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

Philipp Ketz† Adam McCloskey‡

October 12, 2022

Abstract

We introduce adaptive confidence intervals on a parameter of interest in the presence of nuisance parameters, such as coefficients on control variables, with known signs. Our confidence intervals are trivial to compute and can provide significant length reductions relative to standard ones when the nuisance parameters are small. At the same time, they entail minimal length increases at any parameter values. We apply our confidence intervals to the linear regression model, prove their uniform validity 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.

JEL Classification: C10, C12.

*We thank the editor Xiaoxia Shi, three anonymous referees, Timothy Armstrong, Christopher Blattman, Michal Kolesár, Koohyun Kwon, Soonwoo Kwon, Julian Jamison, Margaret Sheridan, Liam Wren-Lewis and Kaspar Wüthrich for helpful comments and suggestions and Chad Brown for excellent research assistance.

†Paris School of Economics - CNRS, 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 from 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.[1] 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 (asymptotic) 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 common practical situations for which 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). This includes regression models estimated by ordinary, generalized and two stage least squares as

[1]In such cases, it is often implicitly assumed that the regression model is equal to the conditional expectation function, which facilitates interpretation of the regression coefficients. We note, however, that even if that assumption is correct the coefficient on a control variable may not equal its causal effect. See the regression example in Section 3 for a discussion of this that includes several empirical examples where sign restrictions are plausible.

well as models with bounded parameter spaces such as (G)ARCH (see e.g., Bollerslev, 1986) and random coefficient models (see e.g., Berry et al., 1995; Andrews, 1999). Even though the standard "constrained" estimator is not normally distributed in large samples when the true parameter vector is at (or close to) the boundary of the parameter space (see e.g., Andrews, 1999), there often exists a "quasi-unconstrained" estimator that is (Ketz, 2018). While noting this generality, we mainly focus on regression models for ease of exposition.

To construct our CIs, we use 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. We show that a particular form of these latter CIs is expected excess length-optimal (among affine CIs—see Proposition 2 for details) when the corresponding control coefficients are equal to zero. It also has low expected excess length when the control coefficients are small but its expected excess length grows without bound as the control coefficients grow larger. On the other hand, we show that standard one-sided CIs based upon the regression including all controls have the minimal maximum expected excess length (among affine CIs) over the parameter space that imposes the sign of the control coefficients.[2] They also have correct coverage and expected excess length that does not depend upon the true values of the control coefficients. 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. Lending further theoretical support to the use of our CIs, we also show numerically that both our one- and two-sided CIs imply "nearly optimal" (NO) hypothesis tests in the sense of Elliott et al. (2015) (EMW).

In particular, we propose a computationally trivial method to find the subset of controls that is able to produce the largest expected (excess) length reductions when using the above intersection

[2]In fact, we show that these two types of CIs are optimal at each quantile of the excess length distribution greater than one minus their nominal coverage probabilities. See Proposition 2 for details.

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 critical values 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 critical values as a function of one or two empirical correlation parameters, depending upon whether they are forming a one- or two-sided CI. A Stata package available in the SSC archive automatically computes the CIs we propose.[3]

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 that of 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 two-sided CIs cannot be more than 2.28% longer than the corresponding standard CI for *any* realization of the data and the expected excess length of our one-sided CIs cannot be more than 3.07% longer than that of the corresponding standard CI.

A leading example of where our proposed CIs should prove useful is in the context of factorial (or cross-cutting) designs in field experiments. Take, for example, the 2×2 factorial design where two treatments are administered independently such that there are three treatment arms, the two "main" treatments (separately) and the combination of the two, and a control arm. The corresponding treatment effects can be consistently estimated by OLS using the "long" regression, i.e., the regression of the outcome variable on a constant and three dummy variables, one for each treatment arm.[4] In many cases, researchers have prior knowledge about the signs of the

[3]The package name is "ssci". Corresponding Matlab code is available on the authors' webpage.

[4]Recently, Muralidharan et al. (2020) have highlighted the importance of using the long

main treatment effects. For example, ethics boards for research grants are unlikely to fund experiments unless they are very likely to entail non-negative average treatment effects. Moreover, experimenters often conduct pilot studies prior to conducting full scale experiments in part to confirm their prior beliefs about the direction of average treatment effects.[5] Such prior knowledge can then be used to obtain sign restrictions on the corresponding regression coefficients. In this context, our proposed CIs will be short if the *estimated* effects of the main treatments are small and/or have the "wrong" sign, which is likely to occur if the unknown *population* effects are small or zero. To illustrate the potential usefulness of our CIs in the context of factorial designs, we revisit Blattman et al. (2017) who study the effect of "therapy" and "cash" on violent and criminal behavior in Liberia using a 2×2 factorial design. Indeed, for some of the treatment effects under study, we find our proposed CIs to be up to 35% shorter than the corresponding standard CIs.

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;

regression, as opposed to the "short" regression, i.e., the regression of the outcome variable on a constant and a dummy for the treatment of interest, to avoid omitted variable biases and accompanying size distortions.

[5]For example, possible negative effects of a treatment may be closely monitored during the pilot phase and, if realized, even lead to early termination of the experiment. Note also that imposing the absence of any negative effects on *individual* participants (that could be associated with the treatment) is stronger than necessary, because our CIs only require sign restrictions on *average* (treatment) effects in this context.

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 empirically-relevant parameter values.

Depending on the application, it may, of course, be possible that a researcher has prior knowledge on the magnitude of the control variables' coefficients rather than their sign. In this case, the recent work by Armstrong et al. (2020) can be employed (see also Li and Müller, 2021). Indeed, Muralidharan et al. (2020) study and suggest (among others) the CI proposed by Armstrong et al. (2020) as a means to improve over standard CIs in the context of factorial designs. In particular, they argue that researchers may, depending on the application, be willing to assume prior knowledge of the maximum (absolute) value of an "interaction effect", i.e., the effect of providing two treatments jointly minus the sum of the two main treatment effects. Here, we provide complementary results to be applied in settings for which it is natural for researchers to know the direction, rather than the magnitude, of control variables' coefficients.

This is certainly not the first paper to produce CIs that adapt to subclasses of DGPs while retaining uniform control of coverage probability. 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 (2020a,b). Given our focus on finite-dimensional models, we are not concerned with the rate of convergence adaptation in this literature but rather finite-sample length adaptation for CIs. Nevertheless, our CIs share some similarities with some of the CIs in this literature. Like the ones we propose, the CIs of Cai and Low (2004) and Kwon and Kwon (2020a,b) are obtained by intersecting CIs that are optimal under different subclasses of DGPs. Within this literature, Kwon and Kwon (2020a,b) are probably the closest studies to ours as they focus on nonparametric regression models with coordinate-wise monotone regression functions. In addition, both studies provide a means of shortening adaptive nonparametric CIs relative to simple Bonferroni corrections in a similar spirit to our CI endpoints computed from response surface regression output. However, in contrast to the existing literature on adaptive CIs for nonparametric models, we prove that our CIs are uniformly asymptotically valid without assuming Gaussian disturbances or fixed regressors.

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 (WAP) maximizing tests such as those of Moreira and Moreira (2013) or EMW 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 and indeed imply NO tests (in the sense of EMW) for certain sets of weights.

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 a general asymptotic setting and providing formal results on their validity. Section 3 then shows how our CIs are constructed in practical finite-sample 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. Proofs as well additional results, including theoretical results that establish the uniform asymptotic validity of our CIs across a wide variety of applications, are collected in the Online Appendix.

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. Let $\mathbb{R}_+ = [0,\infty)$, $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{\infty\}$, $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ and z_ξ denote the ξ^{th} quantile of the standard normal distribution. For a square matrix A, $\mathrm{Diag}(A)$ denotes the diagonal matrix with the same diagonal entries as A.

2 Large Sample Problem

Under standard conditions, estimators of commonly-employed (semi-)parametric models are asymptotically normally distributed with a consistently-estimable covariance matrix. Examples of these models include regression models, instrumental variables models, maximum likelihood models and models estimated by the generalized method of moments.[6] In these settings, estimators can be scale-normalized by consistent estimators of their standard deviations so that inference on a finite-dimensional parameter is asymptotically equivalent to inference on the unknown mean vector θ from a single observation of a Gaussian random variable $Y \overset{d}{\sim} \mathcal{N}(\theta,\Omega)$, 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

[6]The conditions implying asymptotic Gaussianity of estimators require the model to be well-behaved. For example, in the context of instrumental variables models or models estimated by the generalized method of moments, weak instruments or other forms of weak identification have to be ruled out. As alluded to in the Introduction, even in models that may not be defined outside the parameter space, such as the random coefficients logit (Berry et al., 1995) and (G)ARCH models for which variance parameters must be non-negative, it is often possible to construct an asymptotically normal estimator (Ketz, 2018). Indeed, such models provide other natural applications for the CIs we introduce in this paper.

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 finite-dimensional nuisance parameter whose elements are known to be greater than or equal to zero.[7,8]

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

[7]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).

[8]At a broad conceptual level, we can think of sign restrictions on nuisance parameters as providing overidentifying restrictions that may yield efficiency gains and more informative inference. In some cases, imposing inequality constraints (such as sign restrictions) can lower the variance and mean squared error of an estimator and "improve" tests and confidence sets based on it; see e.g., Moon and Schorfheide (2009) in the context of overidentifying inequality moment conditions. In the case at hand, it follows from Section 3.4 of Rothenberg (1973) that the maximum likelihood estimator of $\theta = (\beta, \delta')'$ that incorporates the restrictions on the parameter space, say $\tilde{Y}$, has lower (weighted) mean squared error than the unrestricted estimator $Y = (Y_\beta, Y_\delta')'$. However, it is not advisable to restrict oneself to the use of $\tilde{Y}$ in the construction of inference procedures: $\tilde{Y}$ takes the form of a projection of Y onto the (restricted) parameter space for θ, discarding potentially useful information. See Andrews and Shapiro (2021) for a related discussion.

small or equal 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. Before proceeding, we note that although the CIs we propose have good length properties over the entire $\delta \geq 0$ parameter space, they converge to standard CIs evaluated at slightly higher nominal coverage levels than desired as the elements of δ grow large. We therefore recommend using only nuisance parameters that are expected to be small or moderately-sized as inputs for constructing our CIs and dropping the rest from the analysis.

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.

2.1 Basic Intuition

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 properties (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 $z_{1-\alpha}$.[9] On a more technical level, we show in Proposition 2(i) below that this CI achieves the minimal maximum excess length quantile for all quantiles larger than α across the $\delta \geq 0$ parameter space. That is, the standard CI is minimax for the problem we are interested in. On the other hand, Proposition 2(ii) below shows that the CI that is excess length-optimal for all excess length quantiles larger than α when δ is known to equal zero is equal to

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

This CI 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 $\delta\to\infty$.[10]

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 $\widetilde{CI}_u(Y_\beta,\rho Y_\delta)$, at different confidence levels that ensure property (i) holds, 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:

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

---

[9]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.

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

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

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

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

where the inequality uses the fact that $\rho\delta \geq 0$.

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

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

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:

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

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 again 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.

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 that includes the sign-restricted control variable 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 that omits the sign-restricted control variable 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, for which Y_δ is the t-statistic for testing whether the sign-restricted control coefficient is equal to zero, but with a nonstandard critical value that incorporates both the two-step nature of the inference procedure as well as the dependence between the (scaled) estimators of the coefficient of interest Y_β and the control coefficient Y_δ. Note that as $\rho \to 1$, this nonstandard pretest approaches a standard t-test pretest that compares the t-statistic Y_δ to a standard normal quantile critical value $z_{1-\alpha+\gamma}$. 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).[11]

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. In what follows, we propose an algorithm that is designed to do just that while maintaining particularly low excess length when δ is small or equal to zero.

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

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

[11]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.

where

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

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

Proposition 1

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

[Figure 1 here.]

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 $\gamma = \alpha/10$.[12] This expected excess length is strictly decreasing in ω. Now, let $Y_\delta^{(s)}$ denote an arbitrary 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 $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}$ and $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}\Omega_{\delta^{(s)}\beta}$, are set equal to zero when $Y_\delta^{(s)} = \emptyset$. Next, consider

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

which has correct coverage for all $\delta \geq 0$ as long as all elements of $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}$ are non-negative. Since the expected excess length of $\widehat{CI}_u(\cdot)$ does not depend on β and is smallest at $\delta = 0$ for a subvector $Y_\delta^{(s)}$ that maximizes $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}\Omega_{\delta^{(s)}\beta}$ (cf. Figure 1), we propose the following algorithm.

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 a

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

subvector $Y_\delta^{(s^*)}$ such that $\Omega_{\beta\delta(s)}\Omega_{\delta(s)\delta(s)}^{-1}\Omega_{\delta(s)\beta}$ is maximized at $s=s^*$. Then, construct

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

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 the function $c(\cdot)$ at the single point $\Omega_{\beta\delta(s^*)}\Omega_{\delta(s^*)\delta(s^*)}^{-1}\Omega_{\delta(s^*)\beta}$. Therefore, the algorithm carries very low computational cost. 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. For each $\alpha\in\{0.01,0.05,0.1\}$, we fit a 6^{th} order polynomial on the grid $\bar{\Omega}_u=\{0,0.001,0.002,...,0.999\}$; the corresponding R^2 values are all greater than 0.999.

[Table 1 here.]

Table 1 reports the corresponding coefficients, where the intercepts have been adjusted to yield minimal coverage probabilities of exactly $1-\alpha$ over $\bar{\Omega}_u$.[13] We also note that $c(\Omega_{\beta\delta(s^*)}\Omega_{\delta(s^*)\delta(s^*)}^{-1}\Omega_{\delta(s^*)\beta})$ can always be replaced by $\sqrt{1-\Omega_{\beta\delta(s^*)}\Omega_{\delta(s^*)\delta(s^*)}^{-1}\Omega_{\delta(s^*)\beta}}\,z_{1-\gamma}$ 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).

In addition to the above motivation for the selection of a subvector $Y_\delta^{(s^*)}$ to form $\widehat{CI}_u^*(Y_\beta,Y_\delta,\Omega)$, Algorithm One-Sided is also justified on theoretical grounds as it entails the intersection of CIs that are optimal over two different subclasses of DGPs like those of e.g., Cai and Low (2004) and Kwon and Kwon (2020a,b). More specifically, the following proposition applies a general result of Armstrong and Kolesár (2018) to the current inference setting to formalize what we mean by "optimal" here.

[13]The probabilities underlying the computation of $c(\omega)$ as well as the minimal coverage probabilities are evaluated using 1,000,000 simulation draws; see Appendix F for details.

Proposition 2

For inference on β in (1)*, the following statements hold for $\alpha \in (0,1)$:*

(i) among all upper one-sided CIs with coverage of at least $(1-\alpha)$ for all $\delta \geq 0$, the CI that minimizes all maximum excess length quantiles over the $\delta \geq 0$ parameter space at quantile levels greater than α is equal to

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

(ii) among all upper one-sided CIs with coverage of at least $(1-\alpha)$ for all $\delta \geq 0$, the CI that minimizes all excess length quantiles at the point $\delta = 0$ and quantile levels greater than α is equal to

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

In particular, this proposition implies that for $\alpha < 1/2$ the CIs in (i) and (ii) above minimize maximum *median* excess length across $\delta \geq 0$ and *median* excess length at $\delta = 0$, respectively. Since median and expected excess length coincide for CIs that are affine in the data (Y), we also have that the CIs in (i) and (ii) minimize maximum *expected* excess length across $\delta \geq 0$ and *expected* excess length at $\delta = 0$ among all affine CIs, respectively, if $\alpha < 1/2$. Algorithm One-Sided computes a CI that intersects two optimal CIs of these forms while making a non-conservative multiplicity correction that improves upon the conservative Bonferroni adjustment.

A natural question is whether the resulting CI is also optimal in some sense. Although uniformly most powerful tests do not exist in the context at hand, we may compare the test implied by $\widehat{CI}^*_u(Y_\beta, Y_\delta, \Omega)$ to the NO test of EMW that nearly maximizes WAP. The corresponding hypothesis testing problem that we consider is

$$H_0: \beta = 0,\ \delta \geq 0 \text{ vs. } H_1: \beta > 0,\ \delta \geq 0, \tag{6}$$

where δ is scalar. The weights underlying the NO test we consider have β taking value 2 with probability 1 and δ uniformly distributed over [0,9]. These weights are inspired by the running example in EMW, which considers the two-sided version of the testing problem in (6).

[Figure 2 here.]

Figure 2 shows the rejection frequency of the test implied by $\widehat{CI}_u^*(Y_\beta,Y_\delta,\Omega)$ using $\gamma=\alpha/10$, the NO test and the standard one-sided t-test (for comparison) for testing (6) at $\alpha=0.05$, as a function of β for different values of δ when $\rho=0.7$; see Appendix D for implementation details. The power functions of the test implied by $\widehat{CI}_u^*(Y_\beta,Y_\delta,\Omega)$ and the NO test are visually indistinguishable, making the test implied by $\widehat{CI}_u^*(Y_\beta,Y_\delta,\Omega)$ NO as well.[14]

We now provide a formal theoretical result establishing lower and upper bounds on the coverage probability of the upper one-sided CI we propose in Algorithm One-Sided.

Theorem 1

For any $\delta\geq 0$, $\alpha\in(0,1/2)$ and $\gamma\in(0,\alpha)$,

$$1-\alpha\leq P\left(\beta\in\widehat{CI}_u^*(Y_\beta,Y_\delta,\Omega)\right)\leq 1-\alpha+\gamma.$$

This result shows not only that our proposed CI has correct coverage but also that by choosing γ to be "small" reduces how conservative the CI can be. However, there is a tradeoff in the choice of γ: although a smaller γ leads to a CI that is closer to having exact $1-\alpha$ coverage for any $\delta\geq 0$, it also allows for less length gains when the elements of δ are close or equal to zero.

As can be seen from Figure 1, extreme values of ω 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 expected excess length of the standard CI. At more empirically-relevant values of ω, 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, the expected excess length of our one-sided CI is bounded above by $z_{1-\alpha+\gamma}=$

[14]The underlying upper bound on WAP (evaluated using 100,000 simulation draws) equals 64.35% and the test implied by $\widehat{CI}_u^*(Y_\beta,Y_\delta,\Omega)$ is within $\epsilon=0.0131$ (using the notation in EMW) of this bound.

1.6954 for $\alpha=0.05$, $\gamma=\alpha/10$ and *any* value of ω. This implies that the expected excess length of our recommended CI relative to the standard one-sided CI is bounded above by $z_{1-\alpha+\gamma}/z_{1-\alpha}=1.6954/1.6449\approx1.0307$ for $\alpha=0.05$. That is the expected excess length increase of our recommended CI relative to the standard CI is bounded above by 3.07% for $\alpha=0.05$ and any value of ω.[15]

[Figure 3 here.]

Figure 3 plots the expected excess length of $\widehat{CI}_u(Z_1,\tilde{Z}_2+\sqrt{\omega}\delta;z_{1-\alpha+\gamma},c(\omega))$ using $\gamma=\alpha/10$, the minimax standard one-sided CI ($CI_u(Z_1)$) and the CI that is optimal when $\delta=0$ ($\widetilde{CI}_u(Z_1,\tilde{Z}_2+\sqrt{\omega}\delta)$) as a function of δ, for $\alpha=0.05$ and several values of ω. Similarly to Figure 1, Figure 3 shows that the gains in expected excess length of our one-sided CI compared to the standard one-sided CI are more pronounced for larger values of ω. Furthermore, Figure 3 illustrates the adaptive nature of our one-sided CI: at the endpoints of the parameter space, $\delta=0$ and $\delta=\infty$, its expected excess length approaches those of the optimal CI at $\delta=0$ and the minimax standard one-sided CI.

A different choice of γ from $\alpha/10$ would entail different tradeoffs for our CI over the $\delta\geq0$ parameter space. For example, a larger choice of γ would yield lower expected excess length when δ is small by bringing it "closer" to the optimal CI at $\delta=0$. Conversely, such a choice for γ would yield higher expected excess length at large values of δ. In fact, the user of our CI could choose γ according to how much of an increase in expected excess length they are willing to tolerate relative to the standard CI at large values of δ since the ratio of the expected excess length of our CI relative to the standard CI is bounded above by $z_{1-\alpha+\gamma}/z_{1-\alpha}$. For example, the choice of $\gamma=\alpha/2$ would entail an expected excess length increase relative to the standard CI bounded above by about 19% for $\alpha=0.05$ since $z_{1-\alpha+\gamma}/z_{1-\alpha}=1.96/1.6449\approx1.1916$. However, we note that the near optimality of the test implied

[15]For α equal to 0.01 and 0.1 (and $\gamma=\alpha/10$), the expected excess length of our one-sided CI is bounded above by 2.3656 and 1.3408 and the expected excess length of the standard one-sided CI is equal to 2.3263 and 1.2816, respectively. This implies that the expected excess length increase of our CI relative to the standard CI is bounded above by 1.69% and 4.62% for $\alpha=0.01$ and $\alpha=0.1$, respectively.

by $\widehat{CI}_u^*(\cdot)$ when using $\gamma = \alpha/10$ provides an additional argument for this particular choice of γ.

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.

To illustrate our two-sided CIs, temporarily suppose that δ is two-dimensional and that the correlation matrix of $(Y_\beta, Y_{\delta_1}, Y_{\delta_2})'$,

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_{12} \\ \rho_2 & \rho_{12} & 1 \end{pmatrix},$$

has $\rho_1 \geq 0$ and $\rho_2 \leq 0$. Define[16]

$$\begin{aligned} &\widehat{CI}_t(Y_\beta, \rho_1 Y_{\delta_1}, \rho_2 Y_{\delta_2}; z_{1-(\alpha-\gamma)/2}, c_\ell(\tilde{\omega}), c_u(\tilde{\omega})) \\ &= \left[Y_\beta - \min\left\{z_{1-(\alpha-\gamma)/2}, \rho_1 Y_{\delta_1} + c_\ell(\tilde{\omega})\right\}, Y_\beta + \min\left\{z_{1-(\alpha-\gamma)/2}, -\rho_2 Y_{\delta_2} + c_u(\tilde{\omega})\right\}\right], \end{aligned}$$

where $\tilde{\omega} = (\omega_{12}, \omega_{13}, \omega_{23}) = (\rho_1^2, \rho_2^2, \rho_{12}\rho_1\rho_2)$ and $c_\ell(\tilde{\omega})$ and $c_u(\tilde{\omega})$ minimize the expected length of $\widehat{CI}_t((Y_\beta, \rho_1 Y_{\delta_1}, \rho_2 Y_{\delta_2}; z_{1-(\alpha-\gamma)/2}, \cdot, \cdot)$ at $\delta = 0$ subject to correct coverage for all $\delta_1, \delta_2 \geq 0$; see Appendix A for details on the construction of $c_\ell(\tilde{\omega})$ and $c_u(\tilde{\omega})$. Note that the coverage probability of $\widehat{CI}_t(\cdot)$ is lowest over $\delta_1, \delta_2 \geq 0$ when $\delta_1 = \delta_2 = 0$ since $\rho_1\delta_1 \geq 0$ and $-\rho_2\delta_2 \geq 0$.

Before introducing the construction of our two-sided CIs in the general case for which the

[16]While it is possible for $\widehat{CI}_t(\cdot)$ to be empty, this is of little empirical relevance as the corresponding probability is very small. For $\alpha = 0.05$, for example, the probability that $\widehat{CI}_t(\cdot)$ is empty (evaluated over $\bar{\Omega}_t$, defined below, using 100,000 simulation draws) is less than 0.1%, 0.01% and 0.001% as long as $\omega_{12} = \rho_1^2$ and $\omega_{13} = \rho_2^2$ are both less than 0.9, 0.75 and 0.65, respectively.

dimension of δ is arbitrary, we illustrate several of its features in low-dimensional examples. First, $\widehat{CI}_t(\cdot)$ offers potential improvements over the standard two-sided CI $[Y_\beta \pm z_{1-\alpha/2}]$ even if there is only one sign-restricted nuisance parameter. This can be seen by noting that if there is only one sign-restricted nuisance parameter and the correlation ρ between Y_β and Y_δ is positive (negative), our two-sided CI is equal to $\widehat{CI}_t(Y_\beta,\rho Y_\delta,0;z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\omega}),c_u(\tilde{\omega}))$ ($\widehat{CI}_t(Y_\beta,0,\rho Y_\delta;z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\omega}),c_u(\tilde{\omega}))$). This CI can improve upon the length of the standard CI by using a larger lower bound (smaller upper bound) when the realizations of Y_δ are small or negative. Similarly, in the above case of two sign-restricted nuisance parameters with correlations of opposite sign, $\rho_1 > 0$ and $\rho_2 < 0$, $\widehat{CI}_t(Y_\beta,\rho_1 Y_{\delta_1},\rho_2 Y_{\delta_1};\cdot)$ can improve upon the length of the standard CI by using a larger lower bound and/or a smaller upper bound when the realizations of Y_{δ_1} and/or Y_{δ_2} are small or negative.

[Figure 4 here.]

Figure 4 shows the fitted surface of a 6^{th} order polynomial regression of the expected length of $\widehat{CI}_t(Y_\beta,\rho_1 Y_{\delta_1},\rho_2 Y_{\delta_2};z_{1-(\alpha-\gamma)/2},c_\ell(\tilde{\omega}),c_u(\tilde{\omega}))$ at $\delta_1 = \delta_2 = 0$ 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 $\bar{\Omega}_t = \bar{\mathcal{S}} \cap \mathcal{G}^2 \times -\mathcal{G} \cup \mathcal{G} \cup \{-0.99,-0.98,...,0.99\}$, where

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

and $-\mathcal{G} = \{g : -g \in \mathcal{G}\}$.[17] 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 $\bar{\Omega}_t$ for any given $(\omega_{12},\omega_{13})$ is equal to 0.0289. Furthermore, we note that the fitted expected length in Figure 4 is strictly decreasing in ω_{12} and ω_{13}.

Now, let δ be of arbitrary dimension and let $Y_\delta^{(s_1)}$ and $Y_\delta^{(s_2)}$ denote two arbitrary (possibly

[17] See Appendix A for the definition of $\bar{\mathcal{S}}$.

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 $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\Omega_{\delta^{(s_1)}\delta^{(s_2)}}\Omega$ are set equal to zero when $Y_\delta^{(s_1)} = \emptyset$ ($Y_\delta^{(s_2)} = \emptyset$). Next, consider

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

where

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

which has correct coverage for all $\delta \geq 0$ as long as 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. Since the expected length of $\widehat{CI}_t(\cdot)$ does not depend on β and is approximately smallest at $\delta = 0$ for 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 $\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s_2)}}\Omega_{\delta^{(s_2)}\beta}$ (cf. Figure 4), we propose the following algorithm.

Algorithm Two-Sided

Amongst all pairs of subvectors of Y_δ (including the empty ones) such that the elements of $\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}$ are non-negative and the elements of $\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}$ are non-positive, find a 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 $s_2 = s_2^*$.[18] Then, construct

$$\widehat{CI}_t^*(Y_\beta, Y_\delta, \Omega)$$

[18]Note that the pair of subvectors $Y_\delta^{(s_1^*)}$ and $Y_\delta^{(s_2^*)}$ may contain overlapping components of Y_δ.

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

While one could directly choose a subvector pair $Y_\delta^{(s_1)}$ and $Y_\delta^{(s_2)}$ to minimize the expected length of $\widehat{CI}_t(\cdot)$ at $\delta = 0$, Algorithm Two-Sided has the advantage that it only requires the computation of the expected length of $\widehat{CI}_t(\cdot)$ once, namely to compute $c_\ell(\tilde{\Omega}^{(s_1^*,s_2^*)})$ and $c_u(\tilde{\Omega}^{(s_1^*,s_2^*)})$. And even this computational step can be avoided by using the following polynomial response surface approximation of $c_u(\tilde{\omega})$. For $\alpha \in \{0.01, 0.05, 0.1\}$, we fit a 6^{th} order polynomial on the grid $\bar{\Omega}_t$ in ω_{12} and ω_{13} alone. Again, the corresponding R^2 values are all greater than 0.999, which is remarkable as it implies that $c_u(\cdot)$ is nearly invariant to ω_{23}.

[Table 2 here.]

Table 2 provides the coefficients for the 6^{th} order polynomial approximation of $c_u(\tilde{\omega})$ for $\alpha = 0.05$, where again the intercept has been adjusted to yield a minimal coverage probability exactly equal to $1-\alpha$ over $\bar{\Omega}_t$.[19] Tables 6 and 7 in Appendix E provide the corresponding coefficients for $\alpha = 0.01$ and $\alpha = 0.1$, respectively. The following proposition shows that one can also approximate $c_\ell(\tilde{\omega})$ using Table 2 (and Tables 6 and 7) by simply reversing the roles of ω_{12} and ω_{13} in the computation of $c_u(\tilde{\omega})$.

Proposition 3

The critical value functions $c_\ell(\cdot)$ and $c_u(\cdot)$ are related through the following symmetry condition: $c_\ell(\omega_{13}, \omega_{12}, \omega_{23}) = c_u(\omega_{12}, \omega_{13}, \omega_{23})$.

In the two-sided case, supporting theoretical results analogous to those in Proposition 2 do not exist since the CI that minimizes expected length at the point $\delta = 0$ depends upon the true value of β (see Pratt, 1961). Nevertheless, the test implied by $\widehat{CI}_t^*(Y_\beta, Y_\delta, \Omega)$ is NO for the two-sided version of the testing problem given in (6); see Appendix D.

Similarly to the one-sided case above, Figure 4 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

[19]The probabilities underlying the computation of $c_u(\tilde{\omega})$ as well as the minimal coverage probabilities are evaluated using 100,000 simulation draws; see Appendix F for details.

the same time, for *any* realization of the data, the realized length of our two-sided CI cannot exceed $2\times z_{1-(\alpha-\gamma)/2}=4.009$ 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 $2\times z_{1-\alpha/2}=3.92$ of the standard two-sided CI.[20]

Mirroring Theorem 1 for our one-sided CI, we now provide a formal theoretical result establishing lower and upper bounds on the coverage probability of the two-sided CI we propose in Algorithm Two-Sided.

Theorem 2

For any $\delta\geq 0$, $\alpha\in(0,1/2)$ and $\gamma\in(0,\alpha)$,

$$1-\alpha\leq P\left(\beta\in\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)\right)\leq 1-\alpha+\gamma.$$

Analogous implications for the choice of γ in the one-sided case also apply to our two-sided CI constructions.

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[21]

$$\sqrt{n}\begin{pmatrix}\hat{b}-b\\ \hat{d}-d\end{pmatrix}\xrightarrow{d}\mathcal{N}(0,\Sigma)\quad\text{with}\quad\Sigma=\begin{pmatrix}\Sigma_{bb} & \Sigma_{bd}\\ \Sigma_{db} & \Sigma_{dd}\end{pmatrix},\tag{7}$$

[20]For α equal to 0.01 and 0.1, the length of our two-sided CI cannot exceed 5.224 and 3.391 and the length of the standard two-sided CI is equal to 5.152 and 3.290, respectively. This implies a maximum length increase of 1.41% and 3.07%, respectively.

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

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 reparameterization 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 $i = 1, \ldots, n$

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

where y_i is the dependent variable, z_i is the scalar regressor of interest, $x_i \in \mathbb{R}^k$ are control variables with *known positive partial effects* $d \geq 0$ on y_i, $w_i \in \mathbb{R}^l$ are control variables with unrestricted partial effects c and ε_i is the error term. The ordinary least squares estimator $(\hat{b}, \hat{d}')'$ satisfies (7) under standard assumptions on the linear regression model.

Researchers must be careful when imposing sign restrictions in the linear regression context. In the context of regressions based on randomized experiments, regression coefficients indeed typically represent causal effects and are therefore more straightforward to interpret and sign. See the empirical application in Section 4 for example. On the other hand, in regressions based upon observational data regression coefficients often do not have a causal interpretation. Nevertheless, it may still be possible to determine their sign, particularly in settings where the regression model corresponds to a conditional expectation function. For example, in the context of Mincer-type regressions it may be reasonable to assume that the partial effect of experience is (weakly) concave, i.e., the sign of the linear term is non-negative and the sign of the quadratic term is non-positive. In a related context, Hanushek et al. (2015) provide arguments for why the partial effect of schooling in a regression of wages on their measure of cognitive skills and schooling should be non-negative. Further examples for which researchers may know the direction of partial effects can be found in growth regressions. Squicciarini and Voigtländer (2015), for example, (indirectly) argue that the partial effects of their proxies for average human capital are non-negative in a regression where the variable of interest is a proxy for knowledge elites.

3.1 Implementation

For a consistent covariance matrix estimator $\widehat{\Sigma}$, (7) suggests the following large-sample distributional approximation consistent with (1) and (4):

$$\mathrm{Diag}\left(\widehat{\Sigma}\right)^{-1/2}\sqrt{n}\begin{pmatrix}\hat{b}\\ \hat{d}\end{pmatrix}\overset{a}{\sim}\mathcal{N}\left(\begin{pmatrix}\beta\\ \delta\end{pmatrix},\Omega\right), \tag{8}$$

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

The distributional approximation in (8) and the availability of the consistent estimator $\widehat{\Omega}$ suggest that we can use

$$\begin{aligned}CI_{u,n}\left(\hat{b},\hat{d},\widehat{\Sigma}\right)&=\frac{\sqrt{\widehat{\Sigma}_{bb}}}{\sqrt{n}}\widehat{CI}_u^*\left(\frac{\sqrt{n}\hat{b}}{\sqrt{\widehat{\Sigma}_{bb}}},\mathrm{Diag}(\widehat{\Sigma}_{dd})^{-1/2}\sqrt{n}\hat{d};\widehat{\Omega}\right)\\ &=\left[\hat{b}-\frac{\widehat{\Sigma}_{bb}}{\sqrt{n}}\min\left\{z_{1-\alpha+\gamma},\widehat{\Omega}_{bd^{(\hat{s}^*)}}\widehat{\Omega}^{-1}_{d^{(\hat{s}^*)}d^{(\hat{s}^*)}}\hat{\delta}^{(\hat{s}^*)}+c(\widehat{\Omega}_{bd^{(\hat{s}^*)}}\widehat{\Omega}^{-1}_{d^{(\hat{s}^*)}d^{(\hat{s}^*)}}\widehat{\Omega}_{d^{(\hat{s}^*)}b})\right\},\infty\right)\end{aligned} \tag{9}$$

and

$$
\begin{aligned}
CI_{t,n}\left(\hat{b},\hat{d},\widehat{\Sigma}\right) &= \frac{\sqrt{\widehat{\Sigma}_{bb}}}{\sqrt{n}}\widehat{CI}_t^*\left(\frac{\sqrt{n}\hat{b}}{\sqrt{\widehat{\Sigma}_{bb}}},\mathrm{Diag}(\widehat{\Sigma}_{dd})^{-1/2}\sqrt{n}\hat{d};\widehat{\Omega}\right)\\
&= \left[\hat{b}-\frac{\widehat{\Sigma}_{bb}}{\sqrt{n}}\min\left\{z_{1-\frac{\alpha-\gamma}{2}},\widehat{\Omega}_{bd^{(\hat{s}_1^*)}}\widehat{\Omega}^{-1}_{d^{(\hat{s}_1^*)}d^{(\hat{s}_1^*)}}\hat{\delta}^{(\hat{s}_1^*)}+c_\ell\left(\widehat{\tilde{\Omega}}^{(\hat{s}_1^*,\hat{s}_2^*)}\right)\right\},\right.\\
&\qquad \left.\hat{b}+\frac{\widehat{\Sigma}_{bb}}{\sqrt{n}}\min\left\{z_{1-\frac{\alpha-\gamma}{2}},-\widehat{\Omega}_{bd^{(\hat{s}_2^*)}}\widehat{\Omega}^{-1}_{d^{(\hat{s}_2^*)}d^{(\hat{s}_2^*)}}\hat{\delta}^{(\hat{s}_2^*)}+c_u\left(\widehat{\tilde{\Omega}}^{(\hat{s}_1^*,\hat{s}_2^*)}\right)\right\}\right]
\end{aligned}
\tag{10}
$$

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.

Even though $\widehat{\Omega}$ is consistent for Ω, the signs of the elements of $\widehat{\Omega}_{bd^{(s)}}\widehat{\Omega}^{-1}_{d^{(s)}d^{(s)}}$ are not consistent for the signs of the elements of $\Omega_{bd^{(s)}}\Omega^{-1}_{d^{(s)}d^{(s)}}$ when they are equal to zero. Nevertheless, for one-sided CIs, any choice of subvector s used to construct our CIs in finite samples will lead to asymptotically correct coverage as long as $\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\geq 0$. This means that the use of $\hat{s}^*$ that either converges to a subvector s with this property or does not converge but takes values in the set of subvectors with this property with probability approaching one will lead to a CI with asymptotically correct coverage. This is indeed the case for the choice of $\hat{s}^*$ that we propose. The analogous argument holds for two-sided CIs. Indeed, the theoretical results in Appendix C formally confirm that these CIs attain uniformly correct asymptotic coverage under weak conditions.

3.2 Practical Discussion

We now make some remarks in order to summarize and clarify the properties of our CIs for the applied user. We focus on one-sided CIs in this discussion because analogous remarks apply to our two-sided CIs. First, we note that at any given value of $\omega\in[0,1)$, the value $c(\omega)$ defined by (5) used to construct our CIs can be computed by simulating the bivariate normal random vector $(Z_1,\tilde{Z}_2)'$ or by numerically evaluating the integral that defines the probability on the left hand side of (5). However, our response surface approximation, based upon our own simulation-based calculations of $c(\omega)$ over a fine grid of ω values, allows the user to directly compute a very accurate approximation of $c(\omega)$ as a deterministic function of ω from Table 1. This obviates the need for an applied researcher to perform any simulations or numerical integration themself.

Second, we note that since ω is one-dimensional, the accuracy of our response surface approximation of $c(\omega)$ does not change depending upon the dimensionality of the vector of nuisance parameters d. However, if the dimension of d relative to the sample size n is large, estimation of Σ can become difficult and since our CIs require an accurate estimator of Σ, such a scenario could induce finite-sample coverage distortions. We therefore view our CI construction approaches as applicable to problems involving low- to moderate-dimensional sign-restricted nuisance parameters.

Third, our algorithms involve computing quantities across all subvectors of a vector of dimension equal to the dimension of the sign-restricted nuisance parameter k. The number of such subvectors is equal to 2^k. If k is very large, this computation can become demanding. However, a typical empirical application is unlikely to involve very many sign-restricted nuisance parameters (that are all expected to be small) and we again view our CI construction approaches as applicable to the empirically-relevant cases involving low- to moderate-dimensional sign-restricted nuisance parameters.

Fourth, we note that CIs constructed according to (9) using the value $c(\omega)$ defined by (5) have correct asymptotic coverage uniformly across general empirically-relevant parameter spaces for which the parameter estimator $(\hat{b},\hat{d}')'$ is asymptotically normal and has a consistently estimable covariance matrix. Theorem 3 in Appendix C formally shows this.

Finally, by plotting the expected excess length of our CIs as a function of $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}\Omega_{\delta^{(s)}\beta}$, we find that a subvector of the nuisance parameter that yields the largest value of $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}\Omega_{\delta^{(s)}\beta}$ (subject to $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}$ being non-negative) minimizes the expected excess length of our CIs at the point for which the nuisance parameter is equal to zero. This choice of subvector is further supported by the theoretical results of Proposition 2, which tell us that it is theoretically optimal for minimizing the excess length of the CIs that our CIs intersect the standard CI $[Y_\beta - z_{1-\alpha+\gamma},\infty)$ with. We note, however, that correct coverage of our CIs does not depend upon using this particular subvector choice so long as $c(\cdot)$ is evaluated at $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}\Omega_{\delta^{(s)}\beta}$ for the subvector being used in the construction of our CIs. In other words, Theorem 1 continues to hold when $\delta^{(s^*)}$ is replaced by any subvector $\delta^{(s)}$ such that $\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}$ is non-negative in the construction of a one-sided CI. The analogous statement holds for our two-sided CIs and Theorem 2.

4 Empirical Application of Sign-Restricted Regression

For our proposed CIs to be able 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, the estimator of the coefficient of interest must be (asymptotically) correlated with the estimator of the sign-restricted control variables' coefficients and the true values of the sign-restricted control variables' coefficients must be small. These 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:

$$Y = \alpha_0 + \alpha_1 T_1 + \alpha_2 T_2 + \alpha_3 T_1 \times T_2 + u, \tag{11}$$

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.[22] 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

[22]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 $\alpha_3 = E[Y_{1,1} - Y_{0,0}] - (E[Y_{1,0} - Y_{0,0}] + E[Y_{0,1} - Y_{0,0}])$.

conditionally homoskedastic, then the asymptotic correlation matrix of $\sqrt{n}(\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3)'$ is given by

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

Lastly, the coefficients α_1 and α_2 are likely to be small, motivating the field experiment to begin with.

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 α_1 and α_2 and a two-sided CI for α_3. We note that 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:

$$Y = \alpha_0 + \alpha_1(T_1 - T_1 \times T_2) + \alpha_2(T_2 - T_1 \times T_2) + \alpha_3^* T_1 \times T_2 + u, \tag{12}$$

where $\alpha_3^* = \alpha_3 + \alpha_1 + \alpha_2$ is the effect of "both" treatments provided jointly, relative to a business-as-usual counterfactual.[23] This regression again results in high correlation: under the same conditions as in the example above, the asymptotic correlation matrix of $\sqrt{n}(\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3^*)'$ is given by

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

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. While it may be reasonable to assume that $\alpha_3^* \geq 0$ in some cases, we note that it is beneficial to *not* impose this constraint in the construction of our CIs if there is reason to believe that α_3^* is large. This is likely to be the case if T_1 and T_2 are complements.

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

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 here.]

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 ($\hat{b}$) 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 (12), i.e., they only estimate the effect of "both" treatments and not the "interaction" effect.[24] Column 3 (SSCI—Simple and Short Confidence Interval) 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

[24]In fact, BJS consider the specification given in equation (12) 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.

the treatment effects of "therapy" and "cash" are *a priori* known to be non-negative.[25,26] They are constructed using Algorithms One-Sided and Two-Sided in combination with the response surface approximations.[27] Column 5 (SCI—Standard Confidence Interval) 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 $\hat{b}$ and the CI's lower bound.[28] Column 7 (Ratio) computes the ratio of the ("excess") length of SSCI relative to SCI. We find that, while our proposed CI is marginally longer than the standard CI for C—its "excess" length reaching the bound on expected excess length increase of $\sim$3%, it is much shorter for T, B and I.

---

[25]Given the knowledge that α_1 and α_2 are non-negative, we could alternatively report the intersection of the above one-sided CIs with $[0,\infty)$. Here, we report the "uncensored" CIs for illustrative purposes: a researcher may, for example, only feel comfortable imposing a non-negativity constraint on "therapy" ("cash") but still be interested in obtaining a one-sided CI for "cash" ("therapy").

[26]Table 8 in Appendix E reports the corresponding results when it is assumed that the treatment effect of "both" is also known to be non-negative, i.e., when the constraint $\alpha_3^* \geq 0$ is also imposed.

[27]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 24.

[28]We write "excess" in quotes to emphasize the fact that this is not equal to the true excess length that cannot be computed here in the absence of knowledge of the true value of the regression coefficients.

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 (11) and (12). 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 in combination with the response surface approximations, 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 covariance matrix estimators computed within each bootstrap sample. Since the empirical distribution of the estimator across bootstrap samples is not normally distributed and the covariance matrix is estimated, this simulation exercise captures the effect on CI coverage of departures from the large sample problem of Section 2.

[Table 4 here.]

Table 4 reports the coverage frequencies (CF) of our proposed CI and of the standard CI across bootstrap samples for all four treatment effects, T, C, B and I. It also reports the average (excess) lengths (A(E)L) of these CIs across the bootstrap samples. 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 samples). 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 estimator and estimated covariance matrix. In terms of average (excess) length, most of our proposed CIs offer sizeable improvements over standard CIs, with average (excess) length improvements of up to 17% for this particular data calibration.

References

Andrews, D. W. K. (1999). Estimation When a Parameter Is on a Boundary. *Econometrica*, 67(6):1341–1383.

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.

Andrews, I. and Shapiro, J. M. (2021). A Model of Scientific Communication. *Econometrica*, 89(5):2117–2142.

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.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841–890.

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.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

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.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., and Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73:103–130.

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. (2020a). Adaptive inference in multivariate nonparametric regression models under monotonicity. arXiv:2011.14219v1.

Kwon, K. and Kwon, S. (2020b). Inference in regression discontinuity designs under monotonicity. arXiv:2011.14216v1.

Li, C. and Müller, U. K. (2021). Linear regression with many controls of limited explanatory power. *Quantitative Economics*, 12(2):405–442.

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.

Moon, H. R. and Schorfheide, F. (2009). Estimation with overidentifying inequality moment conditions. *Journal of Econometrics*, 153:136–154.

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.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Pratt, J. W. (1961). Length of confidence intervals. *Journal of the American Statistical Association*, 56:549–567.

Rothenberg, T. J. (1973). *Efficient estimation with a priori information*. Yale University Press.

Squicciarini, M. P. and Voigtländer, N. (2015). Human capital and industrialization: Evidence from the age of enlightenment. *The Quarterly Journal of Economics*, 130(4):1825–1883.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–838.

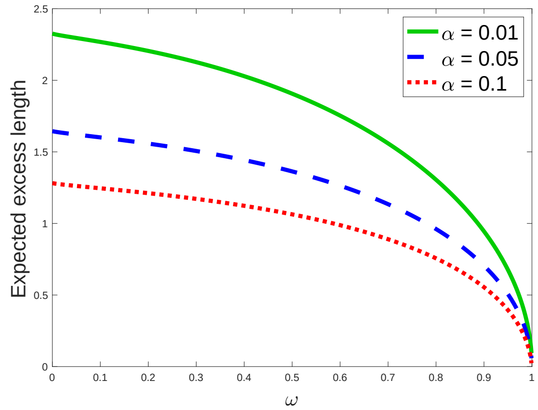

Figure 1: Expected excess length of $\widehat{CI}_u(Z_1,\tilde{Z}_2,z_{1-\alpha+\gamma},c(\omega))$ as a function of ω for $\alpha \in \{0.01,0.05,0.1\}$ and $\gamma = \alpha/10$.

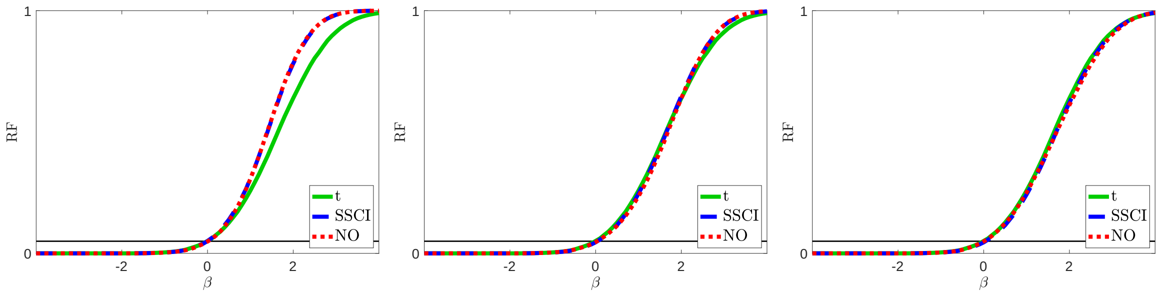

Figure 2: Rejection frequency as a function of β of standard one-sided t-test (t - solid), test implied by $\widehat{CI}_u^*(Y_\beta,Y_\delta,\Omega)$ using $\gamma=\alpha/10$ (SSCI - dashed) and the NO test (NO - dotted) for testing (6) with $d=0,1,2$ from left to right for $\alpha=0.05$ and $\rho=0.7$.

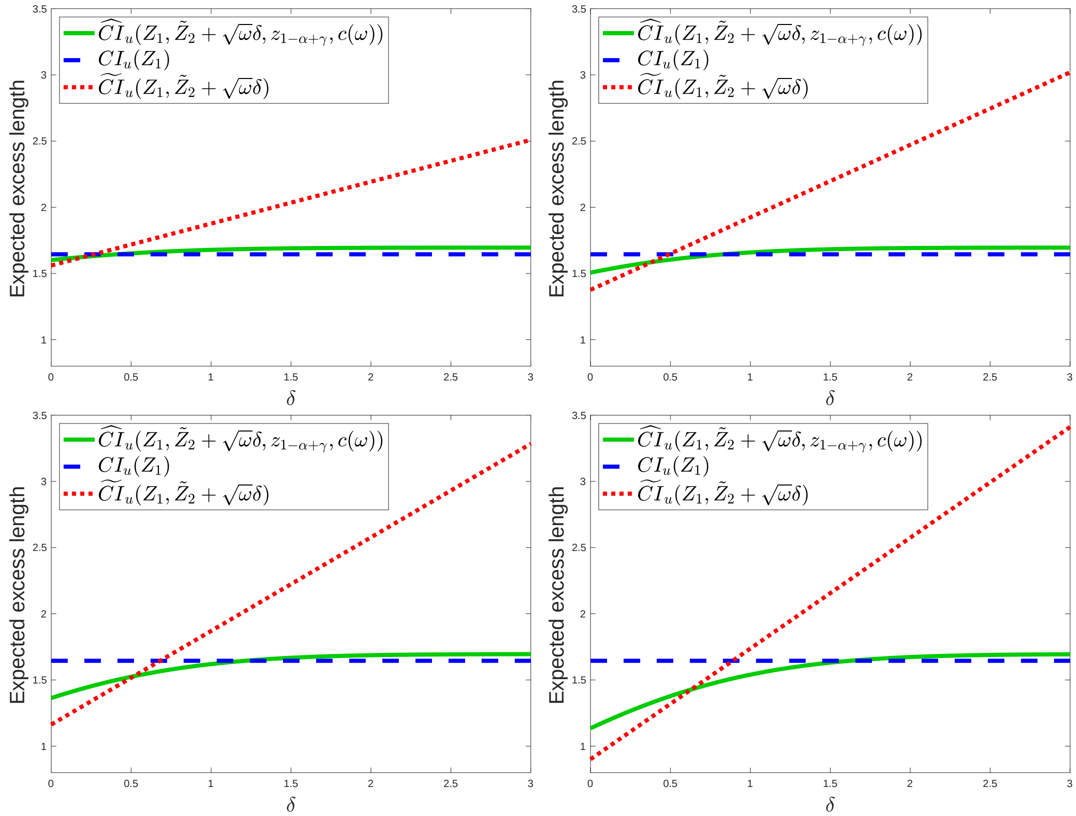

Figure 3: Expected excess length of $\widehat{CI}_u(Z_1,\tilde{Z}_2+\sqrt{\omega}\delta;z_{1-\alpha+\gamma},c(\omega))$ using $\gamma=\alpha/10$, $CI_u(Z_1)$ and $\widetilde{CI}_u(Z_1,\tilde{Z}_2+\sqrt{\omega}\delta)$ as a function of δ for ω equal to 0.1, 0.3, 0.5 and 0.7 from left to right and $\alpha=0.05$.

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	2.3476	2.5073	-19.6229	65.0489	-122.0242	112.9814	-40.9895
0.05	1.6597	2.4813	-16.1007	52.6998	-98.9348	91.7646	-33.3628
0.1	1.2917	2.4250	-14.1041	46.0326	-86.7946	80.8189	-29.4840

The table lists the coefficients for the 6$^{\text{th}}$ order polynomial approximation of $c(\omega)$ for $\alpha \in \{0.01, 0.05, 0.1\}$ and $\gamma=\alpha/10$. For example, the coefficient on the term ω^3 equals 52.6998 for $\alpha=0.05$.

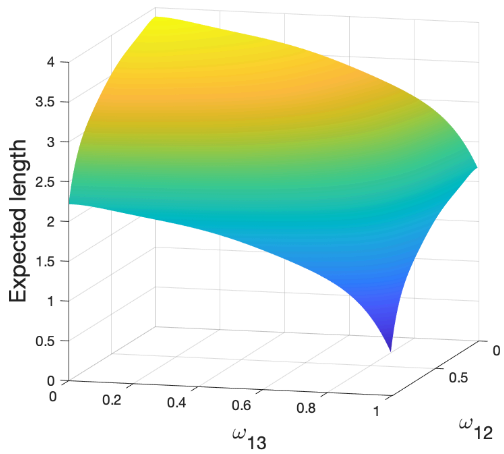

Figure 4: Expected length of $\widehat{CI}_t(Y_\beta, \rho_1 Y_\delta, \rho_2 Y_\delta; z_{1-(\alpha-\gamma)/2}, c_\ell(\tilde{\omega}), c_u(\tilde{\omega}))$ at $\delta_1 = \delta_2 = 0$ as a function of $\omega_{12} = \rho_1^2$ and $\omega_{13} = \rho_2^2$ for $\alpha = 0.05$ and $\gamma = \alpha/10$.

Table 2: Coefficients for 6^{th} order polynomial approximation of $c_u(\tilde{\omega})$ 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	1.9749	1.3388	−4.5110	11.7294	−18.8756	15.5342	−5.2786
ω_{13}	1.1289	−0.8006	1.1262	−1.1742	2.1281	−0.5511	
ω_{13}^2	−12.2929	0.0090	0.9084	−3.2329	0.1723		
ω_{13}^3	45.6505	0.5939	0.8153	1.7625			
ω_{13}^4	−92.3587	−1.0048	−0.9854				
ω_{13}^5	89.5045	0.2851					
ω_{13}^6	−33.3683						

The table lists the coefficients for the 6^{th} order polynomial approximation of $c_u(\tilde{\omega})$ in terms of ω_{12} and ω_{13} for $\alpha{=}0.05$ and $\gamma{=}\alpha/10$. For example, the coefficient on the interaction term $\omega_{12}^4{\times}\omega_{13}^2$ equals 0.1723.

Table 3: 95% confidence intervals for treatment effects on summary index of antisocial behaviors (times minus one)

	$\hat{b}$	SE	SSCI	(E)L	SCI	(E)L	Ratio
T	0.0829	0.0929	$[-0.0168,\infty)$	0.0997	$[-0.0700,\infty)$	0.1529	0.6524
C	-0.1316	0.0969	$[-0.2959,\infty)$	0.1643	$[-0.2910,\infty)$	0.1594	1.0307
B	0.2468	0.0883	[0.0969,0.4238]	0.3269	[0.0737,0.4198]	0.3462	0.9443
I	0.2955	0.1255	[0.0439,0.4127]	0.3688	[0.0495,0.5415]	0.4920	0.7496

The table shows point estimates ($\hat{b}$), standard errors (SE), our proposed CIs (SSCI), standard CIs (SCI), the corresponding ("excess") lengths ((E)Ls), and their ratio (Ratio) for the treatment effects of "therapy" (T), "cash" (C), "both" (B) and "interaction" (I). The CIs are upper one-sided for T and C and two-sided for B and I and it is assumed that the treatment effects of T and C are known to be non-negative.

Table 4: Coverage frequencies and average (excess) lengths of proposed CI and standard CI across bootstrap samples

		T	C	B	I	B0	I0
CF	SSCI	94.27	94.49	95.34	94.81	95.08	94.10
	SCI	94.30	93.96	94.78	94.49	94.78	94.49
A(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	0.9357	1.0088	0.9764	0.8954	0.9264	0.8283

The table shows the coverage frequencies (CF) of our proposed CI (SSCI) and the standard CI (SCI) across bootstrap samples, the corresponding average (excess) lengths (A(E)L) and the ratio of the latter (Ratio) for the treatment effects of "therapy" (T), "cash" (C), "both" (B), "interaction" (I), "both" and "interaction" when the treatment effect of T is 0 (B0 and I0, respectively). The CIs are upper one-sided for T and C and two-sided for B, I, B0 and I0 and it is assumed that the treatment effects of T and C are known to be non-negative.

Online Appendix for

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

Philipp Ketz Adam McCloskey

October 12, 2022

This Online Appendix contains supplemental material for the paper "Short and Simple Confidence Intervals when the Directions of Some Effects are Known." Appendix A details the construction of the critical value functions that enter our two-sided confidence intervals (CIs). Appendix B provides the mathematical proofs of our theoretical results in the general asymptotic setting. Appendix C provides theoretical results that establish the uniform asymptotic validity of our CIs across a wide variety of applications and specifies a parameter space for the standard linear regression model that satisfies the requirements for uniform asymptotic validity to hold. Appendix D contains implementation details for the "nearly optimal" (NO) test of Elliott et al. (2015) (EMW) and the near optimality results for the two-sided version of the testing problem given in (6). Appendix E contains additional tables referenced in the paper. Lastly, Appendix F provides details on the numerical computations underlying some of the results in the paper.

A Construction of Critical Values for Two-Sided Confidence Interval

Let

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

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 $\mathcal{S}=\{(x,y,z)\in\mathbb{R}^3:x,y\in(0,1),-z^2+2xyz+xy-x^2y-xy^2>0\}$,[29] where $\underline{c_u}:\bar{\mathcal{S}}\to\mathbb{R}$ is implicitly defined by

$$P(-\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3+\underline{c_u}(\tilde{\omega})\}\le Z_1\le z_{1-(\alpha-\gamma)/2})=1-\alpha. \tag{14}$$

For $\gamma\in(0,\alpha)$, consider the function $\tilde{c}:\widetilde{\mathcal{C}}\to\mathbb{R}_\infty$ implicitly defined by

$$P(-\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3+c_u\}\le Z_1\le\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2+\tilde{c}(c_u,\tilde{\omega})\})=1-\alpha \tag{15}$$

at points $(c_u,\tilde{\omega})\in\widetilde{\mathcal{C}}$ for which $\omega_{12},\omega_{13}\neq 0$. 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 (14) so that for any given $\tilde{\omega}$, the solution to (15) 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 (15). 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 $\tilde{c}(c_u,\tilde{\omega})=\lim_{\bar{\omega}_{13}\to 0}\tilde{c}(c_u,\omega_{12},\bar{\omega}_{13},\omega_{23})$.[30] Finally, define the correspondence $\tilde{c}_u:\bar{\mathcal{S}}\rightrightarrows\mathbb{R}_\infty$ as

$$\tilde{c}_u(\tilde{\omega})=\operatorname{argmin}_{c_u\in[\underline{c_u}(\tilde{\omega}),\infty]}E[\max\{\min\{z_{1-(\alpha-\gamma)/2},\tilde{Z}_2+\tilde{c}(c_u,\tilde{\omega})\}+\min\{z_{1-(\alpha-\gamma)/2},-\tilde{Z}_3+c_u\},0\}]. \tag{16}$$

[29] In terms of arguments (x,y,z), the definition of $\mathcal{S}$ is equivalent to the positive definiteness of the matrix

$$\begin{pmatrix} 1 & x & y \\ x & x & z \\ y & z & y \end{pmatrix}.$$

[30] 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 4 in Appendix B. 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 (15) when $\omega_{12}=0$ and we wish to treat ω_{12} and ω_{13} symmetrically in light of Proposition 3.

Note that $\tilde{c}_u(0) = z_{1-\alpha/2}$. The following proposition ensures that $\tilde{c}_u : \bar{\mathcal{S}} \rightrightarrows \mathbb{R}_\infty$ is well-defined and possesses some desirable properties.[31]

Proposition 4

For any $\tilde{\omega} \in \bar{\mathcal{S}}$, $\tilde{c}_u(\tilde{\omega}) \subset \mathbb{R}_\infty$ defined in (16) is non-empty and compact and $\tilde{c}_u : \bar{\mathcal{S}} \rightrightarrows \mathbb{R}_\infty$ is upper hemicontinuous.

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

B Proofs

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 $\gamma \in (0, \alpha)$.

---

[31] In our numerical work, we have found the solution to (16) to be a singleton and $\tilde{c}_u$ to be a continuous function when $\omega_{12}, \omega_{13} \neq 0$.

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. $\square$

The next lemmata are used to prove Proposition 2.

Lemma 1

For conformable matrices E, F, G, H, J and K, let

$$X = \begin{bmatrix} E & F & G \\ F' & H & J \\ G' & J' & K \end{bmatrix}.$$

Then, assuming the relevant inverse matrices exist, we have

$$X^{-1} = \begin{bmatrix} E^{-1} + E^{-1}[FA^{-1}F' + US^{-1}U']E^{-1} & -E^{-1}[F - US^{-1}B']A^{-1} & -E^{-1}US^{-1} \\ -A^{-1}[F' - BS^{-1}U']E^{-1} & A^{-1} + A^{-1}BS^{-1}B'A^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}U'E^{-1} & -S^{-1}B'A^{-1} & S^{-1} \end{bmatrix},$$

where $A = H - F'E^{-1}F$, $B = J - F'E^{-1}G$, $D = K - G'E^{-1}G$, $S = D - B'A^{-1}B$ and $U = G - FA^{-1}B$.

Proof. The proof follows from repeated application of the formula for blockwise inversion of a matrix. $\square$

Lemma 2

For conformable matrices Y and Z, assuming the relevant inverse matrices exist, we have

$$(Y+Z)^{-1}=Y^{-1}-Y^{-1}Z(Y+Z)^{-1}.$$

Proof. The proof follows directly from the Woodbury identity. □

Lemma 3

Let $\delta^{(s)}$ and $\delta^{(-s)}$ be two arbitrary subvectors of δ such that $\delta=(\delta^{(s)},\delta^{(-s)})$, where the order of the elements is without loss of generality. Furthermore, let

$$\begin{bmatrix} 1 & \Omega_{\beta\delta^{(s)}} & \Omega_{\beta\delta^{(-s)}} \\ \Omega_{\delta^{(s)}\beta} & \Omega_{\delta^{(s)}\delta^{(s)}} & \Omega_{\delta^{(s)}\delta^{(-s)}} \\ \Omega_{\delta^{(-s)}\beta} & \Omega_{\delta^{(-s)}\delta^{(s)}} & \Omega_{\delta^{(-s)}\delta^{(-s)}} \end{bmatrix} \text{ and } \begin{bmatrix} \Omega^{\beta\beta} & \Omega^{\beta\delta^{(s)}} & \Omega^{\beta\delta^{(-s)}} \\ \Omega^{\delta^{(s)}\beta} & \Omega^{\delta^{(s)}\delta^{(s)}} & \Omega^{\delta^{(s)}\delta^{(-s)}} \\ \Omega^{\delta^{(-s)}\beta} & \Omega^{\delta^{(-s)}\delta^{(s)}} & \Omega^{\delta^{(-s)}\delta^{(-s)}} \end{bmatrix}$$

be conformable partitions of Ω and Ω^{-1}, respectively. Then, we have

(i) $\Omega^{\beta\delta^{(s)}} - \Omega^{\beta\delta^{(-s)}}(\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1}\Omega^{\delta^{(-s)}\delta^{(s)}} = -(\Omega^{\beta\beta} - \Omega^{\beta\delta^{(-s)}}(\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1}\Omega^{\delta^{(-s)}\beta})\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}$

and

(ii) $1=(\Omega^{\beta\beta}-\Omega^{\beta\delta^{(-s)}}(\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1}\Omega^{\delta^{(-s)}\beta})(1-\Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}\Omega_{\delta^{(s)}\beta})$.

Proof. (i) Using Lemma 1, we show the equivalent result that

$$X^{12}-X^{13}(X^{33})^{-1}X^{32}+(X^{11}-X^{13}(X^{33})^{-1}X^{31})X_{12}X_{22}^{-1}=0,$$

where we use the same notational convention concerning sub- and superscripts as for Ω and Ω^{-1}. We have

$$\begin{aligned} & X^{12}-X^{13}(X^{33})^{-1}X^{32}+(X^{11}-X^{13}(X^{33})^{-1}X^{31})X_{12}X_{22}^{-1} \\ =& -E^{-1}[F-US^{-1}B']A^{-1}-E^{-1}US^{-1}B'A^{-1} \end{aligned}$$

$$+\big[E^{-1}+E^{-1}[FA^{-1}F'+US^{-1}U']E^{-1}-E^{-1}US^{-1}U'E^{-1}\big]FH^{-1}$$
$$=E^{-1}FH^{-1}-E^{-1}FA^{-1}+E^{-1}FA^{-1}F'E^{-1}FH^{-1}$$
$$=E^{-1}F[H^{-1}-A^{-1}+A^{-1}F'E^{-1}FH^{-1}]=0,$$

where the last equality follows from the fact that $H^{-1}-A^{-1}+A^{-1}F'E^{-1}FH^{-1}=0$ which, in turn, follows from Lemma 2 using $Y=A=H-F'E^{-1}F$ and $Z=F'E^{-1}F$.

(ii) Using Lemma 1, we show the equivalent result that

$$(X^{11}-X^{13}(X^{33})^{-1}X^{31})(X_{11}-X_{12}X_{22}^{-1}X_{21})=I,$$

where we again use the same notational convention as for Ω and Ω^{-1}. We have

$$
\begin{aligned}
&(X^{11}-X^{13}(X^{33})^{-1}X^{31})(X_{11}-X_{12}X_{22}^{-1}X_{21})\\
=&[E^{-1}+E^{-1}FA^{-1}F'E^{-1}][E-FH^{-1}F']\\
=&I-E^{-1}FH^{-1}F'+E^{-1}FA^{-1}F'-E^{-1}FA^{-1}F'E^{-1}FH^{-1}F'\\
=&I-E^{-1}F[H^{-1}-A^{-1}+A^{-1}F'E^{-1}FH^{-1}]F'=I,
\end{aligned}
$$

where we used again that $H^{-1}-A^{-1}+A^{-1}F'E^{-1}FH^{-1}=0$. $\square$

Proof of Proposition 2. Note that the problem of forming a CI for β when we observe $Y\sim\mathcal{N}(\theta,\Omega)$ with $\theta=(\beta,\delta')'$, $\delta\geq 0$ and known Ω is equivalent to forming a CI for β in the setting of Armstrong and Kolesár (2018) (AK):

$$\widetilde{Y}=X\theta+\varepsilon,\quad \varepsilon\sim\mathcal{N}(0,I_{k+1})$$

where $\widetilde{Y}=\Omega^{-1/2}Y$, $X=\Omega^{-1/2}$ and k is the dimension of δ. In what follows, we appeal to Theorem 3.1 of AK.

(i) In order to form the CI in this theorem, we must first form the affine estimator $\hat{L}_{\tilde{\delta},\mathcal{F},\mathcal{G}}$ in (23) of AK, for $\mathcal{F}=\mathcal{G}=\{(\beta,\delta')'\in\mathbb{R}^{k+1}:\delta\geq 0\}$. The modulus of continuity defined on p. 667 of AK specialized to our setting is

$$\omega(\tilde{\delta};\mathcal{F},\mathcal{G})=\sup_{\gamma_1,\theta}\{\gamma_1-\beta\}$$

$$\text{s.t. } (\gamma_1-\beta,\gamma_1'-\delta')\Omega^{-1}(\gamma_1-\beta,\gamma_{-1}'-\delta')'\le\tilde{\delta}^2, \quad \gamma_{-1}\ge 0, \quad \delta\ge 0,$$

where we use the partition $\gamma\equiv(\gamma_1,\gamma_{-1}')'$. Let

$$\Omega^{-1}=\begin{pmatrix}\Omega^{\beta\beta} & \Omega^{\beta\delta}\\ \Omega^{\delta\beta} & \Omega^{\delta\delta}\end{pmatrix}$$

so that the constraints for the modulus problem can be written as

$$(\gamma_1-\beta)^2\Omega^{\beta\beta}+2(\gamma_1-\beta)\Omega^{\beta\delta}(\gamma_{-1}-\delta)+(\gamma_{-1}-\delta)'\Omega^{\delta\delta}(\gamma_{-1}-\delta)\le\tilde{\delta}^2, \quad \gamma_{-1}\ge 0, \quad \delta\ge 0.$$

For any θ and γ that solve this optimization problem when dropping the final two constraints, we may simply add $(c,...,c)'$ for a large constant c to both θ and γ and obtain the same value without dropping the final two constraints on γ_{-1} and δ. Thus, these final two constraints do not affect the optimal procedure and we may instead focus on the modulus problem that drops them with the understanding that the solutions in γ_{-1} and δ must be large and positive.

After dropping these constraints, the first order condition wrt δ in the modulus problem is

$$-2\lambda[(\gamma_1-\beta)\Omega^{\beta\delta}+(\gamma_{-1}-\delta)'\Omega^{\delta\delta}]=0,$$

where $\lambda>0$ is the KKT multiplier associated with the remaining constraint. Using the formula for blockwise inversion of a matrix, the optimal solution to the modulus problem must therefore satisfy

$$(\gamma_{-1}-\delta)'=-(\gamma_1-\beta)\Omega^{\beta\delta}(\Omega_{\delta\delta}-\Omega_{\delta\beta}\Omega_{\beta\delta}).$$

The modulus problem thus simplifies to

$$\omega(\tilde{\delta};\mathcal{F},\mathcal{G}) = \sup_{\gamma_1,\beta}\{\gamma_1 - \beta\}$$
$$\text{s.t. } (\gamma_1-\beta)^2\Omega^{\beta\beta} - (\gamma_1-\beta)^2\Omega^{\beta\delta}(\Omega_{\delta\delta}-\Omega_{\delta\beta}\Omega_{\beta\delta})\Omega^{\delta\beta} \le \tilde{\delta}^2,$$

where the constraint further simplifies to

$$\text{s.t. } (\gamma_1-\beta)^2 \le \tilde{\delta}^2,$$

by the formula for blockwise inversion of a matrix. Thus, we have

$$\omega(\tilde{\delta};\mathcal{F},\mathcal{G}) = \tilde{\delta}$$

with a solution given by

$$\gamma^*_{\tilde{\delta},\mathcal{F},\mathcal{G}} = \begin{pmatrix} \tilde{\delta}/2 \\ \delta^* + \tilde{\delta}\Omega_{\delta\beta} \end{pmatrix}, \quad \theta^*_{\tilde{\delta},\mathcal{F},\mathcal{G}} = \begin{pmatrix} -\tilde{\delta}/2 \\ \delta^* \end{pmatrix}$$

and midpoint

$$\theta^*_{M,\tilde{\delta},\mathcal{F},\mathcal{G}} = (\theta^*_{\tilde{\delta},\mathcal{F},\mathcal{G}} + \gamma^*_{\tilde{\delta},\mathcal{F},\mathcal{G}})/2 = \begin{pmatrix} 0 \\ \delta^* + \tilde{\delta}\Omega_{\delta\beta}/2 \end{pmatrix}$$

for some large and positive δ^*, where we use the fact that $\Omega^{\beta\delta}(\Omega_{\delta\delta} - \Omega_{\delta\beta}\Omega_{\beta\delta}) = -\Omega_{\beta\delta}$ by the formula for blockwise inversion of a matrix. Formula (23) of AK thus yields

$$\begin{aligned}
\hat{L}_{\tilde{\delta},\mathcal{F},\mathcal{G}} &= \tilde{\delta}^{-1}(\gamma^*_{\tilde{\delta},\mathcal{F},\mathcal{G}} - \theta^*_{\tilde{\delta},\mathcal{F},\mathcal{G}})'\Omega^{-1}(Y - \theta^*_{M,\tilde{\delta},\mathcal{F},\mathcal{G}}) \\
&= (1,\Omega_{\beta\delta})\begin{pmatrix} \Omega^{\beta\beta} & \Omega^{\beta\delta} \\ \Omega^{\delta\beta} & \Omega^{\delta\delta} \end{pmatrix}\begin{pmatrix} Y_\beta \\ Y_\delta - \delta^* - \tilde{\delta}\Omega_{\delta\beta}/2 \end{pmatrix} \\
&= (\Omega^{\beta\beta} + \Omega_{\beta\delta}\Omega^{\delta\beta})Y_\beta + (\Omega^{\beta\delta} + \Omega_{\beta\delta}\Omega^{\delta\delta})Y_\delta - (\Omega^{\beta\delta} + \Omega_{\beta\delta}\Omega^{\delta\delta})(\delta^* + \tilde{\delta}/2)\Omega_{\delta\beta}
\end{aligned}$$

$$=Y_\beta,$$

where the final equality follows from the facts $\Omega^{\beta\beta}+\Omega_{\beta\delta}\Omega^{\delta\beta}=1$ and $\Omega^{\beta\delta}+\Omega_{\beta\delta}\Omega^{\delta\delta}=0$ by the formula for blockwise inversion of a matrix. Theorem 3.1 of AK then provides that among all upper one-sided CIs with coverage of at least $(1-\alpha)$ for all $\delta\geq 0$

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

minimizes all maximum excess length quantiles over the $\delta\geq 0$ parameter space at quantile levels greater than α.

(ii) We first form the affine estimator $\hat{L}_{\tilde{\delta},\mathcal{F},\mathcal{G}}$ in (23) of AK, for $\mathcal{F}=\{(\beta,\delta')'\in\mathbb{R}^{k+1}:\delta\geq 0\}$ and $\mathcal{G}=\{(\beta,\delta')'\in\mathbb{R}^{k+1}:\delta=0\}$. The modulus of continuity in this setting is

$$\begin{aligned}\omega(\tilde{\delta};\mathcal{F},\mathcal{G}) &= \sup_{\gamma_1,\theta}\{\gamma_1-\beta\} \\ \text{s.t. } (\gamma_1-\beta)^2\Omega^{\beta\beta}-2(\gamma_1-\beta)\Omega^{\beta\delta}\delta+\delta'\Omega^{\delta\delta}\delta &\leq \tilde{\delta}^2, \quad \delta\geq 0. \end{aligned} \tag{17}$$

Here, the first order condition wrt δ_i is

$$-2\lambda[(\gamma_1-\beta)\Omega_i^{\beta\delta}-\Omega_{i,\cdot}^{\delta\delta}\delta]-\mu_i=0,$$

where $\lambda>0$ is the KKT multiplier associated with the first constraint, $\mu_i\geq 0$ is the KKT multiplier associated with the constraint $\delta_i\geq 0$ that satisfies the complementary slackness condition $\mu_i\delta_i=0$ and $\Omega_{i,\cdot}^{\delta\delta}$ denotes the i^{th} row of $\Omega^{\delta\delta}$. The solution to the modulus problem must therefore satisfy

$$\Omega_{i,\cdot}^{\delta\delta}\delta=(\gamma_1-\beta)\Omega_i^{\delta\beta}+\tilde{\mu}_i \tag{18}$$

for $i=1,...,k$ and some constants $\tilde{\mu}_i\geq 0$ such that $\tilde{\mu}_i\delta_i=0$. The solution to the modulus problem thus maximizes $(\gamma_1-\beta)$ amongst all γ_1,θ values that satisfy (17) and (18) for $i=1,...,k$ and some constants $\tilde{\mu}_i \geq 0$ such that $\tilde{\mu}_i \delta_i = 0$.

Next, we consider the candidate solutions to the modulus problem. Let $\delta^{(s)}$ ($\delta^{(-s)}$) denote a (possibly empty) subvector of δ that satisfies $\delta^{(s)} = 0$ ($\delta^{(-s)} \geq 0$) with $\tilde{\mu}^{(s)} > 0$ ($\tilde{\mu}^{(-s)} = 0$). Then, using the notational conventions introduced in Lemma 3 in what follows, the set of equations given in (18) implies

$$\delta^{(-s)} = (\gamma_1 - \beta)(\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1}\Omega^{\delta^{(-s)}\beta}, \tag{19}$$

where we use the convention that $(\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1}\Omega^{\delta^{(-s)}\beta} = 0$ for $\delta^{(s)} = \delta$, and the modulus problem simplifies to

$$\omega(\tilde{\delta};\mathcal{F},\mathcal{G}) = \sup_{\gamma_1,\beta}\{\gamma_1 - \beta\}$$
$$\text{s.t. } (\gamma_1 - \beta)^2(\Omega^{\beta\beta} - \Omega^{\beta\delta^{(-s)}}(\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1}\Omega^{\delta^{(-s)}\beta}) \leq \tilde{\delta}^2.$$

Recall that, given the definition of $\delta^{(s)}$ and $\delta^{(-s)}$, the constraint $\delta \geq 0$ is satisfied. Thus, we have

$$\omega(\tilde{\delta};\mathcal{F},\mathcal{G}) = \tilde{\delta}/\sqrt{\Omega^{\beta\beta} - \Omega^{\beta\delta^{(-s^{**})}}(\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}})^{-1}\Omega^{\delta^{(-s^{**})}\beta}},$$

where s^{**} is such that $\delta^{(-s^{**})}$ maximizes $\Omega^{\beta\delta^{(-s)}}(\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1}\Omega^{\delta^{(-s)}\beta}$ (subject to $\delta \geq 0$), with a solution given by

$$\gamma^*_{\tilde{\delta},\mathcal{F},\mathcal{G}} = \begin{pmatrix} \tilde{\delta}/(2\sqrt{\Omega^{\beta\beta} - \Omega^{\beta\delta^{(-s^{**})}}(\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}})^{-1}\Omega^{\delta^{(-s^{**})}\beta}}) \\ 0_{k\times 1} \end{pmatrix},$$
$$\theta^*_{\tilde{\delta},\mathcal{F},\mathcal{G}} = \begin{pmatrix} -\tilde{\delta}/(2\sqrt{\Omega^{\beta\beta} - \Omega^{\beta\delta^{(-s^{**})}}(\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}})^{-1}\Omega^{\delta^{(-s^{**})}\beta}}) \\ \delta^{**} \end{pmatrix}$$

and midpoint

$$\theta^*_{M,\tilde{\delta},\mathcal{F},\mathcal{G}} = (\theta^*_{\tilde{\delta},\mathcal{F},\mathcal{G}} + \gamma^*_{\tilde{\delta},\mathcal{F},\mathcal{G}})/2 = (0, \delta^{**\prime}/2)',$$

where δ^{**} has elements $\delta^{(s^{**})}$ and $\delta^{(-s^{**})}$. Formula (23) of AK thus yields

$$\begin{aligned}
\hat{L}_{\tilde{\delta},\mathcal{F},\mathcal{G}} &= \left(\tilde{\delta}\sqrt{\Omega^{\beta\beta}-\Omega^{\beta\delta^{(-s^{**})}}(\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}})^{-1}\Omega^{\delta^{(-s^{**})}\beta}}\right)^{-1}(\gamma^{*}_{\tilde{\delta},\mathcal{F},\mathcal{G}}-\theta^{*}_{\tilde{\delta},\mathcal{F},\mathcal{G}})'\Omega^{-1}(Y-\theta^{*}_{M,\tilde{\delta},\mathcal{F},\mathcal{G}}) \\
&= \left(\Omega^{\beta\beta}-\Omega^{\beta\delta^{(-s^{**})}}(\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}})^{-1}\Omega^{\delta^{(-s^{**})}\beta}\right)^{-1}(1,\ddot{\delta}')\begin{pmatrix}\Omega^{\beta\beta} & \Omega^{\beta\delta} \\ \Omega^{\delta\beta} & \Omega^{\delta\delta}\end{pmatrix}\begin{pmatrix} Y_{\beta} \\ Y_{\delta}-\delta^{**}/2\end{pmatrix} \\
&= Y_{\beta}+\frac{\Omega^{\beta\delta^{(s^{**})}}-\Omega^{\beta\delta^{(-s^{**})}}(\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}})^{-1}\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}}}{\Omega^{\beta\beta}-\Omega^{\beta\delta^{(-s^{**})}}(\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}})^{-1}\Omega^{\delta^{(-s^{**})}\beta}}Y_{\delta}^{(s^{**})} \\
&= Y_{\beta}-\Omega_{\beta\delta^{(s^{**})}}\Omega^{-1}_{\delta^{(s^{**})}\delta^{(s^{**})}}Y_{\delta}^{(s^{**})},
\end{aligned}$$

where $\ddot{\delta}=\sqrt{\Omega^{\beta\beta}-\Omega^{\beta\delta^{(-s^{**})}}(\Omega^{\delta^{(-s^{**})}\delta^{(-s^{**})}})^{-1}\Omega^{\delta^{(-s^{**})}\beta}}\,\delta^{**}/\tilde{\delta}$ and the last equality follows from Lemma 3. Similarly, Lemma 3 implies that

$$\omega(\tilde{\delta};\mathcal{F},\mathcal{G})=\tilde{\delta}\sqrt{1-\Omega_{\beta\delta^{(s^{**})}}\Omega^{-1}_{\delta^{(s^{**})}\delta^{(s^{**})}}\Omega_{\delta^{(s^{**})}\beta}}.$$

Theorem 3.1 of AK then provides that among all upper one-sided CIs with coverage of at least $(1-\alpha)$ for all $\delta\geq 0$

$$[Y_{\beta}-\Omega_{\beta\delta^{(s^{**})}}\Omega^{-1}_{\delta^{(s^{**})}\delta^{(s^{**})}}Y_{\delta}-z_{1-\alpha}\sqrt{1-\Omega_{\beta\delta^{(s^{**})}}\Omega^{-1}_{\delta^{(s^{**})}\delta^{(s^{**})}}\Omega_{\delta^{(s^{**})}\beta}},\infty) \tag{20}$$

minimizes all excess length quantiles at $\delta=0$ and quantile levels greater than α.

Next, we show that $s^{**}=s^{*}$. First, note that the excess length of any CI of the form given in (20)—for some $\delta^{(s^{**})}$—at $\delta=0$ and any quantile greater than α is equal to

$$c\sqrt{1-\Omega_{\beta\delta^{(s^{**})}}\Omega^{-1}_{\delta^{(s^{**})}\delta^{(s^{**})}}\Omega_{\delta^{(s^{**})}\beta}}$$

for some $c>0$. Furthermore, recall (from the discussion in the main text) that for any CI of this form to have coverage of at least $(1-\alpha)$ for all $\delta\geq 0$ we need $\Omega^{-1}_{\delta^{(s^{**})}\delta^{(s^{**})}}\Omega_{\delta^{(s^{**})}\beta}\geq 0$.[32] Therefore,

[32]Note that the condition $\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}\Omega_{\delta^{(s)}\beta}\geq 0$ can also be derived from the modulus problem. To

the CI that minimizes excess length at $\delta = 0$ and all quantiles greater than α, among all upper one-sided CIs with coverage of at least $(1-\alpha)$ for all $\delta \geq 0$, must also be equal to

$$[Y_\beta - \Omega_{\beta\delta^{(s^*)}} \Omega^{-1}_{\delta^{(s^*)}\delta^{(s^*)}} Y_\delta^{(s^*)} - \sqrt{1 - \Omega_{\beta\delta^{(s^*)}} \Omega^{-1}_{\delta^{(s^*)}\delta^{(s^*)}} \Omega_{\delta^{(s^*)}\beta}}\, z_{1-\alpha}, \infty),$$

where $\delta^{(s^*)}$ is a subvector of δ that maximizes $\Omega_{\beta\delta^{(s)}} \Omega^{-1}_{\delta^{(s)}\delta^{(s)}} \Omega_{\delta^{(s)}\beta}$ amongst all subvectors for which $\Omega^{-1}_{\delta^{(s)}\delta^{(s)}} \Omega_{\delta^{(s)}\beta} \geq 0$. □

Proof of Theorem 1. To prove the lower bound, note that

$$\begin{aligned} & P\Big(\beta \in \widehat{CI}^*_u(Y_\beta, Y_\delta, \Omega)\Big) \\ &= P\Big(\beta \geq Y_\beta - \min\Big\{ z_{1-\alpha+\gamma}, \Omega_{\beta\delta^{(s^*)}} \Omega^{-1}_{\delta^{(s^*)}\delta^{(s^*)}} Y_\delta^{(s^*)} + c\big(\Omega_{\beta\delta^{(s^*)}} \Omega^{-1}_{\delta^{(s^*)}\delta^{(s^*)}} \Omega_{\delta^{(s^*)}\beta}\big)\Big\}\Big) \\ &= P\Big(Z_1 \leq \min\Big\{ z_{1-\alpha+\gamma}, \Omega_{\beta\delta^{(s^*)}} \Omega^{-1}_{\delta^{(s^*)}\delta^{(s^*)}} Y_\delta^{(s^*)} + c\big(\Omega_{\beta\delta^{(s^*)}} \Omega^{-1}_{\delta^{(s^*)}\delta^{(s^*)}} \Omega_{\delta^{(s^*)}\beta}\big)\Big\}\Big), \end{aligned}$$

---

see this, note that the set of equations in (18) implies

$$\Omega^{\delta^{(s)}\delta^{(-s)}} \delta^{(-s)} = (\gamma_1 - \beta)\Omega^{\delta^{(s)}\beta} + \tilde{\mu}^{(s)}.$$

Plugging in the formula for $\delta^{(-s)}$ in equation (19), we get

$$\begin{aligned} & (\gamma_1-\beta)\Omega^{\delta^{(s)}\delta^{(-s)}} (\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1} \Omega^{\delta^{(-s)}\beta} && = (\gamma_1-\beta)\Omega^{\delta^{(s)}\beta} + \tilde{\mu}^{(s)} \\ \Leftrightarrow \quad & -\frac{\tilde{\mu}^{(s)}}{\gamma_1-\beta} && = \Omega^{\delta^{(s)}\beta} - \Omega^{\delta^{(s)}\delta^{(-s)}} (\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1} \Omega^{\delta^{(-s)}\beta} \\ \Leftrightarrow \quad & \frac{\tilde{\mu}^{(s)}}{(\gamma_1-\beta)(\Omega^{\beta\beta} - \Omega^{\beta\delta^{(-s)}} (\Omega^{\delta^{(-s)}\delta^{(-s)}})^{-1} \Omega^{\delta^{(-s)}\beta})} && = \Omega^{-1}_{\delta^{(s)}\delta^{(s)}} \Omega_{\delta^{(s)}\beta}, \end{aligned}$$

where the last step uses Lemma 3 (with sub- and superscripts interchanged). As the left hand side is non-negative, we conclude that $\Omega^{-1}_{\delta^{(s)}\delta^{(s)}} \Omega_{\delta^{(s)}\beta} \geq 0$.

where

$$\begin{pmatrix} Z_1 \\ Y_\delta^{(s^*)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \delta^{(s^*)} \end{pmatrix}, \begin{pmatrix} 1 & \Omega_{\beta\delta(s^*)} \\ \Omega_{\delta(s^*)\beta} & \Omega_{\delta(s^*)\delta(s^*)} \end{pmatrix} \right).$$

Since $\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\delta^{(s^*)} \geq 0$,

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

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

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

yielding the lower bound in the statement of the theorem for $\widehat{CI}_u^*(\cdot)$.

To prove the upper bound, note that for the probability to the left of the inequality in (29),

$$\begin{aligned} &P\Big(Z_1 \leq \min\Big\{z_{1-\alpha+\gamma}, \Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\delta^{(s^*)} + \tilde{Z}_2 + c\Big(\Omega_{\beta\delta(s^*)}\Omega^{-1}_{\delta(s^*)\delta(s^*)}\Omega^*_{\delta(s^*)\beta}\Big)\Big\}\Big) \\ &\leq P(Z_1 \leq z_{1-\alpha+\gamma}) = 1-\alpha+\gamma. \end{aligned}$$

$\square$

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

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

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 $(Z_1^*,\tilde{Z}_2^*,\tilde{Z}_3^*)'\sim(Z_1,\tilde{Z}_3,\tilde{Z}_2)'$. □

The following lemmata are used in the proofs of Theorems 2 and 3 and Proposition 4.

Lemma 4

The function $\tilde{c}:\widetilde{\mathcal{C}}\to\mathbb{R}_\infty$ 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 (13),

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

For $(\tilde{c},c_u,\tilde{\omega})\in\mathbb{R}_\infty\times\widetilde{\mathcal{C}}$ with $\omega_{12},\omega_{13}\neq 0$,

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

$$
\begin{aligned}
&-\int_{-\infty}^{\infty}\Bigg[\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}})\\
&\qquad +\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)\Bigg]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\\
&+\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\\
&+\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\\
&-\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\\
&-\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\\
&-\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),
\end{aligned}
$$

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,

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

$$+\int_{c_u-z_{1-\frac{\alpha-\gamma}{2}}}^{\infty}\int_{\tilde{z}_3-c_u-\tilde{c}}^{z_{1-\frac{\alpha-\gamma}{2}}-\tilde{c}}\frac{1}{\sigma(\tilde{\omega})}\phi\left(\frac{\tilde{z}_2+\tilde{c}-\mu(\tilde{z}_2,\tilde{z}_3)}{\sigma(\tilde{\omega})}\right)g(\tilde{z}_2,\tilde{z}_3)d\tilde{z}_2d\tilde{z}_3>0$$

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 (14) 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 5

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 (15),

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

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 4 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 Theorem 2. To prove the lower bound, note that

$$
\begin{aligned}
P\Big(\beta\in\widehat{CI}_t^{*}(Y_\beta,Y_\delta,\Omega)\Big)=P\Big(&Y_\beta-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}Y_\delta^{(s_1^*)}+c_\ell\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\}\le\beta \\
&\le Y_\beta+\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}Y_\delta^{(s_2^*)}+c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\}\Big) \\
=P\Big(&-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}Y_\delta^{(s_2^*)}+c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\}\le Y_\beta-\beta \\
&\le\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}Y_\delta^{(s_1^*)}+\tilde c\Big(c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big),\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\}\Big).
\end{aligned}
\tag{22}
$$

Since $\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}\delta^{(s_1^*)}\ge0$ and $\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}\delta^{(s_2^*)}\le0$, (22) can be bounded:

$$
\begin{aligned}
&P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}Y_\delta^{(s_2^*)}+c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\} \\
&\quad\le Z_1\le\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}Y_\delta^{(s_1^*)}+\tilde c\Big(c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big),\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\}\Big) \\
&=P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}\delta^{(s_2^*)}-\tilde Z_3+c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\} \\
&\quad\le Z_1\le\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}\delta^{(s_1^*)}+\tilde Z_2+\tilde c\Big(c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big),\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\}\Big) \\
&\ge P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde Z_3+c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\} \\
&\quad\le Z_1\le\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\tilde Z_2+\tilde c\Big(c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big),\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\}\Big)=1-\alpha
\end{aligned}
\tag{23}
$$

by the definition of $\tilde c(\cdot)$ in (15) and Lemma 5, where

$$
\begin{pmatrix} Z_1\\ \tilde Z_2\\ \tilde Z_3\end{pmatrix}\sim\mathcal N\left(\begin{pmatrix}0\\0\\0\end{pmatrix},\begin{pmatrix}1 & \Omega^*_{\beta\delta^{(s_1^*)}} & \Omega^*_{\beta\delta^{(s_2^*)}}\\ \Omega^*_{\delta^{(s_1^*)}\beta} & \Omega^*_{\delta^{(s_1^*)}\delta^{(s_1^*)}} & \Omega^*_{\delta^{(s_1^*)}\delta^{(s_2^*)}}\\ \Omega^*_{\delta^{(s_2^*)}\beta} & \Omega^*_{\delta^{(s_2^*)}\delta^{(s_1^*)}} & \Omega^*_{\delta^{(s_2^*)}\delta^{(s_2^*)}}\end{pmatrix}\right).
$$

To prove the upper bound, note that for the probability to the left of the equality in (23),

$$
\begin{aligned}
&P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega_{\beta\delta^{(s_2^*)}}\Omega^{-1}_{\delta^{(s_2^*)}\delta^{(s_2^*)}}Y_\delta^{(s_2^*)}+c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\} \\
&\le Z_1\le\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\Omega_{\beta\delta^{(s_1^*)}}\Omega^{-1}_{\delta^{(s_1^*)}\delta^{(s_1^*)}}Y_\delta^{(s_1^*)}+\tilde c\Big(c_u\Big(\tilde\Omega^{(s_1^*,s_2^*)}\Big),\tilde\Omega^{(s_1^*,s_2^*)}\Big)\Big\}\Big)
\end{aligned}
$$

$$\leq P\left(-z_{1-\frac{\alpha-\gamma}{2}} \leq Z_1 \leq z_{1-\frac{\alpha-\gamma}{2}}\right) = 1-\alpha+\gamma.$$

$\square$

Proof of Proposition 4. Very similar arguments to those given in the proof of Lemma 4 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 (16) is

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

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 $\tilde{c}:\widetilde{\mathcal{C}} \to \mathbb{R}_\infty$ is continuous by Lemma 4. The maximum theorem then implies the statement of the proposition. $\square$

C Theoretical Uniformity Results

We now present theoretical results ensuring the uniformly correct asymptotic coverage of both the one- and two-sided finite-sample CIs defined in (9) and (10), 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.[33] We assume that we have a consistent estimator $\widehat{\Sigma}_n$ of Σ at our disposal. In what follows, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of the matrix A, respectively.

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 assumptions hold.

Assumption 1

$b\in\mathbb{R}$ and $d\in\mathbb{R}_+^k$ for some positive integer k.

Assumption 2

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

Assumption 3

Under any sequence of parameters $\{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}=(b_{n,\mathfrak{b}},d_{n,\mathfrak{d}},\Sigma_{n,\Sigma^*},F_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}):n\geq1\}$ in Λ such that

$$\sqrt{n}(b_{n,\mathfrak{b}},d_{n,\mathfrak{d}})\rightarrow(\mathfrak{b},\mathfrak{d}), \tag{24}$$

$$\Sigma_{n,\Sigma^*}\rightarrow\Sigma^* \tag{25}$$

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

(i) $\widehat{\Sigma}_n$ exists and $\lambda_{\min}(\widehat{\Sigma}_n)>0$ with probability 1 for all $n\geq1$ and $\widehat{\Sigma}_n\xrightarrow{p}\Sigma^*$;

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

Assumptions 1–3 are a set of high-level conditions on the underlying DGP that are straightforward to verify in particular estimation contexts. More specifically, Assumptions 1 and 2 are standard

[33]We note that it is technically redundant to specify Σ as a separate element of the parameter λ from F since Σ is a function of F. However, we maintain this convention for notational convenience.

parameter space assumptions while Assumption 3 can typically be verified under standard dependence and moment conditions on the underlying data via laws of large numbers and central limit theorems. Importantly, although Assumption 3 is stated in terms of parameter sequences, it can be verified in terms of primitive conditions that do not involve parameter sequences. We refer the interested reader to Appendix C.1 for details in the context of the standard linear regression model.

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

Theorem 3

For $\alpha \in (0,1/2)$ and $\gamma \in (0,\alpha)$ and a parameter space Λ satisfying Assumptions 1–3,

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

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 $CI_{t,n}(\cdot)$.

Proof. Under Assumptions 1 and 2, standard subsequencing arguments in the uniform inference literature (see e.g., Andrews and Guggenberger, 2010) provide that showing

$$1-\alpha \leq \lim_{n\to\infty} P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(b_{n,\mathfrak{b}} \in CI_{\cdot,n}(\hat{b}_n,\hat{d}_n,\widehat{\Sigma}_n)\right) \leq 1-\alpha+\gamma \tag{26}$$

under all sequences $\{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*} : n \geq 1\}$ in Λ satisfying (24)–(25) such that the limit in (26) exists is sufficient for proving the statement of the theorem.

First, we verify (26) for $CI_{u,n}(\cdot)$. Let $\Omega^* = \mathrm{Diag}(\Sigma^*)^{-1/2}\Sigma^*\mathrm{Diag}(\Sigma^*)^{-1/2}$. For any $\{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*} : n \geq 1\}$ in Λ satisfying (24)–(25) such that the limit in (26) exists, the latter is equal to

$$\lim_{n\to\infty} P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Bigg(b_{n,\mathfrak{b}} \geq \hat{b}_n - \frac{\sqrt{\widehat{\Sigma}_{n,bb}}}{\sqrt{n}}\min\{z_{1-\alpha+\gamma},$$

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

$$=\lim_{n\to\infty}P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(\frac{\sqrt{n}(\hat{b}_n-b_{n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{n,bb}}}\le\min\{z_{1-\alpha+\gamma},\right.$$

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

Note that if $\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\not\geq 0$, $P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}(\hat{s}^*=s)\to 0$ since, by Assumption 3(i), $\hat{s}^*$ maximizes

$$\widehat{\Omega}_{n,bd^{(s)}}\widehat{\Omega}^{-1}_{n,d^{(s)}d^{(s)}}\widehat{\Omega}_{n,d^{(s)}b}\mathbf{1}(\widehat{\Omega}_{n,bd^{(s)}}\widehat{\Omega}^{-1}_{n,d^{(s)}d^{(s)}}\geq 0)$$

over $s\subseteq\{1,...,k\}$. Thus, we can bound the term in the middle of (26) from below by

$$\lim_{n\to\infty}\min_{s\subseteq\{1,...,k\}:\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\geq 0}P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(\frac{\sqrt{n}(\hat{b}_n-b_{n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{n,bb}}}\le\min\{z_{1-\alpha+\gamma},\right.$$

$$\left.\widehat{\Omega}_{n,bd^{(s)}}\widehat{\Omega}^{-1}_{n,d^{(s)}d^{(s)}}\mathrm{Diag}(\widehat{\Sigma}_{n,d^{(s)}d^{(s)}})^{-1/2}\sqrt{n}\hat{d}_n^{(s)}+c\Big(\widehat{\Omega}_{n,bd^{(s)}}\widehat{\Omega}^{-1}_{n,d^{(s)}d^{(s)}}\widehat{\Omega}_{n,d^{(s)}b}\Big)\Big\}\right). \tag{27}$$

Now, note that for any $s\subseteq\{1,...,k\}$ such that $\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\geq 0$,

$$P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(\frac{\sqrt{n}(\hat{b}_n-b_{n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{n,bb}}}\le\min\{z_{1-\alpha+\gamma},\right.$$

$$\left.\widehat{\Omega}_{n,bd^{(s)}}\widehat{\Omega}^{-1}_{n,d^{(s)}d^{(s)}}\mathrm{Diag}(\widehat{\Sigma}_{n,d^{(s)}d^{(s)}})^{-1/2}\sqrt{n}\hat{d}_n^{(s)}+c\Big(\widehat{\Omega}_{n,bd^{(s)}}\widehat{\Omega}^{-1}_{n,d^{(s)}d^{(s)}}\widehat{\Omega}_{n,d^{(s)}b}\Big)\Big\}\right)$$

$$\to\begin{cases}P\Big(Z_1\le\min\Big\{z_{1-\alpha+\gamma},\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}Y_\delta^{(s)}+c\big(\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\Omega^*_{d^{(s)}b}\big)\Big\}\Big) & \text{if } \|\mathfrak{d}^{(s)}\|<\infty,\\ P(Z_1\le z_{1-\alpha+\gamma}) & \text{if } \|\mathfrak{d}^{(s)}\|=\infty\end{cases} \tag{28}$$

by Assumptions 2 and 3 and Proposition 1, where

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

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 Assumptions 1, 2 and 3(i),

$$\begin{aligned} &P\Big(Z_1 \leq \min\Big\{z_{1-\alpha+\gamma}, \Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}Y^{(s)}_\delta + c\big(\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\Omega^*_{d^{(s)}b}\big)\Big\}\Big) \\ &= P\Big(Z_1 \leq \min\Big\{z_{1-\alpha+\gamma}, \Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\delta^{(s)} + \tilde{Z}_2 + c\big(\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\Omega^*_{d^{(s)}b}\big)\Big\}\Big) \\ &\geq P\Big(Z_1 \leq \min\Big\{z_{1-\alpha+\gamma}, \tilde{Z}_2 + c\big(\Omega^*_{bd^{(s)}}\Omega^{*-1}_{d^{(s)}d^{(s)}}\Omega^*_{d^{(s)}b}\big)\Big\}\Big) = 1-\alpha \end{aligned} \tag{29}$$

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,

$$P(Z_1 \leq z_{1-\alpha+\gamma}) = 1-\alpha+\gamma > 1-\alpha. \tag{30}$$

Together, (27)–(30) yield the lower bound in (26) for $CI_{u,n}(\cdot)$.

To prove the upper bound in (26) for $CI_{u,n}(\cdot)$, note that the term in the middle of (26) is bounded above by

$$\lim_{n\to\infty} P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(\frac{\sqrt{n}(\hat{b}_n - b_{n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{n,bb}}} \leq z_{1-\alpha+\gamma},\right) = 1-\alpha+\gamma.$$

Next, we show (26) holds for $CI_{t,n}(\cdot)$. Using analogous reasoning to the $CI_{u,n}(\cdot)$ case above, for any $\{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*} : n \geq 1\}$ in Λ satisfying (24)–(25) such that the limit in (26) exists, the latter is equal to

$$\begin{aligned} &\lim_{n\to\infty} P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Bigg(\hat{b}_n - \frac{\sqrt{\widehat{\Sigma}_{n,bb}}}{\sqrt{n}}\min\Big\{z_{1-\frac{\alpha-\gamma}{2}}, \\ &\quad \widehat{\Omega}_{n,bd^{(\hat{s}^*_1)}}\widehat{\Omega}^{-1}_{n,d^{(\hat{s}^*_1)}d^{(\hat{s}^*_1)}}\mathrm{Diag}(\widehat{\Sigma}_{n,d^{(\hat{s}^*_1)}d^{(\hat{s}^*_1)}})^{-1/2}\sqrt{n}\hat{d}^{(\hat{s}^*_1)}_n + c_\ell\Big(\widehat{\widetilde{\Omega}}^{(\hat{s}^*_1,\hat{s}^*_2)}_n\Big)\Big\} \leq b_{n,\mathfrak{b}} \leq \hat{b}_n \end{aligned}$$

$$+\frac{\sqrt{\widehat{\Sigma}_{n,bb}}}{\sqrt{n}}\min\left\{z_{1-\frac{\alpha-\gamma}{2}},-\widehat{\Omega}_{n,bd^{(\hat{s}_2^*)}}\widehat{\Omega}^{-1}_{n,d^{(\hat{s}_2^*)}d^{(\hat{s}_2^*)}}\mathrm{Diag}(\widehat{\Sigma}_{n,d^{(\hat{s}_2^*)}d^{(\hat{s}_2^*)}})^{-1/2}\sqrt{n}\hat{d}_n^{(\hat{s}_2^*)}+c_u\left(\widehat{\tilde{\Omega}}_n^{(\hat{s}_1^*,\hat{s}_2^*)}\right)\right\}\Bigg)$$

$$=\lim_{n\to\infty}P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Bigg(-\min\left\{z_{1-\frac{\alpha-\gamma}{2}},-\widehat{\Omega}_{n,bd^{(\hat{s}_2^*)}}\widehat{\Omega}^{-1}_{n,d^{(\hat{s}_2^*)}d^{(\hat{s}_2^*)}}\mathrm{Diag}(\widehat{\Sigma}_{n,d^{(\hat{s}_2^*)}d^{(\hat{s}_2^*)}})^{-1/2}\sqrt{n}\hat{d}_n^{(\hat{s}_2^*)}+c_u\left(\widehat{\tilde{\Omega}}_n^{(\hat{s}_1^*,\hat{s}_2^*)}\right)\right\}$$
$$\leq\frac{\sqrt{n}(\hat{b}_n-b_{n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{n,bb}}} \tag{31}$$
$$\leq\min\left\{z_{1-\frac{\alpha-\gamma}{2}},\widehat{\Omega}_{n,bd^{(\hat{s}_1^*)}}\widehat{\Omega}^{-1}_{n,d^{(\hat{s}_1^*)}d^{(\hat{s}_1^*)}}\mathrm{Diag}(\widehat{\Sigma}_{n,d^{(\hat{s}_1^*)}d^{(\hat{s}_1^*)}})^{-1/2}\sqrt{n}\hat{d}_n^{(\hat{s}_1^*)}+\tilde{c}\left(c_u\left(\widehat{\tilde{\Omega}}_n^{(\hat{s}_1^*,\hat{s}_2^*)}\right),\widehat{\tilde{\Omega}}_n^{(\hat{s}_1^*,\hat{s}_2^*)}\right)\right\}\Bigg)$$

and can be bounded below by

$$\lim_{n\to\infty}\min_{s_1,s_2\subseteq\{1,\ldots,k\}:\Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}}\geq 0,\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}}\leq 0}P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\Bigg(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},$$
$$-\widehat{\Omega}_{n,bd^{(s_2)}}\widehat{\Omega}^{-1}_{n,d^{(s_2)}d^{(s_2)}}\mathrm{Diag}(\widehat{\Sigma}_{n,d^{(s_2)}d^{(s_2)}})^{-1/2}\sqrt{n}\hat{d}_n^{(s_2)}+c_u\left(\widehat{\tilde{\Omega}}_n^{(s_1,s_2)}\right)\Big\}\leq\frac{\sqrt{n}(\hat{b}_n-b_{n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{n,bb}}} \tag{32}$$
$$\leq\min\left\{z_{1-\frac{\alpha-\gamma}{2}},\widehat{\Omega}_{n,bd^{(s_1)}}\widehat{\Omega}^{-1}_{n,d^{(s_1)}d^{(s_1)}}\mathrm{Diag}(\widehat{\Sigma}_{n,d^{(s_1)}d^{(s_1)}})^{-1/2}\sqrt{n}\hat{d}_n^{(s_1)}+\tilde{c}\left(c_u\left(\widehat{\tilde{\Omega}}_n^{(s_1,s_2)}\right),\widehat{\tilde{\Omega}}_n^{(s_1,s_2)}\right)\right\}\Bigg).$$

Since $\widehat{\tilde{\Omega}}_n^{(s_1,s_2)}\xrightarrow{p}\tilde{\Omega}^{*(s_1,s_2)}$ under $\{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}:n\geq 1\}$ as $n\to\infty$ by Assumptions 2 and 3(i), there exists a subsequence $\{l_n:n\geq 1\}$ of $\{n:n\geq 1\}$ such that $\widehat{\tilde{\Omega}}_{l_n}^{(s_1,s_2)}\xrightarrow{a.s.}\tilde{\Omega}^{*(s_1,s_2)}$ 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 4, 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 4, this implies that

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

$$\leq \frac{\sqrt{h_n}(\hat{b}_{h_n} - b_{h_n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{h_n,bb}}} \tag{33}$$

$$\leq \min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, \widehat{\Omega}_{h_n,bd^{(s_1)}} \widehat{\Omega}^{-1}_{h_n,d^{(s_1)}d^{(s_1)}} \mathrm{Diag}(\widehat{\Sigma}_{h_n,d^{(s_1)}d^{(s_1)}})^{-1/2} \sqrt{h_n} \hat{d}^{(s_1)}_{h_n} + \tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right), \tilde{\Omega}^{*(s_1,s_2)} \right) \right\} \Big).$$

For $\Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}} \geq 0$ and $\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}} \leq 0$, $\Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}}\delta^{(s_1)} \geq 0$ and $\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}}\delta^{(s_2)} \leq 0$ so that if $\|\mathfrak{d}^{(s_1)}\|, \|\mathfrak{d}^{(s_2)}\| < \infty$, by Assumptions 1, 2 and 3(i), (33) is equal to

$$\begin{aligned}
& P\Big(-\min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, -\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}} Y_\delta^{(s_2)} + c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right) \right\} \\
& \quad \leq Z_1 \leq \min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, \Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}} Y_\delta^{(s_1)} + \tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right), \tilde{\Omega}^{*(s_1,s_2)} \right) \right\} \Big) \\
& = P\Big(-\min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, -\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}} \delta^{(s_2)} - \tilde{Z}_3 + c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right) \right\} \\
& \quad \leq Z_1 \leq \min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, \Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}} \delta^{(s_1)} + \tilde{Z}_2 + \tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right), \tilde{\Omega}^{*(s_1,s_2)} \right) \right\} \Big) \\
& \geq P\Big(-\min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, -\tilde{Z}_3 + c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right) \right\} \\
& \quad \leq Z_1 \leq \min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, \tilde{Z}_2 + \tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right), \tilde{\Omega}^{*(s_1,s_2)} \right) \right\} \Big) = 1-\alpha
\end{aligned} \tag{34}$$

using the definition of $\tilde{c}(\cdot)$ in (15) and Lemma 5, where

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

For $\Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}} \geq 0$ and $\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}} \leq 0$, $\Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}}\delta^{(s_1)} \geq 0$ so that if $\|\mathfrak{d}^{(s_1)}\| < \infty$ and $\|\mathfrak{d}^{(s_2)}\| = \infty$, by Assumptions 1, 2 and 3(i), (33) is equal to

$$\begin{aligned}
& P\Big(-z_{1-\frac{\alpha-\gamma}{2}} \leq Z_1 \leq \min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, \Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}} Y_\delta^{(s_1)} + \tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right), \tilde{\Omega}^{*(s_1,s_2)} \right) \right\} \Big) \\
& \quad \geq P\Big(-z_{1-\frac{\alpha-\gamma}{2}} \leq Z_1 \leq \min\left\{ z_{1-\frac{\alpha-\gamma}{2}}, \tilde{Z}_2 + \tilde{c}\left(c_u^*\left(\tilde{\Omega}^{*(s_1,s_2)}\right), \tilde{\Omega}^{*(s_1,s_2)} \right) \right\} \Big)
\end{aligned} \tag{35}$$

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

by the definition of $\tilde{c}(\cdot)$ in (15) and Lemma 5. For $\Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}}\geq 0$ and $\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}}\leq 0$, $\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}}\delta^{(s_2)}\leq 0$ so that if $\|\mathfrak{d}^{(s_1)}\|=\infty$ and $\|\mathfrak{d}^{(s_2)}\|<\infty$, by Assumptions 1, 2 and 3(i), (33) is equal to

$$
\begin{aligned}
&P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}}Y_\delta^{(s_2)}+c_u^*\Big(\tilde{\Omega}^{*(s_1,s_2)}\Big)\Big\}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}}\Big)\\
&\qquad\geq P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{Z}_3+c_u^*\Big(\tilde{\Omega}^{*(s_1,s_2)}\Big)\Big\}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}}\Big) \qquad (36)\\
&\geq P\Big(-\min\Big\{z_{1-\frac{\alpha-\gamma}{2}},-\tilde{Z}_3+c_u^*\Big(\tilde{\Omega}^{*(s_1,s_2)}\Big)\Big\}\leq Z_1\leq \min\Big\{z_{1-\frac{\alpha-\gamma}{2}},\tilde{Z}_2+\tilde{c}\Big(c_u^*\Big(\tilde{\Omega}^{*(s_1,s_2)}\Big),\tilde{\Omega}^{*(s_1,s_2)}\Big)\Big\}\Big)=1-\alpha
\end{aligned}
$$

by the definition of $\tilde{c}(\cdot)$ in (15) and Lemma 5. Finally, if $\|\mathfrak{d}^{(s_1)}\|,\|\mathfrak{d}^{(s_2)}\|=\infty$, for $\Omega^*_{bd^{(s_1)}}\Omega^{*-1}_{d^{(s_1)}d^{(s_1)}}\geq 0$ and $\Omega^*_{bd^{(s_2)}}\Omega^{*-1}_{d^{(s_2)}d^{(s_2)}}\leq 0$, (33) is equal to

$$
P\Big(z_{1-\frac{\alpha-\gamma}{2}}\leq Z_1\leq z_{1-\frac{\alpha-\gamma}{2}}\Big)=1-\alpha+\gamma>1-\alpha. \qquad (37)
$$

by Assumptions 2 and 3. Together, (31)–(37) yield the lower bound in (26) for $CI_{t,n}(\cdot)$.

To prove the upper bound in (26) for $CI_{t,n}(\cdot)$, note that the term in the middle of (26) is bounded above by

$$
\lim_{n\to\infty}P_{\lambda_{n,\mathfrak{b},\mathfrak{d},\Sigma^*}}\left(z_{1-\frac{\alpha-\gamma}{2}}\leq\frac{\sqrt{n}(\hat{b}_n-b_{n,\mathfrak{b}})}{\sqrt{\widehat{\Sigma}_{n,bb}}}\leq z_{1-\frac{\alpha-\gamma}{2}},\right)=1-\alpha+\gamma.
$$

□

C.1 Parameter Space for the Standard Linear Regression Model

In this section, we provide details for parameter spaces satisfying the high level conditions of Assumptions 1–3 above 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, $x_i \in \mathbb{R}^k$ are control variables with *known positive partial effects* $d \geq 0$ on y_i, $w_i \in \mathbb{R}^l$ 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 $k+1$ entries of $(\sum_{i=1}^n h_i h_i')^{-1} \sum_{i=1}^n h_i y_i$. 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 assumed to include parameters $\tilde{\lambda} = (b, d, c, \mathcal{V}, Q, F)$ such that for some finite $\kappa > 0$, the following conditions hold.

Assumption Reg1

$b \in \mathbb{R}$, $d \in \mathbb{R}_+^k$ and $c \in \mathbb{R}^l$.

Assumption Reg2

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

Assumption Reg3

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

Assumption Reg4

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 (24) holds, $\mathcal{V}_{n,\mathcal{V}^*} \to \mathcal{V}^*$ and $Q_{n,Q^*} \to Q^*$ for some $\mathcal{V}^*, Q^* \in \Phi$, the following conditions hold:

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

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

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

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

Assumption Reg1 imposes known sign restrictions for the nuisance coefficients d while letting the coefficient of interest b and the other nuisance coefficients c remain unrestricted. Assumptions Reg2 and Reg3 are standard conditions ensuring the existence of asymptotic covariance matrices while Assumptions Reg4(i)–(iii) are high level conditions that guarantee consistent estimators of these covariance matrices are available, typically shown via application of a law of large numbers.

Finally, Assumption Reg4(iv) 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.

Note that for Σ equal to the upper left $(k+1)\times(k+1)$ submatrix of $Q^{-1}\mathcal{V}Q^{-1}$ and $\widehat{\Sigma}_n$ equal to the upper left $(k+1)\times(k+1)$ submatrix of $\widehat{Q}_n^{-1}\widehat{\mathcal{V}}_n\widehat{Q}_n^{-1}$, Assumptions Reg1–Reg4 on the parameter space $\widetilde{\Lambda}$ imply Assumptions 1–3 for the parameter space Λ. More specifically, Assumption Reg1 implies Assumption 1, Assumptions Reg2 and Reg3 imply Assumption 2, Assumptions Reg4(i)–(iii) imply Assumption 3(i) and Assumptions Reg4(iii) and (iv) imply Assumption 3(ii).

Assumptions Reg1–Reg4 on the parameter space $\widetilde{\Lambda}$ are 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. For completeness, we reproduce here two sets of weak low-level sufficient conditions from Section 3 of McCloskey (2020) that guarantee the high-level Assumption Reg4 holds 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. We refer the interested reader to McCloskey (2020) for a discussion of how these conditions imply that Assumption Reg4 holds.

For the first set of conditions, applicable to randomly sampled data, suppose that $\widehat{\mathcal{V}}_n$ is equal to the heteroskedasticity-robust estimator of White (1980) and that F satisfies the following conditions for some fixed constants $C<\infty$ and $\vartheta>0$.

Assumption RS1

$\{(h_i',\varepsilon_i)':\ i\geq 1\}$ are i.i.d. under F.

Assumption RS2

$E_F|h_{i,j}h_{i,m}\varepsilon_i^2|^{1+\vartheta}\leq C$ for $j,m=1,\ldots,k+l+1$.

Assumption RS3

$E_F|h_{i,j}|^{4+\vartheta},E_F|\varepsilon_i|^{2+\vartheta}\leq C$ for $j=1,\ldots,k+l+1$.

For the second set of conditions, applicable to dependent data, suppose that $\widehat{\mathcal{V}}_n$ is equal to the heteroskedasticity and autocorrelation-robust estimator of Newey and West (1987) and that (along with the conditions in Newey and West (1987) imposed on user-chosen parameters) F satisfies the following conditions for some fixed constants $C<\infty$, $\vartheta>\tau>2$, $r\geq 1$ and $\delta>0$.

Assumption TS1

$\{(h_i',e_i)' : i\geq 1\}$ are stationary strong mixing under F with strong mixing numbers $\alpha_F(m)$.

Assumption TS2

$\alpha_F(m)=O(m^{-\vartheta\tau/(\vartheta-\tau)})$ as $m\to\infty$ and $E_F|h_{i,j}|^{\vartheta}, E_F|h_{i,j}e_i|^{\vartheta}\leq C$ for $j=1,...,k+l+1$.

Assumption TS3

$\alpha_F(m)=O(m^{-2r/(r-1)})$ as $m\to\infty$ and $E_F|h_{i,j}|^{2+\delta}, E_F|h_{i,j}e_i|^{4(r+\delta)}\leq C$ for $j=1,...,k+l+1$.

D Near optimality results

The NO (switching) tests whose power functions are displayed in Figures 2, 5 and 6 are obtained using the algorithm given in Section A.2.2. of EMW.[34] We closely follow the implementation of their running example for which the testing problem is given by

$$H_0:\beta=0,\ \delta\geq 0 \text{ vs. } H_1:\beta\neq 0,\ \delta\geq 0, \tag{38}$$

where δ is scalar. In particular, we use the same discretization of the null parameter space (given by uniform distributions on the intervals $\{[0,0.04],[0,0.5],[0.5,1],[1,1.5],...,[12,12.5]\}$), the same number of simulation draws (i.e., $N_0=20{,}000$ and $N_1=100{,}000$), the same switching rule (i.e., "switch to the standard test" when $Y_\delta>6$) and we also choose $\epsilon=0.005$.[35]

For the two-sided testing problem given in (38), we first compare the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ to the NO (switching) test that uses the same weights as EMW, i.e., β takes values -2 and 2 with equal probability and δ is uniformly distributed on [0,9].

[34] The rejection frequencies in Figures 2, 5 and 6 are obtained using 20,000 simulation draws. The underlying grid of β values is $\{-4,-3.9,...,3.9,4\}$.

[35] Here, we borrow the notation from EMW for ease of reference.

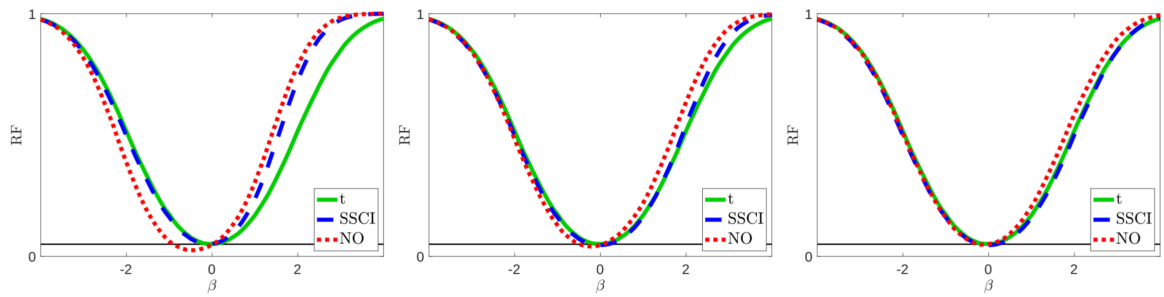

Figure 5: Rejection frequency as a function of β of standard two-sided t-test (t - solid), test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ using $\gamma=\alpha/10$ (SSCI - dashed) and the NO test (NO - dotted) for testing (38) with d=0,1,2 from left to right for α=0.05 and ρ=0.7.

Figure 5 shows the rejection frequency of the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ using $\gamma=\alpha/10$, the aforementioned NO test and the standard two-sided t-test (for comparison) for testing (38) at $\alpha=0.05$, as a function of β for different values of d when $\rho=0.7$. Here, in contrast to the one-sided case depicted in Figure 2, the power functions of the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ and the NO test are somewhat different. The test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ is almost unbiased, while the NO test sacrifices power at small negative values of β when δ is (close to) zero, even introducing some bias, for greater power at positive positive values of β. Furthermore, given the above weights, the weighted average power (WAP) of the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ only comes within $\epsilon=0.0278$ of the upper bound, which equals 53.34%. Therefore, we cannot claim the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ to be NO *with respect to the above weights*.

Although the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ is not NO with respect to the weights used by EMW, the power function of the NO test using these weights may not be desirable for practical use. For example, a researcher may desire more symmetric power over the β alternatives or a less biased test. For this reason, we construct an "alternative" set of weights and show that the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ *is* NO with respect to these weights. These "alternative" weights specify a discrete distribution over "support points" in the alternative space and are given in Table 5.

Figure 6 is analogous to Figure 5 except that the NO test uses the "alternative" weights given in Table 5. Here, the power functions of the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ and the NO test are visually indistinguishable. Furthermore, the WAP of the test implied by $\widehat{CI}_t^*(Y_\beta,Y_\delta,\Omega)$ is within

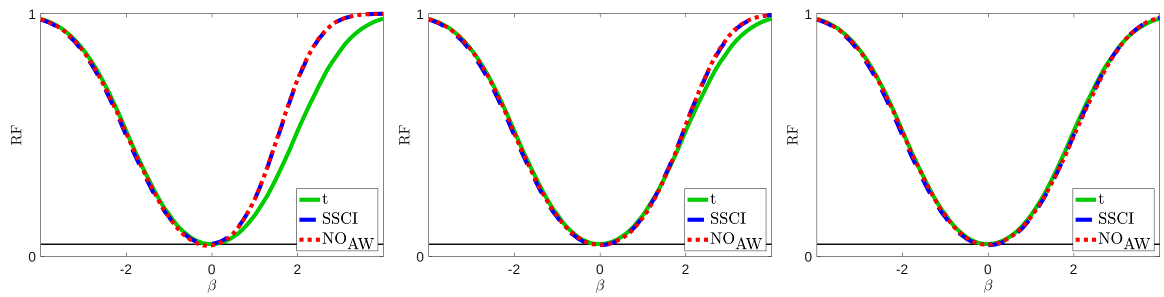

Figure 6: Rejection frequency as a function of β of standard two-sided t-test (t - solid), test implied by $\widehat{CI}^*_t(Y_\beta,Y_\delta,\Omega)$ using $\gamma=\alpha/10$ (SSCI - dashed) and the NO test using "alternative" weights (NO$_{\text{AW}}$ - dotted) for testing (38) with $d=0,1,2$ from left to right for $\alpha=0.05$ and $\rho=0.7$.

$\epsilon=0.0066$ of the upper bound, which equals 17.85%.[36] Thus, the test implied by $\widehat{CI}^*_t(Y_\beta,Y_\delta,\Omega)$ is NO with respect to these "alternative" weights that yield arguably a more desirable power function than the weights examined by EMW.

[36]Note that the upper bound on WAP is somewhat smaller with respect to the "alternative" weights because the latter give non-zero weight to values of β that are closer to the null hypothesis, presumably an empirically important region of the parameter space.

Table 5: Support points and probabilities for F (using the notation in EMW)

β	δ	prob
-2.0000	0.0000	0.0114
-1.0000	0.0000	0.0486
-0.5000	0.0000	0.1072
-0.3000	0.0000	0.1201
-0.1000	0.0000	0.0935
0.1000	0.0000	0.0666
0.3000	0.0000	0.0500
0.5000	0.0000	0.0292
1.0000	0.0000	0.0042
2.0000	0.0000	0.0024
-2.0000	1.0000	0.0085
-1.0000	1.0000	0.0172
1.0000	1.0000	0.0319
2.0000	1.0000	0.0031
-2.0000	2.0000	0.0043
-1.0000	2.0000	0.0133
1.0000	2.0000	0.0307
2.0000	2.0000	0.0315
-2.0000	3.0000	0.0068
-1.0000	3.0000	0.0112
1.0000	3.0000	0.0153
2.0000	3.0000	0.0151
-2.0000	4.0000	0.0101
-1.0000	4.0000	0.0151
1.0000	4.0000	0.0112
2.0000	4.0000	0.0041
-2.0000	5.0000	0.0150
-1.0000	5.0000	0.0226
1.0000	5.0000	0.0168
2.0000	5.0000	0.0042
-2.0000	6.0000	0.0063
-1.0000	6.0000	0.0187
1.0000	6.0000	0.0171
2.0000	6.0000	0.0074
-2.0000	7.0000	0.0065
-1.0000	7.0000	0.0123
1.0000	7.0000	0.0163
2.0000	7.0000	0.0066
-2.0000	8.0000	0.0072
-1.0000	8.0000	0.0135
1.0000	8.0000	0.0164
2.0000	8.0000	0.0063
-2.0000	9.0000	0.0072
-1.0000	9.0000	0.0145
1.0000	9.0000	0.0162
2.0000	9.0000	0.0063

The table lists support points, indexed by β and δ, and corresponding probabilities (prob) for the weighting function F.

E Additional Tables

Table 6: Coefficients for 6[th] order polynomial approximation of $c_u(\tilde{\omega})$ for $\alpha = 0.01$ and $\gamma = \alpha/10$

	1	ω_{12}	ω_{12}^2	ω_{12}^3	ω_{12}^4	ω_{12}^5	ω_{12}^6
1	2.6091	1.4378	−4.7977	12.2591	−20.5823	18.2815	−6.5866
ω_{13}	1.1854	−1.1672	3.6035	−2.5234	0.2467	0.6751	
ω_{13}^2	−16.4621	−2.1843	−2.6765	0.8411	−0.6847		
ω_{13}^3	63.1856	8.4153	1.0849	0.7850			
ω_{13}^4	−128.0372	−9.2032	−0.3625				
ω_{13}^5	123.3096	3.1479					
ω_{13}^6	−45.5050						

The table lists the coefficients for the 6[th] order polynomial approximation of $c_u(\tilde{\omega})$ in terms of ω_{12} and ω_{13} for $\alpha = 0.01$ and $\gamma = \alpha/10$. For example, the coefficient on the interaction term $\omega_{12}^4 \times \omega_{13}^2$ equals -0.6847.

Table 7: Coefficients for 6[th] order polynomial approximation of $c_u(\tilde{\omega})$ for $\alpha = 0.1$ and $\gamma = \alpha/10$

	1	ω_{12}	ω_{12}^2	ω_{12}^3	ω_{12}^4	ω_{12}^5	ω_{12}^6
1	1.6552	1.2890	−4.8501	14.0485	−23.9082	20.3891	−7.0186
ω_{13}	1.2271	0.0224	−0.6555	0.7875	1.0308	−0.5813	
ω_{13}^2	−11.7243	−2.0585	3.7550	−5.0051	1.5399		
ω_{13}^3	43.6253	3.2898	−1.7097	1.1221			
ω_{13}^4	−87.8291	−2.6854	0.6640				
ω_{13}^5	84.6893	0.5102					
ω_{13}^6	−31.4176						

The table lists the coefficients for the 6[th] order polynomial approximation of $c_u(\tilde{\omega})$ in terms of ω_{12} and ω_{13} for $\alpha = 0.1$ and $\gamma = \alpha/10$. For example, the coefficient on the interaction term $\omega_{12}^4 \times \omega_{13}^2$ equals 1.5399.

Table 8: 95% confidence intervals for treatment effects on summary index of antisocial behaviors (times minus one) imposing $\alpha_3^* \geq 0$

	$\hat{b}$	SE	SSCI	EL	SCI	EL	Ratio
T	0.0829	0.0929	$[-0.0747, \infty)$	0.1576	$[-0.0700, \infty)$	0.1529	1.0307
C	−0.1316	0.0969	$[-0.2959, \infty)$	0.1643	$[-0.2910, \infty)$	0.1594	1.0307
B	0.2468	0.0883	$[0.1025, \infty)$	0.1442	$[0.1015, \infty)$	0.1453	0.9929

The table shows point estimates ($\hat{b}$), standard errors (SE), our proposed CIs (SSCI), standard CIs (SCI), the corresponding ("excess") lengths ((E)L), and their ratio (Ratio) for the treatment effects of "therapy" (T), "cash" (C) and "both" (B). All CIs are upper one-sided and it is assumed that all treatment effects are known to be non-negative.

F Computational details

F.1 One-sided CI

Some of the results in the paper rely on the function $c(\omega)$ implicitly defined by (5) and the expected excess length of our proposed CI. Next, we give a detailed description of how we numerically compute them.

F.1.1 $c(\omega)$

In order to compute $c(\omega)$, we need to compute the probability $P(Z_1 > \min\{z_{1-\alpha+\gamma}, \tilde{Z}_2 + c\})$ for some $c \in \mathbb{R}_+$. Instead of simulating from a bivariate normal we use a "conditioning argument" and simulate from a univariate normal. This serves two purposes: First, it makes the numerical approximation of $P(Z_1 > \min\{z_{1-\alpha+\gamma}, \tilde{Z}_2 + c\})$ smooth, as a function of c. Second, it improves the accuracy of the approximation for a given number of simulation draws. Note that

$$P(Z_1 > \min\{z_{1-\alpha+\gamma}, \tilde{Z}_2 + c\}) = \int_{-\infty}^{\infty} P(Z_1 > \min\{z_{1-\alpha+\gamma}, \tilde{z}_2 + c\} | \tilde{Z}_2 = \tilde{z}_2) \phi(\tilde{z}_2 | 0, \omega) d\tilde{z}_2,$$

where $\phi(\cdot|\mu,\sigma^2)$ denotes the pdf of a normal with mean μ and variance σ^2. Furthermore, note that $Z_1 | \tilde{Z}_2 = \tilde{z}_2 \sim N(\tilde{z}_2, 1-\omega)$ such that

$$\begin{aligned}
P(Z_1 > \min\{z_{1-\alpha+\gamma}, \tilde{z}_2 + c\} | \tilde{Z}_2 = \tilde{z}_2) &= \begin{cases} P(Z_1 > z_{1-\alpha+\gamma} | \tilde{Z}_2 = \tilde{z}_2) & \text{if } \tilde{z}_2 + c > z_{1-\alpha+\gamma} \\ P(Z_1 > \tilde{z}_2 + c | \tilde{Z}_2 = \tilde{z}_2) & \text{if } \tilde{z}_2 + c < z_{1-\alpha+\gamma} \end{cases} \\
&= \begin{cases} \Phi\left(\frac{\tilde{z}_2 - z_{1-\alpha+\gamma}}{\sqrt{1-\omega}}\right) & \text{if } \tilde{z}_2 + c > z_{1-\alpha+\gamma} \\ \Phi\left(-\frac{c}{\sqrt{1-\omega}}\right) & \text{if } \tilde{z}_2 + c < z_{1-\alpha+\gamma} \end{cases}.
\end{aligned}$$

Therefore, $P(Z_1 > \min\{z_{1-\alpha+\gamma}, \tilde{Z}_2 + c\})$ is equal to

$$\int_{-\infty}^{z_{1-\alpha+\gamma}-c} \Phi\left(-\frac{c}{\sqrt{1-\omega}}\right) \phi(\tilde{z}_2|0,\omega) d\tilde{z}_2 + \int_{z_{1-\alpha+\gamma}-c}^{\infty} \Phi\left(\frac{\tilde{z}_2 - z_{1-\alpha+\gamma}}{\sqrt{1-\omega}}\right) \phi(\tilde{z}_2|0,\omega) d\tilde{z}_2. \tag{39}$$

We compute $P(Z_1 > \min\{z_{1-\alpha+\gamma}, \tilde{Z}_2 + c\})$ based on (39) using 1,000,000 draws from a univariate normal.

F.1.2 Expected excess length

Let $\psi = \Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}$ and $\omega = \Omega_{\beta\delta^{(s)}}\Omega^{-1}_{\delta^{(s)}\delta^{(s)}}\Omega_{\delta^{(s)}\beta}$. Then, the expected excess length of $[Y_\beta - \psi Y_\delta^{(s)} - c, \infty)$ is given by

$$
\begin{aligned}
&-E[Y_\beta - \min\{z_{1-\alpha+\gamma}, \psi Y_\delta^{(s)} + c\} - \beta] \\
&= E[\min\{z_{1-\alpha+\gamma}, \psi Y_\delta^{(s)} + c\}] \\
&= z_{1-\alpha+\gamma} P(\psi Y_\delta^{(s)} + c > z_{1-\alpha+\gamma}) + E[\psi Y_\delta^{(s)} + c | \psi Y_\delta^{(s)} + c \leq z_{1-\alpha+\gamma}] \\
&\times P(\psi Y_\delta^{(s)} + c \leq z_{1-\alpha+\gamma}) \\
&= z_{1-\alpha+\gamma}\left[1 - \Phi\left(\frac{z_{1-\alpha+\gamma} - \psi\delta^{(s)} - c}{\sqrt{\omega}}\right)\right] + \left[\psi\delta^{(s)} + c - \sqrt{\omega}\lambda\left(\frac{z_{1-\alpha+\gamma} - \psi\delta^{(s)} - c}{\sqrt{\omega}}\right)\right] \\
&\times \Phi\left(\frac{z_{1-\alpha+\gamma} - \psi\delta^{(s)} - c}{\sqrt{\omega}}\right),
\end{aligned}
$$

where $\lambda(x) \equiv \phi(x)/\Phi(x)$.

F.2 Two-sided CI

Some of the results in the paper rely on the probability underlying (15) and the expected length of our proposed CI. Next, we give a detailed description of how we numerically compute them.

F.2.1 Probability underlying (15)

We need to compute $P(-\min\{z_{1-(\alpha-\gamma)/2}, -\tilde{Z}_3 + c_u\} \leq Z_1 \leq \min\{z_{1-(\alpha-\gamma)/2}, \tilde{Z}_2 + c_\ell\})$ for some $c_u, c_\ell \in \mathbb{R}_+$. In what follows, let $z_. = z_{1-(\alpha-\gamma)/2}$. As before, note that

$$
\begin{aligned}
&P(-\min\{z_., -\tilde{Z}_3 + c_u\} \leq Z_1 \leq \min\{z_., \tilde{Z}_2 + c_\ell\}) \\
&= \int_{\mathbb{R}^2} P(-\min\{z_., -\tilde{z}_3 + c_u\} \leq Z_1 \leq \min\{z_., \tilde{z}_2 + c_\ell\} | \tilde{Z}_2 = \tilde{z}_2, \tilde{Z}_3 = \tilde{z}_3) \\
&\qquad \phi(\tilde{z}_2, \tilde{z}_3 | 0, 0, \omega_{12}, \omega_{13}, \omega_{23}) d\tilde{z}_2 d\tilde{z}_3, \qquad (40)
\end{aligned}
$$

where $\phi(\cdot,\cdot|\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\sigma_{12})$ denotes the pdf of a bivariate normal. Furthermore, note that

$$Z_1|\tilde{Z}_2=\tilde{z}_2,\tilde{Z}_3=\tilde{z}_3\sim N(\psi_1\tilde{z}_2+\psi_2\tilde{z}_3,1-\omega^*),$$

where

$$(\psi_1\ \psi_2)=(\omega_{12}\ \omega_{13})\begin{pmatrix}\omega_{12} & \omega_{23}\\ \omega_{23} & \omega_{13}\end{pmatrix}^{-1} \text{ and } \omega^*=(\omega_{12}\ \omega_{13})\begin{pmatrix}\omega_{12} & \omega_{23}\\ \omega_{23} & \omega_{13}\end{pmatrix}^{-1}\begin{pmatrix}\omega_{12}\\ \omega_{13}\end{pmatrix}.$$

Therefore, we have

$$\begin{aligned}&P(-\min\{z.,-\tilde{z}_3+c_u\}\leq Z_1\leq\min\{z.,\tilde{z}_2+c_\ell\}|\tilde{Z}_2=\tilde{z}_2,\tilde{Z}_3=\tilde{z}_3)\\ =&\begin{cases}P(-z.\leq Z_1\leq z.|\tilde{Z}_2=\tilde{z}_2,\tilde{Z}_3=\tilde{z}_3) & \text{if } -\tilde{z}_3+c_u>z. \text{ and } \tilde{z}_2+c_\ell>z.\\ P(\tilde{z}_3-c_u\leq Z_1\leq z.|\tilde{Z}_2=\tilde{z}_2,\tilde{Z}_3=\tilde{z}_3) & \text{if } -z.<-\tilde{z}_3+c_u<z. \text{ and } \tilde{z}_2+c_\ell>z.\\ P(-z.\leq Z_1\leq\tilde{z}_2+c_\ell|\tilde{Z}_2=\tilde{z}_2,\tilde{Z}_3=\tilde{z}_3) & \text{if } -\tilde{z}_3+c_u>z. \text{ and } -z.<\tilde{z}_2+c_\ell<z.\\ P(\tilde{z}_3-c_u\leq Z_1\leq\tilde{z}_2+c_\ell|\tilde{Z}_2=\tilde{z}_2,\tilde{Z}_3=\tilde{z}_3) & \text{if } -z_2-c_\ell<-\tilde{z}_3+c_u<z. \text{ and } \tilde{z}_2+c_\ell<z.\\ 0 & \text{otherwise,}\end{cases}\end{aligned}\tag{41}$$

where, for example,

$$P(-z.\leq Z_1\leq z.|\tilde{Z}_2=\tilde{z}_2,\tilde{Z}_3=\tilde{z}_3)=\Phi\left(\frac{z.-(\psi_1\tilde{z}_2+\psi_2\tilde{z}_3)}{\sqrt{1-\omega^*}}\right)-\Phi\left(\frac{-z.-(\psi_1\tilde{z}_2+\psi_2\tilde{z}_3)}{\sqrt{1-\omega^*}}\right).$$

We compute $P(-\min\{z.,-\tilde{Z}_3+c_u\}\leq Z_1\leq\min\{z.,\tilde{Z}_2+c_\ell\})$ based on (40) and (41) using 100,000 draws from a bivariate normal.

F.2.2 Expected length

Let $\psi_1=\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}$ and $\omega_{12}=\Omega_{\beta\delta^{(s_1)}}\Omega^{-1}_{\delta^{(s_1)}\delta^{(s_1)}}\Omega_{\delta^{(s_1)}\beta}$, as well as $\psi_2=\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}$ and $\omega_{13}=\Omega_{\beta\delta^{(s_2)}}\Omega^{-1}_{\delta^{(s_2)}\delta^{(s_2)}}\Omega_{\delta^{(s_2)}\beta}$. Furthermore, let $\tilde{Y}_2=\psi_1 Y_\delta^{(s_1)}$, $\tilde{Y}_3=\psi_2 Y_\delta^{(s_2)}$ and $\omega_{23}=\psi_1\Omega_{\delta^{(s_1)}\delta^{(s_2)}}\psi_2'$.

The expected length of $[Y_\beta - \min\{z_\cdot, \tilde{Y}_2 + c_\ell\}, Y_\beta + \min\{z_\cdot, -\tilde{Y}_3 + c_u\}]$ is given by

$$
\begin{aligned}
&E[\max\{\min\{z_\cdot, \tilde{Y}_2 + c_\ell\} + \min\{z_\cdot, -\tilde{Y}_3 + c_u\}, 0\}] \\
=&E[E[\max\{\min\{z_\cdot, \tilde{y}_2 + c_\ell\} + \min\{z_\cdot, -\tilde{Y}_3 + c_u\}, 0\} | \tilde{Y}_2 = \tilde{y}_2]].
\end{aligned}
$$

Note that

$$
\tilde{Y}_3 | \tilde{Y}_2 = \tilde{y}_2 \sim N(m_3(\tilde{y}_2), \sigma_3^2),
$$

where $m_3(\tilde{y}_2) = \psi_2 \delta^{(s_2)} + \frac{\omega_{23}}{\omega_{12}}(\tilde{y}_2 - \psi_1 \delta^{(s_1)})$ and $\sigma_3^2 = \omega_{13} - \frac{\omega_{23}^2}{\omega_{12}}$. In what follows, we use that

$$
E[Z | a < Z < b] = -\frac{\phi(b) - \phi(a)}{\Phi(b) - \Phi(a)},
$$

where $Z \sim N(0,1)$. Next, note that

$$
E[\max\{\min\{z_\cdot, \tilde{y}_2 + c_\ell\} + \min\{z_\cdot, -\tilde{Y}_3 + c_u\}, 0\} | \tilde{Y}_2 = \tilde{y}_2]
$$

is equal to

$$
E[\max\{z_\cdot + \min\{z_\cdot, -\tilde{Y}_3 + c_u\}, 0\} | \tilde{Y}_2 = \tilde{y}_2],
$$

if $\tilde{y}_2 + c_\ell > z_\cdot$, which, in turn, is equal to

$$
\begin{aligned}
&2z_\cdot \times P(-\tilde{Y}_3 + c_u > z_\cdot | \tilde{Y}_2 = \tilde{y}_2) \\
&+ \left[z_\cdot + c_u - E[\tilde{Y}_3 | -z_\cdot < -\tilde{Y}_3 + c_u < z_\cdot, \tilde{Y}_2 = \tilde{y}_2] \right] \\
&\times P(-z_\cdot < -\tilde{Y}_3 + c_u < z_\cdot | \tilde{Y}_2 = \tilde{y}_2) \\
=&2z_\cdot \times \Phi\left(\frac{-z_\cdot + c_u - m_3(\tilde{y}_2)}{\sigma_3} \right) \\
&+ \left[z_\cdot + c_u - m_3(\tilde{y}_2) + \sigma_3 \frac{\phi\left(\frac{z_\cdot + c_u - m_3(\tilde{y}_2)}{\sigma_3} \right) - \phi\left(\frac{-z_\cdot + c_u - m_3(\tilde{y}_2)}{\sigma_3} \right)}{\Phi\left(\frac{z_\cdot + c_u - m_3(\tilde{y}_2)}{\sigma_3} \right) - \Phi\left(\frac{-z_\cdot + c_u - m_3(\tilde{y}_2)}{\sigma_3} \right)} \right] \\
&\times \left[\Phi\left(\frac{z_\cdot + c_u - m_3(\tilde{y}_2)}{\sigma_3} \right) - \Phi\left(\frac{-z_\cdot + c_u - m_3(\tilde{y}_2)}{\sigma_3} \right) \right]. \qquad (42)
\end{aligned}
$$

Similarly,

$$E[\max\{\min\{z.,\tilde{y}_2+c_\ell\}+\min\{z.,-\tilde{Y}_3+c_u\},0\}|\tilde{Y}_2=\tilde{y}_2]$$

is equal to

$$E[\max\{\tilde{y}_2+c_\ell+\min\{z.,-\tilde{Y}_3+c_u\},0\}|\tilde{Y}_2=\tilde{y}_2],$$

if $\tilde{y}_2+c_\ell<z.$, which, in turn, is equal to $\mathbb{1}(-\tilde{y}_2-c_\ell<z.)$ times

$$\begin{aligned}
&[\tilde{y}_2+c_\ell+z.]\times P(-\tilde{Y}_3+c_u>z.|\tilde{Y}_2=\tilde{y}_2)\\
&+\left[\tilde{y}_2+c_\ell+c_u-E[\tilde{Y}_3|-\tilde{y}_2-c_\ell<-\tilde{Y}_3+c_u<z.,\tilde{Y}_2=\tilde{y}_2]\right]\\
&\times P(-\tilde{y}_2-c_\ell<-\tilde{Y}_3+c_u<z.|\tilde{Y}_2=\tilde{y}_2)\\
&=[\tilde{y}_2+c_\ell+z.]\times\Phi\left(\frac{-z.+c_u-m_3(\tilde{y}_2)}{\sigma_3}\right)\\
&+\left[\tilde{y}_2+c_\ell+c_u-m_3(\tilde{y}_2)+\sigma_3\frac{\phi\left(\frac{\tilde{y}_2+c_\ell+c_u-m_3(\tilde{y}_2)}{\sigma_3}\right)-\phi\left(\frac{-z.+c_u-m_3(\tilde{y}_2)}{\sigma_3}\right)}{\Phi\left(\frac{\tilde{y}_2+c_\ell+c_u-m_3(\tilde{y}_2)}{\sigma_3}\right)-\Phi\left(\frac{-z.+c_u-m_3(\tilde{y}_2)}{\sigma_3}\right)}\right]\\
&\times\left[\Phi\left(\frac{\tilde{y}_2+c_\ell+c_u-m_3(\tilde{y}_2)}{\sigma_3}\right)-\Phi\left(\frac{-z.+c_u-m_3(\tilde{y}_2)}{\sigma_3}\right)\right].
\end{aligned}\tag{43}$$

We compute $E[\max\{\min\{z.,\tilde{Y}_2+c_\ell\}+\min\{z.,-\tilde{Y}_3+c_u\},0\}]$ based on (42) and (43) using 100,000 draws from a univariate normal (equal to the subset of draws used to compute $P(-\min\{z.,-\tilde{Z}_3+c_u\}\leq Z_1\leq\min\{z.,\tilde{Z}_2+c_\ell\})$ that corresponds to $\tilde{Z}_2$).