Short and Simple Confidence Intervals when the Directions of Some Effects are Known

Philipp Ketz* Adam McCloskey†

June 10, 2021

Abstract

We provide adaptive confidence intervals on a parameter of interest in the presence of nuisance parameters when some of the nuisance parameters have known signs. The confidence intervals are adaptive in the sense that they tend to be short at and near the points where the nuisance parameters are equal to zero. We focus our results primarily on the practical problem of inference on a coefficient of interest in the linear regression model when it is unclear whether or not it is necessary to include a subset of control variables whose partial effects on the dependent variable have known directions (signs). Our confidence intervals are trivial to compute and can provide significant length reductions relative to standard confidence intervals in cases for which the control variables do not have large effects. At the same time, they entail minimal length increases at any parameter values. We prove that our confidence intervals are asymptotically valid uniformly over the parameter space and illustrate their length properties in an empirical application to a factorial design field experiment and a Monte Carlo study calibrated to the empirical application.

KEYWORDS: CONFIDENCE INTERVALS, ADAPTIVE INFERENCE, UNIFORM INFERENCE, SIGN RESTRICTIONS, BOUNDARY PROBLEMS

*Paris School of Economics, philipp.ketz@psemail.eu

†Department of Economics, University of Colorado, adam.mccloskey@colorado.edu

1 Introduction

Consider the common empirical setting for which a researcher is interested in estimating the causal effect of one variable on another via a linear regression in the presence of one or more observed potential control variables. The researcher believes that a regression including this full set of controls should not suffer omitted variables bias but is uncertain whether it is necessary to include them all to overcome this bias. To obtain more informative inference, the researcher would prefer not to include controls unnecessarily and knows that if some of these controls indeed influence the outcome variable, it must be in a known positive or negative direction. Indeed, typical heuristic explanations for the potential inclusion of a control variable to mitigate omitted variables bias involve a known "direction" for the effect of the omitted variable on the outcome of interest. In this paper, we develop confidence intervals (CIs) with desirable properties for these types of settings.

More specifically, we develop CIs for a parameter of interest in the presence of nuisance parameters with a known sign. Our CIs are designed to have uniformly correct coverage and desirable length properties across the entire parameter space while becoming particularly short when these nuisance parameters are small or zero. In the regression context, this latter property is motivated by the fact that the researcher believes the regression coefficients on a subset of control variables with known partial effect directions are likely to be small or zero. In general, our CIs can be used for inference on a parameter in any well-behaved finite-dimensional model with a large-sample normally distributed estimator when some nuisance parameters are restricted above or below by zero (possibly after a location shift). These include regression models estimated by ordinary, generalized and two stage least squares as well as models with bounded parameter spaces such as (G)ARCH and random coefficient models. While noting this generality, we mainly focus on regression models with control coefficients of known sign since it is natural to desire CIs that shorten when these coefficients are small or zero.

To construct our CIs, we use the the fact that knowledge of the signs of control variable coefficients, in addition to a standard consistent estimator of the covariance matrix of the underlying coefficient estimates, can be used to determine the sign of the corresponding omitted variables biases incurred by omitting the corresponding control variables. In turn, standard one-sided CIs for the coefficient of interest based on regressions that omit some of these control variables maintain correct coverage. These latter CIs are optimal when the corresponding control coefficients are equal to zero and have low expected excess length when they are close to zero but their expected excess length grows without bound as the control coefficients grow larger. On the other hand, standard one-sided CIs based upon the regression including all controls have constant expected excess length and correct coverage. We propose adaptive one-sided CIs that utilize the strengths of both of these types of CIs by intersecting them. We make use of the same logic for constructing two-sided CIs essentially by intersecting our lower- and upper- one-sided CIs.

In particular, we propose a computationally trivial method to find the subset of controls that is able to produce the largest expected length reductions when using this intersection principle. In addition, the restricted parameter space implies that the coverage of these intersected CIs is lowest at its boundary. This feature allows us to provide the user a simple means to compute the smallest CI endpoints that yield correct coverage uniformly across the parameter space via response surface regression output, rather than using a conservative Bonferroni correction. Using our reported response surface regression coefficients, the user can immediately compute these CI endpoints as a function of one or two empirical correlation parameters, depending upon whether they are forming a one- or two-sided CI.

We show that our proposed CIs are uniformly asymptotically valid and characterize their length properties. The latter depend upon the correlation structure of the underlying data and the true values of the unknown control coefficients. For extreme values of correlation between the estimators of the coefficient of interest and sign-restricted controls, the expected (excess) length of our CIs can be close to 100% smaller than standard CIs based upon the regression including all controls. For correlation values more likely to be encountered in practice, these expected (excess) length reductions can still exceed 30% for commonly used confidence levels. On the other hand, for a confidence level of 95%, for example, our proposed one-sided (two-sided) CIs cannot be more than 3% (2.28%) longer than the corresponding standard CIs for *any* realization of the data.

1.1 Relationship with the Literature

Several results in the statistics and econometrics literatures provide bounds on the ability for CIs to simultaneously maintain uniformly correct coverage over a class of data-generating processes (DGPs) while adapting to a given subclass. Here we develop CIs with this very goal in mind: our CIs maintain correct coverage for the parameter of interest uniformly across the parameter space for the nuisance parameters while becoming shorter when these nuisance parameters are equal to zero. Although most of this literature is devoted to nonparametric methods (e.g., Low, 1997; Cai and Low, 2004), the recent work of

Armstrong and Kolesár (2018) has produced similar implications for parametric models like those in the asymptotic versions of the problems we study. Indeed, Armstrong et al. (2020) provide bounds on the ability to shorten CIs while maintaining correct coverage for regression coefficients at points for which potential control coefficients are zero. However, all of the aforementioned results rely upon an assumption of symmetry about zero for the underlying parameter space (among others). Because we are interested in problems with sign-restricted nuisance parameters, the underlying parameter space is asymmetric and these results do not apply, allowing for us to achieve the goal of constructing CIs that become significantly shorter at the parameter values we are interested in.

This is certainly not the first paper attempting to produce CIs that adapt to subclasses of DGPs while retaining uniform size control. Several authors have provided such adaptive CIs for various smoothness classes and shape constraints in the nonparametric literature. See, e.g., Cai and Low (2004), Cai et al. (2013), Armstrong (2015) and Kwon and Kwon (2020). Like the ones we propose, the CIs of Cai and Low (2004) and Kwon and Kwon (2020) are obtained by intersecting CIs that are shorter under different subclasses of DGPs. Within this literature, Kwon and Kwon (2020) is probably the closest to our analysis as they focus on nonparametric regression models with a coordinate-wise monotone regression function. Since our focus is on finite-dimensional models, we are not concerned with the rate of convergence adaptation that this literature is focused upon but rather finite-sample length adaptation for CIs.[1] Moreover, we prove our CIs are uniformly asymptotically valid without assuming Gaussian disturbances or fixed regressors as does much of this nonparametric literature.

Finally, our work is related to the literature on uniform inference when nuisance parameters may be at or near a boundary, e.g., Andrews and Guggenberger (2009), McCloskey (2017) and Ketz (2018). While CIs with uniform asymptotic validity could in principle be computed by inverting the tests in this literature, this is often computationally prohibitive, especially when the nuisance parameter exceeds one or two dimensions. Similarly, inverting weighted average power maximizing tests such as those of Moreira and Moreira (2013) or Elliott et al. (2015) is computationally intractable for most realistic applications. In contrast, our CIs are direct and trivial to compute since they do not rely on test inversion. Moreover, our CIs are designed to have length properties that are desirable from a practical perspective without requiring the user to specify weights or tuning parameters to optimize over.

[1] Nevertheless, we note that Kwon and Kwon (2020) provide a means of shortening adaptive nonparametric CIs relative to simple Bonferroni corrections in a similar spirit to our CI enpoints computed from response surface regression output.

1.2 Outline of Paper

The remainder of this paper is organized as follows. Section 2 imparts the basic intuition of our CI constructions in a stylized asymptotic version of the inference problem we consider before providing computationally trivial algorithms for constructing the one- and two-sided CIs we propose in the general asymptotic setting. Section 3 then shows how our CIs are contructed in practical finite-sample applications and provides theoretical results establishing their uniform asymptotic validity across a wide variety of applications. In Section 4, we illustrate the usefulness of our CIs in an empirical application of inference on treatment effects in a factorial design field experiment while Section 5 examines their finite-sample properties in a simulation study calibrated to the empirical application. Appendix A provides the mathematical proofs of our theoretical results and Appendix B specifies a parameter space for the standard linear regression model that satisfies the requirements for some of our theoretical results.

Throughout this paper, we use the following notational conventions. For any two column vectors a and b, we sometimes write (a,b) instead of $(a',b')'$ and let $a \geq b$ denote the element-by-element inequality. $\mathbb{R}_+$ denotes the positive orthant of the real line, $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{\infty\}$, $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ and z_ξ denotes the ξ^{th} quantile of the standard normal distribution. For a square matrix A, $\text{Diag}(A)$ denotes the diagonal matrix with the same diagonal entries as A and $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its smallest and largest eigenvalues.

2 Normal Means Large Sample Problem

LeCam's Limits of Experiments Theory provides that inference on the parameter of a well-behaved model is equivalent to inference on the mean of a Gaussian random vector with known variance matrix in large samples. This powerful result incorporates regression models, instrumental variables models, maximum likelihood models and models estimated by the generalized method of moments.[2] See Chapter 9 of van der Vaart (1998) and Chapter 13 of Lehmann and Romano (2005) for textbook treatments of this theory.[3] Since the variance matrix is known in this setting, each element of the Gaussian random vector can

[2]This result holds under assumptions ensuring the model is well-behaved. For example, in the context of instrumental variables models or models estimated by the generalized method of moments, this does not allow for weak instruments or other forms of weak identification.

[3]Ketz (2018) also provides a means of obtaining this limit experiment even when the true parameter may be at or near the boundary of its parameter space via a quadratic expansion of the underlying objective function used for estimation. Indeed, models for which parameters are necessarily bounded, such as coefficients in (G)ARCH models or variance parameters in models for consumer demand, provide natural applications for the inference approach we take in this paper.

be scale-normalized so that the large sample inference problem reduces to inference on the mean vector h from a single observation $Y \overset{d}{\sim} \mathcal{N}(h,\Omega)$, where Ω is a known correlation matrix.

It is often the case in econometric applications that the researcher is interested in constructing a CI for a scalar parameter of interest in the presence of nuisance parameters. In addition, the researcher often has knowledge about the sign of the nuisance parameters. For example, when performing inference on a single coefficient in the linear regression model when "control" variables may be included in the regression to mitigate potential omitted variable bias, the researcher often knows the direction of the partial effects of some of the controls from economic theory or logical reasoning. In the large sample problem, this corresponds to conducting inference on a scalar β from a single observation

$$\begin{pmatrix} Y_\beta \\ Y_\delta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, \Omega\right), \tag{1}$$

where Ω is a known positive-definite correlation matrix and δ is a nuisance parameter whose elements are known to be greater than or equal to zero.[4]

In many contexts, it is natural for the researcher to desire a CI with the following properties: (i) correct coverage $1-\alpha$ (coverage of at least $1-\alpha$) across the entire $\delta \geq 0$ parameter space, (ii) good length properties across the entire $\delta \geq 0$ parameter space and (iii) shortness when δ is equal or close to zero. For example, if it is not obvious whether a regressor should enter as a control variable or not, it is sensible to desire an especially short CI when the unknown population regression coefficient is equal to or near zero (reflecting the researcher's uncertainty about whether it is an important variable) while maintaining correct coverage and decent length no matter the coefficient's magnitude. In this section, we provide CI constructions for the large sample problem with this very goal in mind. We begin by describing the intuition for the CIs in the simplest version of the problem and subsequently provide general formulations for both one– and two–sided CIs.

[4] The restriction $\delta \geq 0$ is without loss of generality because parameters without sign restrictions may be dropped from the analysis in the limiting problem and limiting Gaussian random variables corresponding to parameters restricted to be greater/less than or equal to a known number may be linearly transformed to conform to (1).

2.1 Basic Intution

To communicate the basic intuition for our CIs, we specialize the large sample problem (1) to the case for which δ is one-dimensional and the correlation between Y_β and Y_δ is positive:

$$\begin{pmatrix} Y_\beta \\ Y_\delta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\rho > 0$, β is unrestricted, and $\delta \geq 0$. Consider the formation of an upper one-sided CI for β with the goal of satisfying properties (i)–(iii) above. To illustrate the tension between properites (ii) and (iii), note that the standard CI that ignores the information in Y_δ, i.e.,

$$CI_u(Y_\beta) = [Y_\beta - z_{1-\alpha}, \infty),$$

satisfies (ii) but not (iii) since its expected excess length is always simply equal to $z_{1-\alpha}$.[5] On the other hand, the CI that is optimal when δ is known to equal zero, i.e.,

$$\widetilde{CI}_u(Y_\beta, \rho Y_\delta) = \left[Y_\beta - \rho Y_\delta - \sqrt{1-\rho^2} z_{1-\alpha}, \infty\right),$$

satisfies (iii) but not (ii) since its expected excess length is equal to $\rho\delta + \sqrt{1-\rho^2} z_{1-\alpha}$, which diverges as $\delta \to \infty$.[6]

In order to attain property (iii) but not at the expense of property (ii), we propose CIs with length performance designed to adapt to the data. Consider intersecting the two CIs $CI_u(Y_\beta)$ and $\widetilde{CI}_u(Y_\beta, \rho Y_\delta)$ to simultaneously retain property (ii) of the former and property (iii) of the latter:

$$\begin{aligned} \widehat{CI}_u\left(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, \sqrt{1-\rho^2} z_{1-\gamma}\right) &= [Y_\beta - z_{1-\alpha+\gamma}, \infty) \cap \left[Y_\beta - \rho Y_\delta - \sqrt{1-\rho^2} z_{1-\gamma}, \infty\right) \\ &= \left[Y_\beta - \min\left\{z_{1-\alpha+\gamma}, \rho Y_\delta + \sqrt{1-\rho^2} z_{1-\gamma}\right\}, \infty\right) \end{aligned}$$

for some $\gamma \in (0, \alpha)$. Note that $\widehat{CI}_u\left(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, \sqrt{1-\rho^2} z_{1-\gamma}\right)$ maintains correct coverage probability over the parameter space:

$$P\left(\beta \in \widehat{CI}_u\left(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, \sqrt{1-\rho^2} z_{1-\gamma}\right)\right)$$

[5] Expected excess length of an upper one-sided CI for β is defined as $E[\beta - \text{lb}]$, where lb denotes the lower bound of the CI.

[6] Both CIs satisfy (i) in this context since we have assumed $\rho > 0$, see equation (2).

$$
\begin{aligned}
&= P\left(\beta \geq Y_\beta - \min\left\{z_{1-\alpha+\gamma}, \rho Y_\delta + \sqrt{1-\rho^2} z_{1-\gamma}\right\}\right) \\
&= 1 - P\left(\beta < Y_\beta - \min\left\{z_{1-\alpha+\gamma}, \rho Y_\delta + \sqrt{1-\rho^2} z_{1-\gamma}\right\}\right) \\
&\geq 1 - P(\beta < Y_\beta - z_{1-\alpha+\gamma}) - P\left(\beta < Y_\beta - \rho Y_\delta - \sqrt{1-\rho^2} z_{1-\gamma}\right) \\
&\geq 1 - (\alpha - \gamma) - \gamma = 1 - \alpha \qquad (2)
\end{aligned}
$$

for all $(\beta,\delta) \in \mathbb{R} \times \mathbb{R}_+$, where the first inequality follows from the Bonferroni inequality and the second inequality uses the fact that

$$
\begin{aligned}
P\left(\beta < Y_\beta - \rho Y_\delta - \sqrt{1-\rho^2} z_{1-\gamma}\right) &= P\left(\beta < \beta - \rho\delta + \tilde{Z}_\rho - \sqrt{1-\rho^2} z_{1-\gamma}\right) \\
&= P\left(\tilde{Z}_\rho < -\rho\delta - \sqrt{1-\rho^2} z_{1-\gamma}\right) \leq P\left(\tilde{Z}_\rho < -\sqrt{1-\rho^2} z_{1-\gamma}\right) = \gamma
\end{aligned}
$$

with

$$
\tilde{Z}_\rho = Y_\beta - \rho Y_\delta - (\beta - \rho\delta) \overset{d}{\sim} \mathcal{N}(0, 1-\rho^2).
$$

Since $\widehat{CI}_u\left(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, \sqrt{1-\rho^2} z_{1-\gamma}\right)$ makes use of a multiplicity correction based upon the Bonferroni bound, for similar reasons used to motivate the adjusted Bonferroni critical values of McCloskey (2017), it is possible to decrease the excess length of $\widehat{CI}_u\left(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, \sqrt{1-\rho^2} z_{1-\gamma}\right)$ while retaining uniform control of coverage probability. In particular, fix $\gamma \in (0,\alpha)$ and find the constant $c^* \in [0, \sqrt{1-\rho^2} z_{1-\gamma}]$ that solves

$$
P(Z_1 > \min\{z_{1-\alpha+\gamma}, \rho Z_2 + c\}) = \alpha \qquad (3)
$$

in c, where

$$
\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).
$$

The CI $\widehat{CI}_u(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, c^*)$ is contained in $\widehat{CI}_u\left(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, \sqrt{1-\rho^2} z_{1-\gamma}\right)$ and maintains correct coverage probability over the parameter space:

$$
\begin{aligned}
P\left(\beta \in \widehat{CI}_u(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, c^*)\right) &= P(\beta \geq Y_\beta - \min\{z_{1-\alpha+\gamma}, \rho Y_\delta + c^*\}) \\
&= P(Z_1 \leq \min\{z_{1-\alpha+\gamma}, \rho\delta + \rho Z_2 + c^*\}) \\
&\geq P(Z_1 \leq \min\{z_{1-\alpha+\gamma}, \rho Z_2 + c^*\}) \\
&= 1 - P(Z_1 > \min\{z_{1-\alpha+\gamma}, \rho Z_2 + c^*\}) = 1 - \alpha
\end{aligned}
$$

for all $(\beta,\delta) \in \mathbb{R} \times \mathbb{R}_+$, where the second equality follows from the fact that $(Y_\beta, Y_\delta) \overset{d}{\sim} (\beta,\delta) + (Z_1, Z_2)$ and the inequality uses the fact that $\rho\delta \geq 0$. The problem (3) is computationally straightforward and can, for example, be solved by means of Monte Carlo simulations. A heuristic approach to choosing the tuning parameter γ makes use of similar reasoning to that used to compute the adjusted Bonferroni critical values in McCloskey (2017): a "small" γ such as $\gamma = \alpha/10$ yields only slightly higher expected excess length when δ is "large" but significantly lower expected excess length when δ is "small".

Finally, it is interesting to note that $\widehat{CI}_u(Y_\beta, \rho Y_\delta; z_{1-\alpha+\gamma}, c^*)$ can be viewed as a CI that results from a model selection procedure *designed for inference*. In the context of the regression model example, we can view the model selection procedure as follows:

1. If $Y_\delta > (z_{1-\alpha+\gamma} - c^*)/\rho$, construct the CI for β from the "full" regression using the critical value $z_{1-\alpha+\gamma}$.

2. If $Y_\delta \leq (z_{1-\alpha+\gamma} - c^*)/\rho$, construct the CI for β from the "short" regression using the critical value c^*.

The model selection pretest rule $Y_\delta > (z_{1-\alpha+\gamma} - c^*)/\rho$ is analogous to using a t-test as a pretest but with a nonstandard critical value that incorporates both the two-step nature of the inference procedure as well as the dependence between Y_β and Y_δ. Note that as $\rho \to 1$, this nonstandard pretest approaches a standard t-test pretest. Unlike standard model selection procedures, this procedure is designed for inference in the sense that (i) it uniformly controls coverage probability by directly incorporating the model selection uncertainty in its construction and (ii) it is designed to yield low excess length rather than a different notion of risk (such as mean-squared error). Though some recent post-selection inference procedures (e.g., Belloni et al., 2014; McCloskey, 2017) uniformly control coverage probability/size, the selection procedures used in their construction are not designed to yield CIs with desirable length properties.

2.2 One-Sided Confidence Intervals

In this section we focus on forming analogous adaptive one-sided CIs, but now allowing $\delta \geq 0$ to be multidimensional so that the large sample problem corresponds to (1), where

$$\Omega = \begin{pmatrix} 1 & \Omega_{\beta\delta} \\ \Omega_{\delta\beta} & \Omega_{\delta\delta} \end{pmatrix}. \tag{4}$$

Without loss of generality, we focus on upper one-sided CIs for β since lower one-sided CIs may be attained analogously upon multiplying Y_β by negative one. The optimal $(1-\alpha)$-level upper one-sided CI for β when $\delta=0$ is equal to

$$\left[Y_\beta-\Omega_{\beta\delta}\Omega_{\delta\delta}^{-1}Y_\delta-z_{1-\alpha}\sqrt{1-\Omega_{\beta\delta}\Omega_{\delta\delta}^{-1}\Omega_{\delta\beta}},\infty\right).$$

The CI that intersects this CI with the standard CI for β that ignores the information in Y_δ will not maintain coverage in general. More specifically, the argument in (2) for showing correct coverage only generalizes when all of the elements of $\Omega_{\beta\delta}\Omega_{\delta\delta}^{-1}$ are non-negative. In the case that this condition does not hold, we can still find adaptive CIs with potential length improvements by "dropping" elements of Y_δ from consideration. The following algorithm is designed to do just that while maintaining particularly low excess length when δ is equal or close to zero.

For $\gamma\in(0,\alpha)$, consider the function $c:[0,1)\rightarrow[0,z_{1-\gamma}]$ such that

$$P(Z_1>\min\{z_{1-\alpha+\gamma},\tilde{Z}_2+c(\omega)\})=\alpha, \tag{5}$$

where

$$\begin{pmatrix} Z_1 \\ \tilde{Z}_2 \end{pmatrix}\sim\mathcal{N}\left(0,\begin{pmatrix} 1 & \omega \\ \omega & \omega \end{pmatrix}\right).$$

The following result ensures that $c:[0,1)\rightarrow[0,z_{1-\gamma}]$ is well-defined and continuous.

Proposition 1
If $\alpha<1/2$, $c:[0,1)\rightarrow[0,z_{1-\gamma}]$ as defined in (5) *exists and is continuous.*

Note that $c(0)=z_{1-\alpha}$. Let $Y_\delta^{(s)}$ denote an arbitray subvector of Y_δ, including the empty one, with

$$\begin{pmatrix} Y_\beta \\ Y_\delta^{(s)} \end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix} \beta \\ \delta^{(s)} \end{pmatrix},\begin{pmatrix} 1 & \Omega_{\beta\delta^{(s)}} \\ \Omega_{\delta^{(s)}\beta} & \Omega_{\delta^{(s)}\delta^{(s)}} \end{pmatrix}\right),$$

where by convention $\delta^{(s)}$, $\Omega_{\beta\delta^{(s)}}$ and $\Omega_{\delta^{(s)}\delta^{(s)}}$ (as well as $\Omega_{\beta\delta^{(s)}}\Omega_{\delta^{(s)}\delta^{(s)}}^{-1}$ and $\Omega_{\beta\delta^{(s)}}\Omega_{\delta^{(s)}\delta^{(s)}}^{-1}\Omega_{\delta^{(s)}\beta}$) are set equal to zero when $Y_\delta^{(s)}=\emptyset$.

Algorithm One-Sided
Amongst all subvectors of Y_δ (including the empty one) such that the elements of $\Omega_{\beta\delta^{(s)}}\Omega_{\delta^{(s)}\delta^{(s)}}^{-1}$ are non-negative, find the subvector $Y_\delta^{(s^*)}$ such that the expected excess length of

$$\widehat{CI}_u\left(Y_\beta,\Omega_{\beta\delta^{(s)}}\Omega_{\delta^{(s)}\delta^{(s)}}^{-1}Y_\delta^{(s)};z_{1-\alpha+\gamma},c\left(\Omega_{\beta\delta^{(s)}}\Omega_{\delta^{(s)}\delta^{(s)}}^{-1}\Omega_{\delta^{(s)}\beta}\right)\right)$$

at $\delta=0$ is minimized at $s=s^*$. Then, construct

$$\widehat{CI}_u(Y_\beta,\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}Y^{(s^*)}_\delta;z_{1-\alpha+\gamma},c(\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\Omega_{\delta(s^*)\beta})). \quad \square$$

The goal of this algorithm is to generate short CIs when the user is agnostic about which elements of δ are more likely to be (close to) zero. Figure 1 shows the expected excess length of $\widehat{CI}_u(Z_1,\tilde{Z}_2,z_{1-\alpha+\gamma},c(\omega))$ as a function of ω, for $\alpha\in\{0.1,0.05,0.01\}$ and our recommended value of $\gamma=\alpha/10$.[7] This expected excess length is strictly decreasing in ω (at least for the considered choices of α and γ). Since it does not depend upon β, this implies that the expected excess length of $\widehat{CI}_u(Y_\beta,\Omega_{\beta\delta(s)}\Omega^{-1}_{\delta(s)\delta(s)}Y^{(s)}_\delta;z_{1-\alpha+\gamma},c(\Omega_{\beta\delta(s)}\Omega^{-1}_{\delta(s)\delta(s)}\Omega_{\delta(s)\beta}))$ evaluated at $\delta=0$ is smallest for the subvector $Y^{(s)}_\delta$ that maximizes $\Omega_{\beta\delta(s)}\Omega^{-1}_{\delta(s)\delta(s)}\Omega_{\delta(s)\beta}$, leading us to the following simplified algorithm.

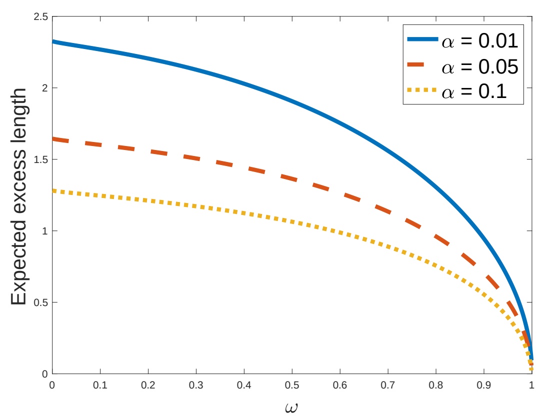

Figure 1: Expected excess length of $\widehat{CI}_u(Z_1,\tilde{Z}_2,z_{1-\alpha+\gamma},c(\omega))$ as a function of ω, for $\alpha\in\{0.01,0.05,0.1\}$ and $\gamma=\alpha/10$.

Algorithm One-Sided*

Amongst all subvectors of Y_δ such that the elements of $\Omega_{\beta\delta(s)}\Omega^{-1}_{\delta(s)\delta(s)}$ are non-negative, find the subvector $Y^{(s^*)}_\delta$ such that $\Omega_{\beta\delta(s)}\Omega^{-1}_{\delta(s)\delta(s)}\Omega_{\delta(s)\beta}$ is maximized at $s=s^*$. Then, construct

$$\widehat{CI}^*_u(Y_\beta,Y_\delta,\Omega)\equiv\widehat{CI}_u(Y_\beta,\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}Y^{(s^*)}_\delta;z_{1-\alpha+\gamma},c(\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\Omega_{\delta(s^*)\beta})). \quad \square$$

It is worth noting that (i) $c(\omega)$ for $\omega\in(0,1)$ is very simple to compute via Monte Carlo simulation, while $c(0)=z_{1-\alpha}$, and (ii) Algorithm One-Sided* only requires one to evaluate

[7] Expected excess length is obtained numerically on the following grid of values: $\omega\in\{0,0.001,0.002,...,0.999\}$. See the online appendix for details.

the function $c(\cdot)$ at the single point $\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\Omega_{\delta(s^*)\beta}$. Therefore, the algorithm carries very low computational cost. We also note that $c\big(\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\Omega_{\delta(s^*)\beta}\big)$ can always be replaced by $\sqrt{1-\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\Omega_{\delta(s^*)\beta}}z_{1-\gamma}$ in the algorithm to yield a CI with correct coverage but worse excess length (in analogy with the CI using the Bonferroni correction in the previous section).

As can be seen from Figure 1, extreme values of $\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\Omega_{\delta(s^*)\beta}$ such as 0.999 can lead to expected excess lengths of our one-sided CI close to zero, entailing expected excess length reductions of nearly 100% relative to the fixed excess length of the standard CI. At more empirically-relevant values of $\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\Omega_{\delta(s^*)\beta}$, such as say 0.7, Figure 1 still implies expected excess length reductions of more than 30% for $\alpha=0.05$. On the other hand, for *any* realization of the data, the excess length of our one-sided CI cannot exceed $z_{1-\alpha+\gamma}=1.695$ for $\alpha=0.05$ and $\gamma=\alpha/10$. This implies that the excess length increase of our recommended CI cannot exceed roughly 3% relative to the standard one-sided CI for $\alpha=0.05$, which has a fixed excess length of $z_{1-\alpha}=1.645$.

In order to make practical implementation computationally trivial for the user, requiring no Monte Carlo simulation, we approximate $c(\omega)$ via a polynomial response surface regression. Table 1 provides the estimated coefficients for a 6$^{\text{th}}$ order polynomial approximation of $c(\omega)$ for $\alpha\in\{0.01,0.05,0.1\}$ and $\gamma=\alpha/10$. For each value of α, the R^2 is greater than 0.999.[8]

Table 1: Coefficients for 6$^{\text{th}}$ order polynomial approximations of $c(\omega)$ for $\gamma=\alpha/10$

α	1	ω	ω^2	ω^3	ω^4	ω^5	ω^6
0.01	2.324	2.507	-19.623	65.049	-122.024	112.981	-40.989
0.05	1.638	2.481	-16.101	52.700	-98.935	91.765	-33.363
0.1	1.273	2.425	-14.104	46.033	-86.795	80.819	-29.484

2.3 Two-Sided Confidence Intervals

In this section we focus on forming analogous adaptive two-sided CIs allowing $\delta\geq 0$ to be multidimensional in the large sample problem characterized by (1) and (4). These two-sided CIs use the same basic logic as the one-sided CIs of the previous section but work to shorten each side of the CI separately while maintaining correct coverage. The following algorithms provide the details.

[8]The regression is performed on the same grid that underlies Figure 1, i.e., $\omega\in\{0,0.001,0.002,...,0.999\}$.

Let

$$\begin{pmatrix} Z_1 \\ \tilde{Z}_2 \\ \tilde{Z}_3 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \omega_{12} & \omega_{13} \\ \omega_{12} & \omega_{12} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{13} \end{pmatrix}\right) \tag{6}$$

and $\widetilde{\mathcal{C}} = \{(c_u,\tilde{\omega}) \in \mathbb{R}_\infty \times \bar{\mathcal{S}} : c_u \in [\underline{c_u}(\tilde{\omega}),\infty]\}$, where $\bar{\mathcal{S}} = \mathcal{S} \cup \{(x,y,z) \in \mathbb{R}^3 : x \in [0,1), y = z = 0\} \cup \{(x,y,z) \in \mathbb{R}^3 : y \in (0,1), x = z = 0\}$ with $\mathcal{S} = \{(x,y,z) \in \mathbb{R}^3 : \ x,y \in (0,1),\ xy < z^2 < xy + 2xyz - xy^2 - x^2y\}$,[9] where $\underline{c_u} : \bar{\mathcal{S}} \to \mathbb{R}$ is implicitly defined by

$$P(-\min\{z_{1-(\alpha-\gamma)/2}, -\tilde{Z}_3 + \underline{c_u}(\tilde{\omega})\} \leq Z_1 \leq z_{1-(\alpha-\gamma)/2}) = 1 - \alpha. \tag{7}$$

For $\gamma \in (0,\alpha)$, consider the function $\tilde{c} : \widetilde{\mathcal{C}} \to \mathbb{R}_\infty$ implicitly defined at points $(c_u,\tilde{\omega}) \in \widetilde{\mathcal{C}}$ for which $\omega_{12},\omega_{13} \neq 0$ by

$$P(-\min\{z_{1-(\alpha-\gamma)/2}, -\tilde{Z}_3 + c_u\} \leq Z_1 \leq \min\{z_{1-(\alpha-\gamma)/2}, \tilde{Z}_2 + \tilde{c}(c_u,\tilde{\omega})\}) = 1 - \alpha. \tag{8}$$

The domain $\widetilde{\mathcal{C}}$ of $\tilde{c}(\cdot)$ is defined in terms of the lower bound $\underline{c_u}(\tilde{\omega})$ on c_u in (7) so that for any given $\tilde{\omega}$, the solution to (8) exists. More specifically, the lower bound $\underline{c_u}(\tilde{\omega})$ rules out c_u values that are too small to admit a solution to (8). Next, for $(c_u,\tilde{\omega}) \in \widetilde{\mathcal{C}}$ with $\omega_{12} = 0$, define $\tilde{c}(c_u,\tilde{\omega}) = \lim_{\bar{\omega}_{12} \to 0} \tilde{c}(c_u,\bar{\omega}_{12},\omega_{13},\omega_{23})$ and for $(c_u,\tilde{\omega}) \in \widetilde{\mathcal{C}}$ with $\omega_{13} = 0$, define $\tilde{c}(c_u,\tilde{\omega}) = \lim_{\bar{\omega}_{13} \to 0} \tilde{c}(c_u,\omega_{12},\bar{\omega}_{13},\omega_{23})$.[10] Finally, define the correspondence $\tilde{c}_u : \bar{\mathcal{S}} \rightrightarrows \mathbb{R}$ as

$$\tilde{c}_u(\tilde{\omega}) = \operatorname{argmin}_{c_u \in [\underline{c_u}(\tilde{\omega}),\infty]} E[\max\{\min\{z_{1-(\alpha-\gamma)/2}, \tilde{Z}_2 + \tilde{c}(c_u,\tilde{\omega})\} + \min\{z_{1-(\alpha-\gamma)/2}, -\tilde{Z}_3 + c_u\}, 0\}]. \tag{9}$$

Note that $\tilde{c}_u(0) = z_{1-\alpha/2}$. The following proposition ensures that $\tilde{c}_u : \bar{\mathcal{S}} \rightrightarrows \mathbb{R}$ is well-defined and possesses some desirable properties.[11]

Proposition 2

For any $\tilde{\omega} \in \bar{\mathcal{S}}$, $\tilde{c}_u(\tilde{\omega}) \subset \mathbb{R}_\infty$ defined in (9) is non-empty and compact and $\tilde{c}_u : \bar{\mathcal{S}} \rightrightarrows \mathbb{R}_\infty$ is

[9]In terms of arguments (x,y,z), the definition of $\mathcal{S}$ is equivalent to the positive definiteness of the matrix $\begin{pmatrix} 1 & x & y \\ x & x & z \\ y & z & y \end{pmatrix}$.

[10]The limits in these definitions exist by the continuity of $\tilde{c}(c_u,\tilde{\omega})$ at all $(c_u,\tilde{\omega}) \in \widetilde{\mathcal{C}}$ with $\omega_{12},\omega_{13} \neq 0$. See Lemma 1 in the Appendix. We define $\tilde{c}(c_u,\tilde{\omega})$ at $\omega_{12} = 0$ and $\omega_{13} = 0$ in terms of limits because multiple values of $\tilde{c}(c_u,\tilde{\omega})$ satisfy (8) when $\omega_{12} = 0$ and we wish to treat ω_{12} and ω_{13} symmetrically in light of Proposition 3 below.

[11]In our numerical work, we have found the solution to (9) to be a singleton and $\tilde{c}_u$ to be a continuous function when $\omega_{12},\omega_{13} \neq 0$.

upper hemicontinuous.

Now, let $Y_\delta^{(s_1)}$ and $Y_\delta^{(s_2)}$ denote two arbitrary (possibly empty) subvectors of Y_δ with

$$\begin{pmatrix} Y_\beta \\ Y_\delta^{(s_1)} \\ Y_\delta^{(s_2)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta \\ \delta^{(s_1)} \\ \delta^{(s_2)} \end{pmatrix}, \begin{pmatrix} 1 & \Omega_{\beta\delta^{(s_1)}} & \Omega_{\beta\delta^{(s_2)}} \\ \Omega_{\delta^{(s_1)}\beta} & \Omega_{\delta^{(s_1)}\delta^{(s_1)}} & \Omega_{\delta^{(s_1)}\delta^{(s_2)}} \\ \Omega_{\delta^{(s_2)}\beta} & \Omega_{\delta^{(s_2)}\delta^{(s_1)}} & \Omega_{\delta^{(s_2)}\delta^{(s_2)}} \end{pmatrix}\right),$$

where by convention, $\delta^{(s_1)}$ ($\delta^{(s_2)}$), $\Omega_{\beta\delta^{(s_1)}}$ ($\Omega_{\beta\delta^{(s_2)}}$), $\Omega_{\delta^{(s_1)}\delta^{(s_1)}}$ ($\Omega_{\delta^{(s_2)}\delta^{(s_2)}}$), and $\Omega_{\delta^{(s_1)}\delta^{(s_2)}}$, as well as $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}$ ($\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}$), $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\Omega_{\delta^{(s_1)}\beta}$ ($\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}\Omega_{\delta^{(s_2)}\beta}$), and $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\Omega_{\delta^{(s_1)}\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}\Omega_{\delta^{(s_2)}\beta}$, are set equal to zero when $Y_\delta^{(s_1)}=\emptyset$ ($Y_\delta^{(s_2)}=\emptyset$). Define

$$\begin{aligned}&\widehat{CI}_t(Y_\beta,\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}Y_\delta^{(s_1)},\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}Y_\delta^{(s_2)};z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\Omega}^{(s_1,s_2)}),c_u(\tilde{\Omega}^{(s_1,s_2)}))\\ &=\Big[Y_\beta-\min\Big\{z_{1-(\alpha-\gamma)/2},\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}Y_\delta^{(s_1)}+c_\ell(\tilde{\Omega}^{(s_1,s_2)})\Big\},\\ &\quad Y_\beta+\min\Big\{z_{1-(\alpha-\gamma)/2},-\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}Y_\delta^{(s_2)}+c_u(\tilde{\Omega}^{(s_1,s_2)})\Big\}\Big],\end{aligned}$$

where $c_u(\tilde{\omega})\in\tilde{c}_u(\tilde{\omega})$, $c_\ell(\tilde{\omega})=\tilde{c}(c_u(\tilde{\omega}),\tilde{\omega})$ and

$$\tilde{\Omega}^{(s_1,s_2)}=(\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\Omega_{\delta^{(s_1)}\beta},\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}\Omega_{\delta^{(s_2)}\beta},\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\Omega_{\delta^{(s_1)}\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}\Omega_{\delta^{(s_2)}\beta}).$$

For any given pair of subvectors $Y_\delta^{(s_1)}$ and $Y_\delta^{(s_2)}$ such that all elements of $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}$ are non-negative and all elements of $\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}$ are non-positive, $c_\ell(\tilde{\Omega}^{(s_1,s_2)})$ and $c_u(\tilde{\Omega}^{(s_1,s_2)})$ minimize the expected length at $\delta=0$ of CIs of the form

$$\widehat{CI}_t(Y_\beta,\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}Y_\delta^{(s_1)},\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}Y_\delta^{(s_2)};z_{1-(\alpha-\gamma)/2},c_\ell,c_u)$$

amongst all (c_ℓ,c_u) values that have valid coverage for all $\delta\geq 0$.

Algorithm Two-Sided

Amongst all pairs of subvectors of Y_δ (including the empty ones) such that all elements of $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}$ are non-negative and all elements of $\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}$ are non-positive, find the subvector pair $Y_\delta^{(s_1^*)}$ and $Y_\delta^{(s_2^*)}$ such that the expected length of

$$\widehat{CI}_t(Y_\beta,\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}Y_\delta^{(s_1)},\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}Y_\delta^{(s_2)};z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\Omega}^{(s_1,s_2)}),c_u(\tilde{\Omega}^{(s_1,s_2)}))$$

at $\delta=0$ is minimized at $s_1=s_1^*$ and $s_2=s_2^*$. Then, construct

$$\widehat{CI}_t(Y_\beta,\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}Y_\delta^{(s_1^*)},\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}Y_\delta^{(s_2^*)};z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\Omega}^{(s_1^*,s_2^*)}),c_u(\tilde{\Omega}^{(s_1^*,s_2^*)})). \quad \square$$

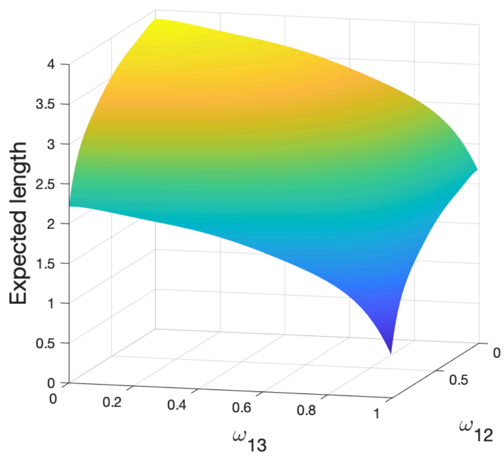

Figure 2: Expected length as a function of ω_{12} and ω_{13}, for $\alpha=0.05$ and $\gamma=\alpha/10$.

Figure 2 shows the fitted surface of a 6$^{\text{th}}$ order polynomial regression of the expected length of $\widehat{CI}_t(Z_1,\tilde{Z}_2,\tilde{Z}_3;z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\omega}),c_u(\tilde{\omega}))$ on ω_{12} and ω_{13} alone, for $\alpha=0.05$ and $\gamma=\alpha/10$. The values of $\tilde{\omega}$ on which this regression is based are given by $\bar{\Omega}=\bar{\mathcal{S}}\cap\mathcal{G}^2\times-\mathcal{G}\cup\mathcal{G}\cup\{-0.99,-0.98,...,0.99\}$, where

$$\mathcal{G}=\{0,0.005,0.01,0.02,...,0.1,0.15,...,0.9,0.91,...,0.99,0.995\} \tag{10}$$

and $-\mathcal{G}=\{g:-g\in\mathcal{G}\}$. The corresponding R^2 is greater than 0.999, implying that the expected length is nearly invariant to ω_{23}. Similarly, the maximum difference between the largest and the smallest expected length over the set $\bar{\Omega}$ for any given $(\omega_{12},\omega_{13})$ is equal to 0.0289. Note also that the fitted expected length in Figure 2 is strictly decreasing in ω_{12} and ω_{13}. Since the expected length does not depend upon β, this implies that the expected length of $\widehat{CI}_t(Y_\beta,\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}Y_\delta^{(s_1)},\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}Y_\delta^{(s_2)};z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\Omega}^{(s_1^*,s_2^*)}),c_u(\tilde{\Omega}^{(s_1,s_2)}))$ evaluated at $\delta=0$ is approximately smallest for the subvectors $Y_\delta^{(s_1)}$ and $Y_\delta^{(s_2)}$ that maximize $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\Omega_{\delta^{(s)}\beta}$ and $\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s_2)}}\Omega_{\delta^{(s_2)}\beta}$, motivating the following simplified algorithm.

Algorithm Two-Sided*

Amongst all pairs of subvectors of Y_δ (including the empty ones) such that all elements of $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}$ are non-negative and all elements of $\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}$ are non-positive, find the subvector pair $Y_\delta^{(s_1^*)}$ and $Y_\delta^{(s_2^*)}$ such that $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\Omega_{\delta^{(s_1)}\beta}$ and $\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}\Omega_{\delta^{(s_2)}\beta}$ are maximized at $s_1=s_1^*$ and $s_2=s_2^*$. Then, construct

$$\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$$
$$\equiv\widehat{CI}_t(Y_\beta,\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}Y_\delta^{(s_1^*)},\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}Y_\delta^{(s_2^*)};z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\Omega}^{(s_1^*,s_2^*)}),c_u(\tilde{\Omega}^{(s_1^*,s_2^*)})). \quad \square$$

Similarly to the one-sided CI case above, Figure 2 shows that the expected length of our two-sided CI can be very small for extreme values of $\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}\Omega_{\delta^{(s_1^*)}\beta}$ and $\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}\Omega_{\delta^{(s_2^*)}\beta}$. At the same time, for *any* realization of the data, the length of our two-sided CI cannot exceed $2*z_{1-(\alpha-\gamma)/2}=4.009$ for $\alpha=0.05$ and $\gamma=\alpha/10$. This implies that the length increase of our recommended CI cannot exceed 2.28% relative to the fixed length $2*z_{1-\alpha/2}=3.92$ of the standard two-sided CI.

Table 2 provides the estimated coefficients for a 6$^{\text{th}}$ order polynomial approximation of $c_u(\tilde{\omega})$ for $\alpha=0.05$ and $\gamma=\alpha/10$ in terms of ω_{12} and ω_{13}. The corresponding R^2 is greater than 0.999.[12]

Table 2: Coefficients for 6$^{\text{th}}$ order polynomial approximation of $c_u(\omega)$ for $\alpha=0.05$ and $\gamma=\alpha/10$

	1	ω_{12}	ω_{12}^2	ω_{12}^3	ω_{12}^4	ω_{12}^5	ω_{12}^6
1	1.954	1.339	-4.511	11.729	-18.876	15.534	-5.279
ω_{13}	1.129	-0.801	1.126	-1.174	2.128	-0.551	
ω_{13}^2	-12.293	0.009	0.908	-3.233	0.172		
ω_{13}^3	45.650	0.594	0.815	1.762			
ω_{13}^4	-92.359	-1.005	-0.985				
ω_{13}^5	89.504	0.285					
ω_{13}^6	-33.368						

The following proposition also enables one to directly compute an approximation to $c_\ell(\tilde{\omega})$ from Table 2 by simply reversing the roles of ω_{12} and ω_{13} in the computation of $c_u(\tilde{\omega})$.

Proposition 3
$c_\ell(\omega_{13},\omega_{12},\omega_{23})=c_u(\omega_{12},\omega_{13},\omega_{23})$.

If one relies on the polynomial approximation of $c_u(\cdot)$ and $c_\ell(\cdot)$ in terms of ω_{12} and ω_{13} when computing $\widehat{CI}_t(\cdot)$, the latter is, of course, no longer guaranteed to cover the true

[12] The regression is performed on the same grid that underlies Figure 2, i.e., $\omega\in\bar{\Omega}$.

unknown parameter with at least 1-α probability since $c_u(\cdot)$ and $c_\ell(\cdot)$ also depend upon ω_{23}. However, we find numerically that $c_u(\cdot)$ and $c_\ell(\cdot)$ are nearly invariant to ω_{23}. Indeed, a grid search over $\bar{\Omega}$ reveals a minimum coverage probability of 94.78% for the CI using the simple polynomial approximation when $\alpha=0.05$. We therefore expect minimal size distortions from employing the polynomial approximations of $c_u(\cdot)$ and $c_\ell(\cdot)$ in practice. This view is also supported by the finite-sample coverage probabilities of the corresponding CIs, which are very close to those of standard CIs. See Section 5.

3 Finite Sample Problem of Restricted Nuisance Parameters

Consider inference on a scalar parameter of interest $b\in\mathbb{R}$ in a well-behaved model with a vector nuisance parameter $d\in\mathbb{R}^k_+$ for some $k\geq 1$ that is known to have all elements greater than or equal to zero. For a standard parameter estimator $(\hat{b},\hat{d}')'$, as the number of observations n in the sample grows, standard assumptions imply[13]

$$\sqrt{n}\begin{pmatrix} \widehat{b}-b \\ \widehat{d}-d \end{pmatrix} \xrightarrow{d} \mathcal{N}(0,\Sigma) \quad \text{with} \quad \Sigma=\begin{pmatrix} \Sigma_{bb} & \Sigma_{bd} \\ \Sigma_{db} & \Sigma_{dd} \end{pmatrix}, \tag{11}$$

where Σ is a consistently estimable covariance matrix. Note that this setting accommodates regression models, instrumental variables models, maximum likelihood models and models estimated by the generalized method of moments under standard assumptions when some nuisance parameters are known to be greater or less than a given bound via simple reparamterization of the nuisance parameters. For example, say that the researcher knows from economic theory that the nuisance parameter $\tilde{d}$ is less than or equal to $\tilde{c}$ for some known constant $\tilde{c}\in\mathbb{R}$. Then $d=-(\tilde{d}-\tilde{c})$ is the simple reparameterization that fits this setting.

Example: Regression with Sign-Restricted Control Coefficients

One of the most common examples that fits this setting is inference on a regression coefficient of interest b in the standard linear regression model for observations $i=1,...,n$

$$y_i=bz_i+x_i'd+w_i'c+\varepsilon_i,$$

where y_i is the dependent variable, z_i is the scalar regressor of interest, $x_i\in\mathbb{R}^{\mathcal{D}_x}$ are control variables with *known positive partial effects* $d\geq 0$ on y_i, $w_i\in\mathbb{R}^{\mathcal{D}_w}$ are control variables with unrestricted partial effects c and ε_i is the error term. The ordinary least squares estimator

[13] For simplicity of notation, we suppress the dependence of certain finite-sample quantities on the sample size n until Section 3.2.

$(\hat{b},\hat{d}')'$ satisfies (11) under standard assumptions on the linear regression model. $\square$

3.1 Implementation

For a consistent covariance matrix estimator $\widehat{\Sigma}$, (11) suggests the following large-sample distributional approximation consistent with (1) and (4):

$$\operatorname{Diag}\left(\widehat{\Sigma}\right)^{-1/2}\sqrt{n}\begin{pmatrix}\widehat{b}\\ \widehat{d}\end{pmatrix}\overset{a}{\sim}\mathcal{N}\left(\begin{pmatrix}\beta\\ \delta\end{pmatrix},\Omega\right), \tag{12}$$

where $\beta=\sqrt{n}b/\sqrt{\Sigma_{bb}}$, $\delta=\operatorname{Diag}(\Sigma_{dd})^{-1/2}\sqrt{n}d$ and $\Omega=\operatorname{Diag}(\Sigma)^{-1/2}\Sigma\operatorname{Diag}(\Sigma)^{-1/2}$. Note, however, that Ω is not typically known in practice but can be consistently estimated by $\widehat{\Omega}=\operatorname{Diag}(\widehat{\Sigma})^{-1/2}\widehat{\Sigma}\operatorname{Diag}(\widehat{\Sigma})^{-1/2}$. Let $\hat{\delta}^{(s)}=\operatorname{Diag}(\widehat{\Sigma}_{dd}^{(s)})^{-1/2}\sqrt{n}\hat{d}^{(s)}$, $\hat{s}^*$ denote the subset of the set of indices $\{1,...,k\}$ that maximizes $\widehat{\Omega}_{bd^{(s)}}\widehat{\Omega}^{-1}_{d^{(s)}d^{(s)}}\widehat{\Omega}_{d^{(s)}b}$ amongst all subsets of indices $s\subset\{1,...,k\}$ such that the elements of $\widehat{\Omega}_{bd^{(s)}}\widehat{\Omega}^{-1}_{d^{(s)}d^{(s)}}$ are non-negative and $(\hat{s}_1^*,\hat{s}_2^*)$ denote the subsets of the set of indices $\{1,...,k\}$ that maximize $\widehat{\Omega}_{\beta\delta^{(s_1)}}\widehat{\Omega}^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\widehat{\Omega}_{\delta^{(s_1)}\beta}$ and $\widehat{\Omega}_{\beta\delta^{(s_2)}}\widehat{\Omega}^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}\widehat{\Omega}_{\delta^{(s_2)}\beta}$ amongst all subsets of indices $s_1,s_2\subset\{1,...,k\}$ such that the elements of $\widehat{\Omega}_{\beta\delta^{(s_1)}}\widehat{\Omega}^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}$ are non-negative and the elements of $\widehat{\Omega}_{\beta\delta^{(s_2)}}\widehat{\Omega}^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}$ are non-positive.

The distributional approximation in (12) and the availability of the consistent estimator $\widehat{\Omega}$ suggest that we can use

$$\begin{aligned}CI_{u,n}\left(\hat{b},\hat{d};\widehat{\Sigma}\right)&=\frac{\sqrt{\widehat{\Sigma}_{bb}}}{\sqrt{n}}\widehat{CI}_u^*\left(\frac{\sqrt{n}\hat{b}}{\sqrt{\widehat{\Sigma}_{bb}}},\operatorname{Diag}(\widehat{\Sigma}_{dd})^{-1/2}\sqrt{n}\hat{d};\widehat{\Omega}\right)\\ &=\left[\hat{b}-\frac{\widehat{\Sigma}_{bb}}{\sqrt{n}}\min\left\{z_{1-\alpha+\gamma},\widehat{\Omega}_{bd^{(\hat{s}^*)}}\widehat{\Omega}^{-1}_{d^{(\hat{s}^*)}d^{(\hat{s}^*)}}\hat{\delta}^{(\hat{s}^*)}+c(\widehat{\Omega}_{bd^{(\hat{s}^*)}}\widehat{\Omega}^{-1}_{d^{(\hat{s}^*)}d^{(\hat{s}^*)}}\widehat{\Omega}_{d^{(\hat{s}^*)b}})\right\},\infty\right)\end{aligned} \tag{13}$$

and

$$\begin{aligned}CI_{t,n}\left(\widehat{b},\widehat{d};\widehat{\Sigma}\right)&=\frac{\sqrt{\widehat{\Sigma}_{bb}}}{\sqrt{n}}\widehat{CI}_t^*\left(\frac{\sqrt{n}\widehat{b}}{\sqrt{\widehat{\Sigma}_{bb}}},\operatorname{Diag}(\widehat{\Sigma}_{dd})^{-1/2}\sqrt{n}\widehat{d};\widehat{\Omega}\right)\\ &=\left[\hat{b}-\frac{\widehat{\Sigma}_{bb}}{\sqrt{n}}\min\left\{z_{1-\frac{\alpha-\gamma}{2}},\widehat{\Omega}_{bd^{(\hat{s}_1^*)}}\widehat{\Omega}^{-1}_{d^{(\hat{s}_1^*)}d^{(\hat{s}_1^*)}}\hat{\delta}^{(\hat{s}_1^*)}+c_\ell\left(\widehat{\widehat{\Omega}}^{(\hat{s}_1^*,\hat{s}_2^*)}\right)\right\},\right.\\ &\qquad\left.\hat{b}+\frac{\widehat{\Sigma}_{bb}}{\sqrt{n}}\min\left\{z_{1-\frac{\alpha-\gamma}{2}},-\widehat{\Omega}_{bd^{(\hat{s}_2^*)}}\widehat{\Omega}^{-1}_{d^{(\hat{s}_2^*)}d^{(\hat{s}_2^*)}}\hat{\delta}^{(\hat{s}_2^*)}+c_u\left(\widehat{\widehat{\Omega}}^{(\hat{s}_1^*,\hat{s}_2^*)}\right)\right\}\right]\end{aligned} \tag{14}$$

as upper one-sided and two-sided CIs for the parameter b, where $\widehat{CI}_u^*(\cdot)$ and $\widehat{CI}_t^*(\cdot)$ are defined in Algorithms One-Sided* and Two-Sided*. The theoretical results of the following section formally confirm that these CIs attain uniformly correct asymptotic coverage under weak conditions.

3.2 Asymptotic Properties

We now present theoretical results ensuring the uniformly correct asymptotic coverage of both the one- and two-sided finite-sample CIs defined in (13) and (14), as well as a uniform upper bound on their asymptotic coverage, under a set of widely-applicable sufficient conditions on the parameter space. In particular, let the parameter λ index the true distribution of the observations used to construct the CIs and decompose λ as follows: $\lambda=(b,d,\Sigma,F)$, where b is the scalar parameter of interest, d is the nuisance parameter known to have all elements greater than zero, Σ is the asymptotic variance corresponding to the parameter estimator $(\hat{b}_n,\hat{d}_n')'$ used by the researcher and F is a (potentially) infinite-dimensional parameter that, along with (b,d), determines the distribution of the observed data. We assume that we have a consistent estimator $\widehat{\Sigma}_n$ of Σ at our disposal.

The parameter space Λ for λ is defined to include parameters $\lambda=(b,d,\Sigma,F)$ such that for some finite $\kappa>0$, the following conditions hold:

(i) $b\in\mathbb{R}$ and $d\in\mathbb{R}_+^k$ for some positive integer k;

(ii) $\Sigma\in\Phi$, $\lambda_{\min}(\Sigma)\geq\kappa$ and $\lambda_{\max}(\Sigma)\leq\kappa^{-1}$, where Φ denotes the set of all positive definite covariance matrices.

In addition, under any sequence of parameters $\{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}=(b_{n,\mathfrak{b}},d_{n,\mathfrak{d}},\Sigma_{n,\Sigma^*},F_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}):n\geq 1\}$ in Λ such that

$$\sqrt{n}(b_{n,\mathfrak{b}},d_{n,\mathfrak{d}})\rightarrow(\mathfrak{b},\mathfrak{d}), \tag{15}$$

$$\Sigma_{n,\Sigma^*}\rightarrow\Sigma^* \tag{16}$$

for some $(\mathfrak{b},\mathfrak{d},\Sigma^*)\in\mathbb{R}_\infty\times\mathbb{R}_{+,\infty}^k\times\Phi$, the following remaining conditions hold:

(iii) $\widehat{\Sigma}_n$ exists and $\lambda_{\min}(\widehat{\Sigma}_n)>0$ with probability 1 for all $n\geq 1$ and $\widehat{\Sigma}_n\xrightarrow{p}\Sigma^*$;

(iv) $\sqrt{n}(\hat{b}_n-b_{n,\mathfrak{b}},\hat{d}_n'-d_{n,\mathfrak{d}}')'\xrightarrow{d}\mathcal{N}(0,\Sigma^*)$;

(v) for any sequence $\{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}\}$ in Λ and any subsequence $\{s_n:n\geq 1\}$ of $\{n:n\geq 1\}$ for which (15)–(16) hold along the subsequence, conditions (iii)–(iv) also hold along the subsequence.

In conjunction with a particular model, parameter estimator $(\hat{b}_n,\hat{d}_n')'$ and covariance matrix estimator $\widehat{\Sigma}_n$, this definition of the parameter space Λ effectively serves as a set of high-level assumptions on the underlying DGP. More specifically, (i)–(ii) are standard parameter space assumptions while (iii)–(v) can typically be verified under standard dependence and moment conditions on the underlying data via laws of large numbers and central limit theorems. We refer the interested reader to Appendix B for details in the context of the standard linear regression model.

With the relevant parameter space defined, we may now state the main theoretical result of this paper that establishes lower and upper bounds on the uniform asymptotic coverage probability of the CIs we propose.

Theorem 1
For $\alpha \in (0,1/2)$ *and* $\gamma \in (0,\alpha)$,

$$\liminf_{n\to\infty} \inf_{\lambda\in\Lambda} P_\lambda\left(b \in CI_{\cdot,n}(\hat{b}_n,\hat{d}_n;\widehat{\Sigma}_n)\right) \geq 1-\alpha$$

and

$$\limsup_{n\to\infty} \sup_{\lambda\in\Lambda} P_\lambda\left(b \in CI_{\cdot,n}(\hat{b}_n,\hat{d}_n;\widehat{\Sigma}_n)\right) \leq 1-\alpha+\gamma,$$

where $CI_{\cdot,n}(\cdot)$ *is equal to either* $CI_{u,n}(\cdot)$ *or* $CI_{t,n}(\cdot)$.

These results show not only that our proposed CIs have correct asymptotic coverage in a strong sense but also that by choosing γ to be "small" reduces how conservative the CIs can be. However, there is a tradeoff in the choice of γ: although a smaller γ leads to CIs that are closer to being similar across the parameter space, it also allows for less length gains when the elements of d are close or equal to zero.

4 Empirical Application of Sign-Restricted Regression

For our proposed CIs to improve upon the length of standard CIs in the standard linear regression context, the researcher must know the sign of at least one of the control variables' coefficients and the estimator of the coefficient of interest must be (asymptotically) correlated with the estimator of the sign-restricted control variables' coefficients. Both conditions are often satisfied in the context of treatment effect regressions for cross-cutting/factorial designs in field experiments. Take, for example, the 2×2 factorial design:

$$Y = \alpha_0 + \alpha_1 T_1 + \alpha_2 T_2 + \alpha_3 T_1 \times T_2 + u, \tag{17}$$

where $E[u|T_1,T_2]=0$ and T_1 and T_2 denote two independent, randomly assigned treatments with $T_i \in \{0,1\}$ for $i\in\{1,2\}$. Here, α_1 and α_2 are the treatment effects of T_1 and T_2 "relative to a business-as-usual counterfactual" (Muralidharan et al., 2020) and α_3 is the "interaction effect", i.e., the treatment effect of jointly providing both treatments minus the sum of the treatment effects of T_1 and T_2.[14] If Y is a "positive" outcome, it is often reasonable to assume that $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$. For example, a research ethics committee is unlikely to clear an experimental design if this is not the case. Furthermore, the OLS estimators of the three treatment effects are likely to be highly correlated in this setting. For example, if each treatment is assigned with probability 1/2 and the error term u is conditionally homoskedastic, then the asymptotic correlation matrix of $\sqrt{n}(\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3)'$ is given by

$$\begin{bmatrix} 1 & 1/2 & -1/\sqrt{2} \\ 1/2 & 1 & -1/\sqrt{2} \\ -1/\sqrt{2} & -1/\sqrt{2} & 1 \end{bmatrix}.$$

Any of the three treatment effects may be of interest and, under the assumption that $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$, it is reasonable to be interested in upper one-sided CIs for (the true values of) α_1 and α_2 and a two-sided CI for (the true value of) α_3. The above correlation structure implies that our upper one-sided CIs for α_1 and α_2 have the potential to improve upon the length of standard upper one-sided CIs. Similarly, our two-sided CI for α_3 has the potential to improve upon the length of the standard two-sided CI through a smaller upper bound.

Sometimes researchers are interested in the following alternative specification of the above regression:

$$Y = \alpha_0 + \alpha_1(T_1 - T_1 \times T_2) + \alpha_2(T_2 - T_1 \times T_2) + \alpha_3^* T_1 \times T_2 + u, \tag{18}$$

where $\alpha_3^* = \alpha_3 - \alpha_1 - \alpha_2$ is the effect of "both" treatments provided jointly, relative to a business-as-usual counterfactual.[15] This regression again results in high correlation between OLS estimators: under the same conditions as the example above, the asymptotic correlation matrix of $\sqrt{n}(\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3^*)'$ is given by

$$\begin{bmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{bmatrix}.$$

[14] Using the potential outcomes notation, where Y_{t_1,t_2} is the potential outcome of Y when $T_1 = t_1$ and $T_2 = t_2$, the three treatment effects can be written as $\alpha_1 = E[Y_{1,0} - Y_{0,0}]$, $\alpha_2 = E[Y_{0,1} - Y_{0,0}]$, and $\alpha_3 = E[Y_{1,1} - Y_{0,0}] - (E[Y_{1,0} - Y_{0,0}] + E[Y_{0,1} - Y_{0,0}])$.

[15] That is $\alpha_3^* = E[Y_{1,1} - Y_{0,0}]$.

In this case, our two-sided CI for α_3^* has the potential to improve upon the length of the standard two-sided CI through a larger lower bound.

To illustrate the usefulness of our proposed CIs, we apply them in the context of a field experiment where a 2×2 factorial design was used. In particular, we revisit Blattman et al. (2017) (BJS) who recruited 999 poor young men in Liberia who exhibited "high rates of violence, crime, and other antisocial behaviors" to participate in an experiment. The two treatments are "therapy", an eight-week program of group cognitive behavior therapy, and "cash", a $200 grant corresponding to roughly three months' wages. In simple terms, the main research question is whether "therapy" and "cash" can help reduce violent, criminal, and other antisocial behaviors. The hypothesized channels are improved noncognitive skills such as self-control ("therapy") and an increase in legal work ("cash"). BJS conducted two follow-up surveys, the first 2–5 weeks and the second 12–13 months after the intervention to elicit "short-term" and "long-term" impacts, respectively.

Table 3: Empirical results

	$\hat{\beta}$	SE	SSCI	(E)L	SCI	(E)L	Ratio
T	0.0829	0.0929	$[-0.0148,\infty)$	0.0977	$[-0.0700,\infty)$	0.1529	0.6391
C	-0.1316	0.0969	$[-0.2959,\infty)$	0.3788	$[-0.2910,\infty)$	0.3739	1.0131
B	0.2468	0.0883	[0.0988,0.4238]	0.3250	[0.0737,0.4198]	0.3462	0.9390
I	0.2955	0.1255	[0.0439,0.4101]	0.3662	[0.04950.5415]	0.4920	0.7443

Table 3 reproduces the results concerning the treatments' long-term impact on a summary index of antisocial behaviors (times minus one) (cf. the first row of Panel B of Table 2 in BJS). The table includes one of the main findings of BJS: while the two treatments do not have statistically significant long-term effects in isolation, they do have a *joint* positive long-term effect on the index of antisocial behaviors. Column 1 ($\hat{\beta}$) shows the OLS point estimates for "therapy" (T), "cash" (C), "both" (B), and "interaction" (I) as defined above and column 2 (SE) reports the corresponding (heteroskedasticity–robust) standard errors. Note that BJS only consider the specification given in equation (18), i.e., they only estimate the effect of "both" treatments and not the "interaction" effect.[16] Column 3 (SSCI) shows our proposed CIs for $\alpha = 0.05$, which are upper one-sided for T and C and two-sided for B and I, when assuming that the treatment effects of "therapy" and

[16] In fact, BJS consider the specification given in equation (18) augmented by a set of additional controls. For the purpose of this analysis, we take the signs of these additional controls as unknown. See BJS for more information on the additional controls.

"cash" are *a priori* known to be nonnegative. Algorithm Two-Sided and Algorithm Two-Sided* yield the same CIs for both B and I in this application.[17] The reported CIs for B and I rely on the polynomial approximation of $c_u(\cdot)$ and $c_l(\cdot)$ given in Table 2.[18] Column 5 (SCI) shows the corresponding standard CIs. Columns 4 and 6 (both E(L)) give the ("excess") lengths of SSCI and SCI, where the "excess" length of one-sided CIs here is computed as the difference between $\hat{\beta}$ and the CI's lower bound. Column 7 (Ratio) computes the ratio of the ("excess") length of SSCI relative to SCI. We find that our proposed CIs are much shorter than standard CIs for T, B, and I and only marginally longer for C.

5 Calibrated Simulations for Sign-Restricted Regression

To illustrate the finite-sample properties of our proposed CIs, we perform a Monte Carlo study calibrated to the BJS factorial design regression of the previous section. In particular, we create 10,000 bootstrap samples by drawing with replacement from the sample of $n = 947$ men underlying the regression results in Table 3. In each bootstrap sample, we estimate the regressions (17) and (18). Since the expected value of the treatment effect of "cash" under the empirical distribution is equal to the point estimate in the original sample, -0.1316, it is outside of the sign-restricted parameter space $\alpha_2 \geq 0$. We therefore recenter the estimates of the treatment effect of "cash" over the bootstrap samples to have mean zero (by adding 0.1316). For each bootstrap sample, we construct our proposed CIs, using Algorithms One-Sided* and Two-Sided* with the polynomial approximation of $c_u(\cdot)$ and $c_l(\cdot)$ for B and I, standard CIs, the (excess) length of each CI and whether they cover the true parameter value, i.e., the corresponding (re-centered) point estimate in the original sample. All CIs are constructed using standard heteroskedasticity-robust variance-covariance matrix estimators computed within each bootstrap sample. Since the empirical distribution from which the bootstrap samples are drawn is not normally distributed, this simulation exercise captures the effect on CI coverage of departures from the large sample normal means problem of Section 2.

[17] The estimated (asymptotic) correlation matrices for the estimator of the effects of i) T, C, and B and ii) T, C, and I are given by

$$\begin{bmatrix} 1.0000 & 0.5238 & 0.6104 \\ 0.5238 & 1.0000 & 0.5543 \\ 0.6104 & 0.5543 & 1.0000 \end{bmatrix} \text{ and } \begin{bmatrix} 1.0000 & 0.5238 & -0.7154 \\ 0.5238 & 1.0000 & -0.7699 \\ -0.7154 & -0.7699 & 1.0000 \end{bmatrix},$$

respectively. We augmented the corresponding regressions by the same set of controls as BJS, cf. footnote 16.

[18] The corresponding CIs that calculate the CI bounds directly are given by $[0.0977, 0.4238]$ and [0.0439,0.4117] for B and I, respectively.

Table 4: Monte Carlo results

		T	C	B	I	B0	I0
CP	SSCI	94.26	94.48	95.34	94.80	95.07	93.97
	SCI	94.30	93.96	94.78	94.49	94.78	94.49
E(E)L	SSCI	0.14	0.16	0.35	0.45	0.33	0.41
	SCI	0.15	0.16	0.36	0.50	0.36	0.50
	Ratio	0.9301	1.0070	0.9743	0.8920	0.9230	0.8241

Table 4 reports the coverage probability (CP) computed across bootstrap realizations of our proposed CIs and of the standard CIs for all four treatment effects, T, C, B, and I. Table 4 also reports the expected (excess) length (E(E)L) of these CIs across the bootstrap realizations. In addition to the above DGP, we also consider a modification where the true value of the treatment effect of "therapy" is set equal to zero (by subtracting the point estimate in the original sample, 0.0829, from the corresponding estimates in the bootstrap iterations). The corresponding results for the effect of "both" treatments and the "interaction" effect are given in the last two columns, B0 and I0.

We observe that our proposed CIs have good finite sample coverage, comparable to that of the standard CIs, with little coverage distortion despite the non-normally distributed data. In terms of expected (excess) length, most of our proposed CIs offer sizeable improvements over standard CIs, with expected length improvements up to nearly 18% for this particular data calibration.

A Technical Appendix

Proof of Proposition 1: Consider the function $f:[0,1)\times[0,z_{1-\gamma}]$ such that for $(Z_1,\tilde{Z}_2)$ defined in (5),

$$f(\omega,c)=P(Z_1>\min\{z_{1-\alpha+\gamma},\tilde{Z}_2+c\})-\alpha.$$

For $\omega\in(0,1)$ and $c\in[0,z_{1-\gamma}]$,

$$\begin{aligned}
f(\omega,c)&=\int_{-\infty}^{\infty}P(Z_1>\min\{z_{1-\alpha+\gamma},\tilde{Z}_2+c\}|\tilde{Z}_2=\tilde{z}_2)\frac{1}{\sqrt{\omega}}\phi(\tilde{z}_2/\sqrt{\omega})d\tilde{z}_2-\alpha\\
&=\int_{-\infty}^{\infty}\Phi\left(\frac{\tilde{z}_2-\min\{z_{1-\alpha+\gamma},\tilde{z}_2+c\}}{\sqrt{1-\omega^2}}\right)\frac{1}{\sqrt{\omega}}\phi(\tilde{z}_2/\sqrt{\omega})d\tilde{z}_2-\alpha\\
&=\int_{-\infty}^{z_{1-\alpha+\gamma}-c}\Phi\left(-\frac{c}{\sqrt{1-\omega^2}}\right)\frac{1}{\sqrt{\omega}}\phi(\tilde{z}_2/\sqrt{\omega})d\tilde{z}_2+\int_{z_{1-\alpha+\gamma}-c}^{\infty}\Phi\left(\frac{\tilde{z}_2-z_{1-\alpha+\gamma}}{\sqrt{1-\omega^2}}\right)\frac{1}{\sqrt{\omega}}\phi(\tilde{z}_2/\sqrt{\omega})d\tilde{z}_2-\alpha\\
&=\Phi\left(-\frac{c}{\sqrt{1-\omega^2}}\right)\Phi\left(\frac{z_{1-\alpha+\gamma}-c}{\sqrt{\omega}}\right)+\int_{z_{1-\alpha+\gamma}-c}^{\infty}\Phi\left(\frac{\tilde{z}_2-z_{1-\alpha+\gamma}}{\sqrt{1-\omega^2}}\right)\frac{1}{\sqrt{\omega}}\phi(\tilde{z}_2/\sqrt{\omega})d\tilde{z}_2-\alpha.
\end{aligned}$$

Clearly, $f(\omega,c)$ is continuously differentiable for all $\omega\in(0,1)$ and $c\in[0,z_{1-\gamma}]$. In addition,

$$\frac{\partial f(\omega,c)}{\partial c}=-\frac{1}{\sqrt{1-\omega^2}}\phi\left(-\frac{c}{\sqrt{1-\omega^2}}\right)\Phi\left(\frac{z_{1-\alpha+\gamma}-c}{\sqrt{\omega}}\right)<0$$

for all $\omega\in(0,1)$ and $c\in[0,z_{1-\gamma}]$ since $\gamma\in(0,\alpha)$.

Finally, note that for any $\omega\in(0,1)$, there exists $c\in[0,z_{1-\gamma}]$ such that $f(\omega,c)=0$ since $f(\omega,\cdot)$ is continuously strictly decreasing,

$$f(\omega,0)=P(Z_1>\min\{z_{1-\alpha+\gamma},\tilde{Z}_2\})-\alpha>P(Z_1-\tilde{Z}_2>0)-\alpha=1/2-\alpha>0$$

and

$$f(\omega,z_{1-\gamma})=P(Z_1>\min\{z_{1-\alpha+\gamma},\tilde{Z}_2+z_{1-\gamma}\})-\alpha\leq P(Z_1>z_{1-\alpha+\gamma})-\alpha=-\gamma<0.$$

In conjunction with the fact that $c(0)=z_{1-\alpha}=\lim_{\omega\to 0}c(\omega)$, the statement of the proposition then follows from the implicit function theorem. ■

The following lemmata are used in the proofs of Proposition 2 and Theorem 1.

Lemma 1
The function $\tilde{c}:\tilde{\mathcal{C}}\to\mathbb{R}_\infty$ *exists and is continuous.*

Proof: Consider the function $f:\mathbb{R}_\infty\times\widetilde{\mathcal{C}}\to[\alpha-1,\alpha]$ such that for $(Z_1,\tilde{Z}_2,\tilde{Z}_3)$ defined in (6),

$$f(\tilde{c},c_u,\tilde{\omega})=P(-\min\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{Z}_3+c_u\}\leq Z_1\leq\min\{z_{1-\frac{\alpha-\gamma}{2}},\tilde{Z}_2+\tilde{c}\})-(1-\alpha).$$

For $(\tilde{c},c_u,\tilde{\omega})\in\mathbb{R}_\infty\times\widetilde{\mathcal{C}}$ with $\omega_{12},\omega_{13}\neq0$,

$$\begin{aligned}
&f(\tilde{c},c_u,\tilde{\omega})\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P(-\min\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{Z}_3+c_u\}\leq Z_1\leq\min\{z_{1-\frac{\alpha-\gamma}{2}},\tilde{Z}_2+\tilde{c}\}|\tilde{Z}_2=\tilde{z}_2,\tilde{Z}_3=\tilde{z}_3)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3-(1-\alpha)\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\Phi\left(\frac{\min\{z_{1-\frac{\alpha-\gamma}{2}},\tilde{z}_2+\tilde{c}\}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)-\Phi\left(\frac{-\min\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{z}_3+c_u\}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)\right]\\
&\qquad\times\mathbf{1}(\min\{z_{1-\frac{\alpha-\gamma}{2}},\tilde{z}_2+\tilde{c}\}\geq-\min\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{z}_3+c_u\})g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3-(1-\alpha)\\
&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}\Phi\left(\frac{\tilde{z}_2+\tilde{c}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)\mathbf{1}(\tilde{z}_2+\tilde{c}\geq-\min\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{z}_3+c_u\})\right.\\
&\qquad\left.+\int_{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}^{\infty}\Phi\left(\frac{z_{1-\frac{\alpha-\gamma}{2}}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)\mathbf{1}(z_{1-\frac{\alpha-\gamma}{2}}\geq-\min\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{z}_3+c_u\})\right]g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3\\
&\quad-\int_{-\infty}^{\infty}\left[\int_{-\infty}^{c_u-z_{1-\frac{\alpha-\gamma}{2}}}\Phi\left(\frac{-z_{1-\frac{\alpha-\gamma}{2}}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)\mathbf{1}(\min\{z_{1-\frac{\alpha-\gamma}{2}},\tilde{z}_2+\tilde{c}\}\geq-z_{1-\frac{\alpha-\gamma}{2}})\right.\\
&\qquad\left.+\int_{c_u-z_{1-\frac{\alpha-\gamma}{2}}}^{\infty}\Phi\left(\frac{\tilde{z}_3-c_u-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)\mathbf{1}(\min\{z_{1-\frac{\alpha-\gamma}{2}},\tilde{z}_2+\tilde{c}\}\geq\tilde{z}_3-c_u)\right]g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_3d\tilde{z}_2-(1-\alpha)\\
&=\int_{-\infty}^{c_u-z_{1-\frac{\alpha-\gamma}{2}}}\int_{-z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}^{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}\Phi\left(\frac{\tilde{z}_2+\tilde{c}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2\tilde{z}_3\\
&\quad+\int_{c_u-z_{1-\frac{\alpha-\gamma}{2}}}^{\infty}\int_{\tilde{z}_3-c_u-\tilde{c}}^{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}\Phi\left(\frac{\tilde{z}_2+\tilde{c}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2\tilde{z}_3\\
&\quad+\int_{-\infty}^{c_u+z_{1-\frac{\alpha-\gamma}{2}}}\int_{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}^{\infty}\Phi\left(\frac{z_{1-\frac{\alpha-\gamma}{2}}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2\tilde{z}_3\\
&\quad-\int_{-\infty}^{c_u-z_{1-\frac{\alpha-\gamma}{2}}}\int_{-z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}^{\infty}\Phi\left(\frac{-z_{1-\frac{\alpha-\gamma}{2}}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2\tilde{z}_3\\
&\quad-\int_{c_u-z_{1-\frac{\alpha-\gamma}{2}}}^{c_u+z_{1-\frac{\alpha-\gamma}{2}}}\int_{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}^{\infty}\Phi\left(\frac{\tilde{z}_3-c_u-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2\tilde{z}_3
\end{aligned}$$

$$-\int_{c_u-z_{1-\frac{\alpha-\gamma}{2}}}^{\infty}\int_{\tilde{z}_3-c_u-\tilde{c}}^{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}\Phi\left(\frac{\tilde{z}_3-c_u-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2\tilde{z}_3-(1-\alpha),$$

where $g(\cdot)$ denotes the probability density function of $(\tilde{Z}_2,\tilde{Z}_3)$, $\mu(\tilde{z}_2,\tilde{z}_3)=(\omega_{12},\omega_{13})\Sigma_{22}^{-1}(\tilde{z}_2,\tilde{z}_3)'$ and $\sigma(\tilde{\omega})=\sqrt{1-(\omega_{12},\omega_{13})\Sigma_{22}^{-1}(\omega_{12},\omega_{13})'}$ with

$$\Sigma_{22}=\begin{pmatrix}\omega_{12} & \omega_{23}\\ \omega_{23} & \omega_{13}\end{pmatrix}.$$

This function is clearly continuously differentiable. In addition,

$$\begin{aligned}\frac{\partial f(\tilde{c},c_u,\tilde{\omega})}{\partial\tilde{c}}=&\int_{c_u+z_{1-\frac{\alpha-\gamma}{2}}}^{\infty}\left[\Phi\left(\frac{\tilde{z}_3-c_u-\mu(z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c},\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)\right.\\&\left.-\Phi\left(\frac{z_{1-\frac{\alpha-\gamma}{2}}-\mu(z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c},\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)\right]g(z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c},\tilde{z}_3)d\tilde{z}_3\\&+\int_{-\infty}^{c_u-z_{1-\frac{\alpha-\gamma}{2}}}\int_{-z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}^{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}\frac{1}{\sigma(\tilde{\omega})}\phi\left(\frac{\tilde{z}_2+\tilde{c}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3\\&+\int_{c_u-z_{1-\frac{\alpha-\gamma}{2}}}^{\infty}\int_{\tilde{z}_3-c_u-\tilde{c}}^{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}\frac{1}{\sigma(\tilde{\omega})}\phi\left(\frac{\tilde{z}_2+\tilde{c}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3>0\end{aligned}$$

for all $(\tilde{c},c_u,\tilde{\omega})\in\mathbb{R}\times\widetilde{\mathcal{C}}$ with $\omega_{12},\omega_{13}\neq 0$ since all three integrals are stricly positive.

Next, note that for any $(c_u,\tilde{\omega})\in\widetilde{\mathcal{C}}$ with $\omega_{12},\omega_{13}\neq 0$, there exists $\tilde{c}\in\mathbb{R}_\infty$ such that $f(\tilde{c},c_u,\tilde{\omega})=0$ since $f(\cdot,c_u,\tilde{\omega})$ is continuously strictly increasing,

$$\lim_{\tilde{c}\to-\infty}f(\tilde{c},c_u,\tilde{\omega})=-(1-\alpha)<0$$

and

$$\lim_{\tilde{c}\to\infty}f(\tilde{c},c_u,\tilde{\omega})=P(-\min\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{Z}_3+c_u\}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}})-(1-\alpha)\geq 0$$

by (7) and the fact that $P(-\min\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{Z}_3+c_u\}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}})$ is increasing in c_u.

Thus, the implicit function theorem implies that $\tilde{c}(c_u,\tilde{\omega})$ is continuous at all $(c_u,\tilde{\omega})\in\widetilde{\mathcal{C}}$ with $\omega_{12},\omega_{13}\neq 0$ and is therefore continuous at all $(c_u,\tilde{\omega})\in\widetilde{\mathcal{C}}$ by the definition of $\tilde{c}(c_u,\tilde{\omega})$ at $(c_u,\tilde{\omega})\in\widetilde{\mathcal{C}}$ with $\omega_{12}=0$ or $\omega_{13}=0$. ■

Lemma 2

For any $(c_u,\tilde{\omega})\in\widetilde{\mathcal{C}}$ *with* $\omega_{12}=0$ *or* $\omega_{13}=0$,

$$P(-\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3+c_u\}\leq Z_1\leq\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2+\tilde{c}(c_u,\tilde{\omega})\})=1-\alpha.$$

Proof: By (8),

$$P(-\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3+c_u\}\leq Z_1\leq\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2+\tilde{c}(c_u,\tilde{\omega})\})=1-\alpha$$

for all $(c_u,\tilde{\omega})\in\widetilde{\mathcal{C}}$ with $\omega_{12},\omega_{13}\neq 0$. Since the probability on the left hand side of this equality is continuous in $\tilde{\omega}$ by Lemma 1 and the continuity of the density function of $(Z_1,\widetilde{Z}_2,\widetilde{Z}_3)$ in $\tilde{\omega}$, the result immediately follows. ■

Proof of Proposition 2: Very similar arguments to those given in the proof of Lemma 1 provide that $\underline{c_u}:\bar{\mathcal{S}}\to\mathbb{R}$ exists and is continuous. Thus, $[\underline{c_u}(\cdot),\infty]$ is nonempty, compact-valued and continuous when treated as a correspondence from $\bar{\mathcal{S}}$ into $\mathbb{R}_\infty$ (see above).

Next, note that the minimand in (9) is

$$\begin{aligned}
&E[\max\{\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2+\tilde{c}(c_u,\tilde{\omega})\}+\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3+c_u\},0\}]\\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\max\{\min\{z_{1-(\alpha-\gamma)/2},\tilde{z}_2+\tilde{c}(c_u,\tilde{\omega})\}+\min\{z_{1-(\alpha-\gamma)/2},-\tilde{z}_3+c_u\},0\}g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3\\
&=\int_{-\infty}^{c_u-z_{1-(\alpha-\gamma)/2}}\int_{-z_{1-(\alpha-\gamma)/2}-\tilde{c}(c_u,\tilde{\omega})}^{z_{1-(\alpha-\gamma)/2}-\tilde{c}(c_u,\tilde{\omega})}(\tilde{z}_2+\tilde{c}(c_u,\tilde{\omega})+z_{1-(\alpha-\gamma)/2})g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3\\
&+\int_{c_u-z_{1-(\alpha-\gamma)/2}}^{\infty}\int_{\tilde{z}_3-\tilde{c}(c_u,\tilde{\omega})-c_u}^{z_{1-(\alpha-\gamma)/2}-\tilde{c}(c_u,\tilde{\omega})}(\tilde{z}_2-\tilde{z}_3+\tilde{c}(c_u,\tilde{\omega})+c_u)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3\\
&+\int_{-\infty}^{c_u-z_{1-(\alpha-\gamma)/2}}\int_{z_{1-(\alpha-\gamma)/2}-\tilde{c}(c_u,\tilde{\omega})}^{\infty}2z_{1-(\alpha-\gamma)/2}g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3\\
&+\int_{c_u-z_{1-(\alpha-\gamma)/2}}^{c_u+z_{1-(\alpha-\gamma)/2}}\int_{z_{1-(\alpha-\gamma)/2}-\tilde{c}(c_u,\tilde{\omega})}^{\infty}(z_{1-(\alpha-\gamma)/2}-\tilde{z}_3+c_u)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3,
\end{aligned}$$

where $g(\cdot)$ denotes the probability density function of $(\tilde{Z}_2,\tilde{Z}_3)$. When treated as a function from $\widetilde{\mathcal{C}}$ into $\mathbb{R}_+$, this expression is clearly continuous in $(c_u,\tilde{\omega})$ since $\tilde{c}:\widetilde{\mathcal{C}}\to\mathbb{R}$ is continuous by Lemma 1. The maximum theorem then implies the statement of the proposition. ■

Proof of Proposition 3: Let $(Z_1^*,\tilde{Z}_2^*,\tilde{Z}_3^*)'=(-Z_1,-\tilde{Z}_3,-\tilde{Z}_2)'$ and note that

$$P(-\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3+c_u(\tilde{\omega}_{12},\tilde{\omega}_{13},\tilde{\omega}_{23})\}\leq Z_1\leq\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2+c_\ell(\tilde{\omega}_{12},\tilde{\omega}_{13},\tilde{\omega}_{23})\})=1-\alpha$$

and

$$E[\max\{\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2+c_\ell(\tilde{\omega}_{12},\tilde{\omega}_{13},\tilde{\omega}_{23})\}+\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3+c_u(\tilde{\omega}_{12},\tilde{\omega}_{13},\tilde{\omega}_{23})\},0\}]$$

are equivalent to

$$P(-\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3^*+c_\ell(\tilde{\omega}_{12},\tilde{\omega}_{13},\tilde{\omega}_{23})\}\leq Z_1^*\leq\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2^*+c_u(\tilde{\omega}_{12},\tilde{\omega}_{13},\tilde{\omega}_{23})\})=1-\alpha$$

and

$$E[\max\{\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2^*+c_u(\tilde{\omega}_{12},\tilde{\omega}_{13},\tilde{\omega}_{23})\}+\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3^*+c_\ell(\tilde{\omega}_{12},\tilde{\omega}_{13},\tilde{\omega}_{23})\},0\}].$$

The result then follows by noting that $(Z_1^*,\tilde{Z}_2^*,\tilde{Z}_3^*)'\sim(Z_1,\tilde{Z}_3,\tilde{Z}_2)'$. ■

Proof of Theorem 1:

First, we provide the proof of the statement of the theorem for $CI_{u,n}(\cdot)$. Under (i)–(ii), standard subsequencing arguments in the uniform inference literature (see e.g., Andrews and Guggenberger, 2010; McCloskey, 2017; Andrews et al., 2020) provide that

$$\liminf_{n\to\infty}\inf_{\lambda\in\Lambda}P_\lambda\left(b\in CI_{u,n}(\hat{b}_n,\hat{d}_n;\widehat{\Sigma}_n)\right)=\lim_{n\to\infty}P_{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(b_{k_n,\mathfrak{b}}\in CI_{u,k_n}(\hat{b}_{k_n},\hat{d}_{k_n};\widehat{\Sigma}_{k_n})\right) \tag{19}$$

for a subsequence $\{k_n:n\geq 1\}$ of $\{n:n\geq 1\}$ such that $\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}\in\Lambda$ for all $n\geq 1$, $\sqrt{k_n}(b_{k_n,\mathfrak{b}},d_{k_n,\mathfrak{d}})\to(\mathfrak{b},\mathfrak{d})$ and $\Sigma_{k_n,\Sigma^*}\to\Sigma^*$ for some $(\mathfrak{b},\mathfrak{d},\Sigma^*)\in\mathbb{R}_\infty\times\mathbb{R}^k_{+,\infty}\times\Phi$ with $\lambda_{\min}(\Sigma^*)\geq\kappa$ and $\lambda_{\max}(\Sigma^*)\leq\kappa^{-1}$.

Let $\Omega^*=\mathrm{Diag}(\Sigma^*)^{-1/2}\Sigma^*\mathrm{Diag}(\Sigma^*)^{-1/2}$ and $s^*(\Omega^*)$ denote the subset of the set of indices $\{1,...,k\}$ that maximizes $\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\Omega^*_{d^{(s)}b}$ amongst all subsets of indices $s\subset\{1,...,k\}$ such that the elements of $\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}$ are non-negative. Since $\widehat{\Omega}_{k_n}\xrightarrow{p}\Omega^*$ under $\{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}:n\geq 1\}$ by (ii), (iii) and (v), note that $s^*(\widehat{\Omega}_{k_n})\xrightarrow{p}s^*(\Omega^*)$ under $\{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}:n\geq 1\}$. Thus, (19) implies

$$\liminf_{n\to\infty}\inf_{\lambda\in\Lambda}P_\lambda\left(b\in CI_{u,n}(\hat{b}_n,\hat{d}_n;\widehat{\Sigma}_n)\right)=\lim_{n\to\infty}P_{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(b_{k_n,\mathfrak{b}}\geq\hat{b}_{k_n}-\frac{\sqrt{\widehat{\Sigma}_{k_n,bb}}}{\sqrt{k_n}}\min\{z_{1-\alpha+\gamma},\right.$$
$$\left.\widehat{\Omega}_{k_n,bd^{(s^*)}}\widehat{\Omega}^{-1}_{k_n,d^{(s^*)}d^{(s^*)}}\mathrm{Diag}(\widehat{\Sigma}_{k_n,d^{(s^*)}d^{(s^*)}})^{-1/2}\sqrt{k_n}\hat{d}^{(s^*)}_{k_n}+c\left(\widehat{\Omega}_{k_n,bd^{(s^*)}}\widehat{\Omega}^{-1}_{k_n,d^{(s^*)}d^{(s^*)}}\widehat{\Omega}_{k_n,d^{(s^*)}b}\right)\}\right)$$
$$=\lim_{n\to\infty}P_{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(\frac{\sqrt{k_n}(\hat{b}_{k_n}-b_{k_n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{k_n,bb}}}\leq\min\{z_{1-\alpha+\gamma},\right.$$

$$
\widehat{\Omega}_{k_n,bd^{(s^*)}}\widehat{\Omega}^{-1}_{k_n,d^{(s^*)}d^{(s^*)}}\mathrm{Diag}(\widehat{\Sigma}_{k_n,d^{(s^*)}d^{(s^*)}})^{-1/2}\sqrt{k_n}\hat{d}^{(s^*)}_{k_n}+c\Big(\widehat{\Omega}_{k_n,bd^{(s^*)}}\widehat{\Omega}^{-1}_{k_n,d^{(s^*)}d^{(s^*)}}\widehat{\Omega}_{k_n,d^{(s^*)}b}\Big)\Big\}\Big)
$$

$$
=\begin{cases} P\Big(Z_1\leq\min\Big\{z_{1-\alpha+\gamma},\Omega^*_{bd^{(s^*)}}\Omega^{-1}_{d^{(s^*)}d^{(s^*)}}Y^{(s^*)}_\delta+c\big(\Omega^*_{bd^{(s^*)}}\Omega^{*-1}_{d^{(s^*)}d^{(s^*)}}\Omega^*_{d^{(s^*)}b}\big)\Big\}\Big) & \text{if } \|\mathfrak{d}^{(s^*)}\|<\infty, \\ P(Z_1\leq z_{1-\alpha+\gamma}) & \text{if } \|\mathfrak{d}^{(s^*)}\|=\infty \end{cases} \tag{20}
$$

by (ii)–(v) and Proposition 1, where we use s^* as shorthand for $s^*(\Omega^*)$ and

$$
\begin{pmatrix} Z_1 \\ Y^{(s^*)}_\delta \end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix} 0 \\ \delta^{(s^*)} \end{pmatrix},\begin{pmatrix} 1 & \Omega_{bd^{(s^*)}} \\ \Omega_{d^{(s^*)}b} & \Omega_{d^{(s^*)}d^{(s^*)}} \end{pmatrix}\right)
$$

with $\delta^{(s^*)}=\mathrm{Diag}(\Sigma^*_{d^{(s^*)}d^{(s^*)}})^{-1/2}\mathfrak{d}^{(s^*)}$. Now for the $\|\mathfrak{d}^{(s^*)}\|<\infty$ case, since $\Omega^*_{bd^{(s^*)}}\Omega^{-1}_{d^{(s^*)}d^{(s^*)}}\delta^{(s^*)}\geq 0$ by (i), (15) and (v),

$$
\begin{aligned}
&P\Big(Z_1\leq\min\Big\{z_{1-\alpha+\gamma},\Omega^*_{bd^{(s^*)}}\Omega^{-1}_{d^{(s^*)}d^{(s^*)}}Y^{(s^*)}_\delta+c\big(\Omega^*_{bd^{(s^*)}}\Omega^{*-1}_{d^{(s^*)}d^{(s^*)}}\Omega^*_{d^{(s^*)}b}\big)\Big\}\Big)\\
&=P\Big(Z_1\leq\min\Big\{z_{1-\alpha+\gamma},\Omega^*_{bd^{(s^*)}}\Omega^{-1}_{d^{(s^*)}d^{(s^*)}}\delta^{(s^*)}+\tilde{Z}_2+c\big(\Omega^*_{bd^{(s^*)}}\Omega^{*-1}_{d^{(s^*)}d^{(s^*)}}\Omega^*_{d^{(s^*)}b}\big)\Big\}\Big)\\
&\geq P\Big(Z_1\leq\min\Big\{z_{1-\alpha+\gamma},\tilde{Z}_2+c\big(\Omega^*_{bd^{(s^*)}}\Omega^{*-1}_{d^{(s^*)}d^{(s^*)}}\Omega^*_{d^{(s^*)}b}\big)\Big\}\Big)=1-\alpha
\end{aligned} \tag{21}
$$

by the definition of $c(\cdot)$ in (5), where

$$
\begin{pmatrix} Z_1 \\ \tilde{Z}_2 \end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix},\begin{pmatrix} 1 & \Omega^*_{bd^{(s^*)}}\Omega^{*-1}_{d^{(s^*)}d^{(s^*)}}\Omega^*_{d^{(s^*)}b} \\ \Omega^*_{bd^{(s^*)}}\Omega^{*-1}_{d^{(s^*)}d^{(s^*)}}\Omega^*_{d^{(s^*)}b} & \Omega^*_{bd^{(s^*)}}\Omega^{*-1}_{d^{(s^*)}d^{(s^*)}}\Omega^*_{d^{(s^*)}b} \end{pmatrix}\right).
$$

On the other hand, for the $\|\mathfrak{d}^{(s^*)}\|=\infty$ case,

$$
P(Z_1\leq z_{1-\alpha+\gamma})=1-\alpha+\gamma>1-\alpha. \tag{22}
$$

Together, (20)–(22) yield the lower bound in the statement of the theorem for $CI_{u,n}(\cdot)$.

To prove the upper bound, note that by nearly identical arguments to those used to establish (19),

$$
\limsup_{n\to\infty}\sup_{\lambda\in\Lambda}P_\lambda\Big(b\in CI_{u,n}(\hat{b}_n,\hat{d}_n;\widehat{\Sigma}_n)\Big)=\lim_{n\to\infty}P_{\lambda_{m_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Big(b_{m_n,\mathfrak{b}}\in CI_{u,m_n}(\hat{b}_{m_n},\hat{d}_{m_n};\widehat{\Sigma}_{m_n})\Big)
$$

for a subsequence $\{m_n:n\geq 1\}$ of $\{n:n\geq 1\}$ such that $\lambda_{m_n,\mathfrak{b},\mathfrak{d},\Sigma^*}\in\Lambda$ for all $n\geq 1$, $\sqrt{m_n}(b_{m_n,\mathfrak{b}},d_{m_n,\mathfrak{d}})\to(\mathfrak{b},\mathfrak{d})$ and $\Sigma_{m_n,\Sigma^*}\to\Sigma^*$ for some $(\mathfrak{b},\mathfrak{d},\Sigma^*)\in\mathbb{R}_\infty\times\mathbb{R}^k_{+,\infty}\times\Phi$ with

$\lambda_{\min}(\Sigma^*) \geq \kappa$ and $\lambda_{\max}(\Sigma^*) \leq \kappa^{-1}$. Note that for the probability to the left of the inequality in (21),

$$P\Big(Z_1 \leq \min\Big\{z_{1-\alpha+\gamma}, \Omega^*_{bd^{(s^*)}} \Omega^{-1}_{d^{(s^*)}d^{(s^*)}} \delta^{(s^*)} + \tilde{Z}_2 + c\big(\Omega^*_{bd^{(s^*)}} \Omega^{*-1}_{d^{(s^*)}d^{(s^*)}} \Omega^*_{d^{(s^*)}b}\big)\Big\}\Big)$$
$$\leq P(Z_1 \leq z_{1-\alpha+\gamma}) = 1-\alpha+\gamma.$$

Then, nearly identical reasoning used to establish (20)–(22), replacing "$\liminf_{n\to\infty}\inf_{\lambda\in\Lambda}$" with "$\limsup_{n\to\infty}\sup_{\lambda\in\Lambda}$" and the subsequences $\{k_n : n\geq 1\}$ and $\{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*} \in \Lambda : n \geq 1\}$ with $\{\lambda_{m_n,\mathfrak{b},\mathfrak{d},\Sigma^*} \in \Lambda : n \geq 1\}$, yields the upper bound in the statement of the theorem for $CI_{u,n}(\cdot)$.

Next, we provide the proof of the statement of the theorem for $CI_{t,n}(\cdot)$. The same arguments used to establish (19) also apply to $CI_{t,n}(\hat{b}_n, \hat{d}_n; \widehat{\Sigma}_n)$ so that it suffices to consider $\lim_{n\to\infty} P_{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}(b_{k_n,\mathfrak{b}} \in CI_{T,k_n}(\hat{b}_{k_n}, \hat{d}_{k_n}; \widehat{\Sigma}_{k_n}))$ under the same subsequences $\{k_n : n \geq 1\}$ of $\{n : n \geq 1\}$ and sequences of parameters $\{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*} : n \geq 1\}$ described in the proof of the statement of the theorem for $CI_{u,n}(\cdot)$.

Using analogous notation and reasoning to the proof above,

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{\lambda\in\Lambda} P_\lambda\Big(b \in CI_{t,n}(\hat{b}_n, \hat{d}_n; \widehat{\Sigma}_n)\Big) = \lim_{n\to\infty} P_{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Bigg(\hat{b}_{k_n} - \frac{\sqrt{\widehat{\Sigma}_{k_n,bb}}}{\sqrt{k_n}} \min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\\
&\quad \widehat{\Omega}_{k_n,bd^{(s_1^*)}} \widehat{\Omega}^{-1}_{k_n,d^{(s_1^*)}d^{(s_1^*)}} \mathrm{Diag}(\widehat{\Sigma}_{k_n,d^{(s_1^*)}d^{(s_1^*)}})^{-1/2} \sqrt{k_n} \hat{d}^{(s_1^*)}_{k_n} + c_\ell\Big(\widehat{\tilde{\Omega}}^{(s_1^*,s_2^*)}_{k_n}\Big)\Big\} \leq b_{k_n,\mathfrak{b}} \leq \hat{b}_{k_n}\\
&+ \frac{\sqrt{\widehat{\Sigma}_{k_n,bb}}}{\sqrt{k_n}} \min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, -\widehat{\Omega}_{k_n,bd^{(s_2^*)}} \widehat{\Omega}^{-1}_{k_n,d^{(s_2^*)}d^{(s_2^*)}} \mathrm{Diag}(\widehat{\Sigma}_{k_n,d^{(s_2^*)}d^{(s_2^*)}})^{-1/2} \sqrt{k_n} \hat{d}^{(s_2^*)}_{k_n} + c_u\Big(\widehat{\tilde{\Omega}}^{(s_1^*,s_2^*)}_{k_n}\Big)\Big\}\Bigg)\\
&= \lim_{n\to\infty} P_{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Bigg(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, -\widehat{\Omega}_{k_n,bd^{(s_2^*)}} \widehat{\Omega}^{-1}_{k_n,d^{(s_2^*)}d^{(s_2^*)}} \mathrm{Diag}(\widehat{\Sigma}_{k_n,d^{(s_2^*)}d^{(s_2^*)}})^{-1/2} \sqrt{k_n} \hat{d}^{(s_2^*)}_{k_n} + c_u\Big(\widehat{\tilde{\Omega}}^{(s_1^*,s_2^*)}_{k_n}\Big)\Big\}\\
&\quad \leq \frac{\sqrt{k_n}(\hat{b}_{k_n} - b_{k_n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{k_n,bb}}}\\
&\quad \leq \min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, \widehat{\Omega}_{k_n,bd^{(s_1^*)}} \widehat{\Omega}^{-1}_{k_n,d^{(s_1^*)}d^{(s_1^*)}} \mathrm{Diag}(\widehat{\Sigma}_{k_n,d^{(s_1^*)}d^{(s_1^*)}})^{-1/2} \sqrt{k_n} \hat{d}^{(s_1^*)}_{k_n} + \tilde{c}\Big(c_u\Big(\widehat{\tilde{\Omega}}^{(s_1^*,s_2^*)}_{k_n}\Big), \widehat{\tilde{\Omega}}^{(s_1^*,s_2^*)}_{k_n}\Big)\Big\}\Bigg).
\end{aligned} \tag{23}$$

Since $\widehat{\tilde{\Omega}}^{(s_1^*,s_2^*)}_{k_n} \xrightarrow{p} \tilde{\Omega}^{*(s_1^*,s_2^*)}$ under $\{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*} : n \geq 1\}$ as $k_n \to \infty$ by (ii), (iii) and (v), there exists a subsequence $\{l_n : n \geq 1\}$ of $\{k_n : n \geq 1\}$ such that $\widehat{\tilde{\Omega}}^{(s_1^*,s_2^*)}_{l_n} \xrightarrow{a.s.} \tilde{\Omega}^{*(s_1^*,s_2^*)}$ under $\{\lambda_{l_n,\mathfrak{b},\mathfrak{d},\Sigma^*} : n \geq 1\}$ as $l_n \to \infty$. Next, by the properties of $\tilde{c}_u : \bar{\mathcal{S}} \rightrightarrows \mathbb{R}$ given in Proposition

2, there exists a subsequence $\{h_n : n \geq 1\}$ of $\{l_n : n \geq 1\}$ for which the subsequence $\left\{c_u\left(\widehat{\tilde{\Omega}}_{h_n}^{(s_1^*,s_2^*)}\right):n\geq 1\right\}$ of $\left\{c_u\left(\widehat{\tilde{\Omega}}_{l_n}^{(s_1^*,s_2^*)}\right):n\geq 1\right\}$ is such that $c_u\left(\widehat{\tilde{\Omega}}_{h_n}^{(s_1^*,s_2^*)}\right) \in \tilde{c}_u\left(\widehat{\tilde{\Omega}}_{h_n}^{(s_1^*,s_2^*)}\right)$ for all $n\geq 1$ and $c_u\left(\widehat{\tilde{\Omega}}_{h_n}^{(s_1^*,s_2^*)}\right) \xrightarrow{a.s.} c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)$ for some $c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right) \in \tilde{c}_u\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)$ as $n\to\infty$. In conjunction with Lemma 1, this implies that (23) is equal to

$$\begin{aligned}\lim_{n\to\infty} P_{\lambda_{h_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Bigg(&-\min\left\{z_{1-\frac{\alpha-\gamma}{2}},-\widehat{\Omega}_{h_n,bd^{(s_2^*)}}\widehat{\Omega}^{-1}_{h_n,d^{(s_2^*)}d^{(s_2^*)}}\mathrm{Diag}(\widehat{\Sigma}_{h_n,d^{(s_2^*)}d^{(s_2^*)}})^{-1/2}\sqrt{h_n}\hat{d}_{h_n}^{(s_2^*)}+c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\\
&\leq \frac{\sqrt{h_n}(\hat{b}_{h_n}-b_{h_n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{h_n,bb}}}\\
&\leq \min\left\{z_{1-\frac{\alpha-\gamma}{2}},\widehat{\Omega}_{h_n,bd^{(s_1^*)}}\widehat{\Omega}^{-1}_{h_n,d^{(s_1^*)}d^{(s_1^*)}}\mathrm{Diag}(\widehat{\Sigma}_{h_n,d^{(s_1^*)}d^{(s_1^*)}})^{-1/2}\sqrt{h_n}\hat{d}_{h_n}^{(s_1^*)}+\tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right),\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\Bigg).\end{aligned} \tag{24}$$

If $\|\mathfrak{d}^{(s_1^*)}\|,\|\mathfrak{d}^{(s_2^*)}\| < \infty$, since $\Omega^*_{bd^{(s_1^*)}}\Omega^{*-1}_{d^{(s_1^*)}d^{(s_1^*)}}\delta^{(s_1^*)} \geq 0$ and $\Omega^*_{bd^{(s_2^*)}}\Omega^{*-1}_{d^{(s_2^*)}d^{(s_2^*)}}\delta^{(s_2^*)} \leq 0$ by (i)–(v) and (15), (24) is equal to

$$\begin{aligned}&P\Bigg(-\min\left\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega^*_{bd^{(s_2^*)}}\Omega^{*-1}_{d^{(s_2^*)}d^{(s_2^*)}}Y_\delta^{(s_2^*)}+c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\\
&\quad\leq Z_1\leq \min\left\{z_{1-\frac{\alpha-\gamma}{2}},\Omega^*_{bd^{(s_1^*)}}\Omega^{*-1}_{d^{(s_1^*)}d^{(s_1^*)}}Y_\delta^{(s_1^*)}+\tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right),\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\Bigg)\\
&=P\Bigg(-\min\left\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega^*_{bd^{(s_2^*)}}\Omega^{*-1}_{d^{(s_2^*)}d^{(s_2^*)}}\delta^{(s_2^*)}-\tilde{Z}_3+c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\\
&\quad\leq Z_1\leq \min\left\{z_{1-\frac{\alpha-\gamma}{2}},\Omega^*_{bd^{(s_1^*)}}\Omega^{*-1}_{d^{(s_1^*)}d^{(s_1^*)}}\delta^{(s_1^*)}+\tilde{Z}_2+\tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right),\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\Bigg)\\
&\geq P\Bigg(-\min\left\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{Z}_3+c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\\
&\quad\leq Z_1\leq \min\left\{z_{1-\frac{\alpha-\gamma}{2}},\tilde{Z}_2+\tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right),\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\Bigg)=1-\alpha\end{aligned} \tag{25}$$

by the definition of $\tilde{c}(\cdot)$ in (8) and Lemma 2, where

$$\begin{pmatrix} Z_1\\ \tilde{Z}_2\\ \tilde{Z}_3\end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix},\begin{pmatrix}1 & \Omega^*_{\beta\delta^{(s_1^*)}} & \Omega^*_{\beta\delta^{(s_2^*)}}\\ \Omega^*_{\delta^{(s_1^*)}\beta} & \Omega^*_{\delta^{(s_1^*)}\delta^{(s_1^*)}} & \Omega^*_{\delta^{(s_1^*)}\delta^{(s_2^*)}}\\ \Omega^*_{\delta^{(s_2^*)}\beta} & \Omega^*_{\delta^{(s_2^*)}\delta^{(s_1^*)}} & \Omega^*_{\delta^{(s_2^*)}\delta^{(s_2^*)}}\end{pmatrix}\right).$$

If $\|\mathfrak{d}^{(s_1^*)}\|<\infty$, $\|\mathfrak{d}^{(s_2^*)}\|=\infty$, since $\Omega^*_{bd^{(s_1^*)}}\Omega^{*-1}_{d^{(s_1^*)}d^{(s_1^*)}}\delta^{(s_1^*)}\geq 0$ by (i)–(v) and (15), (24) is equal to

$$P\left(-z_{1-\frac{\alpha-\gamma}{2}}\leq Z_1\leq \min\left\{z_{1-\frac{\alpha-\gamma}{2}},\Omega^*_{bd^{(s_1^*)}}\Omega^{*-1}_{d^{(s_1^*)}d^{(s_1^*)}}Y_\delta^{(s_1^*)}+\tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1^*,s_2^*)}\right),\tilde{\Omega}^{*(s_1^*,s_2^*)}\right)\right\}\right)$$

$$\geq P\Big(-z_{1-\frac{\alpha-\gamma}{2}} \leq Z_1 \leq \min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, \tilde{Z}_2 + \tilde{c}\Big(c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big), \tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}\Big)$$

$$\geq P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, -\tilde{Z}_3 + c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\} \leq Z_1 \leq \min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, \tilde{Z}_2 + \tilde{c}\Big(c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big), \tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}\Big) = 1-\alpha \quad (26)$$

by the definition of $\tilde{c}(\cdot)$ in (8) and Lemma 2. If $\|\mathfrak{d}^{(s_1^*)}\| = \infty$, $\|\mathfrak{d}^{(s_2^*)}\| < \infty$, since $\Omega^*_{bd^{(s_2^*)}}\Omega^{*-1}_{d^{(s_2^*)}d^{(s_2^*)}}\delta^{(s_2^*)} \leq 0$ by (i)–(v) and (15), (24) is equal to

$$P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, -\Omega^*_{bd^{(s_2^*)}}\Omega^{*-1}_{d^{(s_2^*)}d^{(s_2^*)}}Y_\delta^{(s_2^*)} + c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\} \leq Z_1 \leq z_{1-\frac{\alpha-\gamma}{2}}\Big)$$

$$\geq P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, -\tilde{Z}_3 + c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\} \leq Z_1 \leq z_{1-\frac{\alpha-\gamma}{2}}\Big)$$

$$\geq P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, -\tilde{Z}_3 + c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\} \leq Z_1 \leq \min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, \tilde{Z}_2 + \tilde{c}\Big(c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big), \tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}\Big) = 1-\alpha \quad (27)$$

by the definition of $\tilde{c}(\cdot)$ in (8) and Lemma 2. Finally, if $\|\mathfrak{d}^{(s_1^*)}\|, \|\mathfrak{d}^{(s_2^*)}\| = \infty$, (24) is equal to

$$P\Big(z_{1-\frac{\alpha-\gamma}{2}} \leq Z_1 \leq z_{1-\frac{\alpha-\gamma}{2}}\Big) = 1-\alpha+\gamma > 1-\alpha. \quad (28)$$

by (ii)-(v). Together, (23)–(28) yield the lower bound in the statement of the theorem for $CI_{t,n}(\cdot)$.

To prove the upper bound, note that by nearly identical arguments to those used to establish (24),

$$\limsup_{n\to\infty}\sup_{\lambda\in\Lambda} P_\lambda\Big(b \in CI_{t,n}(\hat{b}_n, \hat{d}_n; \widehat{\Sigma}_n)\Big) =$$
$$\lim_{n\to\infty} P_{\lambda_{m_n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, -\widehat{\Omega}_{m_n,bd^{(s_2^*)}}\widehat{\Omega}^{-1}_{m_n,d^{(s_2^*)}d^{(s_2^*)}}\mathrm{Diag}(\widehat{\Sigma}_{m_n,d^{(s_2^*)}d^{(s_2^*)}})^{-1/2}\sqrt{m_n}\hat{d}_{m_n}^{(s_2^*)} + c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}$$
$$\leq \frac{\sqrt{m_n}(\hat{b}_{m_n} - b_{m_n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{m_n,bb}}}$$
$$\leq \min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, \widehat{\Omega}_{m_n,bd^{(s_1^*)}}\widehat{\Omega}^{-1}_{m_n,d^{(s_1^*)}d^{(s_1^*)}}\mathrm{Diag}(\widehat{\Sigma}_{m_n,d^{(s_1^*)}d^{(s_1^*)}})^{-1/2}\sqrt{m_n}\hat{d}_{m_n}^{(s_1^*)} + \tilde{c}\Big(c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big), \tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}\Big)$$

for a subsequence $\{m_n : n \geq 1\}$ of $\{n : n \geq 1\}$ such that $\lambda_{m_n,\mathfrak{b},\mathfrak{d},\Sigma^*} \in \Lambda$ for all $n \geq 1$, $\sqrt{m_n}(b_{m_n,\mathfrak{b}}, d_{m_n,\mathfrak{d}}) \to (\mathfrak{b},\mathfrak{d})$ and $\Sigma_{m_n,\Sigma^*} \to \Sigma^*$ for some $(\mathfrak{b},\mathfrak{d},\Sigma^*) \in \mathbb{R}_\infty \times \mathbb{R}^k_{+,\infty} \times \Phi$ with $\lambda_{\min}(\Sigma^*) \geq \kappa$ and $\lambda_{\max}(\Sigma^*) \leq \kappa^{-1}$ and some $c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big) \in \tilde{c}_u\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)$. Note that for the probability to the left of the inequality in (25),

$$P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega^*_{bd^{(s_2^*)}}\Omega^{*-1}_{d^{(s_2^*)}d^{(s_2^*)}}Y_\delta^{(s_2^*)}+c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}$$
$$\leq Z_1\leq\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\Omega^*_{bd^{(s_1^*)}}\Omega^{*-1}_{d^{(s_1^*)}d^{(s_1^*)}}Y_\delta^{(s_1^*)}+\tilde{c}\Big(c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big),\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}\Big)$$
$$\leq P\Big(-z_{1-\frac{\alpha-\gamma}{2}}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}}\Big)=1-\alpha+\gamma;$$

for the probability to the left of the first inequality in (26),

$$P\Big(-z_{1-\frac{\alpha-\gamma}{2}}\leq Z_1\leq\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\Omega^*_{bd^{(s_1^*)}}\Omega^{*-1}_{d^{(s_1^*)}d^{(s_1^*)}}Y_\delta^{(s_1^*)}+\tilde{c}\Big(c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big),\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}\Big)$$
$$\leq P\Big(-z_{1-\frac{\alpha-\gamma}{2}}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}}\Big)=1-\alpha+\gamma;$$

and for the probability to the left of the first inequality in (27),

$$P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega^*_{bd^{(s_2^*)}}\Omega^{*-1}_{d^{(s_2^*)}d^{(s_2^*)}}Y_\delta^{(s_2^*)}+c_u^*\Big(\tilde{\Omega}^{*(s_1^*,s_2^*)}\Big)\Big\}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}}\Big)$$
$$\leq P\Big(-z_{1-\frac{\alpha-\gamma}{2}}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}}\Big)=1-\alpha+\gamma.$$

Then, nearly identical reasoning used to establish (23)–(28), replacing "$\liminf_{n\to\infty}\inf_{\lambda\in\Lambda}$" with "$\limsup_{n\to\infty}\sup_{\lambda\in\Lambda}$" and the subsequences $\{k_n:n\geq1\}$ and $\{\lambda_{k_n,\mathfrak{b},\mathfrak{d},\Sigma^*}\in\Lambda:n\geq1\}$ with $\{m_n:n\geq1\}$ and $\{\lambda_{m_n,\mathfrak{b},\mathfrak{d},\Sigma^*}\in\Lambda:n\geq1\}$, yields the upper bound in the statement of the theorem for $CI_{t,n}(\cdot)$. ■

B Parameter Space for the Standard Linear Regression Model

In this section, we provide details for parameter spaces satisfying (i)–(v) in Section 3.2 in the context of the standard linear regression model. Recall in this setting we are interested in conducting inference on a regression coefficient of interest b in the standard linear regression model for observations $i=1,...,n$

$$y_i=bz_i+x_i'd+w_i'c+\varepsilon_i,$$

where y_i is the dependent variable, z_i is the scalar regressor of interest, $x_i\in\mathbb{R}^{\mathcal{D}_x}$ are control variables with *known positive partial effects* $d\geq0$ on y_i, $w_i\in\mathbb{R}^{\mathcal{D}_w}$ are control variables with unrestricted partial effects c and ε_i is the error term.

Define $h_i=(z_i,x_i',w_i')'$ so that the ordinary least squares estimator of $(b,d')'$, $(\hat{b}_n,\hat{d}_n')'$, is equal to the first $\mathcal{D}_x+1$ entries of $(\sum_{i=1}^n h_ih_i')^{-1}\sum_{i=1}^n h_iy_i$. Let F denote the joint distribution of the stationary random vectors $\{(h_i',\varepsilon_i)':i\geq 1\}$ and define the parameter $\tilde{\lambda}=(b,d,c,\mathcal{V},Q,F)$. The parameter space $\widetilde{\Lambda}$ for $\tilde{\lambda}$ is defined to include parameters $\tilde{\lambda}=(b,d,c,\mathcal{V},Q,F)$ such that for some finite $\kappa>0$, the following conditions hold:

(i') $b\in\mathbb{R}$, $d\in\mathbb{R}_+^{\mathcal{D}_x}$ and $c\in\mathbb{R}^{\mathcal{D}_w}$;

(ii') $\lim_{n\to\infty}n^{-1}\sum_{i=1}^n\sum_{j=1}^n E_F[h_ih_j'\varepsilon_i\varepsilon_j]$ exists and equals $\mathcal{V}\in\Phi$ with $\lambda_{\max}(\mathcal{V})\leq\kappa^{-1}$;

(iii') $E_F[h_ih_i']$ exists and equals $Q\in\Phi$ with $\lambda_{\min}(Q)\geq\kappa$.

In addition, under any sequence of parameters $\{\tilde{\lambda}_{n,\mathfrak{b},\mathfrak{d},\mathcal{V}^*,Q^*}=(b_{n,\mathfrak{b}},d_{n,\mathfrak{d}},\mathcal{V}_{n,\mathcal{V}^*},Q_{n,Q^*},F_{n,\mathfrak{b},\mathfrak{d},\mathcal{V}^*,Q^*}):n\geq 1\}$ in $\widetilde{\Lambda}$ such that (15) holds,

$$\mathcal{V}_{n,\mathcal{V}^*}\to\mathcal{V}^* \tag{29}$$

and

$$Q_{n,Q^*}\to Q^* \tag{30}$$

for some $\mathcal{V}^*,Q^*\in\Phi$, the following remaining conditions hold:

(iv') $\widehat{\mathcal{V}}_n$ and $\widehat{Q}_n\equiv n^{-1}\sum_{i=1}^n h_ih_i'$ exist and $\lambda_{\min}(n^{-1}\sum_{i=1}^n h_ih_i')>0$ with probability one for all $n\geq 1$;

(v') $\widehat{\mathcal{V}}_n\xrightarrow{p}\mathcal{V}^*$;

(vi') $\widehat{Q}_n\xrightarrow{p}Q^*$;

(vii') $n^{-1/2}\sum_{i=1}^n h_i\varepsilon_i\xrightarrow{d}\mathcal{N}(0,\mathcal{V}^*)$;

(viii') for any sequence $\{\tilde{\lambda}_{n,\mathfrak{b},\mathfrak{d},\mathcal{V}^*,Q^*}\}$ in $\widetilde{\Lambda}$ and any subsequence $\{s_n:n\geq 1\}$ of $\{n:n\geq 1\}$ for which (15), (29)–(30) hold along the subsequence, conditions (iv')–(vii') also hold along the subsequence.

Note that for Σ equal to the upper left $(\mathcal{D}_x+1)\times(\mathcal{D}_x+1)$ submatrix of $Q^{-1}\mathcal{V}Q^{-1}$ and $\widehat{\Sigma}_n$ equal to the upper left $(\mathcal{D}_x+1)\times(\mathcal{D}_x+1)$ submatrix of $\widehat{Q}_n^{-1}\widehat{\mathcal{V}}_n\widehat{Q}_n^{-1}$, the conditions (i')–(viii') on the parameter space $\widetilde{\Lambda}$ imply (i)–(v) for the parameter space Λ. More specifically, (i') implies (i), (ii')–(iii') imply (ii), (iv')–(vi') imply (iii), (vi')–(vii') imply (iv) and (viii') implies (v).

In conjunction with a suitable choice of covariance matrix estimator $\widehat{\mathcal{V}}_n$, the above definition of the parameter space $\widetilde{\Lambda}$ effectively serves as a set of assumptions on the underlying DGP in the context of the standard linear regression model when using ordinary least squares for estimation. Part (i') imposes known sign restrictions for the nuisance coefficients d while letting the coefficient of interest b and the other nuisance coefficients c remain unrestricted. Parts (ii')–(iii') are standard conditions ensuring the existence of asymptotic covariance matrices while parts (iv')–(vi') are high level assumptions that guarantee consistent estimators of these covariance matrices are available, typically shown via application of a law of large numbers. Part (vii') is a high level assumption that directly assumes a central limit theorem holds for the product of the regressors and error term in the regression model, a result that is typically invoked when proving asymptotic normality of ordinary least squares estimators. Finally, part (viii') is a mild technical condition used to establish results under relevant drifting sequences of DGPs.

The definition of the parameter space $\widetilde{\Lambda}$ is written at such a level of generality to allow for heteroskedasticity and/or weak dependence in the data, enabling the use of our CIs in both cross-sectional and time series settings. We refer the interested reader to Section 3 of McCloskey (2020) for two sets of weak low-level sufficient conditions that guarantee the high-level assumptions (iv')–(vii') hold when using standard covariance matrix estimators in the context of estimation robust to heteroskedasticity for randomly sampled data and estimation robust to heteroskedasticity and autocorrelation for time series data.

References

Andrews, D. W. K., Cheng, X., and Guggenberger, P. (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics*, 218:496–531.

Andrews, D. W. K. and Guggenberger, P. (2009). Hybrid and size-corrected subsampling methods. *Econometrica*, 77:721–762.

Andrews, D. W. K. and Guggenberger, P. (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory*, 26:426–468.

Armstrong, T. (2015). Adaptive testing on a regression function at a point. *Annals of Statistics*, 43:2086–2101.

Armstrong, T. B. and Kolesár, M. (2018). Optimal inference in a class of regression models. *Econometrica*, 86:655–683.

Armstrong, T. B., Kolesár, M., and Kwon, S. (2020). Bias-aware inference in regularized regression models. Working Paper.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81:608–650.

Blattman, C., Jamison, J. C., and Sheridan, M. (2017). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia. *American Economic Review*, 107(4):1165–1206.

Cai, T. T. and Low, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Annals of Statistics*, 32:1805–1840.

Cai, T. T., Low, M. G., and Xia, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Annals of Statistics*, 41:722–750.

Elliott, G., Müller, U. K., and Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, 83:771–811.

Ketz, P. (2018). Subvector inference when the true parameter vector may be near or at the boundary. *Journal of Econometrics*, 207:285–306.

Kwon, K. and Kwon, S. (2020). Adaptive inference in multivariate nonparametric regression models under monotonicity. arXiv:2011.14219v1.

Lehmann, E. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, third edition edition.

Low, M. G. (1997). On nonparametric confidence intervals. *Annals of Statistics*, 25:2547–2554.

McCloskey, A. (2017). Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics*, 200:17–35.

McCloskey, A. (2020). Asymptotically uniform tests after consistent model selection in the linear regression model. *Journal of Business and Economic Statistics*, 38:810–825.

Moreira, H. and Moreira, M. J. (2013). Contributions to the theory of optimal tests. FGV Working Paper.

Muralidharan, K., Romero, M., and Wüthrich, K. (2020). Factorial designs, model selection, and (incorrect) inference in randomized experiments. NBER working paper.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.