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

Consider the smallest rectangle that contains this ellipse:

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

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 Assumption 1 and provided that Σ is nonsingular. The band is conservative if $k \geq 2$. Straight-forward algebra shows that the

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

θ-projection band is equal to the band $\widehat{B}(\chi_{k,1-\alpha})$ in the one-parameter class.[32]

μ-PROJECTION. The following result states that the μ-projection band defined in Section 2.2 lies in the one-parameter class, up to asymptotically negligible terms.

Proposition 2. *Under Assumption 1 and positive definiteness of Ω, the μ-projection band equals $\widehat{B}(\chi_{p,1-\alpha})$ up to terms of order $o_p(n^{-1/2})$:*

$$\widehat{C}_{\mu\text{-}proj} = \bigtimes_{j=1}^{k} \left[\hat{\theta}_j - \hat{\sigma}_j \chi_{p,1-\alpha} + o_p(n^{-1/2}), \hat{\theta}_j + \hat{\sigma}_j \chi_{p,1-\alpha} + o_p(n^{-1/2})\right].$$

BOOTSTRAP VERSIONS. Many bootstrap variants of the bands in the one-parameter class are possible. Lemma 1 implies that 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. For example, our analysis covers confidence bands obtained by bootstrapping the standard errors $\hat{\sigma}_j$ instead of using plug-in delta method estimates.

A.2 Comparison of critical values

Here we provide additional analytical and graphical results comparing the critical values listed in Table 1, cf. Section 2.3. Most of these results are well known in the multiple comparisons literature, but it is useful to state them in terms of our notation.

The following lemma states that the pointwise critical value and the Šidák critical value provide extreme bounds on the sup-t critical value $q_{1-\alpha}(\Sigma)$, cf. Definition 2. These bounds are sharp if $k \leq p$, in which case a more precise expression for the sup-t critical value would need to rely on the specific correlation structure of $\hat{\theta}$. Dunn (1958, 1959) conjectured a version of this statement, since proven by Šidák (1967).

Lemma 3. *Define*

$$\mathcal{S}_{p,k} \equiv \left\{\tilde{\Sigma} \in \mathbb{R}^{k\times k} \;\middle|\; \tilde{\Sigma} \text{ symmetric positive semidefinite, } \operatorname{rank}(\tilde{\Sigma}) \leq p,\ \tilde{\Sigma}_{jj} > 0 \text{ for all } j\right\}.$$

[32] Simply maximize/minimize $\tilde{\theta}_j$ subject to the quadratic constraint $n(\hat{\theta}-\tilde{\theta})'\hat{\Sigma}^{-1}(\hat{\theta}-\tilde{\theta}) \leq \chi^2_{k,1-\alpha}$, 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.

For all $\zeta \in (0,1)$,

$$\inf_{\tilde{\Sigma} \in \mathcal{S}_{p,k}} q_\zeta(\tilde{\Sigma}) = \chi_\zeta, \qquad \sup_{\tilde{\Sigma} \in \mathcal{S}_{p,k}} q_\zeta(\tilde{\Sigma}) \leq \chi_{\zeta^{1/k}}.$$

The inequality for the supremum is an equality if $k \leq p$.

Lemma 3 provides sharp bounds on the sup-t critical value when $k \leq p$. The following lemma provides a slightly more informative upper bound in the case $k > p$. It states that the sup-t critical value is also upper-bounded by the μ-projection critical value, although this bound is not sharp if $k > p \geq 2$.

Lemma 4. *Using the same notation as Lemma 3, we have for all* $\zeta \in (0,1)$,

$$\sup_{\tilde{\Sigma} \in \mathcal{S}_{p,k}} q_\zeta(\tilde{\Sigma}) \leq \min\left\{\chi_{\zeta^{1/k}}, \chi_{p,\zeta}\right\}.$$

If $k > p \geq 2$ *and* $\zeta \in (0,1)$, *then*

$$\chi_{\zeta^{1/p}} < \sup_{\tilde{\Sigma} \in \mathcal{S}_{p,k}} q_\zeta(\tilde{\Sigma}) < \chi_{p,\zeta}.$$

The next lemma provides analytical results to complement the visual observations in Figure 2 about the pointwise, Šidák, Bonferroni, and θ-projection critical values. It shows that (i) the Bonferroni critical value always exceeds Šidák, (ii) the θ-projection critical value always exceeds Šidák, and (iii) the Šidák critical value grows at rate $\sqrt{\log k}$ in k. These results are well known in the multiple comparisons literature.

Lemma 5. *(i)* $\chi_{1-\alpha/k} > \chi_{(1-\alpha)^{1/k}}$ *for all* $\alpha \in (0,1)$ *and* $k \geq 2$.

(ii) $\chi_{k,1-\alpha} > \chi_{(1-\alpha)^{1/k}}$ *for all* $\alpha \in (0,1)$ *and* $k \geq 2$.

(iii) There exists $\varepsilon > 0$ *such that, for all* $\alpha \in (0,1)$ *and* $k \geq 1$,

$$\varepsilon\sqrt{\log k} - \sqrt{-2\log(1-\alpha)} \leq \chi_{(1-\alpha)^{1/k}} \leq \sqrt{2\log 2k} + \sqrt{-2\log\alpha}.$$

Figure 6 compares the Šidák and μ-projection critical values. In Section 2.3 we argued that Šidák and Bonferroni bands are both narrower than μ-projection if $\alpha < 0.5$ and $k \leq p$. What if $k > p$? The figure shows the smallest value of k needed for the μ-projection band to be asymptotically weakly narrower than Šidák. Clearly, for this to happen at usual significance levels, either the model dimension p must be very small, or the number k of parameters of interest must be in the 1,000s.

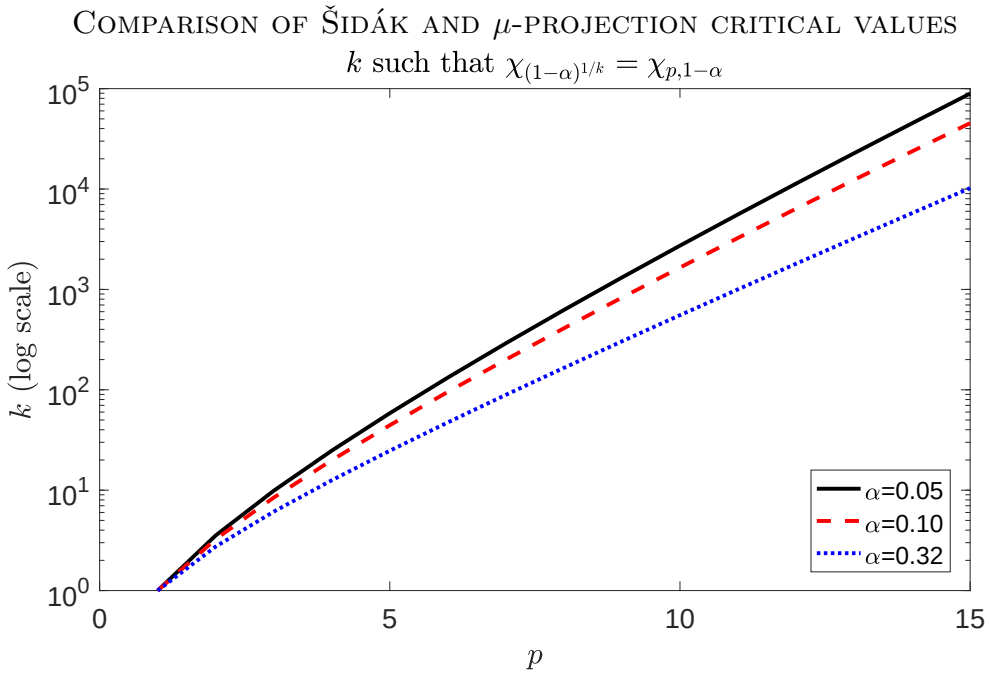

Figure 6: Smallest value of k needed for the Šidák band to be asymptotically weakly wider than the μ-projection band, as a function of p. Horizontal axis: p, vertical axis: k in log scale. The curves correspond to significance levels $\alpha = 0.05$ (black), $\alpha = 0.1$ (red), and $\alpha = 0.32$ (blue).

A.3 Gaussian decision problem

The argument for invariance of the decision problem in Section 3 is standard and we sketch it here for the sake of exposition. See Berger (1985, section 6.2.2) for a definition of invariant decision problems. Let $\mathcal{T} \equiv \{f_\lambda(x) = x + \lambda \mid \lambda \in \mathbb{R}^p\}$ denote the group of translations of the data X by arbitrary vectors $\lambda \in \mathbb{R}^p$. First, we note that the Gaussian statistical model (3) is invariant under $\mathcal{T}$. Second, for any data transformation $f_\lambda \in \mathcal{T}$ and any action $C = \times_{j=1}^k [a_j, b_j] \in \mathcal{R}$, the alternative action given by $\tilde{C} \equiv G\lambda + C = \times_{j=1}^k [g_j'\lambda + a_j, g_j'\lambda + b_j] \in \mathcal{R}$ (where g_j' is the j-th row of G) satisfies $\mathbb{1}\{G\mu \in C\} = \mathbb{1}\{G(\lambda + \mu) \in \tilde{C}\}$ and $L(C) = L(\tilde{C})$ for all $\mu \in \mathbb{R}^p$.

A.4 Implementation of sup-t band

Here we discuss an alternative bootstrap procedure, and we state formal results guaranteeing the validity of the plug-in, bootstrap, and Bayesian bands discussed in Section 4.

ALTERNATIVE BOOTSTRAP PROCEDURE. Algorithm 3 defines an alternative critical-value-based bootstrap band. The procedure first computes the standard deviation $\hat{\sigma}_j^*$ of the bootstrap draws of $\hat{\theta}_j$, for each j. It then computes a bootstrap approximation $\hat{q}_{1-\alpha}$ to the sup-t critical value $q_{1-\alpha}(\Sigma)$. Finally, the band is given by $\widehat{B}(\hat{q}_{1-\alpha})$, except that the bootstrap standard errors $\hat{\sigma}_j^*$ are used in place of the delta method standard errors $\hat{\sigma}_j$. Thus, Algorithm 3 does not require evaluation of the partial derivatives of $h(\cdot)$. Unlike the quantile-based bootstrap band, the critical-value-based band is symmetric around the point estimate $\hat{\theta}$ in any finite sample. Proposition 3 below shows that the critical-value-based band is asymptotically equivalent with the sup-t band $\widehat{B}(q_{1-\alpha}(\Sigma))$ if the bootstrap for $\hat{\mu}$ is valid and the bootstrap standard errors $\hat{\sigma}_j^*$ are consistent.

Algorithm 3 Critical-value-based bootstrap band

Let $\hat{P}$ be the bootstrap distribution of $\hat{\mu}$
Draw N samples $\hat{\mu}^{(1)}, \ldots, \hat{\mu}^{(N)}$ from $\hat{P}$
for $\ell = 1, \ldots, N$ **do**
    $\hat{\theta}^{(\ell)} = h(\hat{\mu}^{(\ell)})$
end for
for $j = 1, \ldots, k$ **do**
    Compute the empirical standard deviation $\hat{\sigma}_j^*$ of draws $\hat{\theta}_j^{(1)}, \ldots, \hat{\theta}_j^{(N)}$
end for
for $\ell = 1, \ldots, N$ **do**
    $\hat{m}^{(\ell)} = \max_{j=1,\ldots,k} \frac{|\hat{\theta}_j^{(\ell)} - \hat{\theta}_j|}{\hat{\sigma}_j^*}$
end for
Let $\hat{q}_{1-\alpha}$ be the $1-\alpha$ empirical quantile of the draws $\hat{m}^{(1)}, \ldots, \hat{m}^{(N)}$
$\widehat{C} = \times_{j=1}^{k} [\hat{\theta}_j - \hat{\sigma}_j^* \hat{q}_{1-\alpha}, \hat{\theta}_j + \hat{\sigma}_j^* \hat{q}_{1-\alpha}]$

The critical-value-based bootstrap band is finite-sample equivalent (up to minor numerical details) with the bootstrap-adjusted Bonferroni or projection ("Wald") bands of Lütkepohl et al. (2015a,b). Lütkepohl et al. view their approach as a method for adjusting downward the critical values used in the Bonferroni or projection approaches, in order to mitigate the conservativeness of the original bands. As our Algorithm 3 makes clear, the "bootstrap-adjusted" procedure is best thought of as a direct bootstrap implementation of the sup-t band. This interpretation is useful from a practical perspective: The purpose of the bootstrap is to deliver good approximations of the bootstrap standard errors and the

bootstrapped sup-t quantile, so the bootstrap procedure—including the number of bootstrap draws—should be designed with these goals in mind.

In principle, Algorithm 3 could also be used to construct a *Bayesian* band with simultaneous credibility $1-\alpha$. However, since the algorithm is based on t-statistics, it appears less well motivated from a finite-sample Bayesian perspective, except perhaps in cases where the posterior distribution is exactly Gaussian (as in Liu, 2011, chapter 2.9).

THEORETICAL RESULTS. First we state the easy result that the sup-t critical value is a continuous function of the (possibly singular) variance-covariance matrix Σ. This result then implies the validity of the plug-in implementation of the sup-t band.

Lemma 6. *For any $\zeta \in (0,1)$, the function $\tilde{\Sigma} \mapsto q_\zeta(\tilde{\Sigma})$ in Definition 2 is continuous on the set $\mathcal{S}_{p,k}$ defined in Lemma 3.*

Next, we state a result guaranteeing that the bootstrap and Bayesian implementations of the sup-t band in Section 4 deliver bands with frequentist asymptotic validity. In the proposition, the auxiliary random variable $\hat{\mu}^*$ should be thought of as a bootstrap draw of $\hat{\mu}$ or a draw from the posterior of μ.

Proposition 3. *Let Assumption 1 hold. Let $\hat{\mu}^* \in \mathbb{R}^p$ be a random vector whose distribution conditional on the data is denoted $\hat{P}$. Let $\hat{P}_M$ denote the distribution of $\sqrt{n}(\hat{\mu}^* - \hat{\mu})$, conditional on the data. Let P_M denote the distribution $N_p(\mathbf{0}_p, \Omega)$. Assume*

$$\rho(\hat{P}_M, P_M) \xrightarrow{p} 0 \quad \text{as } n \to \infty,$$

where $\rho(\cdot,\cdot)$ denotes the Bounded Lipschitz metric or any other metric that metrizes weak convergence of probability measures on $\mathbb{R}^p$.

(i) Assume for each $j = 1, \ldots, k$, there exists a random variable $\hat{\sigma}_j^*$ such that $\sqrt{n}\hat{\sigma}_j^* \xrightarrow{p} \Sigma_{jj}^{1/2}$. Let $\hat{q}_{1-\alpha}$ denote the $1-\alpha$ quantile of the distribution of $\max_j(\hat{\sigma}_j^*)^{-1}|h_j(\hat{\mu}^*) - h_j(\hat{\mu})|$, conditional on the data (and thus also conditional on the $\hat{\sigma}_j^*$). Then

$$\hat{q}_{1-\alpha} \xrightarrow{p} q_{1-\alpha}(\Sigma).$$

(ii) Denote the ζ quantile of $h_j(\hat{\mu}^*)$, conditional on the data, by $\hat{Q}_{j,\zeta}$. Define $\hat{\zeta}$ as the largest value of $\zeta \in [0, 1/2]$ such that $\hat{P}(h(\hat{\mu}^*) \in \times_{j=1}^k [\hat{Q}_{j,\zeta}, \hat{Q}_{j,1-\zeta}]) \geq 1-\alpha$, conditional on

the data. Let $\Phi(\cdot)$ *denote the standard normal CDF. Then*

$$\hat{\zeta} \xrightarrow{p} \zeta^* \equiv \Phi(-q_{1-\alpha}(\Sigma)).$$

(iii) Under the same conditions as in (ii), we have, for any $j = 1, \ldots, k$*,*

$$\hat{Q}_{j,\hat{\zeta}} = \hat{\theta}_j - \hat{\sigma}_j q_{1-\alpha}(\Sigma) + o_p(n^{-1/2}),$$

$$\hat{Q}_{j,1-\hat{\zeta}} = \hat{\theta}_j + \hat{\sigma}_j q_{1-\alpha}(\Sigma) + o_p(n^{-1/2}).$$

A.5 Impulse response function confidence bands

In this section we review the literature on confidence bands for impulse response functions, give additional details of the VAR application, and present a simulation study of the VAR confidence band procedures.

Literature review. Here we briefly review the literature on confidence bands for impulse response functions, as well as the closely related literature that constructs confidence bands for path forecasts.[33] Hymans (1968) constructs path forecast bands using θ-projection. Sims & Zha (1999) propose a procedure for plotting the principal components decomposition of the variance-covariance matrix, although this does not lead to a confidence band in the sense of this paper. Lütkepohl (2005, pp. 115–116) recommends the Bonferroni band. Jordà (2009) and Jordà & Marcellino (2010) develop projection-like confidence bands which control the "Wald coverage", in the terminology of Jordà et al. (2013); however, these bands do not control simultaneous coverage in the usual sense of equation (1) (cf. Wolf & Wunderli, 2015, section 3.3). Lütkepohl et al. (2015a,b) mention the Šidák band and propose bootstrap adjustments of the Bonferroni, μ-projection, and θ-projection procedures to make these less conservative; the adjusted procedures are essentially equivalent with the bootstrap sup-t band in Appendix A.4. Wolf & Wunderli (2015) use a bootstrap sup-t band to construct confidence bands for path forecasts (but not VAR impulse responses). Inoue & Kilian (2016) summarize estimation uncertainty for impulse responses using "shotgun plots", i.e., random samples from a bootstrapped confidence ellipsoid.[34] Lütkepohl et al. (2016) construct

[33] The two problems are equivalent (only) in Gaussian time series models (Wolf & Wunderli, 2015, p. 361).

[34] This deliberately does not generate a rectangular confidence region. The smallest rectangular region containing the Inoue & Kilian (2016) confidence ellipsoid equals the θ-projection confidence band, using the bootstrapped critical value and standard errors.

highest-density rectangular regions from bootstrap draws of the impulse responses, which is asymptotically equivalent to θ-projection under the regularity conditions in Section 4.

VAR MODEL AND IMPULSE RESPONSES. The VAR model assumes that the d-dimensional vector $y_t = (y_{1,t}, \ldots, y_{d,t})'$ of observed time series is driven in an autoregressive manner by a d-dimensional vector $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{d,t})'$ of unobserved economic shocks:

$$y_t = \nu + \sum_{\ell=1}^{\tau} A_\ell y_{t-\ell} + H\varepsilon_t, \quad t = 1, 2, \ldots, T.$$

The intercept vector ν is $d \times 1$, while the lag coefficient matrices A_ℓ and the impact matrix H are each $d \times d$. The VAR lag length τ is assumed finite here. The shocks are a strictly stationary martingale difference sequence with identity variance-covariance matrix:

$$E[\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots] = \mathbf{0}_d, \quad \mathrm{Var}(\varepsilon_t) = I_d.$$

The identified model parameters are $\mu \equiv (\nu', \mathrm{vec}(A_1)', \ldots, \mathrm{vec}(A_\tau)', \mathrm{vech}(\Psi)')'$, where $\Psi \equiv HH'$ is the one-step forecast error variance-covariance matrix.

The impulse response matrix at horizon ℓ is given by $\Theta_\ell \equiv \partial y_{t+\ell}/\partial \varepsilon_t'$. It can be computed by the recursion

$$\Theta_0 = H, \quad \Theta_\ell = \sum_{b=1}^{\min\{\ell,\tau\}} A_b \Theta_{\ell-b}, \ \ell = 1, 2, \ldots$$

We are interested in the impulse response function of the first observed variable to the first shock, from horizon 0 to $k-1$: $\theta \equiv (\Theta_{0,11}, \Theta_{1,11}, \ldots, \Theta_{k-1,11})'$, where $\Theta_{\ell,11}$ denotes the $(1,1)$ element of Θ_ℓ. Since H is only identified up to $\Psi = HH'$, θ is not identified without further assumptions (Stock & Watson, 2016). We may point identify θ by imposing exclusion restrictions on H or on Θ_ℓ for various ℓ (or as $\ell \to \infty$). Alternatively, we may assume that an external instrument z_t is available and satisfies $E[z_t\varepsilon_{1,t}] \neq 0$ and $E[z_t\varepsilon_{i,t}] = 0$ for $i \geq 2$ (Stock & Watson, 2012; Mertens & Ravn, 2013). However point identification is achieved, there exists a function $h(\cdot)$ such that $\theta = h(\mu)$.[35] This function is nonlinear and typically continuously differentiable (Lütkepohl, 2005, chapter 3.7). Lütkepohl (2005) and Kilian & Lütkepohl (2017) review limit theory for the least-squares estimator $\hat{\mu}$, bootstrap methods, and posterior sampling in VARs.[36]

[35] In the case of the external instrument, we augment the vector μ by the parameter vector $\gamma = E[(Y_t - E[Y_t \mid Y_{t-1}, \ldots, Y_{t-\tau}])z_t]$ (Montiel Olea et al., 2016).

[36] In cointegrated models as well as certain stationary models, the asymptotic variance Ω of $\hat{\mu}$ may be

Details of empirical implementation. To simplify comparisons with the bootstrap and Bayes procedures, the asymptotic variance of the VAR estimator $\hat{\mu}$ is calculated under the assumption of homoskedastic shocks ε_t. However, any of our procedures can be extended to allow for heteroskedasticity using standard methods.

The bootstrap is a homoskedastic recursive residual bootstrap. We use 10,000 bootstrap draws. For Bayesian inference we use a maximally diffuse normal-inverse-Wishart prior, and we sample from the posterior using its closed-form expression under Gaussian shocks (Uhlig, 2005, Appendix B). We use 10,000 posterior draws. The bootstrap and Bayesian procedures treat pre-sample observations of y_t as fixed. The plug-in sup-t quantile $q_{1-\alpha}(\hat{\Sigma})$ is approximated using 100,000 normal draws. We adjust for the fact that the sample for the external instrument is smaller than the sample for the VAR variables: The variance-covariance matrix for the VAR least-squares estimator is computed on the larger sample and then stitched together with the remaining variance-covariance on the smaller sample. It takes less than 3 minutes to compute all bootstrap and Bayes bands using Matlab R2016b on a personal laptop (2.60 GHz processor, single core, 8 GB RAM).

Additional empirical results. Figures 7 and 8 compare all common bands in the one-parameter class for the recursive and external instrument specifications, including θ-projection and μ-projection.[37] The μ-projection band is given by the asymptotic approximation $\widehat{B}(\chi_{p,1-\alpha})$. Evidently, both projection bands are substantially wider than the sup-t, Šidák, and Bonferroni bands, as theory predicts. The μ-projection band is wider than the θ-projection band since $p > k$.

Simulation study. Here we present Monte Carlo evidence on the coverage probability and average width of simultaneous confidence bands for VAR impulse response functions.

Following Lütkepohl et al. (2015b), we consider the bivariate VAR

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \sum_{\ell=1}^{\tau} \begin{pmatrix} \varphi \cdot \mathbb{1}(\ell = 1) & 0 \\ 0.5\ell^{-2} & 0.5\ell^{-2} \end{pmatrix} \begin{pmatrix} y_{1,t-\ell} \\ y_{2,t-\ell} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0.3 & \sqrt{1-0.3^2} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix},$$

where $(\varepsilon_{1,t}, \varepsilon_{2,t})' \overset{i.i.d.}{\sim} N_2(\mathbf{0}_2, I_2)$. For lag length $\tau = 1$, this is the data generating process considered by Lütkepohl et al. (2015b). The parameter φ indexes the persistence of the

singular. Our theory and methods allow for singularities.

[37] A caveat is that the asymptotic validity of the projection bands rests on the asymptotic variances Ω and Σ in Assumption 1 being positive definite, which is not necessarily guaranteed in the VAR setting.

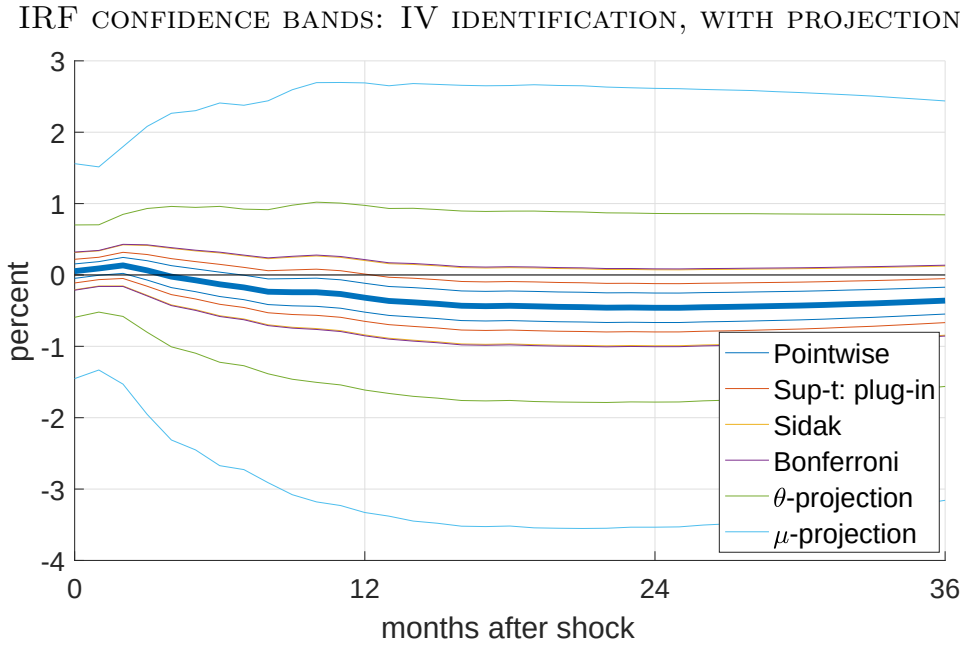

Figure 7: 68% confidence bands for IRF of IP to 1-stdev contractionary monetary policy shock, external instrument identification. See caption for Figure 3.

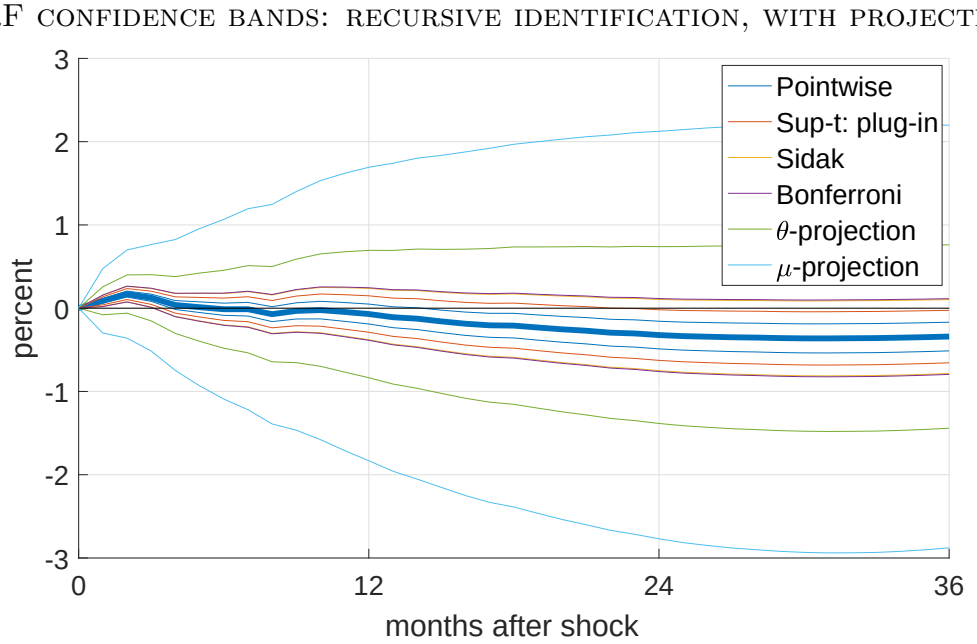

Figure 8: 68% confidence bands for IRF of IP to 1-stdev contractionary monetary policy shock, recursive identification. See caption for Figure 3.

VAR. We consider designs with $\tau \in \{1, 4\}$ and $\varphi \in \{0, 0.5, 0.9, 1\}$. Some of our designs assume the availability of an external instrument

$$z_t = \varepsilon_{1,t} + \sqrt{1/R^2 - 1} \cdot v_t,$$

where $v_t \overset{i.i.d.}{\sim} N(0, 1)$, independent of $(\varepsilon_{1,t}, \varepsilon_{2,t})$. Note that $R^2 = \text{Var}(\varepsilon_{1,t})/\text{Var}(z_t)$. We consider the values $R^2 \in \{0.1, 0.5\}$.

We compute confidence bands for the impulse response function of $y_{2,t}$ to $\varepsilon_{1,t}$. The VAR is either estimated under recursive identification (ordering $\varepsilon_{1,t}$ first, correctly) or using the external IV z_t. Our results consider impulse responses out to horizon 10 or 20 (i.e., $k = 11$ or $k = 21$ parameters of interest, as the impact response is also included). We consider confidence levels $1-\alpha \in \{68\%, 90\%\}$ and sample sizes $T \in \{200, 500\}$. We compute pointwise, Šidák, Bonferroni, μ-projection (the asymptotic approximation $\widehat{B}(\chi_{p,1-\alpha})$), and θ-projection bands. We also compute the plug-in sup-t band, the homoskedastic residual bootstrap sup-t band, and the maximally diffuse normal-inverse-Wishart Bayes band. We run 2,000 Monte Carlo replications per data generating process. The plug-in sup-t band uses 100,000 normal draws, while the bootstrap and Bayes sup-t bands each use 2,000 draws.

Tables 2 and 3 display the simulated finite-sample simultaneous coverage probability and expected sum of component widths for the confidence bands in Sections 2 and 4. The plug-in, bootstrap, and Bayes sup-t bands all perform similarly well for moderately persistent VAR processes ($\phi \in \{0, 0.5\}$). For highly persistent VARs ($\phi \in \{0.9, 1\}$), only the Bayesian band exhibits satisfactory coverage, which comes at the expense of slightly larger width. Šidák and Bonferroni bands have coverage rates that are comparable to the Bayesian band for confidence level $1 - \alpha = 0.90$, but they are very conservative at $1 - \alpha = 0.68$. The Šidák and Bonferroni bands tend to be 10–20% wider than the sup-t bands at confidence level $1 - \alpha = 0.90$, and 30–35% wider at $1 - \alpha = 0.68$. For most data generating processes, the projection bands are highly conservative and on average 60–120% wider than the sup-t bands. In the case of external instrument identification, a sample size of $T = 500$ is required for reasonable coverage of the plug-in and bootstrap sup-t bands.

We conclude that the Bayesian band possesses the best mix of coverage and width properties among the bands considered here. We caution, however, that the present simulation study is of relatively small scale. It is plausible that the plug-in and bootstrap sup-t implementations can be improved using bias reduction techniques (Lütkepohl et al., 2015a, section A.1) or modifications to the bootstrap procedure.

VAR SIMULATIONS: COVERAGE PROBABILITY

DGP				Coverage: sup-t			Coverage: other bands				
τ	φ	R^2	T	Plu	Boo	Bay	Pw	Sid	Bon	θ-p	μ-p
A. Recursive, 90% bands, max horizon 10											
1	0.0	–	200	0.72	0.88	0.90	0.56	0.77	0.77	0.84	0.83
1	0.0	–	500	0.80	0.89	0.89	0.61	0.87	0.87	0.95	0.94
1	0.5	–	200	0.79	0.87	0.91	0.59	0.85	0.85	0.95	0.94
1	0.5	–	500	0.85	0.89	0.91	0.65	0.92	0.92	0.98	0.97
1	0.9	–	200	0.80	0.73	0.88	0.61	0.88	0.89	0.97	0.97
1	0.9	–	500	0.87	0.84	0.89	0.68	0.93	0.94	1.00	0.99
1	1.0	–	200	0.63	0.38	0.76	0.41	0.77	0.77	0.96	0.94
1	1.0	–	500	0.73	0.55	0.79	0.52	0.85	0.85	0.99	0.98
4	0.5	–	200	0.80	0.74	0.87	0.50	0.85	0.86	0.99	1.00
4	0.5	–	500	0.88	0.85	0.90	0.58	0.93	0.93	1.00	1.00
4	0.9	–	200	0.79	0.71	0.86	0.53	0.86	0.86	0.98	0.99
4	0.9	–	500	0.86	0.82	0.88	0.60	0.91	0.91	0.99	1.00
4	1.0	–	200	0.69	0.55	0.81	0.43	0.77	0.77	0.96	0.99
4	1.0	–	500	0.82	0.76	0.87	0.56	0.89	0.89	0.99	1.00
B. Recursive, 68% bands, max horizon 10											
1	0.9	–	200	0.59	0.51	0.65	0.25	0.79	0.81	0.96	0.95
1	0.9	–	500	0.66	0.60	0.69	0.27	0.85	0.87	0.99	0.98
C. Recursive, 90% bands, max horizon 20											
1	0.9	–	200	0.74	0.74	0.89	0.55	0.83	0.83	0.94	0.90
1	0.9	–	500	0.83	0.83	0.90	0.64	0.92	0.92	0.99	0.97
D. External IV, 90% bands, max horizon 10											
1	0.5	0.1	200	0.76	0.83	–	0.59	0.83	0.83	0.93	0.93
1	0.5	0.1	500	0.84	0.88	–	0.66	0.91	0.91	0.98	0.98
1	0.9	0.1	200	0.77	0.70	–	0.60	0.86	0.86	0.97	0.97
1	0.9	0.1	500	0.84	0.82	–	0.66	0.93	0.93	0.99	0.99
1	0.5	0.5	200	0.80	0.87	–	0.63	0.85	0.86	0.95	0.95
1	0.5	0.5	500	0.84	0.88	–	0.67	0.91	0.91	0.98	0.98
1	0.9	0.5	200	0.81	0.74	–	0.61	0.90	0.90	0.98	0.98
1	0.9	0.5	500	0.86	0.83	–	0.67	0.93	0.93	1.00	1.00

Table 2: Simultaneous coverage probability of confidence bands in Monte Carlo study of VAR impulse response functions. First 4 columns: DGP parameters. Last 8 columns: simultaneous coverage rate of confidence bands (respectively: plug-in sup-t, bootstrap sup-t, Bayes sup-t, pointwise, Šidák, Bonferroni, θ-projection, μ-projection).

VAR SIMULATIONS: AVERAGE WIDTH RELATIVE TO POINTWISE BAND

DGP				Rel. width: sup-t			Rel. width: other bands			
τ	φ	R^2	T	Plu	Boo	Bay	Sid	Bon	θ-p	μ-p
A. Recursive, 90% bands, max horizon 10										
1	0.0	–	200	1.34	1.36	1.48	1.58	1.59	2.53	2.33
1	0.0	–	500	1.33	1.34	1.39	1.58	1.59	2.53	2.33
1	0.5	–	200	1.35	1.34	1.49	1.58	1.59	2.53	2.33
1	0.5	–	500	1.35	1.35	1.41	1.58	1.59	2.53	2.33
1	0.9	–	200	1.32	1.31	1.40	1.58	1.59	2.53	2.33
1	0.9	–	500	1.32	1.33	1.35	1.58	1.59	2.53	2.33
1	1.0	–	200	1.31	1.46	1.37	1.58	1.59	2.53	2.33
1	1.0	–	500	1.30	1.44	1.33	1.58	1.59	2.53	2.33
4	0.5	–	200	1.43	1.40	1.55	1.58	1.59	2.53	3.31
4	0.5	–	500	1.42	1.41	1.47	1.58	1.59	2.53	3.31
4	0.9	–	200	1.39	1.41	1.49	1.58	1.59	2.53	3.31
4	0.9	–	500	1.39	1.40	1.43	1.58	1.59	2.53	3.31
4	1.0	–	200	1.38	1.39	1.46	1.58	1.59	2.53	3.31
4	1.0	–	500	1.38	1.39	1.41	1.58	1.59	2.53	3.31
B. Recursive, 68% bands, max horizon 10										
1	0.9	–	200	1.60	1.60	1.65	2.13	2.19	3.57	3.24
1	0.9	–	500	1.60	1.61	1.62	2.13	2.19	3.57	3.24
C. Recursive, 90% bands, max horizon 20										
1	0.9	–	200	1.34	1.23	1.54	1.71	1.72	3.31	2.33
1	0.9	–	500	1.34	1.29	1.41	1.71	1.72	3.31	2.33
D. External IV, 90% bands, max horizon 10										
1	0.5	0.1	200	1.34	1.34	–	1.58	1.59	2.53	2.53
1	0.5	0.1	500	1.33	1.34	–	1.58	1.59	2.53	2.53
1	0.9	0.1	200	1.31	1.31	–	1.58	1.59	2.53	2.53
1	0.9	0.1	500	1.31	1.32	–	1.58	1.59	2.53	2.53
1	0.5	0.5	200	1.35	1.34	–	1.58	1.59	2.53	2.53
1	0.5	0.5	500	1.35	1.35	–	1.58	1.59	2.53	2.53
1	0.9	0.5	200	1.32	1.31	–	1.58	1.59	2.53	2.53
1	0.9	0.5	500	1.32	1.32	–	1.58	1.59	2.53	2.53

Table 3: Average width of confidence bands in Monte Carlo study of VAR impulse response functions. First 4 columns: DGP parameters. Last 7 columns: average sum of component widths of band, divided by same quantity for pointwise band. See abbreviations in caption for Table 2.

B Proofs

B.1 Lemma 1

For each $j = 1, \ldots, k$, define $\hat{d}_j \equiv \frac{\hat{b}_j - \hat{a}_j}{2}$ and $\hat{e}_j \equiv \frac{\sqrt{n}(\hat{a}_j + \hat{b}_j)}{2c}$. Note that $\hat{d} = (\hat{d}_1, \ldots, \hat{d}_k)' = o_p(n^{-1/2})$ and $\hat{e} = (\hat{e}_1, \ldots, \hat{e}_k)' = o_p(1)$. Define also $\hat{\sigma} \equiv (\hat{\sigma}_1, \ldots, \hat{\sigma}_k)'$. For any vector $x \in \mathbb{R}^k$, let $\text{diag}(x)$ denote the $k \times k$ diagonal matrix with the elements of x in order along the diagonal. Then, for any $c > 0$,

$$\begin{aligned} P\left(\theta \in \widehat{B}(c)\right) &= P\left(\| \text{diag}(\sqrt{n}\hat{\sigma} + \hat{e})^{-1}\sqrt{n}(\hat{\theta} - \theta + \hat{d})\|_\infty \leq c\right) \\ &= P\left(\| \text{diag}(\sqrt{n}\hat{\sigma} + o_p(1))^{-1}(\sqrt{n}(\hat{\theta} - \theta) + o_p(1))\|_\infty \leq c\right). \end{aligned}$$

The proposition now follows from the limiting distribution (2) of $\hat{\theta}$, the continuous mapping theorem, and the Portmanteau lemma. To apply the latter, we need to show that the probability measure of $\max_j |\Sigma_{jj}^{-1/2} V_j|$, where $V \sim N_k(\mathbf{0}_k, \Sigma)$, is dominated by Lebesgue measure. This follows from the fact that $P(\max_{j=1,\ldots,k} X_j \in \mathcal{A}) \leq \sum_{j=1}^k P(X_j \in \mathcal{A}) = 0$ for any collection $\{X_j\}_{j=1,\ldots,k}$ of scalar random variables and any Lebesgue null set $\mathcal{A}$. □

B.2 Lemma 2

If $C \in \mathcal{C}_{\text{eq}}$, then $C(x + \lambda) = G\lambda + C(x)$ for any $x, \lambda \in \mathbb{R}^p$. Hence, for any $x \in \mathbb{R}^p$,

$$C(x) = C(\mathbf{0}_p + x) = Gx + C(\mathbf{0}_p).$$

The lemma follows by setting $R = C(\mathbf{0}_p) \in \mathcal{R}$. □

B.3 Lemma 3

Given any $\tilde{\Sigma} \in \mathcal{S}_{p,k}$, if we let $\tilde{V} = (\tilde{V}_1, \ldots, \tilde{V}_k)' \sim N_k(\mathbf{0}_k, \tilde{\Sigma})$, then

$$\zeta = P\left(\max_j |\tilde{\Sigma}_{jj}^{-1/2}\tilde{V}_j| \leq q_\zeta(\tilde{\Sigma})\right) \leq P(|\tilde{\Sigma}_{11}^{-1/2}\tilde{V}_1| \leq q_\zeta(\tilde{\Sigma})) = P(\chi^2(1) \leq q_\zeta^2(\tilde{\Sigma})),$$

so

$$\inf_{\tilde{\Sigma} \in \mathcal{S}_{p,k}} q_\zeta^2(\tilde{\Sigma}) \geq \chi_\zeta^2.$$

On the other hand, let $G^* \equiv (g^*, \ldots, g^*)' \in \mathbb{R}^{k\times p}$, where $g^* \in \mathbb{R}^p$ is any vector satisfying $\|g^*\| = 1$, and define $\Sigma^* \equiv G^*(G^*)' \in \mathcal{S}_{p,k}$. Note that $\Sigma^*_{jj} = 1$ for all j. Then

$$\inf_{\tilde{\Sigma}\in\mathcal{S}_{p,k}} q_\zeta(\tilde{\Sigma}) \leq q_\zeta(\Sigma^*) = \chi_\zeta,$$

since

$$\zeta = P(|(g^*)'Z| \leq \chi_\zeta) = P(\|G^*Z\|_\infty \leq \chi_\zeta) = P(\|N_k(\mathbf{0}_k, \Sigma^*)\|_\infty \leq \chi_\zeta).$$

The inequality for the supremum in the lemma is the famous Šidák (1967) inequality. Finally, if $k \leq p$, then $I_k \in \mathcal{S}_{p,k}$, so $\sup_{\tilde{\Sigma}\in\mathcal{S}_{p,k}} q_\zeta(\tilde{\Sigma}) \geq q_\zeta(I_k) = \chi_{\zeta^{1/k}}$. □

B.4 Lemma 4

Given $\tilde{\Sigma} \in \mathcal{S}_{p,k}$ and $\tilde{V} \sim N_k(\mathbf{0}_k, \tilde{\Sigma})$, we can write $\tilde{V} \sim \tilde{G}Z$, where $Z \sim N_p(\mathbf{0}_p, I_p)$, and $\tilde{G} = (\tilde{g}_1, \ldots, \tilde{g}_k)'$ satisfies $\tilde{G}\tilde{G}' = \tilde{\Sigma}$ and thus $\|\tilde{g}_j\|^2 = \tilde{\Sigma}_{jj}$ for all j. Hence, the first statement of the lemma (a standard projection result) follows from Lemma 3 and

$$\max_j |\tilde{\Sigma}_{jj}^{-1/2}\tilde{V}_j| = \max_j \|\tilde{g}_j\|^{-1}|\tilde{g}_j'Z| \leq \max_j \|\tilde{g}_j\|^{-1}\|\tilde{g}_j\|\|Z\| = \|Z\| \sim \sqrt{\chi^2(p)}.$$

Now consider the second statement. That the supremum is strictly smaller than $\chi_{p,\zeta}$ follows from the above display and the fact that the event $\{Z \propto \tilde{g}_j\}$ has probability zero for any vector $\tilde{g}_j \in \mathbb{R}^p$ (when $p \geq 2$). To show the strict lower bound on the supremum, consider the particular $k \times p$ matrix $G^* \equiv (I_p, \iota/\sqrt{p}, \ldots, \iota/\sqrt{p})'$, where $\iota \equiv (1, \ldots, 1)'$. Then $\Sigma^* \equiv G^*(G^*)'$ satisfies $\Sigma^*_{jj} = 1$ for all j. If we let $Z \sim N_p(\mathbf{0}_p, I_p)$, then

$$\begin{aligned} P(\|N_k(\mathbf{0}_k, \Sigma^*)\|_\infty \leq \chi_{\zeta^{1/p}}) &= P(\|G^*Z\|_\infty \leq \chi_{\zeta^{1/p}}) \\ &= P(\|Z\|_\infty \leq \chi_{\zeta^{1/p}}) P\Big(|\iota'Z|/\sqrt{p} \leq \chi_{\zeta^{1/p}} \;\Big|\; \|Z\|_\infty \leq \chi_{\zeta^{1/p}}\Big) \\ &= \zeta\left\{1 - P\Big(|\iota'Z| > \sqrt{p}\chi_{\zeta^{1/p}} \;\Big|\; \|Z\|_\infty \leq \chi_{\zeta^{1/p}}\Big)\right\}. \end{aligned}$$

The lemma follows if we show that

$$P\Big(|\iota'Z| > \sqrt{p}\chi_{\zeta^{1/p}},\ \|Z\|_\infty \leq \chi_{\zeta^{1/p}}\Big) > 0.$$

Let $\varepsilon > 0$ satisfy $p(\chi_{\zeta^{1/p}} - \varepsilon) > \sqrt{p}\chi_{\zeta^{1/p}}$; such an ε exists because $p \geq 2$. Then

$$P\Big(|\iota'Z| > \sqrt{p}\chi_{\zeta^{1/p}},\ \|Z\|_\infty \leq \chi_{\zeta^{1/p}}\Big)$$

$$
\begin{aligned}
&\geq P\Big(|\iota' Z| > \sqrt{p}\chi_{\zeta^{1/p}},\ \|Z\|_\infty \leq \chi_{\zeta^{1/p}},\ \min_j Z_j \geq \chi_{\zeta^{1/p}} - \varepsilon\Big) \\
&\geq P\Big(p(\chi_{\zeta^{1/p}} - \varepsilon) > \sqrt{p}\chi_{\zeta^{1/p}},\ \|Z\|_\infty \leq \chi_{\zeta^{1/p}},\ \min_j Z_j \geq \chi_{\zeta^{1/p}} - \varepsilon\Big) \\
&= P\Big(\|Z\|_\infty \leq \chi_{\zeta^{1/p}},\ \min_j Z_j \geq \chi_{\zeta^{1/p}} - \varepsilon\Big) \\
&> 0. \quad \square
\end{aligned}
$$

B.5 Lemma 5

Let $U = (U_1, \ldots, U_k)' \sim N_k(\mathbf{0}_k, I_k)$.

(I): The statement is equivalent with $\log(1-(\frac{1}{k}\alpha+\frac{k-1}{k}\times 0)) > \frac{1}{k}\log(1-\alpha)+\frac{k-1}{k}\log(1-0)$. This is Jensen's inequality applied to the concave function $x \mapsto \log(1-x)$.

(II): This standard projection bias result follows from $\|U\|_\infty^2 \leq \|U\|^2 \sim \chi^2(k)$. Note that $\chi^2_{(1-\alpha)^{1/k}}$ is the $1-\alpha$ quantile of $\|U\|_\infty^2$.

(III): By Giné & Nickl (2016, Lemmas 2.3.4 and 2.4.11), there exists $\varepsilon > 0$ such that

$$
\varepsilon\sqrt{\log k} \leq E\|U\|_\infty \leq \sqrt{2\log 2k}.
$$

Hence, using Giné & Nickl (2016, Theorem 2.5.8),

$$
P\left(\|U\|_\infty \geq \sqrt{2\log 2k} + \sqrt{-2\log\alpha}\right) \leq P\left(\|U\|_\infty \geq E\|U\|_\infty + \sqrt{-2\log\alpha}\right) \leq \alpha,
$$

so $\chi_{(1-\alpha)^{1/k}} \leq \sqrt{2\log 2k} + \sqrt{-2\log\alpha}$. Similarly, Giné & Nickl (2016, Theorem 2.5.8) yields

$$
\begin{aligned}
P\left(\|U\|_\infty \leq \varepsilon\sqrt{\log k} - \sqrt{-2\log(1-\alpha)}\right) &\leq P\left(\|U\|_\infty \leq E\|U\|_\infty - \sqrt{-2\log(1-\alpha)}\right) \\
&\leq 1-\alpha,
\end{aligned}
$$

so $\chi_{(1-\alpha)^{1/k}} \geq \varepsilon\sqrt{\log k} - \sqrt{-2\log(1-\alpha)}$. $\square$

B.6 Lemma 6

Let $\tilde{\Sigma} \in \mathcal{S}_{p,k}$. We want to show $q_\zeta(\tilde{\Sigma}^{(\ell)}) \to q_\zeta(\tilde{\Sigma})$ as $\ell \to \infty$ for any sequence $\{\tilde{\Sigma}^{(\ell)}\} \in \mathcal{S}_{p,k}$ tending to $\tilde{\Sigma}$ as $\ell \to \infty$.

First we argue that the distribution $N_k(\mathbf{0}_k, \tilde{\Sigma}^{(\ell)})$ converges weakly to $N_k(\mathbf{0}_k, \tilde{\Sigma})$ as $\ell \to \infty$. This statement is obvious if $k = 1$. It then follows for general k by the Cramér-Wold device.

Now let $\tilde{V} \sim N_k(\mathbf{0}_k, \tilde{\Sigma})$ as well as $\tilde{V}^{(\ell)} \sim N_k(\mathbf{0}_k, \tilde{\Sigma}^{(\ell)})$ for all ℓ. By the continuous mapping theorem, $\tilde{\Sigma}_{jj} > 0$, and the above paragraph, the distribution of $\max_j |(\tilde{\Sigma}_{jj}^{(\ell)})^{-1/2}\tilde{V}_{jj}^{(\ell)}|$ converges weakly to the distribution $\max_j |\tilde{\Sigma}_{jj}^{-1/2}\tilde{V}_{jj}|$ as $\ell \to \infty$.

The statement of the lemma now follows from van der Vaart (1998, Lemma 21.2) if we show that the distribution of $\max_j |\tilde{\Sigma}_{jj}^{-1/2}\tilde{V}_{jj}|$ is absolutely continuous on $\mathbb{R}_+$. Represent this distribution as the distribution of $\|GZ\|_\infty$ where $G \in \mathbb{R}^{k\times p}$ and $Z \sim N_p(\mathbf{0}_p, I_p)$. We showed that the probability measure of $\|GZ\|_\infty$ is dominated by Lebesgue measure in the proof of Lemma 1. Now take an arbitrary non-empty interval (a, b), $0 \le a < b$. Denote elements of G by $g_{j\ell}$. We may assume the first column of G is not identically zero. Select $j^* \in \operatorname{argmax}_j |g_{j1}|$. Let e_1 denote the first p-dimensional unit vector. Then $\|Gz^*\|_\infty = \frac{a+b}{2}$ for $z^* \equiv \frac{a+b}{2g_{j^*1}}e_1$, so there exists a neighborhood S of z^* in $\mathbb{R}^p$ such that $\|Gz\|_\infty \in (a, b)$ for all $z \in S$. Then $P(\|GZ\|_\infty \in (a, b)) > P(Z \in S) > 0$. □

B.7 Proposition 1

We need an auxiliary lemma. It states that the coordinate-wise width of any translation equivariant confidence band of confidence level $1-\alpha$ is bounded from below by the coordinate-wise width of the band that has pointwise confidence level $1 - \alpha$. A similar result is stated by Piegorsch (1984, p. 15). To remind the reader of our notation: R_j denotes the interval $[a_j, b_j]$ (where $b_j > a_j$) and $R = \times_{j=1}^k R_j$. Moreover, g_j' is the j-th row of $G \in \mathbb{R}^{k\times p}$.

Lemma 7. *Let $C(x) = Gx + R \in \mathcal{C}_{1-\alpha} \cap \mathcal{C}_{eq}$. Then $b_j - a_j \ge 2\|g_j\|\chi_{1-\alpha}$ for $j = 1, \dots, k$.*

Proof. Let $Z \sim N_p(\mathbf{0}_p, I_p)$. For any $j = 1, \dots, k$,

$$
\begin{aligned}
P_\mu(G\mu \in C(x)) &= P(GZ \in R_1 \times \cdots \times R_k) \\
&\quad \text{(by the translation equivariance of } C(x)) \\
&\quad \le P(g_j'Z \in [a_j, b_j]) \\
&\quad \text{(by the monotonicity of probability)} \\
&\quad \le P(g_j'Z \in [-(b_j - a_j)/2, (b_j - a_j)/2]) \\
&\quad \text{(by Anderson's lemma)} \\
&\quad = P(\,|N_1(0,1)| \,\le (b_j - a_j)/(2\|g_j\|)).
\end{aligned}
$$

Since $C(x)$ has confidence level $1-\alpha$, we have that the right-hand side of the last equation is greater than or equal $1-\alpha$. This can only happen if $(b_j - a_j)/(2\|g_j\|) \geq \chi_{1-\alpha}$. Note that the second inequality in the above display applies Anderson's lemma.[38] □

The proof of Proposition 1 proceeds in three steps.

STEP 1: We first upper-bound the worst-case regret of the sup-t band. Define $\sigma \equiv (\|g_1\|, \ldots, \|g_k\|)'$. For any $L \in \mathcal{L}_H$, Lemma 7 implies that

$$\begin{aligned}
\frac{L(R_{\sup})}{\inf_{\tilde{R}\in\mathcal{R}_{1-\alpha}} L(\tilde{R})} &\leq \frac{L(R_{\sup})}{L(2\sigma\chi_{1-\alpha})} \\
&\quad \text{(by Lemma 7 and the monotonicity of } L\text{)} \\
&= \frac{L(2\sigma q_{1-\alpha}(GG'))}{L(2\sigma\chi_{1-\alpha})} \\
&\quad \text{(by definition of the sup-t band)} \\
&= \frac{2q_{1-\alpha}(GG')L(\sigma)}{2\chi_{1-\alpha}L(\sigma)} \\
&\quad \text{(by homogeneity of degree 1 of } L\text{)} \\
&= \frac{q_{1-\alpha}(GG')}{\chi_{1-\alpha}}.
\end{aligned}$$

Consequently, Step 1 shows that the worst-case relative regret of the sup-t band is no larger than the ratio of the sup-t critical value and the point-wise critical value:

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

STEP 2: We now find a lower bound on the worst-case regret of an arbitrary rectangle $R = \times_{j=1}^{k} R_j \in \mathcal{R}_{1-\alpha}$. Fix R and let $j_R^* \in \text{argmax}_{j=1,\ldots,k}(b_j - a_j)/\|g_j\|$. Thus, j_R^* is the coordinate at which band R has the largest width relative to the pointwise standard error. Consider now the loss function given by $L_R^*(r) \equiv r_{j_R^*}$ for all $r = (r_1, \ldots, r_k)' \in \mathbb{R}_+^k$. We make three observations: i) this loss function reports, for any vector $(r_1, r_2, \ldots, r_k)'$, the width corresponding to the j_R^*-th entry; ii) $L_R^* \in \mathcal{L}_H$; and iii):

$$\inf_{\tilde{R}\in\mathcal{R}_{1-\alpha}} L_R^*(\tilde{R}) = 2\|g_{j_R^*}\|\chi_{1-\alpha},$$

[38] https://en.wikipedia.org/wiki/Anderson%27s_theorem

where the infimum is achieved by the sequence of bands that equal the Wald interval $g_j'x \pm \|g_j\|(\chi_{1-\alpha} + \varepsilon_n)$ at coordinate j_R^* (with $\varepsilon_n \to 0$) and have interval endpoints tending to plus/minus infinity at all other components. Thus, the worst-case relative regret of any band $R = \times_{j=1}^k [a_j, b_j]$ is bounded below by:

$$\sup_{L \in \mathcal{L}_H} \frac{L(R)}{\inf_{\tilde{R} \in \mathcal{C}_{1-\alpha}} L(\tilde{R})} \geq \frac{L_R^*(R)}{\inf_{\tilde{R} \in \mathcal{R}_{1-\alpha}} L_R^*(\tilde{R})} = \frac{b_{j_R^*} - a_{j_R^*}}{2\|g_{j_R^*}\|\chi_{1-\alpha}} = \frac{1}{2\chi_{1-\alpha}} \max_{j=1,\ldots,k} \frac{b_j - a_j}{\|g_j\|}.$$

STEP 3: Applying Step 2 to $R = R_{\text{sup}}$, the far right-hand side above equals $q_{1-\alpha}(GG')/\chi_{1-\alpha}$. Therefore, Step 1 and 2 imply that

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

Hence, it now suffices to show that $R = \times_{j=1}^k [a_j, b_j] \neq R_{\text{sup}}$ implies $\max_j (b_j - a_j)/\|g_j\| > 2q_{1-\alpha}(GG')$. Suppose to the contrary that there existed a rectangle $R \in \mathcal{R}_{1-\alpha}$ such that $b_j - a_j \leq 2\|g_j\| q_{1-\alpha}(GG')$ for all j, with strict inequality for at least one j. This contradicts the tautness of the sup-t band (Freyberger & Rai, 2017, Corollary 1). Hence, we conclude that for, any $R \in \mathcal{R}_{1-\alpha}$,

$$\sup_{L \in \mathcal{L}_H} \frac{L(R)}{\inf_{\tilde{R} \in \mathcal{R}_{1-\alpha}} L(\tilde{R})} \geq \frac{q_{1-\alpha}(GG')}{\chi_{1-\alpha}},$$

with strict inequality for any $R \neq R_{\text{sup}}$. □

B.8 Proposition 2

Fix $j = 1, \ldots, k$. Continuous differentiability of $h(\cdot)$ at μ implies $h(\tilde{\mu}) - h(\mu) = \dot{h}_j(\mu)'(\tilde{\mu} - \mu) + o(\|\tilde{\mu} - \mu\|)$ as $\|\tilde{\mu} - \mu\| \to 0$. Hence, for any $\tilde{\mu} \in \widehat{W}_\mu$,

$$\begin{aligned} h_j(\tilde{\mu}) &= h_j(\hat{\mu}) + h_j(\tilde{\mu}) - h_j(\mu) - [h_j(\hat{\mu}) - h_j(\mu)] \\ &= h_j(\hat{\mu}) + \dot{h}_j(\mu)'(\tilde{\mu} - \mu) - \dot{h}_j(\mu)'(\hat{\mu} - \mu) + o(\|\tilde{\mu} - \mu\|) + o_p(\|\hat{\mu} - \mu\|) \\ &= h_j(\hat{\mu}) + \dot{h}_j(\mu)'(\tilde{\mu} - \hat{\mu}) + o_p(n^{-1/2}) \quad \text{(uniformly in } \tilde{\mu}), \end{aligned}$$

where the last line uses $\|\hat{\mu} - \mu\| = O_p(n^{-1/2})$—by Assumption 1(ii)—and $\|\tilde{\mu} - \mu\| \leq \|\hat{\mu} - \mu\| + \|\hat{\mu} - \tilde{\mu}\| \leq \|\hat{\mu} - \mu\| + \|\hat{\Omega}^{1/2}\|\|\hat{\Omega}^{-1/2}(\hat{\mu} - \tilde{\mu})\| = O_p(n^{-1/2})$ uniformly for $\tilde{\mu} \in \widehat{W}_\mu$. Thus,

$$\begin{aligned}\sup_{\tilde{\mu} \in \widehat{W}_\mu} h_j(\tilde{\mu}) &= h_j(\hat{\mu}) + \sup_{\tilde{\mu} \in \widehat{W}_\mu} \dot{h}_j(\mu)'(\tilde{\mu} - \hat{\mu}) + o_p(n^{-1/2}) \\ &= h_j(\hat{\mu}) + \frac{\chi_{p,1-\alpha}}{\sqrt{n}} \|\hat{\Omega}^{1/2} \dot{h}_j(\mu)\| + o_p(n^{-1/2}) \\ &= \hat{\theta}_j + \chi_{p,1-\alpha} \hat{\sigma}_j + o_p(n^{-1/2}).\end{aligned}$$

The second equality above follows from the Cauchy-Schwarz inequality and the fact that $\mathcal{M}$ contains a neighborhood of μ, which implies $P(\{\tilde{\mu} \in \mathbb{R}^p \mid \|\hat{\Omega}^{-1/2}(\tilde{\mu} - \hat{\mu})\| = \chi_{p,1-\alpha}/\sqrt{n}\} \subset \widehat{W}_\mu) \to 1$. The result for the infimum follows by substituting $-h(\cdot)$ for $h(\cdot)$. □

B.9 Proposition 3

(I): Let $\{n_\ell\}$ be an arbitrary subsequence of $\{n\}$. We need to show that there exists a further subsequence along which $\hat{q}_{1-\alpha} \xrightarrow{a.s.} q_{1-\alpha}(\Sigma)$. By assumption, $\rho(\hat{P}_M, P_M) \xrightarrow{p} 0$ and $\sqrt{n}\hat{\sigma}_j^* \xrightarrow{p} \Sigma_{jj}^{1/2}$ (for all j) along the subsequence $\{n_\ell\}$. Thus, we can extract a further subsequence $\{n_m\}$ of $\{n_\ell\}$ such that $\rho(\hat{P}_M, P_M) \xrightarrow{a.s.} 0$ and $\sqrt{n}\hat{\sigma}_j^* \xrightarrow{a.s.} \Sigma_{jj}^{1/2}$ (for all j) along $\{n_m\}$. All remaining asymptotic statements in the proof of part (i) are implicitly with respect to this subsequence $\{n_m\}$.

Since $\hat{P}_M$ converges weakly to P_M, almost surely, the continuous differentiability of $h(\cdot)$ and the delta method imply that the conditional distribution of $\max_j(\hat{\sigma}_j^*)^{-1}|h_j(\hat{\mu}^*) - h_j(\hat{\mu})|$ converges weakly to the distribution of $\max_j |\Sigma_{jj}^{-1/2} V_j|$, where $V \sim N_k(\mathbf{0}_k, \Sigma)$, almost surely. Almost sure convergence of the $1-\alpha$ quantile follows as in the proof of Lemma 6.

(II): As above, given a subsequence of $\{n\}$, extract a further subsequence along which $\rho(\hat{P}_M, P_M) \xrightarrow{a.s.} 0$ and $\sqrt{n}\hat{\sigma}_j \xrightarrow{a.s.} \Sigma_{jj}^{1/2}$ (for all j). We need to show

$$\hat{\zeta} \xrightarrow{a.s.} \zeta^* \equiv \Phi(-q_{1-\alpha}(\Sigma))$$

along this latter subsequence. Except where noted, all asymptotic statements in the rest of the proof are with respect to this subsequence.

For each j, let $\hat{Q}_{j,\zeta}^V$ denote the ζ quantile of the distribution of $\sqrt{n}(h_j(\hat{\mu}^*) - h_j(\hat{\mu}))$, conditional on the data. By the monotone transformation preservation property of quantiles, we have $\hat{Q}_{j,\zeta}^V = \sqrt{n}(\hat{Q}_{j,\zeta} - \hat{\theta}_j)$ for all j and ζ. Using a delta method argument as in part (i)

above, the conditional distribution of $\sqrt{n}(h_j(\hat{\mu}^*)-h_j(\hat{\mu}))$ converges weakly to the distribution of $V \sim N_k(\mathbf{0}_k, \Sigma)$, almost surely. Thus, $\hat{Q}^V_{j,\zeta} \overset{a.s.}{\to} \Sigma_{jj}^{1/2}\Phi^{-1}(\zeta)$, almost surely, for any j and ζ.

We first show $\liminf \hat{\zeta} \geq \zeta^*$, almost surely. Suppose to the contrary that for some $\varepsilon > 0$, we have $\hat{\zeta} < \zeta^* - \varepsilon$ along some (further) subsequence $\{\tilde{n}_\ell\}$, with positive probability. Choose $\delta > 0$ and $\tilde{\alpha} \in (0, \alpha)$ so that $\Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^* - \varepsilon) + \delta = -\Sigma_{jj}^{1/2} q_{1-\tilde{\alpha}}(\Sigma)$. By the argument in the previous paragraph, there exists an event E with probability 1 such that $\hat{Q}^V_{j,\zeta^*-\varepsilon} < \Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^* - \varepsilon) + \delta$ and $\hat{Q}^V_{j,1-(\zeta^*-\varepsilon)} > \Sigma_{jj}^{1/2}\Phi^{-1}(1-(\zeta^* - \varepsilon)) - \delta$ for all j, when sufficiently far along $\{\tilde{n}_\ell\}$. Since $\hat{\zeta}$ is defined as the *largest* value of ζ such that $\hat{P}(h(\hat{\mu}^*) \in \times_{j=1}^k [\hat{Q}_{j,\zeta}, \hat{Q}_{j,1-\zeta}]) \geq 1 - \alpha$, we have that

$$\hat{\zeta} < \zeta^* - \varepsilon$$

implies

$$\hat{P}\left(h(\hat{\mu}^*) \in \times_{j=1}^k [\hat{Q}_{j,\zeta^*-\varepsilon}, \hat{Q}_{j,1-(\zeta^*-\varepsilon)}]\right) < 1 - \alpha,$$

which is equivalent with

$$\hat{P}\left(\sqrt{n}(h(\hat{\mu}^*) - h(\hat{\mu})) \in \times_{j=1}^k [\hat{Q}^V_{j,\zeta^*-\varepsilon}, \hat{Q}^V_{j,1-(\zeta^*-\varepsilon)}]\right) < 1 - \alpha,$$

which on the event E further implies

$$\hat{P}\left(\sqrt{n}(h(\hat{\mu}^*) - h(\hat{\mu})) \in \times_{j=1}^k [\Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^* - \varepsilon) + \delta, \Sigma_{jj}^{1/2}\Phi^{-1}(1-(\zeta^* - \varepsilon)) - \delta]\right) < 1 - \alpha$$

when sufficiently far along $\{\tilde{n}_\ell\}$, or equivalently,

$$\hat{P}\left(\sqrt{n}(h(\hat{\mu}^*) - h(\hat{\mu})) \in \times_{j=1}^k [-\Sigma_{jj}^{1/2} q_{1-\tilde{\alpha}}(\Sigma), \Sigma_{jj}^{1/2} q_{1-\tilde{\alpha}}(\Sigma)]\right) < 1 - \alpha,$$

or equivalently

$$\hat{P}\left(\max_j \Sigma_{jj}^{-1/2}\sqrt{n}|h_j(\hat{\mu}^*) - h_j(\hat{\mu})| \leq q_{1-\tilde{\alpha}}(\Sigma)\right) < 1 - \alpha,$$

or equivalently

$$\hat{P}\left(\max_j \Sigma_{jj}^{-1/2}\sqrt{n}|h_j(\hat{\mu}^*) - h_j(\hat{\mu})| \leq q_{1-\tilde{\alpha}}(\Sigma)\right) - (1 - \tilde{\alpha}) < \tilde{\alpha} - \alpha.$$

However, while the right-hand side above is strictly negative, the left-hand side tends to zero along $\{\tilde{n}_\ell\}$ almost surely by the above-mentioned weak convergence of $\sqrt{n}(h_j(\hat{\mu}^*) - h_j(\hat{\mu}))$,

the continuous mapping theorem, and Definition 2. We have arrived at a contradiction, and thus conclude that $\liminf \hat{\zeta} \geq \zeta^*$ almost surely.

We similarly show that $\limsup \hat{\zeta} \leq \zeta^*$, almost surely. Suppose to the contrary that for some $\varepsilon > 0$, we have $\hat{\zeta} > \zeta^* + \varepsilon$ along some (further) subsequence $\{\tilde{n}_\ell\}$, with positive probability. By monotonicity of quantiles in ζ,

$$\hat{\zeta} > \zeta^* + \varepsilon \quad \text{and} \quad \hat{P}\left(h(\hat{\mu}^*) \in \times_{j=1}^k [\hat{Q}_{j,\hat{\zeta}}, \hat{Q}_{j,1-\hat{\zeta}}]\right) \geq 1 - \alpha$$

imply

$$\hat{P}\left(h(\hat{\mu}^*) \in \times_{j=1}^k [\hat{Q}_{j,\zeta^*+\varepsilon}, \hat{Q}_{j,1-(\zeta^*+\varepsilon)}]\right) \geq 1 - \alpha.$$

We can now apply analogous arguments to the previous paragraph to ultimately show that the event that the above inequality holds along $\{\tilde{n}_\ell\}$ must have probability zero.

(III): Continue with the subsequence chosen at the beginning of part (ii). It suffices to show that

$$\sqrt{n}(\hat{Q}_{j,\hat{\zeta}} - \hat{\theta}_j) \xrightarrow{a.s.} -\Sigma_{jj}^{1/2} q_{1-\alpha}(\Sigma) \tag{6}$$

along this subsequence (the argument for $\hat{Q}_{j,1-\hat{\zeta}}$ follows the same way).

Let $\varepsilon > 0$ be arbitrary. Let $\delta > 0$ satisfy $\Phi^{-1}(\zeta^* + \delta) - \Phi^{-1}(\zeta^* - \delta) = \Sigma_{jj}^{-1/2}\varepsilon/2$. Part (ii) implies $|\hat{\zeta} - \zeta^*| < \delta$ and $\hat{Q}^V_{j,\zeta^*+\delta} > \Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^*) > \hat{Q}^V_{j,\zeta^*-\delta}$ when sufficiently far along the subsequence, almost surely. Thus,

$$\begin{aligned}
&\left|\sqrt{n}(\hat{Q}_{j,\hat{\zeta}} - \hat{\theta}_j) + \Sigma_{jj}^{1/2} q_{1-\alpha}(\Sigma)\right| \\
&\quad = \left|\hat{Q}^V_{j,\hat{\zeta}} - \Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^*)\right| \\
&\quad \leq \left(\hat{Q}^V_{j,\zeta^*+\delta} - \Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^*)\right) + \left(\Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^*) - \hat{Q}^V_{j,\zeta^*-\delta}\right) \\
&\quad = \left(\hat{Q}^V_{j,\zeta^*+\delta} - \Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^* + \delta)\right) + \left(\Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^* - \delta) - \hat{Q}^V_{j,\zeta^*-\delta}\right) \\
&\qquad + \Sigma_{jj}^{1/2}\left(\Phi^{-1}(\zeta^* + \delta) - \Phi^{-1}(\zeta^* - \delta)\right) \\
&\quad = \left(\hat{Q}^V_{j,\zeta^*+\delta} - \Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^* + \delta)\right) + \left(\Sigma_{jj}^{1/2}\Phi^{-1}(\zeta^* - \delta) - \hat{Q}^V_{j,\zeta^*-\delta}\right) + \frac{\varepsilon}{2},
\end{aligned}$$

when sufficiently far along the subsequence, almost surely (the inequality above uses monotonicity of quantiles in ζ). By the argument in part (ii), the far right-hand side of the above display is less than ε when sufficiently far along the subsequence, almost surely. Since $\varepsilon > 0$ was arbitrary, we have shown (6). □