Supplementary Appendix For: "Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs"

José Luis Montiel Olea — Columbia University
Mikkel Plagborg-Møller — Princeton University

June 28, 2018

A Appendix

A.1 Asymptotic analysis of one-parameter class

Here we state our formal assumptions for asymptotic analysis and provide detailed analytical comparisons of these bands.

A.1.1 Regularity conditions for asymptotic analysis

For the asymptotic results in this paper, we impose the following regularity conditions on the general model in Section 2.

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. 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. However, we do not require Σ to have full rank, so that we cover cases with $k > p$ as well as the degenerate VAR applications mentioned in Lütkepohl et al. (2015b, p. 807).

A.1.2 Coverage probability

Next, we derive the coverage probability of any band in the one-parameter class as a function of the critical value c. Analogous results are common in the theory of multiple testing (Lehmann & Romano, 2005, chapter 9).

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.

Proof. See Appendix A.7.1. □

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.

A.1.3 μ-projection band

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

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* $o_p(n^{-1/2})$; *that is,*

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

Proof. See Appendix A.7.8. □

A.1.4 Detailed comparisons of popular bands

We here provide a detailed analytical comparison of popular bands in the one-parameter class. For any two bands in the one-parameter class, the ratio of their critical values yields

[1] 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.

the 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 $1 - \alpha$.[2] 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, respectively; cf. Lemma 2 in Appendix A.1.5. 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—regardless of the dimensions k and p—than both the Šidák critical value and the μ-projection critical value, cf. Lemma 3 in Appendix A.1.5 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 2.

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. 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

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

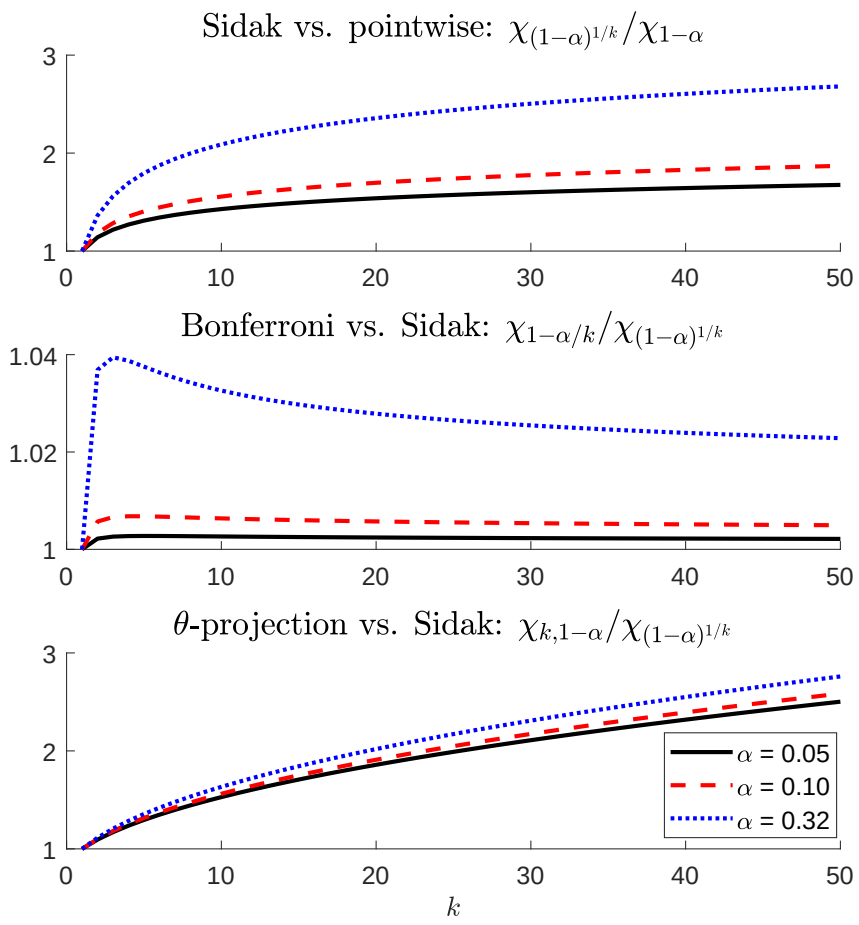

Figure 4: Relative critical values of the pointwise, Šidák, Bonferroni, and θ-projection bands. The dimension $k = \dim(\theta)$ is along the horizontal axis. The three colored curves correspond to the significance levels $\alpha = 0.05$ (black), $\alpha = 0.1$ (red), and $\alpha = 0.32$ (blue).

additional information is needed to compare these different bands.[3]

Figure 4 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 $p > k$.

[3]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.

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, Lemma 4 in Appendix A.1.5 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.[4]

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.1.5 for analytical results supporting the graphical evidence in Figure 4.

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, Figure 5 in Appendix A.1.5 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.

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

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

if $p \geq k$.

A.1.5 Analytical results on one-parameter critical values

Finally, we provide supplementary analytical and graphical results comparing the critical values listed in Table 1. 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. Equation (6). 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 2. *Let $\mathcal{S}_k$ denote the set of $k \times k$ symmetric positive semidefinite matrices. Define*

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

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 $k \leq p$.

Proof. See Appendix A.7.2. □

Lemma 2 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 3. *Using the same notation as Lemma 2, 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*

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

Proof. See Appendix A.7.3. □

The next lemma provides analytical results to complement the visual observations in Figure 4 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 4.

(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* $k \geq 1$,

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

Proof. See Appendix A.7.4. □

Figure 5 compares the Šidák and μ-projection critical values. In Appendix A.1.4 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

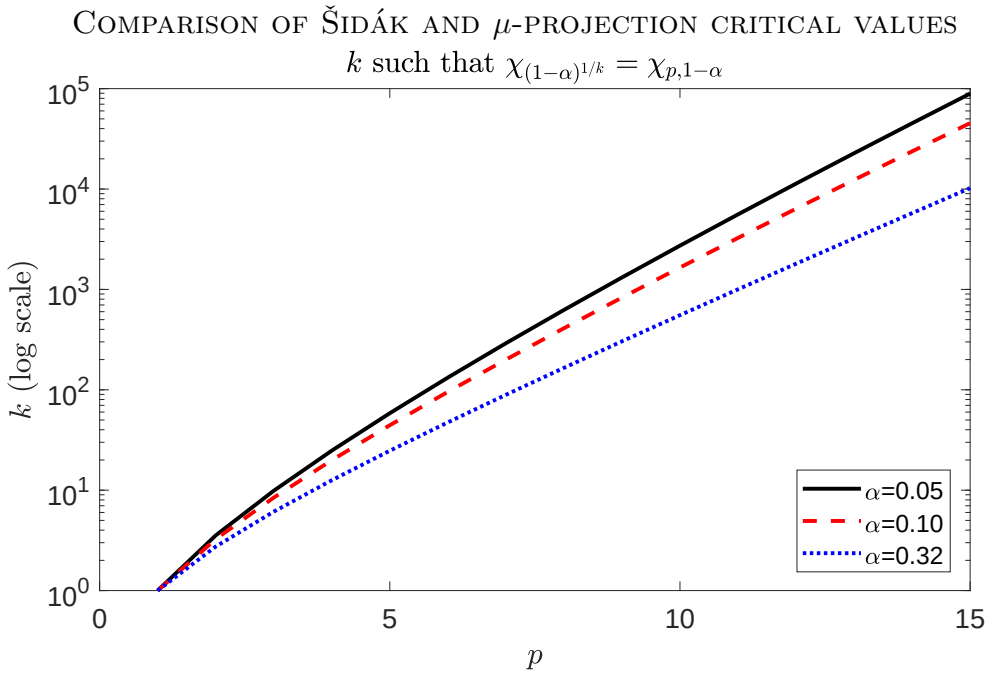

Figure 5: Smallest value of k needed for the Šidák band to be asymptotically weakly wider than the μ-projection band, as a function of p. Horizontal axis: p, vertical axis: k in log scale. The curves correspond to significance levels $\alpha = 0.05$ (black), $\alpha = 0.1$ (red), and $\alpha = 0.32$ (blue).

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.

Finally, we state the simple result that the sup-t critical value is continuous in the variance-covariance matrix Σ. This result implies the validity of the plug-in sup-t band, cf. Section 2.

Lemma 5. *For any $\zeta \in (0,1)$, the function $\tilde{\Sigma} \mapsto q_\zeta(\tilde{\Sigma})$ defined in equation (6) is continuous on the set $\mathcal{S}_{p,k}$ defined in Lemma 2 in Appendix A.1.5.*

Proof. See Appendix A.7.5. □

A.2 Implementing the 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 in Section 2.

Alternative bootstrap procedure. Algorithm 3 defines a well-known alternative critical-value-based bootstrap band, often used in the nonparametric econometrics literature. 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.

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 simulta-

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 $\hat{m}^{(1)}, \ldots, \hat{m}^{(N)}$
$\widehat{C} = \times_{j=1}^{k} [\hat{\theta}_j - \hat{\sigma}_j^* \hat{q}_{1-\alpha}, \hat{\theta}_j + \hat{\sigma}_j^* \hat{q}_{1-\alpha}]$

neous 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. According to Lemma 5 in Appendix A.1.5, the sup-t critical value (6) is a continuous function of the (possibly singular) variance-covariance matrix Σ. This implies the validity of the plug-in implementation of the sup-t band.

Next, we state a result guaranteeing that the bootstrap and Bayesian implementations of the sup-t band in Section 2 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 $j = 1, \ldots, k$*,*

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

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

Proof. See Appendix A.7.9. □

A.3 Decision theoretic details

This section gives technical details for the decision theoretic analysis in Section 4.

GAUSSIAN DECISION PROBLEM. The argument for invariance of the decision problem in Section 4 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 (7) 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 $\mu \in \mathbb{R}^p$.

CHARACTERIZATION OF EQUIVARIANT BANDS. Here we formally state the characterization of translation equivariant bands used in Section 4.

Lemma 6. $\mathcal{C}_{eq} = \{C \colon \mathbb{R}^p \to \mathcal{R} \mid C(x) = Gx + R,\ R \in \mathcal{R}\}$.

Proof. See Appendix A.7.6. □

A.4 Confidence bands for impulse response functions

In this section we review the literature on confidence bands for impulse response functions and give additional details of the VAR application.

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.[5] 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 (2) (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.2. 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.[6] Lütkepohl et al. (2016) construct highest-density rectangular regions from bootstrap draws of the impulse responses, which is asymptotically equivalent to θ-projection under Assumption 1 and $\operatorname{rank}(\Sigma) = k$.

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

[6] 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.

Bruder & Wolf (2018) construct a "balanced bootstrap band" using pre-pivoting; their band is asymptotically equivalent with the sup-t band defined in this paper under our assumptions.

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 $\theta = h(\mu)$.[7] 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.[8]

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).

[7] 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).

[8] In cointegrated models as well as certain stationary models, the asymptotic variance Ω of $\hat{\mu}$ may be singular. Our theory and methods allow for singularities.

ADDITIONAL EMPIRICAL RESULTS. Figures 6 and 7 compare all common bands in the one-parameter class for the recursive and external instrument specifications, including θ-projection and μ-projection.[9] 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 $p > k$.

[9] 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.

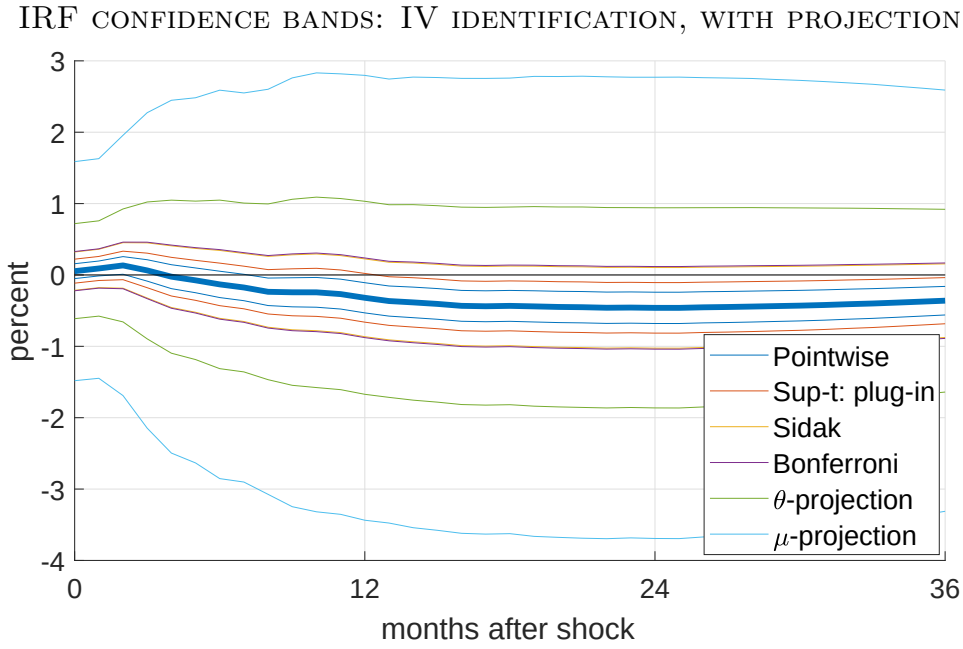

Figure 6: 68% confidence bands for IRF of IP to 1-stdev contractionary monetary policy shock, external instrument identification. See caption for Figure 2.

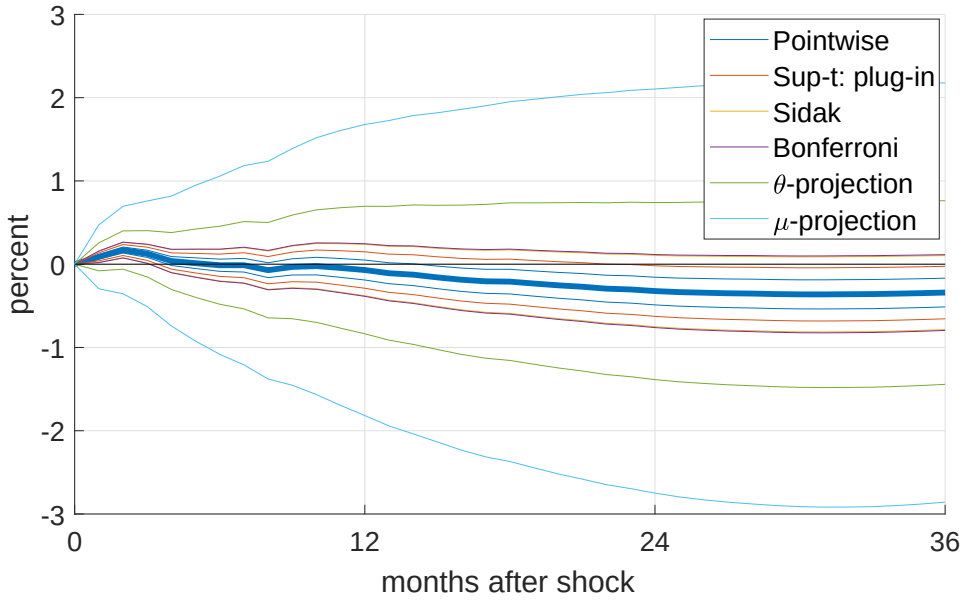

Figure 7: 68% confidence bands for IRF of IP to 1-stdev contractionary monetary policy shock, recursive identification. See caption for Figure 2.

A.5 VAR simulation study

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

DESIGN. 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 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 $R^2 = \operatorname{Var}(\varepsilon_{1,t}) / \operatorname{Var}(z_t)$. 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.

Results. 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 3. The plug-in, bootstrap, and Bayes sup-t bands all perform similarly well for moderately persistent VAR processes ($\varphi \in \{0, 0.5\}$). For highly persistent VARs ($\varphi \in \{0.9, 1\}$), 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.

A.6 Application: 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.[10]

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.[11] We use the sample for the main linear regression specification in Head et al.

[10] Several authors have considered the related issue of adjusting tests to account for "data snooping", e.g., Andrews (1993), Inoue & Kilian (2005), Hansen & Timmermann (2012), and Armstrong & Kolesár (2016).

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

(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 after independence. We consider five different sets of control variables, as described below.[12] All specifications control for time fixed effects. Standard errors, bootstraps replications, and Bayesian inference are clustered by dyad.[13]

Results. Figure 8 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 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 8 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

[12] 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).

[13] 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 variance-covariance matrix for the plug-in band is computed by stacking the scores of the individual regression specifications and imposing independence across clusters.

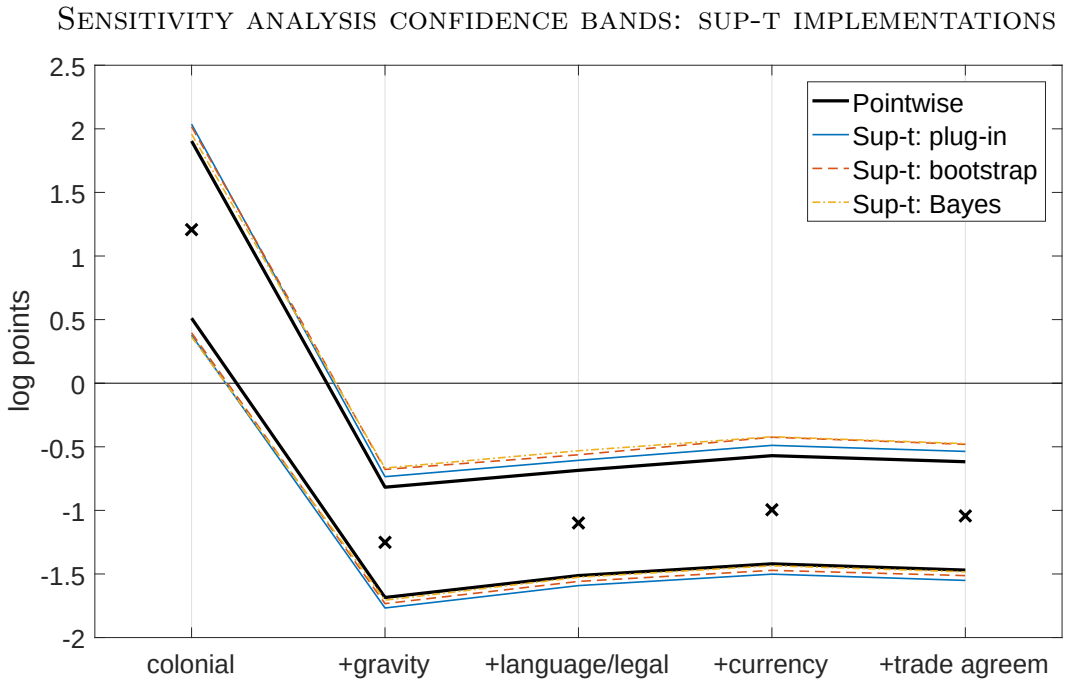

Figure 8: 90% confidence bands for the linear regression coefficient measuring the 40-year effect of independence on log bilateral trade, estimated across five different sets of controls. Crosses: point estimates. Lines: confidence bands. Specifications: "colonial" – colonial history dummy, ongoing colonial 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.

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.

A.7 Proofs

Let $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the Euclidean and maximum norms, respectively.

A.7.1 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 $\mathrm{diag}(x)$ denote the $k \times k$ diagonal matrix with the elements of x in order along the diagonal. Then, for any $c > 0$,

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

The proposition now follows from the limiting distribution (4) 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 $\mathcal{A}$. □

A.7.2 Lemma 2

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 a consequence of the Šidák (1967) inequality. Finally, if $k \leq p$, then $I_k \in \mathcal{S}_{p,k}$, so $\sup_{\tilde{\Sigma} \in \mathcal{S}_{p,k}} q_\zeta(\tilde{\Sigma}) \geq q_\zeta(I_k) = \chi_{\zeta^{1/k}}$.

A.7.3 Lemma 3

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 Šidák (1967)'s inequality 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

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

A.7.4 Lemma 4

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

A.7.5 Lemma 5

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 $\max_j |(\tilde{\Sigma}^{(\ell)}_{jj})^{-1/2}\tilde{V}^{(\ell)}_{jj}|$ 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), $0 \leq a < b$. 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 $P(\|GZ\|_\infty \in (a,b)) > P(Z \in S) > 0$. □

A.7.6 Lemma 6

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 $R = C(\mathbf{0}_p) \in \mathcal{R}$. □

A.7.7 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 $b_j - a_j \geq 2\|g_j\|\chi_{1-\alpha}$ for $j = 1, \ldots, k$.*

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

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

$$= P(\,|N_1(0,1)|\,\leq (b_j - a_j)/(2\|g_j\|)).$$

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.[14] □

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:

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

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

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):

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

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 $R \neq R_{\text{sup}}$. □

A.7.8 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 $\tilde{\mu} \in \widehat{W}_\mu$,

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

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 $h(\cdot)$. □

A.7.9 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, $\rho(\hat{P}_M, P_M) \xrightarrow{p} 0$, $\hat{\theta} \xrightarrow{p} \theta$, 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 $\rho(\hat{P}_M, P_M) \xrightarrow{a.s.} 0$, $\hat{\theta} \xrightarrow{a.s.} \theta$, 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 $\sqrt{n}(h_j(\hat{\mu}^*) - h_j(\hat{\mu}))$ converges weakly to the distribution of $V \sim N_k(\mathbf{0}_k, \Sigma)$, conditionally almost surely (cf. the proof of van der Vaart, 1998, Thm. 23.5). The continuous mapping theorem then implies 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 again $V \sim N_k(\mathbf{0}_k, \Sigma)$, almost surely. Almost sure convergence of the $1-\alpha$ quantile follows as in the proof of Lemma 5.

(II): By Assumption 1, $\sqrt{n}\hat{\sigma}_j \xrightarrow{p} \Sigma_{jj}^{1/2}$. As above, given a subsequence of $\{n\}$, extract a further subsequence along which $\rho(\hat{P}_M, P_M) \xrightarrow{a.s.} 0$, $\hat{\theta} \xrightarrow{a.s.} \theta$, 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. 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 Equation (6). 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

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

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 (1). □

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.

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

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

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. *Annals of Statistics*, *41*(2), 802–837.

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

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

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

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.

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. & 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.

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

Inoue, A. & Kilian, L. (2005). In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? *Econometric Reviews*, *23*(4), 371–402.

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.

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

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.

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.

Piegorsch, W. W. (1984). *Admissible and Optimal Confidence Bands in Linear Regression*. PhD thesis, Cornell University.

Rubin, D. B. (1981). The Bayesian Bootstrap. *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.

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