An Honest Approach to Parallel Trends[*]

Ashesh Rambachan[†]   Jonathan Roth[‡]

July 18, 2021

Abstract

This paper proposes tools for robust inference for difference-in-differences and event-study designs. Instead of requiring that the parallel trends assumption holds exactly, we impose that pre-treatment violations of parallel trends ("pre-trends") are informative about the possible post-treatment violations of parallel trends. Such restrictions allow us to formalize the intuition behind the common practice of testing for pre-existing trends while avoiding issues related to pre-testing. The causal effect of interest is partially identified under such restrictions. We introduce two approaches that guarantee uniformly valid ("honest") inference under the imposed restrictions, and we derive novel results showing that they have good power properties in our context. We recommend that researchers conduct sensitivity analyses to show what conclusions can be drawn under various restrictions on the possible differences in trends.

Keywords: Difference-in-differences, event-study, parallel trends, sensitivity analysis, robust inference, partial identification.

1 Introduction

When using difference-in-differences and related methods, applied researchers are often unsure whether the needed parallel trends assumption holds in practice. It is therefore common to assess the plausibility of the parallel trends assumption by testing for pre-treatment differences in trends ("pre-trends"). There are concerns, however, that such tests may have low power (Freyaldenhoven, Hansen and Shapiro, 2019; Roth, 2019; Kahn-Lang and Lang, 2020; Bilinski and Hatfield, 2020), and relying on them further introduces statistical issues from pre-testing (Roth, 2019). This paper introduces an alternative approach to causal inference in settings where parallel trends may be violated. Our approach formalizes the intuition motivating tests of pre-trends while avoiding the limitations described above.

We consider a setting in which the researcher estimates a vector of "event-study" coefficients $\hat{\beta} = (\hat{\beta}'_{pre}, \hat{\beta}'_{post}) \in \mathbb{R}^{\underline{T}+\bar{T}}$, where $\hat{\beta}_{pre}$ and $\hat{\beta}_{post}$ respectively correspond with estimates for $\underline{T}$ pre-treatment periods and $\bar{T}$ post-treatment periods. The parameter $\beta = \mathbb{E}\left[\hat{\beta}\right]$ can be decomposed as

$$\beta = \underbrace{\begin{pmatrix} 0 \\ \tau_{post} \end{pmatrix}}_{:=\tau} + \underbrace{\begin{pmatrix} \delta_{pre} \\ \delta_{post} \end{pmatrix}}_{:=\delta}, \tag{1}$$

where τ is a causal parameter of interest (assumed to be 0 in the pre-treatment period) and δ is a bias from a difference in trends. For instance, in the canonical (non-staggered) difference-in-differences framework, τ is the vector of period-specific average treatment effects on the treated (ATT) for some policy of interest, and δ is the difference in trends of untreated potential outcomes between the treated and comparison groups. The usual parallel trends assumption is that $\delta_{post} = 0$, which gives point identification of τ_{post}. Researchers frequently test the plausibility of this assumption by testing whether $\delta_{pre} = 0$ (a "pre-trends" test).

Instead of imposing that parallel trends holds exactly, we place restrictions on the possible values of the post-treatment difference in trends δ_{post} given the (identified) value of the pre-trend δ_{pre}. Such restrictions formalize the intuition motivating pre-trends tests, namely that pre-trends are informative about counterfactual post-treatment differences in trends. More formally, we impose that $\delta \in \Delta$ for some researcher-specified set Δ and show that the causal parameter τ_{post} is partially identified under such restrictions.

We show that a variety of commonly expressed intuitions about possible violations of parallel trends can be captured via different choices of the set Δ. For example, applied researchers often have the intuition that any differences in trends evolve smoothly over time (e.g. owing to long-run secular trends), which can be formalized by restricting how quickly the slope of the differential trend can change over time. Likewise, our framework allows

2

researchers to formalize the intuition that the magnitude of the violation of parallel trends in the post-treatment period cannot be too much larger than the worst-case violation over the pre-treatment periods. We adopt a flexible framework that allows researchers to capture these intuitions, as well as a variety of other restrictions that are implied by context-specific knowledge about the possible confounds.

We then introduce methods that provide uniformly valid ("honest") inference for the treatment effect of interest under the restriction that $\delta \in \Delta$. Our approach to inference can be applied whenever Δ can be written as the finite union of polyhedra, which incorporates all of the restrictions described above and many others. Specifically, we introduce two methods for inference with different attractive features depending on the exact form of Δ considered.

We first consider inference based on optimal fixed length confidence intervals (FLCIs) (Donoho, 1994), which have desirable finite-sample guarantees for particular Δs of interest. Results from Armstrong and Kolesar (2018a,b) imply that FLCIs have near-optimal expected length among honest procedures when the class Δ is convex and centrosymmetric, as is the case for our baseline smoothness class. FLCIs are thus an attractive choice for particular choices of Δ. Unfortunately, we show that FLCIs have unattractive properties for other leading choices of Δ: in many cases, they will be inconsistent in the strong sense that power against fixed points outside the identified set need not converge to one asymptotically.

Motivated by this finding, we next introduce a more general inference approach that can accommodate a larger class of restrictions Δ. We show that a wide variety of relevant restrictions Δ can be written as the finite union of polyhedra, in which case testing hypotheses about treatment effects can be cast as a moment inequality problem with nuisance parameters that enter the moments linearly. This formulation allows us to leverage the large econometrics literature on testing for moment inequalities (see Canay and Shaikh (2017); Molinari (2020) for recent reviews). We consider an implementation of this approach based on the conditional test proposed in Andrews, Roth and Pakes (2019, henceforth ARP), which has several desirable features in our setting. First, it is computationally tractable even when the dimension of the nuisance parameters is large, as occurs whenever there are many post-treatment periods. Second, we show that the conditional test has optimal local asymptotic power for parameter configurations satisfying a linear independence constraint qualification (LICQ) condition. When Δ bounds the post-treatment bias by the maximal pre-treatment violation of parallel trends, for example, this condition is satisfied when the pre-treatment maximum is unique. Our optimal local asymptotic power result is novel, and relies on structure in our context not present in the more general setting considered in ARP.

We recommend empirical researchers use our methods to conduct sensitivity analyses in which they report confidence sets under varying restrictions on the possible differences in

trends. For example, one class of restrictions we consider restricts the post-treatment violation of parallel trends to be no more than $\bar{M}$ times larger in magnitude than the maximum pre-treatment violation. It is then natural for the researcher to report confidence sets for different values of $\bar{M}$, which highlights how the results change under different assumptions about how bad the violation of parallel trends could be relative to the pre-trend. Performing such sensitivity analyses makes clear what must be assumed about the possible differences in trends in order to draw specific causal conclusions. We provide an R package, `HonestDiD`, that implements our recommended methods.[1] We illustrate our recommended approach with an application to two recently published papers.

Related literature: Our approach is most closely related to Manski and Pepper (2018), who consider partial identification of treatment effects under direct bounds on the extent to which parallel trends is violated in the post-treatment period. These restrictions are nested as a special case within our framework. We consider a larger class of possible restrictions, however, which allows us to formalize a variety of intuitions expressed in applied work, including the notion that pre-trends are informative about post-treatment differences in trends. Additionally, we develop methods for conducting inference on the causal effects of treatment under these assumptions, whereas Manski and Pepper (2018) only consider identification.

Several other recent papers consider various relaxations of the parallel trends assumption. Keele, Small, Hsu and Fogarty (2019) develop techniques for testing the sensitivity of difference-in-differences designs to violations of the parallel trends assumption, but they do not incorporate information from the observed pre-trends in their sensitivity analysis. Empirical researchers commonly adjust for the extrapolation of a linear trend from the pre-treatment periods when there are concerns about violations of the parallel trends assumption, which is valid if the difference in trends is exactly linear (e.g., Dobkin, Finkelstein, Kluender and Notowidigdo, 2018; Goodman-Bacon, 2018, 2021; Bhuller, Havnes, Leuven and Mogstad, 2013). Our methods nest this approach as a special case, but allow for valid inference under less restrictive assumptions about the class of possible differences in trends. Freyaldenhoven et al. (2019) propose a method that allows for violations of the parallel trends assumption but requires an additional covariate that is affected by the same confounding factors as the outcome but not by the treatment of interest. Ye, Keele, Hasegawa and Small (2020) consider partial identification of treatment effects when there exist two control groups whose outcomes have a bracketing relationship with the outcome of the treated group. Leavitt (2020) proposes an empirical Bayes approach calibrated to pre-treatment differences in trends, and

---

[1]The latest version may be downloaded here.

Bilinski and Hatfield (2020) and Dette and Schumann (2020) propose approaches based on pre-tests for the magnitude of the pre-treatment violations of parallel trends.

Our methods address several concerns related to established empirical practice in difference-in-differences and event-study designs. First, common tests for pre-trends may be underpowered against meaningful violations of parallel trends, potentially leading to severe undercoverage of conventional confidential intervals (Freyaldenhoven et al., 2019; Roth, 2019; Bilinski and Hatfield, 2020; Kahn-Lang and Lang, 2020). Second, statistical distortions from pre-testing for pre-trends may further undermine the performance of conventional inference procedures (Roth, 2019). Third, parametric approaches to controlling for pre-existing trends may be sensitive to functional form assumptions (Wolfers, 2006; Lee and Solon, 2011). We address these issues by providing tools for inference that do not rely on an exact parallel trends assumption and that make clear the mapping between assumptions on the potential differences in trends and the strength of one's conclusions.

Our work complements a growing literature on the causal interpretation of event-study coefficients in two-way fixed effects models in the presence of staggered treatment timing or heterogeneous treatment effects (Borusyak and Jaravel, 2016; Athey and Imbens, 2018; Goodman-Bacon, 2021; Callaway and Sant'Anna, 2020; de Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2020). Several alternative estimators have been proposed that consistently estimate sensible causal estimands under a suitable parallel trends assumption. Our methodology complements these approaches by providing tools to assess the sensitivity of these methods to violations of the corresponding parallel trends assumption; see Remark 1 for additional details.

2 General set-up

We now introduce the assumptions, target parameter, and inferential goal considered in the paper. In the main text, we consider a finite-sample normal model with known covariance matrix, which arises as an asymptotic approximation to a variety of econometric settings of interest. In the supplementary materials, we show how the finite-sample results presented in this model translate to uniform asymptotic statements over a large class of data-generating processes.

2.1 Finite sample normal model

Consider the model

$$\hat{\beta}_n \sim \mathcal{N}\left(\beta, \, \Sigma_n\right), \tag{2}$$

where $\hat{\beta}_n \in \mathbb{R}^{\underline{T}+\bar{T}}$ and $\Sigma_n = \frac{1}{n}\Sigma^*$ for Σ^* a known, positive-definite $(\underline{T} + \bar{T}) \times (\underline{T} + \bar{T})$ matrix. We refer to $\hat{\beta}_n$ as the estimated event-study coefficients, and partition $\hat{\beta}_n$ into vectors corresponding with the pre-treatment and post-treatment periods, $\hat{\beta}_n = (\hat{\beta}'_{n,pre}, \hat{\beta}'_{n,post})'$, where $\hat{\beta}_{n,pre} \in \mathbb{R}^{\underline{T}}$ and $\hat{\beta}_{n,post} \in \mathbb{R}^{\bar{T}}$. We adopt analogous notation to partition other vectors that are the same length as $\hat{\beta}_n$.

The finite sample normal model (2) can be viewed as an asymptotic approximation, since a variety of estimators for difference-in-differences and event study designs will yield asymptotically normally-distributed event-study coefficients, $\sqrt{n}\left(\hat{\beta}_n - \beta\right) \xrightarrow{d} \mathcal{N}(0, \Sigma^*)$, under mild regularity conditions (see Remarks 1-2). This convergence in distribution suggests the finite-sample approximation $\hat{\beta}_n \overset{d}{\approx} \mathcal{N}(\beta, \Sigma_n)$, where $\overset{d}{\approx}$ denotes approximate equality in distribution and $\Sigma_n = \frac{1}{n}\Sigma^*$. We derive results assuming this equality in distribution holds exactly in finite samples. In the supplemental materials, we show that results in the finite sample normal model translate to uniform asymptotic statements for a large class of data-generating processes.

We assume the mean vector β satisfies the following causal decomposition.

Assumption 1. *The parameter vector β can be decomposed as*

$$\beta = \underbrace{\begin{pmatrix} \tau_{pre} \\ \tau_{post} \end{pmatrix}}_{:= \tau} + \underbrace{\begin{pmatrix} \delta_{pre} \\ \delta_{post} \end{pmatrix}}_{:= \delta} \quad \text{with } \tau_{pre} \equiv 0. \tag{3}$$

The first term, τ, represents the dynamic causal effects of interest. We assume the treatment has no causal effect prior to its implementation, so $\tau_{pre} = 0$. The second term, δ, represents the difference in trends between the treated and comparison groups that would have occurred absent treatment. The parallel trends assumption imposes that $\delta_{post} = 0$, and therefore $\beta_{post} = \tau_{post}$ under parallel trends.

Example: Difference-in-differences   We observe an outcome Y_{it} for a sample of individuals $i = 1, \ldots, N$ for three time periods, $t = -1, 0, 1$. Individuals in the treated population $(D_i = 1)$ receive a treatment between period $t = 0$ and $t = 1$.[2] The observed outcome equals $Y_{i,t} = D_i Y_{i,t}(1) + (1 - D_i)Y_{i,t}(0)$, where $Y_{i,t}(1)$ and $Y_{i,t}(0)$ are the potential outcomes for individual i in period t associated with the treatment and control conditions. Assume the treatment has no causal effect prior to implementation, meaning $Y_{i,t}(1) = Y_{i,t}(0)$ for $t < 1$. The causal estimand of interest is the average treatment effect on the treated (ATT),

---

[2]For the purposes of this example, we think of the observed sample as consisting of N_1 independent draws from the treated $(D_i = 1)$ population and N_0 independent draws from the control population $(D_i = 0)$ with $N = N_0 + N_1$, as in Abadie and Imbens (2006).

$\tau_{ATT} = \mathbb{E}\left[Y_{i,1}(1) - Y_{i,1}(0) \mid D_i = 1\right]$. In this setting, researchers commonly estimate the "dynamic event study regression"

$$Y_{it} = \lambda_i + \phi_t + \sum_{s \neq 0} \beta_s \times \mathbb{1}[t = s] \times D_i + \epsilon_{it}. \tag{4}$$

The estimated coefficient $\hat{\beta}_1$ is the "difference-in-differences" of sample means across treated and untreated groups between period $t = 0$ and $t = 1$, $\hat{\beta}_1 = (\bar{Y}_{1,1} - \bar{Y}_{1,0}) - (\bar{Y}_{0,1} - \bar{Y}_{0,0})$, where $\bar{Y}_{d,t}$ is the sample mean of Y_{it} for treatment group d in period t. The "pre-period" coefficient $\hat{\beta}_{-1}$ can likewise be written as $\hat{\beta}_{-1} = (\bar{Y}_{1,-1} - \bar{Y}_{1,0}) - (\bar{Y}_{0,-1} - \bar{Y}_{0,0})$.

Taking expectations and re-arranging, we see that

$$\mathbb{E}\left[\hat{\beta}_1\right] = \tau_{ATT} + \underbrace{\mathbb{E}\left[Y_{i,1}(0) - Y_{i,0}(0) \mid D_i = 1\right] - \mathbb{E}\left[Y_{i,1}(0) - Y_{i,0}(0) \mid D_i = 0\right]}_{\text{Post-treatment differential trend} =: \delta_1},$$

$$\mathbb{E}\left[\hat{\beta}_{-1}\right] = \underbrace{\mathbb{E}\left[Y_{i,-1}(0) - Y_{i,0}(0) \mid D_i = 1\right] - \mathbb{E}\left[Y_{i,-1}(0) - Y_{i,0}(0) \mid D_i = 0\right]}_{\text{Pre-treatment differential trend} =: \delta_{-1}}.$$

The parameter $\beta = \mathbb{E}\left[\hat{\beta}\right]$ thus satisfies the decomposition (3), where $\tau_{post} = \tau_{ATT}$ is the ATT, $\delta_{post} = \delta_1$ is the difference in trends in *untreated* potential outcomes between $t = 0$ and $t = 1$, and $\delta_{pre} = \delta_{-1}$ is the analogous difference in trends for untreated potential outcomes between $t = -1$ and $t = 0$. Under suitable regularity conditions, $\hat{\beta}$ will also satisfy a central limit theorem, so that (2) will hold approximately in large samples. ▲

Remark 1 (Staggered timing). As discussed above, several recent papers have noted that the estimand of specification (4) does not have an intuitive causal interpretation when treatment is staggered and there are heterogeneous treatment effects. However, Sun and Abraham (2020), Callaway and Sant'Anna (2020), and de Chaisemartin and D'Haultfoeuille (2021) propose alternative methods for forming "event-studies" with a sensible interpretation in such settings. Since these estimators are asymptotically normally distributed, they fit into our framework, where now τ_{post} corresponds with a weighted average of causal effects at each lag since treatment, and δ corresponds with a weighted average of differences in untreated potential outcomes.

Remark 2 (Other event-study estimators). Other examples of estimators that yield asymptotically normal event-study estimates (under suitable regularity conditions) include the GMM procedure proposed by Freyaldenhoven et al. (2019), instrumental variables event-studies (Hudson, Hull and Liebersohn, 2017), as well as a range of procedures that flexibly control for differences in covariates between treated and comparison groups (e.g., Heckman,

Ichimura, Smith and Todd, 1998; Abadie, 2005; Sant'Anna and Zhao, 2020).

Remark 3 (Anticipatory effects). In some cases, there may be changes in behavior in anticipation of the policy of interest, and therefore, β_{pre} may reflect the causal anticipatory effects of the policy (Malani and Reif, 2015). This violates Assumption 1, which assumes pre-treatment coefficients do not reflect causal effects. A simple solution is available if one is willing to assume that anticipatory effects only occur in a fixed window prior to the policy change. Under such an assumption, the researcher may re-normalize the definition of the "pre-treatment" period to be the period prior to when anticipatory effects can occur, in which case β_{pre} is determined only based on untreated potential outcomes.

Remark 4 (Design-based Uncertainty). Rambachan and Roth (2020) shows that the normal model (2) also arises from a design-based model that treats the finite population of observed units as fixed and views the assignment of treatment as the source of randomness in the data. This perspective may be preferred to the usual sampling-based approach to uncertainty in settings where the super-population is not clear, such as when all 50 US states are observed (Manski and Pepper, 2018; Abadie, Athey, Imbens and Wooldridge, 2020).

2.2 Target parameter and identification

The parameter of interest is a linear combination of the post-treatment causal effects, $\theta := l'\tau_{post}$ for some known $\bar{T}$-vector l. For example, θ equals the t-th period causal effect τ_t when the vector l equals the t-th standard basis vector. Similarly, θ equals the average causal effect across all post-treatment periods when $l = \left(\frac{1}{T}, ..., \frac{1}{T}\right)'$.

We relax the parallel trends assumption by assuming that δ lies in a set of possible differences in trends Δ, which is specified by the researcher. This nests the usual parallel trends assumption as a special case with $\Delta = \{\delta \ : \ \delta_{post} = 0\}$. Since $\delta_{pre} = \mathbb{E}\left[\hat{\beta}_{pre}\right]$ is identified, the assumption that $\delta = (\delta'_{pre}, \delta'_{post})' \in \Delta$ restricts the possible values of δ_{post} given the (identified) value of the pre-treatment difference in trends δ_{pre}. It is natural to place restrictions on the relationship between δ_{pre} and δ_{post}, since researchers frequently test the null hypothesis that $\delta_{pre} = 0$ as a way of assessing the plausibility of the assumption that $\delta_{post} = 0$.

Under the assumption that $\delta \in \Delta \neq \{\delta \ : \ \delta_{post} = 0\}$, the parameter θ will typically be set-identified. For a given value of β, the *identified set* for θ under the assumption $\delta \in \Delta$ is

$$\mathcal{S}(\beta, \Delta) := \left\{ \theta \ : \ \exists \delta \in \Delta, \tau_{post} \in \mathbb{R}^{\bar{T}} \text{ s.t. } l'\tau_{post} = \theta, \beta = \delta + \begin{pmatrix} 0 \\ \tau_{post} \end{pmatrix} \right\}, \tag{5}$$

8

i.e. the set of values of θ consistent with β under the restriction that $\delta \in \Delta$. When Δ is a closed and convex set, the identified set has a simple characterization.

Lemma 2.1. *If Δ is closed and convex, then $\mathcal{S}(\beta, \Delta)$ is an interval in $\mathbb{R}$, $\mathcal{S}(\beta, \Delta) = [\theta^{lb}(\beta, \Delta), \theta^{ub}(\beta, \Delta)]$, where*

$$\theta^{lb}(\beta, \Delta) := l'\beta_{post} - \underbrace{\left(\max_{\delta} l'\delta_{post}, \ s.t. \ \delta \in \Delta, \delta_{pre} = \beta_{pre} \right)}_{=:b^{max}(\beta_{pre};\Delta)}, \tag{6}$$

$$\theta^{ub}(\beta, \Delta) := l'\beta_{post} - \underbrace{\left(\min_{\delta} l'\delta_{post}, \ s.t. \ \delta \in \Delta, \delta_{pre} = \beta_{pre} \right)}_{=:b^{min}(\beta_{pre};\Delta)}. \tag{7}$$

Proof. Re-arranging terms in (5), the identified set can be equivalently written as $\mathcal{S}(\beta, \Delta) = \{\theta \ : \ \exists \delta \in \Delta \ \text{s.t.} \ \delta_{pre} = \beta_{pre}, \theta = l'\beta_{post} - l'\delta_{post}\}$. The result is then immediate. $\square$

Example: Difference-in-differences (continued) Point identification of the ATT in the difference-in-differences design is typically obtained by assuming that the counterfactual post-treatment difference in trends δ_1 is exactly zero. Instead, we assume $\delta = (\delta_{-1}, \delta_1)' \in \Delta$ for some set Δ. When Δ is closed and convex, the identified set for the ATT will be $[\beta_1 - b^{max}, \beta_1 - b^{min}]$, where $b^{max} = \max_{\delta} \delta_1$ s.t $(\delta_{-1}, \delta_1)' \in \Delta$ is the maximum possible bias of $\hat{\beta}_1$ given δ_1 and b^{min} is defined analogously.▲

Additionally, it is immediate from the definition of the identified set in (5) that if Δ is the finite union of sets, $\Delta = \bigcup_{k=1}^{K} \Delta_k$, then its identified set is the union of the identified sets for its subcomponents,

$$\mathcal{S}(\beta, \Delta) = \bigcup_{k=1}^{K} \mathcal{S}(\beta, \Delta_k). \tag{8}$$

This fact will be useful, since several Δs of interest in empirical practice can be written as the finite union of convex sets, as we will see below.

2.3 Possible choices of Δ

The class of possible differences in trends Δ must be specified by the researcher, and the choice of Δ will depend on the economic context. We highlight several possible choices of Δ that may be reasonable in empirical applications and formalize intuitive arguments that are commonly made by applied researchers regarding possible violations of parallel trends.

2.3.1 Smoothness restrictions

Researchers often worry about confounding factors that lead to different secular trends among the treated and comparison groups. When the researcher expects the secular trends to evolve smoothly over time, it is common to control for a linear group-specific time trend.[3] This approach is valid if the difference in trends is linear, i.e. $\Delta = \{\delta \colon \delta_t = \gamma \cdot t, \gamma \in \mathbb{R}\}$, where we adopt the convention that periods $t < 0$ and $t > 0$ respectively correspond with the elements of δ_{pre} and δ_{post}, and $\delta_0 \equiv 0$.[4] There are often concerns, however, that the linear specification is not exactly correct (Wolfers, 2006; Lee and Solon, 2011). A natural relaxation is therefore to impose only that the differential trends evolve smoothly over time – say with slope changing by no more than M between consecutive periods. This can be formalized by requiring that δ lie in the set

$$\Delta^{SD}(M) := \{\delta \,:\, |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leqslant M, \forall t\}. \tag{9}$$

The parameter $M \geqslant 0$ governs the amount by which the slope of δ can change between consecutive periods, and thus bounds the discrete analog of the second derivative (we use the abbreviation SD for "second differences" or "second derivative").[5] In the special case where $M = 0$, $\Delta^{SD}(0)$ requires that the difference in trends be exactly linear.

It is worth highlighting that the common practice of testing for pre-trends is intuitively based on the notion that differences in trends evolve smoothly over time. Indeed, a pre-trends test would not be very informative about the bias in a difference-in-differences design if the difference in trends could be close to zero in the pre-treatment period and then change sharply around the time of treatment. The restriction that $\delta \in \Delta^{SD}(M)$ is thus one way of formalizing this intuition.

Example: Difference-in-differences (continued) In the three-period difference-in-differences model, assuming the differential trend is exactly linear is equivalent to assuming $\Delta = \{\delta \,:\, \delta_1 = -\delta_{-1}\}$. Assuming $\delta \in \Delta^{SD}(M)$ requires only that the linear extrapolation be *approximately* correct, $\delta_1 \in [-\delta_{-1} - M, -\delta_{-1} + M]$.

---

[3]Specifically, researchers often augment specification (4) with group-specific linear trends, an approach Dobkin et al. (2018) refer to as a "parametric event-study." An analogous approach is to estimate a linear trend using only observations prior to treatment, and then subtract out the estimated linear trend from the observations after treatment (Bhuller et al., 2013; Goodman-Bacon, 2018, 2021).

[4]Setting $\delta_0 = 0$ corresponds with the common practice of normalizing $\beta_0 = 0$, as in specification (4).

[5]Restrictions on the second derivative of the conditional expectation function or density have been used in regression discontinuity settings (Kolesar and Rothe, 2018; Frandsen, 2016; Noack and Rothe, 2020). Smoothness restrictions are also used to obtain partial identification in Kim, Kwon, Kwon and Lee (2018).

2.3.2 Bounding Relative Magnitudes

A second related approach bounds the worst-case post-treatment violation of parallel trends in terms of the worst-case violation in the pre-treatment period. For instance, the restriction

$$\Delta^{RM}(\bar{M}) = \{\delta \; : \; \forall t \geqslant 0, \; |\delta_{t+1} - \delta_t| \leqslant \bar{M} \cdot \max_{s<0} |\delta_{s+1} - \delta_s|\}$$

bounds the maximum post-treatment violation of parallel trends (between consecutive periods) by $\bar{M}$ times the maximum pre-treatment violation of parallel trends. (We use the abbreviation RM for "relative magnitudes".) Likewise, the restriction

$$\Delta^{SDRM}(\bar{M}) = \{\delta \; : \; \forall t \geqslant 0, \; |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leqslant \bar{M} \cdot \max_{s<0} |(\delta_{s+1} - \delta_s) - (\delta_s - \delta_{s-1})|\}$$

bounds the maximum deviation from a linear trend in the post-treatment period by $\bar{M}$ times the equivalent maximum in the pre-treatment period. The set $\Delta^{SDRM}(\bar{M})$ is thus similar to $\Delta^{SD}(M)$ introduced above, except it allows the magnitude of the possible non-linearity to explicitly depend on the observed pre-trends.

Example: Difference-in-differences (continued) Assuming $\delta \in \Delta^{RM}(\bar{M})$ bounds the magnitude of δ_1 based on the magnitude of δ_{-1}, i.e. $\Delta^{RM}(\bar{M}) = \{(\delta_{-1}, \delta_1)' \; : \; |\delta_1| \leqslant \bar{M}|\delta_{-1}|\}$. The larger the magnitude of the observed pre-period violation in parallel trends, $|\delta_{-1}|$, the wider the range of possible post-period violations of parallel trends.▲

2.3.3 Sign and monotonicity restrictions

Context-specific knowledge may sometimes also suggest sign or monotonicity restrictions on the differential trend. For instance, if the policy of interest occurs at the same time as a confounding policy change that we expect to have a positive effect on the outcome, we might restrict the post-treatment bias to be positive, $\delta \in \Delta^{PB} := \{\delta \; : \; \delta_t \geqslant 0 \; \forall t \geqslant 0\}$. Likewise, there may be secular pre-existing trends that we expect would have continued following the treatment date.[6] We may then wish to impose that the differential trend be increasing, $\delta \in \Delta^I := \{\delta \; : \; \delta_t \geqslant \delta_{t-1} \; \forall t\}$, or monotone with unknown sign, $\delta \in \Delta^{Mon} := \Delta^I \cup (-\Delta^I)$. Sign and monotonicity restrictions may be combined with the previously discussed restrictions, such as $\Delta^{SDPB}(M) := \Delta^{SD}(M) \cap \Delta^{PB}$, $\Delta^{SDI}(M) := \Delta^{SD}(M) \cap \Delta^I$, and $\Delta^{RMI}(\bar{M}) :=$

---

[6]Monotone violations of parallel trends are often discussed in applied work. For example, Lovenheim and Willen (2019) argue that violations of parallel trends cannot explain their results because "pre-[treatment] trends are either zero or in the wrong direction (i.e., opposite to the direction of the treatment effect)." Greenstone and Hanna (2014) estimate upward-sloping pre-existing trends and argue that "if the pre-trends had continued" their estimates would be upward biased.

$\Delta^{RM}(\bar{M}) \cap \Delta^I$.

2.3.4 Polyhedral restrictions

Although the restrictions described above will be sensible in many empirical contexts, researchers will often have context-specific knowledge that motivates alternative restrictions. To accommodate such cases, we consider the broad class of Δs that can be written as polyhedra (sets defined by linear inequalities), or the finite union of polyhedra.

Definition 1 (Polyhedral restriction)**.** The class Δ is polyhedral if it takes the form $\Delta = \{\delta \ : \ A\delta \leqslant d\}$ for some known matrix A and vector d, where the matrix A has no all-zero rows.

All of the examples described above can be written either as polyhedral restrictions or finite unions of such restrictions. For instance, $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$ can be written directly as polyhedra.[7] Likewise, $\Delta^{RM}(\bar{M})$ or $\Delta^{SDRM}(\bar{M})$ can be written as the finite union of polyhedra, where each polyhedron corresponds with a different location for the maximum pre-treatment violation.[8] The class of (finite unions of) polyhedra is quite broad, and allows for a variety of other restrictions that may be relevant in empirical work.

Remark 5 (Bounded variation assumptions)**.** Manski and Pepper (2018, henceforth MP) consider identification of treatment effects under "bounded variation assumptions" which can be expressed in the polyhedral form introduced in Definition 1. In the ongoing difference-in-differences example, MP's "bounded difference-in-differences variation" assumption directly bounds the magnitude of $|\delta_1|$ when $\hat{\beta}_1$ is the coefficient from specification (4). MP also consider "bounded time" and "bounded state" variation assumptions, which correspond with bounds on the magnitudes of $|\mu_{11} - \mu_{10}|$ and $|\mu_{11} - \mu_{01}|$, where $\mu_{ds} := \mathbb{E}\left[Y(0)|D = d, t = s\right]$. These restrictions can be accommodated by augmenting the vector $\hat{\beta}$ to include the sample means corresponding with estimates of the differences in outcomes for the appropriate treatment-group by time period cells.[9] ∎

---

[7]In our ongoing three-period difference-in-differences example, $\Delta^{SD}(M) = \{\delta \ : \ A^{SD}\delta \leqslant d^{SD}\}$ for $A^{SD} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ and $d^{SD} = (M, M)'$. This generalizes naturally when there are multiple pre-periods and multiple post-periods.

[8]For example, define the polyhedra $\Delta^{RM}_{s,+}(\bar{M}) = \{\delta \ : \ \forall t \geqslant 0, |\delta_{t+1} - \delta_t| \leqslant \bar{M}(\delta_{s+1} - \delta_s)\}$ and $\Delta^{RM}_{s,-} = \{\delta \ : \ \forall t \geqslant 0, |\delta_{t+1} - \delta_t| \leqslant -\bar{M}(\delta_{s+1} - \delta_s)\}$. Then $\Delta^{RM}(\bar{M}) = \bigcup_{s<0}\left(\Delta^{RM}_{s,+}(\bar{M}) \cup \Delta^{RM}_{s,-}(\bar{M})\right)$.

[9]After augmenting the vector for the event-study coefficients, Equation (3) must be re-written to replace $(0, \tau'_{post})'$ with $C\tau_{post}$, where C is a matrix that accounts for the fact that elements of τ enter both the event-study coefficients and the augmented terms. Our proposed methods and results do not rely on the structure that $C = (0, I)'$ and thus easily accommodate this modification.

Remark 6 (Ashenfelter's dip). Researchers studying labor market training and related programs may be concerned about Ashenfelter's dip (Ashenfelter, 1978), in which earnings for the treated group trend downwards (relative to control) before treatment and upwards afterwards. In this type of setting, researchers might naturally use a polyhedral Δ to impose i) restrictions on the signs of the pre-treatment and post-treatment biases, as well as ii) restrictions on the magnitude of the rebound effect relative to the pre-treatment shock.

2.4 Inferential Goal

Given a particular choice of Δ, we construct confidence sets $\mathcal{C}_n$ that are uniformly valid for all parameter values θ in the identified set,

$$\inf_{\delta \in \Delta, \tau} \inf_{\theta \in \mathcal{S}(\Delta, \delta + \tau)} \mathbb{P}_{(\delta, \tau, \Sigma_n)} \left(\theta \in \mathcal{C}_n \right) \geq 1 - \alpha. \tag{10}$$

We subscript the probability operator by (δ, τ, Σ_n) to make explicit that the distribution of $\hat{\beta}_n$ (and hence $\mathcal{C}_n$) depends on these parameters. In the supplemental materials, we show that the coverage requirement (10) in the normal model translates to uniform asymptotic coverage over a large class of data-generating processes. Confidence sets satisfying this criterion are referred to as "honest" (Li, 1989).

We will primarily focus our attention on constructing confidence sets for the case where Δ is a polyhedron. A valid confidence set for the case where Δ is the finite union of polyedra can then be constructed by taking the union of the confidence sets for each of its components, as formalized in the following lemma.

Lemma 2.2. *Suppose that for each $k = 1, ..., K$, the confidence set $\mathcal{C}_{n,k}$ satisfies (10) with $\Delta = \Delta_k$. Then the confidence set $\mathcal{C}_n = \bigcup_{k=1}^{K} \mathcal{C}_{n,k}$ satisfies (10) with $\Delta = \bigcup_{k=1}^{K} \Delta_k$.*

In the next two sections, we introduce two approaches to obtain confidence sets satisfying (10). The first approach, fixed length confidence intervals, provide particularly attractive properties for specific forms of Δ, such as $\Delta^{SD}(M)$. The second approach, based on moment inequalities, can accommodate a much wider range of restrictions.

3 Inference using Fixed Length Confidence Intervals

We first consider fixed length confidence intervals (FLCIs) based on affine estimators. FLCIs deliver attractive finite-sample guarantees for certain choices of Δ, including our baseline smoothness class $\Delta^{SD}(M)$, but may perform poorly for other types of restrictions.

3.1 Constructing FLCIs

Following Donoho (1994) and Armstrong and Kolesar (2018a, 2020), we consider fixed length confidence intervals based on an affine estimator for θ, denoted by $\mathcal{C}_{\alpha,n}(a, v, \chi) := \left(a + v'\hat{\beta}_n\right) \pm \chi$, where a and χ are scalars and $v \in \mathbb{R}^{T+\bar{T}}$. We minimize the half-length of the confidence interval, χ, subject to the constraint that $\mathcal{C}_{\alpha,n}(a, v, \chi)$ satisfies the coverage requirement (10).

To do so, note that $a + v'\hat{\beta}_n \sim \mathcal{N}\left(a + v'\beta, v'\Sigma_n v\right)$, and hence $|a + v'\hat{\beta}_n - \theta| \sim |\mathcal{N}(b, v'\Sigma_n v)|$, where $b = a + v'\beta - \theta$ is the affine estimator's bias for θ. Observe further that $\theta \in \mathcal{C}_n(a, v, \chi)$ if and only if $|a + v'\hat{\beta}_n - \theta| \leqslant \chi$. For fixed values a and v, the smallest value of χ that satisfies (10) is therefore the $1 - \alpha$ quantile of the $|\mathcal{N}\left(\bar{b}, v'\Sigma_n v\right)|$ distribution, where $\bar{b}$ is the affine estimator's worst-case bias

$$\bar{b}(a, v) := \sup_{\delta \in \Delta, \tau_{post} \in \mathbb{R}^{\bar{T}}} \left| a + v' \left(\delta + \begin{pmatrix} 0 \\ \tau_{post} \end{pmatrix}\right) - l'\tau_{post} \right|. \tag{11}$$

Let $cv_\alpha(t)$ denote the $1 - \alpha$ quantile of the folded normal distribution $|\mathcal{N}(t, 1)|$.[10] For fixed a and v, the smallest value of χ satisfying the coverage requirement (10) is thus

$$\chi_n(a, v; \alpha) = \sigma_{v,n} \cdot cv_\alpha(\bar{b}(a, v)/\sigma_{v,n}), \tag{12}$$

where $\sigma_{v,n} := \sqrt{v'\Sigma_n v}$. The optimal (i.e., minimum-length) FLCI is constructed by choosing the values of a and v to minimize (12). When Δ is convex, this minimization can be solved as a nested optimization problem, where both the inner and outer minimizations are convex (Low, 1995; Armstrong and Kolesar, 2018a, 2020). We denote the $1 - \alpha$ level, optimal FLCI by $\mathcal{C}_{\alpha,n}^{FLCI} := \left(a_n + v'_n\hat{\beta}_n\right) \pm \chi_n$, where $\chi_n := \inf_{a,v} \chi_n(a, v; \alpha)$ and a_n, v_n are the optimal values in the minimization.

Example: $\Delta^{SD}(M)$. Suppose $\theta = \tau_1$. For $\Delta^{SD}(M)$, the affine estimator used by the optimal FLCI takes the form $a + v'\hat{\beta}_n = \hat{\beta}_{n,1} - \sum_{s=-\underline{T}+1}^{0} w_s \left(\hat{\beta}_{n,s} - \hat{\beta}_{n,s-1}\right)$, where the weights w_s sum to one (but may be negative). This estimator adjusts the event-study coefficient for $t = 1$ by an estimate of the differential trend between $t = 0$ and $t = 1$ formed by taking a weighted average of the differential trends in periods prior to treatment. The worst-case bias will be smaller if more weight is placed on pre-treatment periods closer to the treatment date, but it may reduce variance to place more weight on earlier pre-periods. The weights w_s are optimally chosen to balance this tradeoff. ▲

---

[10]If $t = \infty$, we define $cv_\alpha = \infty$.

3.2 Finite-sample near optimality

In particular cases of interest, such as when $\Delta = \Delta^{SD}(M)$, the FLCIs introduced above have near-optimal expected length in the finite-sample normal model. The following result, which is a consequence of results in Armstrong and Kolesar (2018a, 2020), bounds the ratio of the expected length of the shortest possible confidence interval that controls size relative to the length of the optimal FLCI.

Assumption 2. *Assume i) Δ is convex and centrosymmetric (i.e. $\delta \in \Delta$ implies $-\delta \in \Delta$), and ii) $\delta_A \in \Delta$ is such that $(\delta - \delta_A) \in \Delta$ for all $\delta \in \Delta$.*

Proposition 3.1. *Suppose δ_A and Δ satisfy Assumption 2.[11] Let $\mathcal{I}_\alpha(\Delta, \Sigma_n)$ denote the class of confidence sets that satisfy the coverage criterion (10) at the $1 - \alpha$ level. Then, for any $\tau_A \in \mathbb{R}^{\bar{T}}$, Σ^* positive definite, and $n > 0$,*

$$\frac{\inf_{\mathcal{C}_{\alpha,n} \in \mathcal{I}_\alpha(\Delta, \Sigma_n)} \mathbb{E}_{(\delta_A, \tau_A, \Sigma_n)} \left[\lambda(\mathcal{C}_{\alpha,n}) \right]}{2\chi_n} \geqslant \frac{z_{1-\alpha}(1 - \alpha) - \tilde{z}_\alpha \Phi(\tilde{z}_\alpha) + \phi(z_{1-\alpha}) - \phi(\tilde{z}_\alpha)}{z_{1-\alpha/2}},$$

where $\lambda(\cdot)$ denotes the length (Lebesgue measure) of a set and $\tilde{z}_\alpha = z_{1-\alpha} - z_{1-\alpha/2}$.

Part i) of Assumption 2 is satisfied for $\Delta^{SD}(M)$ but not for our other ongoing examples. For example, $\Delta^{SDPB}(M)$ is convex but not centrosymmetric, and $\Delta^{RM}(\bar{M})$ is neither convex nor centrosymmetric. Part ii) of Assumption 2 is satisfied whenever parallel trends holds in both the pre-treatment and post-treatment periods ($\delta_A = 0$) and whenever δ_A is a linear trend for the case of $\Delta^{SD}(M)$.

FLCIs thus offer attractive guarantees for the case of $\Delta^{SD}(M)$. When $\alpha = 0.05$, the lower bound in Proposition 3.1 evaluates to 0.72, so the expected length of the shortest possible confidence set that satisfies the coverage requirement (10) is at most 28% shorter than the length of the optimal FLCI when the conditions of the proposition hold.

3.3 (In)Consistency of FLCIs

The finite-sample guarantees discussed above do not apply for several types of restrictions Δ of importance, including those that construct bounds using the maximum pre-treatment violation or that incorporate sign and shape restrictions. We now show that the FLCIs can perform poorly under such restrictions. We first provide two illustrative examples, and then state a formal inconsistency result.

---

[11]We use δ_A for the null value of δ, rather than δ_0, since we use the notation δ_t to refer to the component of δ corresponding with period t.

Example: $\Delta^{SDPB}(M)$ **and** $\Delta^{SDI}(M)$. Suppose $\theta = \tau_1$. The worst-case bias of an affine estimator over $\Delta^{SDPB}(M)$ or Δ^{SDI} is the same as the worst-case bias for that estimator over $\Delta^{SD}(M)$.[12] Since the construction of the optimal FLCI depends only on the worst-case bias and variance of the affine estimator, it follows that the optimal FLCI constructed using $\Delta^{SDPB}(M)$ or $\Delta^{SDI}(M)$ is the same as the one constructed using $\Delta^{SD}(M)$. Therefore, the optimal FLCI does not adapt to additional sign or monotonicity restrictions. ▲

Example: $\Delta^{RM}(\bar{M})$. Suppose $\theta = \tau_1$. If $\Delta = \Delta^{RM}(\bar{M})$ and $\bar{M} > 0$, then all affine estimators for τ_1 have infinite worst-case bias, since $\delta \in \Delta^{RM}(\bar{M})$ can have $|\delta_1|$ arbitrarily large if $|\delta_{-1}|$ is also sufficiently large. Thus, the only valid FLCI is the entire real line. ▲

We next provide a formal result on the (in)consistency of the FLCIs. We consider "small-Σ" asymptotics wherein the sampling uncertainty grows small relative to the length of the identified set, and provide necessary and sufficient conditions under which the FLCIs include fixed points outside of the identified set with non-vanishing probability.[13] Recall from Lemma 2.1 that the identified set $\mathcal{S}(\beta, \Delta)$ is an interval when Δ is convex, with length equal to $\theta^{ub}(\beta, \Delta) - \theta^{lb}(\beta, \Delta) = b^{max}(\beta_{pre}, \Delta) - b^{min}(\beta_{pre}, \Delta)$. Since the length of the identified set only depends on Δ and β_{pre}, denote it by $LID(\beta_{pre}, \Delta)$. Our next result shows that $\mathcal{C}_{\alpha,n}^{FLCI}$ is consistent if and only if $LID(\beta_{pre}, \Delta)$ is its maximum possible value, provided that the identified set is not the entire real line (in which case any procedure is trivially consistent).

Assumption 3 (Identified set maximal length and finite). *Suppose $\delta_A \in \Delta$ is such that $LID(\delta_{A,pre}, \Delta) = \sup_{\delta_{pre} \in \Delta_{pre}} LID(\delta_{pre}, \Delta) < \infty$, where $\Delta_{pre} = \{\delta_{pre} \in \mathbb{R}^T : \exists \delta_{post} \ s.t. \ (\delta'_{pre}, \delta'_{post})' \in \Delta\}$ is the set of possible values for δ_{pre}.*

Proposition 3.2. *Suppose Δ is convex and $\alpha \in (0, 0.5]$. Fix $\delta_A \in \Delta$ and $\tau_A \in \mathbb{R}^{\bar{T}}$, and suppose $\mathcal{S}(\delta_A + \tau_A, \Delta) \neq \mathbb{R}$. Then (δ_A, Δ) satisfy Assumption 3 if and only if $\mathcal{C}_{\alpha,n}^{FLCI}$ is consistent, meaning that*

$$\lim_{n \to \infty} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left(\theta^{out} \in \mathcal{C}_{\alpha,n}^{FLCI} \right) = 0 \ for \ all \ \theta^{out} \notin \mathcal{S}(\delta_A + \tau_A, \Delta).$$

Thus, if Assumption 3 fails, then $\mathcal{C}_{\alpha,n}^{FLCI}$ is inconsistent in the strong sense that it includes fixed points outside of the identified set with non-vanishing probability. It follows that there

---

[12]Suppose the vector δ maximizes the bias for an affine estimator (a, v) over $\Delta^{SD}(M)$. The vector that adds a constant slope to δ, say $\tilde{\delta}_c = \delta + c \cdot (-\underline{T}, ..., \bar{T})'$, also lies in $\Delta^{SD}(M)$, and for c sufficiently large, $\tilde{\delta}_c$ will lie in $\Delta^{SDPB}(M)$. Moreover, the worse-case bias will be the same for δ and $\tilde{\delta}_c$, since if (a, v) has finite worst-case bias it must subtract out a weighted average of the pre-treatment slopes.

[13]See, e.g., Kadane (1971) and Moreira and Ridder (2019) for other uses of small-Σ asymptotics.

will be some $\delta_A \in \Delta$ such that the FLCI is inconsistent under δ_A unless the identified set is always the same length.[14]

Remark 7. In the three-period difference-in-differences example, Assumption 3 holds everywhere for $\Delta^{SD}(M)$ (since the identified set is always the same length, $2M$), for values of δ where the sign restrictions do not bind for $\Delta^{SDPB}(M)$, and nowhere for $\Delta^{RMI}(\bar{M})$. The restrictiveness of Assumption 3 thus depends greatly on Δ. ∎

Remark 8. Proposition 3.2 implies that FLCIs can potentially be inconsistent when Δ is convex and centrosymmetric if $\delta \neq 0$. For example, if $\Delta = \{\delta \in \Delta^{SD}(M) : |\delta_1| \leqslant M\}$, then the FLCI is inconsistent whenever $\delta_{-1} \neq 0$, even though Proposition 3.1 implies that the FLCI is near-optimal for $\delta = 0$. As discussed above, however, such inconsistency does not arise for our baseline smoothness class $\Delta^{SD}(M)$.

Remark 9. In Appendix A.1, we further show that if Assumption 3 along with an additional condition (Assumption 4 introduced below) hold, then the FLCI also has local asymptotic power approaching the power envelope under the same asymptotics considered in Proposition 3.2. ∎

The results in this section establish that when certain conditions on Δ are satisfied, the FLCIs are consistent and have desirable finite-sample guarantees in terms of expected length. These conditions hold for our baseline smoothness class $\Delta^{SD}(M)$, but fail for choices of Δ that may be of interest in empirical applications such as those that construct bounds using a pre-treatment maximum or incorporate sign and monotonicity restrictions. This motivates us to next consider an alternative method for inference that can accommodate a larger range of restrictions.

4 Inference using Moment Inequalities

In this section, we introduce a more general approach for inference that has good asymptotic properties over a large class of possible restrictions Δ. We show that inference on the partially identified parameter $\theta = l'\tau_{post}$ in this setting is equivalent to testing a system of moment inequalities with a potentially large number of nuisance parameters that enter the moments linearly. We consider an implementation based on the conditional approach developed in ARP, which allows us to obtain computationally tractable confidence sets with desirable power properties for many parameter configurations.

---

[14]We also show in Lemma C.26 in the Appendix that the conditions of Proposition 3.1 imply that Assumption 3 holds. Thus, the FLCIs obtain finite sample near-optimality in a subset of the cases where they are consistent.

4.1 Representation as a moment inequality problem with linear nuisance parameters

Consider testing the null hypothesis, $H_0 : \theta = \bar{\theta}, \delta \in \Delta$ when $\Delta = \{\delta : A\delta \leqslant d\}$. We now show that testing H_0 is equivalent to testing a system of moment inequalities with linear nuisance parameters.

The model (2) implies $\mathbb{E}_{(\delta,\tau,\Sigma_n)}\left[\hat{\beta}_n - \tau\right] = \delta$, and hence $\delta \in \Delta$ if and only if $\mathbb{E}_{(\delta,\tau,\Sigma_n)}\left[A\hat{\beta}_n - A\tau\right] \leqslant d$. Defining $Y_n = A\hat{\beta}_n - d$ and $M_{post} = [0, I]'$ to be the matrix such that $\tau = M_{post}\tau_{post}$, it is immediate that the null hypothesis H_0 is equivalent to the composite null

$$H_0 : \exists \tau_{post} \in \mathbb{R}^{\bar{T}} \text{ s.t. } l'\tau_{post} = \bar{\theta} \text{ and } \mathbb{E}_{(\delta,\tau,\Sigma_n)}\left[Y_n - AM_{post}\tau_{post}\right] \leqslant 0. \tag{13}$$

In this equivalent form, $\tau_{post} \in \mathbb{R}^{\bar{T}}$ is a vector of nuisance parameters that must satisfy the linear constraint $l'\tau_{post} = \bar{\theta}$.

By applying a change of basis, we can further re-write H_0 as an equivalent composite null hypothesis with an unconstrained nuisance parameter. Re-write the expression $AM_{post}\tau_{post}$ as $\tilde{A}\begin{pmatrix} \theta \\ \tilde{\tau} \end{pmatrix}$, where $\tilde{A}$ is the matrix that results from applying a suitable change of basis to the columns of AM_{post}, and $\tilde{\tau} \in \mathbb{R}^{\bar{T}-1}$.[15] The null H_0 is then equivalent to

$$H_0 : \exists \tilde{\tau} \in \mathbb{R}^{\bar{T}-1} \text{ s.t. } \mathbb{E}\left[\tilde{Y}_n(\bar{\theta}) - \tilde{X}\tilde{\tau}\right] \leqslant 0, \tag{14}$$

where $\tilde{Y}(\bar{\theta}) = Y_n - \tilde{A}_{(\cdot,1)}\bar{\theta}$ and $\tilde{X} = \tilde{A}_{(\cdot,-1)}$. Since $\tilde{Y}_n(\bar{\theta})$ is normally distributed with covariance matrix $\tilde{\Sigma}_n = A\Sigma_n A'$ under the finite-sample normal model (2), testing $H_0 : \theta = \bar{\theta}, \delta \in \Delta$ is equivalent to testing a system of moment inequalities with linear nuisance parameters.

Remark 10. Testing the hypothesis (14) is a special case of the problem studied in ARP, which focuses on testing null hypotheses of the form $H_0 : \exists \tau \text{ s.t. } \mathbb{E}\left[Y(\theta) - X\tau \mid X\right] \leqslant 0$. Our setting is a special case of this framework in which: i) the variable X takes the degenerate distribution $X = \tilde{X}$, and ii) $Y(\theta) = \tilde{Y}(\theta)$ is linear in θ. The first feature plays an important role in developing our consistency and local asymptotic power results presented later in this section: if i) fails and X is continuously distributed, then the tests proposed by ARP will generally not be consistent, as they do not allow for the number of moments to grow with n. The current proof of the optimal local asymptotic result also exploits the geometry of

---

[15]Let Γ be a square matrix with the vector l' in the first row and remaining rows chosen so that Γ has full rank. Define $\tilde{A} := AM_{post}\Gamma^{-1}$. Then $AM_{post}\tau = \tilde{A}\Gamma\tau_{post} = \tilde{A}\begin{pmatrix} \theta \\ \underbrace{\Gamma_{(-1,\cdot)}\tau_{post}}_{:=\tilde{\tau}} \end{pmatrix}$. If $\bar{T} = 1$, then $\tilde{\tau}$ is 0-dimensional and should be interpreted as 0.

feature ii), although we conjecture that this could be relaxed to allow $Y(\theta)$ to vary smoothly in θ. ∎

4.2 Constructing conditional confidence sets

A practical challenge to testing the hypothesis (14) in our setting is that the dimension of the nuisance parameter $\tilde{\tau} \in \mathbb{R}^{\bar{T}-1}$ grows linearly with the number of post-periods $\bar{T}$ and may be large in practice. For instance, in Section 6 we revisit an empirical paper in which $\bar{T} = 23$. Moreover, 5 of the 12 recent event-study papers reviewed in Roth (2019) have $\bar{T} > 10$. This renders many moment inequality methods, especially those which rely on test inversion over a grid for the full parameter vector, practically infeasible. To tractably deal with the nuisance parameter, we leverage the conditional approach of ARP, which directly exploits the linear structure of the hypothesis (14) and delivers computationally tractable and powerful tests even when the number of post-periods $\bar{T}$ is large.[16] We briefly sketch the construction of the conditional testing approach in our testing problem, and refer the reader to ARP for full details.

Suppose we wish to test (14) for some fixed $\bar{\theta}$. The conditional testing approach considers tests based on the profiled test statistic

$$\hat{\eta} := \min_{\eta, \tilde{\tau}} \eta \text{ s.t. } \tilde{Y}_n(\bar{\theta}) - \tilde{X}\tilde{\tau} \leqslant \tilde{\sigma}_n \cdot \eta, \tag{15}$$

where $\tilde{\sigma}_n = \sqrt{\operatorname{diag}(\tilde{\Sigma}_n)}$. This linear program selects the value of the nuisance parameters $\tilde{\tau} \in \mathbb{R}^{\bar{T}-1}$ that makes the maximum studentized moment the smallest. Duality results from linear programming (e.g. Schrijver (1986), Section 7.4) imply that the value $\hat{\eta}$ obtained from the primal program (15) equals the optimal value of the dual program,[17]

$$\hat{\eta} = \max_{\gamma} \gamma' \tilde{Y}_n(\bar{\theta}) \text{ s.t. } \gamma' \tilde{X} = 0, \gamma' \tilde{\sigma}_n = 1, \gamma \geqslant 0. \tag{16}$$

If a vector γ_* is optimal in the dual problem above, then it is a vector of Lagrange multipliers for the primal problem. We denote by $\hat{V}_n$ the set of optimal vertices of the dual program.[18]

---

[16]Other moment inequality methods have been proposed for subvector inference, but typically do not exploit the linear structure of our setting — see, e.g, Romano and Shaikh (2008); Chernozhukov, Newey and Santos (2015); Bugni, Canay and Shi (2017); Chen, Christensen and Tamer (2018); Kaido, Molinari and Stoye (2019). Cho and Russell (2019), Gafarov (2019), and Flynn (2019) also provide methods for subvector inference with linear moment inequalities, but in contrast to our approach require a linear independence constraint qualification (LICQ) assumption for size control.

[17]Technically, the duality results require that $\hat{\eta}$ be finite. However, one can show that $\hat{\eta}$ is finite with probability 1, unless the span of $\tilde{X}$ contains a vector with all negative entries, in which case the identified set for θ is the real line. We therefore trivially define our test never to reject if $\hat{\eta} = -\infty$.

[18]In general, there may not be a unique solution to the dual program. However, Lemma 11 of ARP shows

To construct critical values, Lemma 9 of ARP shows that conditional on the event $\gamma_* \in \hat{V}_n$ and a sufficient statistic S_n for the nuisance parameters, the test statistic $\hat{\eta}$ follows a truncated normal distribution,

$$\hat{\eta} \mid \{\gamma_* \in \hat{V}_n, S_n = s\} \sim \xi \mid \xi \in [v^{lo}, v^{up}],$$

where $\xi \sim \mathcal{N}\left(\gamma_*'\tilde{\mu}, \gamma_*'\tilde{\Sigma}_n\gamma_*\right)$, $\tilde{\mu} = \mathbb{E}\left[\tilde{Y}_n(\bar{\theta})\right]$, $S_n = (I - \frac{\tilde{\Sigma}_n\gamma_*}{\gamma_*'\tilde{\Sigma}_n\gamma_*}\gamma_*')\tilde{Y}_n(\bar{\theta})$, and v^{lo}, v^{up} are known functions of $\tilde{\Sigma}_n, s, \gamma_*$.[19] All quantiles of the conditional distribution of $\hat{\eta}$ in the previous display are increasing in $\gamma_*'\tilde{\mu}$, and the null hypothesis (14) implies $\gamma_*'\tilde{\mu} \leqslant 0$. Therefore, the critical value for the conditional test is the $1 - \alpha$ quantile of the truncated normal distribution $\xi | \xi \in [v^{lo}, v^{up}]$ under the worst-case assumption that $\gamma_*'\tilde{\mu} = 0$. Let $\psi_\alpha^C(\tilde{Y}_n(\bar{\theta}), \tilde{\Sigma}_n)$ denote an indicator for whether the conditional test rejects at the $1 - \alpha$ level. Proposition 6 in ARP implies that the conditional test controls size in the normal model (2). A confidence set satisfying the uniform coverage criterion (10) can thus be constructed via test inversion, $\mathcal{C}_{\alpha,n}^C := \{\bar{\theta} : \psi_\alpha^C(\tilde{Y}_n(\bar{\theta}), \tilde{\Sigma}_n) = 0\}$. Such confidence sets are easy to compute, because they only require test inversion for the scalar parameter $\bar{\theta}$, and not for the higher-dimensional nuisance $\tilde{\tau}$.

ARP provide high-level conditions under which coverage in the normal model translates to uniform asymptotic coverage over a large class of data-generating processes. In the supplementary material, we provide analogous uniform asymptotic results under weaker, lower-level conditions applicable to the difference-in-differences setting.

4.3   Consistency and optimal local asymptotic power of conditional confidence sets

We now provide two results on the asymptotic power of the conditional test in our setting. First, the conditional test is consistent, meaning that any fixed point outside of the identified set is rejected with probability approaching one as the sample size $n \to \infty$.

Proposition 4.1. *The conditional test is consistent for all polyhedral* Δ. *For any* $\delta_A \in \Delta$, $\tau_A \in \mathbb{R}^{\bar{T}}$, *and* $\theta^{out} \notin \mathcal{S}(\delta_A + \tau_A, \Delta)$,

$$\lim_{n \to \infty} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)}\left(\theta^{out} \notin \mathcal{C}_{\alpha,n}^C\right) = 1.$$

---

that conditional on any one vertex of the dual program's feasible set being optimal, every other vertex is optimal with either probability 0 or 1. It thus suffices to condition on the event that a vector $\gamma_* \in \hat{V}$.

[19]The cutoffs v^{lo} and v^{up} are the maximum and minimum of the set $\{x : x = \max_{\gamma \in F_n} \gamma'(s + \frac{\tilde{\Sigma}_n\gamma_*}{\gamma_*'\tilde{\Sigma}_n\gamma_*}x)\}$ when $\gamma_*'\tilde{\Sigma}_n\gamma_* \neq 0$, where F_n is the feasible set of the dual program (16). When $\gamma_*'\tilde{\Sigma}_n\gamma_* = 0$, we define $v^{lo} = -\infty$ and $v^{up} = \infty$, so the conditional test rejects if and only if $\hat{\eta} > 0$.

Thus, in contrast to the optimal FLCI, the conditional test is consistent for all polyhedral Δ. Moreover, this result extends immediately to confidence sets for the case where Δ is the finite union of polyhedra of the form considered in Lemma 2.2. In the supplemental materials, we provide a uniform asymptotic version of this consistency result for points bounded away from the boundary of the identified set.

We next consider the local asymptotic power of the conditional test. We provide a condition under which the power of the conditional test against local alternatives converges to the power envelope. This condition guarantees that the binding and non-binding moments are sufficiently well-separated at points close to the boundary of the identified set.

Assumption 4. *Let $\Delta = \{\delta : A\delta \leqslant d\}$ and fix $\delta_A \in \Delta$. Consider the optimizations:*

$$b^{max}(\delta_{A,pre}) = \max_{\delta} l'\delta_{post} \ s.t. \ A\delta \leqslant d, \delta_{pre} = \delta_{A,pre}$$

$$b^{min}(\delta_{A,pre}) = \min_{\delta} l'\delta_{post} \ s.t. \ A\delta \leqslant d, \delta_{pre} = \delta_{A,pre}$$

Assume there exists a solution δ^* to b^{max} such that that the rank of $A_{(B(\delta^*),post)}$ is equal to $|B(\delta^*)|$, where $B(\delta^*)$ denotes the index of the binding moments at δ^*.[20] Likewise, assume there exists a solution δ^{**} to b^{min} such that the rank of $A_{(B(\delta^{**}),post)}$ is equal to $|B(\delta^{**})|$.

Assumption 4 considers the problem of finding the differential trend $\delta \in \Delta$ that is consistent with the pre-trend identified from the data and causes $l'\hat{\beta}_{post}$ to be maximally (or minimally) biased for $\theta := l'\tau_{post}$. It requires that the "right" number of moments bind when we do this optimization.

Remark 11 (Connection to LICQ)**.** Assumption 4 is slightly weaker than linear independence constraint qualification (LICQ), which has been used recently in the moment inequality settings of Gafarov (2019), Cho and Russell (2019), Flynn (2019), and Kaido and Santos (2014); see Kaido, Molinari and Stoye (2020) for a synthesis. We discuss this connection formally in Appendix A.2. We note, however, that many of the aforementioned papers require LICQ for asymptotic size control, whereas we impose Assumption 4 only for our results on local asymptotic power. ∎

Remark 12. In the case with one post-treatment period ($\underline{T} = 1$), so that there are no nuisance parameters, Assumption 4 is satisfied when there is one moment binding at the edge of the identified set. This assumption holds everywhere for $\Delta^{SD}(M)$ when $M > 0$. It holds almost everywhere for $\Delta^{SDPB}(M)$ when $M > 0$, although it fails when both the sign restrictions and smoothness restrictions are simultaneously binding. When $M = 0$, both the

---

[20]That is, $A_{(B(\delta^*),\cdot)}\delta^* = d_{B(\delta^*)}$ and $A_{(-B(\delta^*),\cdot)}\delta^* - d_{-B(\delta^*)} = -\epsilon_{-B(\delta^*)} < 0$.

upper and lower bounds for $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$ are binding, so the assumption fails. More generally, one can show that Assumption 4 does not hold if θ is point identified. ∎

With this definition in hand, we can now formalize the sense in which the conditional test has optimal local asymptotic power under Assumption 4. Again let $\mathcal{I}_\alpha(\Delta, \Sigma_n)$ denote the class of confidence sets that satisfy the coverage criterion in (10) at the $1 - \alpha$ level. Under Assumption 4, the power of the conditional test against local alternatives converges to the power envelope over $\mathcal{I}_\alpha(\Delta, \Sigma_n)$ as $n \to \infty$.

Proposition 4.2. *Fix $\delta_A \in \Delta, \tau_A$, and suppose Σ^* is positive definite. Let $\theta_A^{ub} = \sup_\theta \mathcal{S}(\delta_A + \tau_A, \Delta)$ be the upper bound of the identified set. Suppose Assumption 4 holds. Then, for any $x > 0$,*

$$\lim_{n \to \infty} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}^C \right) = \lim_{n \to \infty} \sup_{\mathcal{C}_{\alpha,n} \in \mathcal{I}_\alpha(\Delta, \Sigma_n)} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n} \right)$$
$$= \Phi(c^* x - z_{1-\alpha}),$$

for a positive constant c^*. The analogous result holds replacing $\theta_A^{ub} + \frac{1}{\sqrt{n}}x$ with $\theta_A^{lb} - \frac{1}{\sqrt{n}}x$, for θ_A^{lb} the lower bound of the identified set (although the constant c^* may differ).

In the supplemental materials, we provide a uniform asymptotic analog to this result under a uniform version of Assumption 4.

The proof of our novel local asymptotic optimality result proceeds in two steps. First, we show that the local asymptotic power of any test that controls size is bounded above by that of a particular one-sided t-test under Assumption 4. More specifically, Assumption 4 implies that there is a unique set of Lagrange multipliers $\bar{\gamma}$ in the "population version" of the test statistic $\hat{\eta}(\theta^{ub})$ that replaces $\tilde{Y}(\theta^{ub})$ with its expectation $\tilde{\mu}(\theta^{ub})$ in (15). We show that the optimal test is a one-sided t-test in the direction of $\bar{\gamma}$ for alternatives sufficiently close to θ^{ub}. Second, we show that the conditional test converges in probability to this optimal one-sided t-test.

An immediate corollary is that when $\Delta = \bigcup_{k=1}^K \Delta_k$, the conditional test based on the union of confidence sets has optimal local asymptotic power when the Δ_k that determines the identified set bounds is unique and satisfies the conditions of Proposition 4.2.

Corollary 4.1. *Fix δ_A, τ_A. Suppose that (δ_A, Δ_1) satisfy Assumption 4. If $\Delta_2, .., \Delta_K$ are polyhedra such that $\mathcal{S}(\delta_A + \tau_A, \Delta_k) \subsetneq \mathcal{S}(\delta_A + \tau_A, \Delta_1)$ for all $k > 1$, then the conclusion of Proposition 4.2 holds for $\Delta = \bigcup_{k=1}^K \Delta_k$ and $\mathcal{C}_{\alpha,n}^C$ the union of the conditional confidence sets for $\Delta_1, ..., \Delta_K$.*

This implies, for instance, that the conditional test has optimal local asymptotic power for $\Delta^{RM}(\bar{M})$ when there is a unique (non-zero) pre-treatment maximum violation, i.e. when $\max_{s<0} |\delta_{s+1} - \delta_s| > 0$ has a unique solution. Likewise, the conditional test is optimal for $\Delta^{SDRM}(\bar{M})$ when there is a unique maximum non-linearity in the pre-treatment period.

Remark 13 (Relationship to other moment inequality methods)**.** We are not aware of results analogous to Proposition 4.2 for any other moment inequality procedure that controls size in the finite sample normal model. Observe that if Assumption 4 holds, then it also holds if Δ is augmented to include a moment that is non-binding at both endpoints of the identified set. Hence, for Proposition 4.2 to hold, the local asymptotic power of the test needs to be unaffected by the inclusion of such slack moments. For example, although relatively insensitive to the inclusion of slack moments, the procedures of Romano, Shaikh and Wolf (2014) and Andrews and Barwick (2012) are still affected by the inclusion of slack moments via the changes to the first-stage critical value and size-adjustment factor, respectively.[21]

Remark 14 (Finite sample power of the conditional test)**.** The argument for the optimality of the conditional approach relies on a unique vector of Lagrange multipliers $\bar{\gamma}$ being dual-optimal with probability approaching 1 asymptotically. The asymptotic guarantees of Proposition 4.2 thus may not translate to good finite-sample performance in settings where multiple vectors of Lagrange multipliers are optimal with nontrivial probability. Since a vector of Lagrange multipliers corresponds with a set of active moments in the primal problem (15), this will tend to occur in cases where the set of binding and non-binding moments are not "well-separated" relative to the sampling variation in the data. ∎

Remark 15 (Hybridization)**.** To mitigate the poor power of the conditional test when the binding and non-binding moments are not "well-separated," ARP recommend the use of a Bonferroni-like hybrid test that combines a first-stage test using least favorable (LF) critical values with the conditional test. In Appendix B.1, we show that a similar hybrid test can be constructed using FLCIs as well. We also show that when the size used for the first-stage test is small, these hybrid approaches have near-optimal local asymptotic power under Assumption 4.2. We evaluate these hybrid approaches in our Monte Carlo simulations below.

5 Simulation study

In this section, we conduct a simulation study to investigate the performance of the discussed confidence sets across a range of relevant data-generating processes. We find good size control

---

[21]In concurrent work, Cox and Shi (2020) propose a new method for testing moment inequalities with nuisance paramaters, which like the ARP test is strongly insensitive to slack moments. It is thus possible that similar results could be obtained for their test as well.

for all of the procedures, and therefore focus in the main text on a comparison of power to provide concrete recommendations on the best approach in practice. In the supplementary material, we present results on size control and other additional simulation results.

5.1 Simulation Design

Our simulations are calibrated using the estimated covariance matrix from the 12 recently-published papers surveyed in Roth (2019). For any given paper in the survey, we denote by $\hat{\Sigma}$ the estimated variance-covariance matrix from the event-study in the paper, calculated using the clustering scheme specified by the authors. For a chosen mean vector β, we simulate event-study coefficients $\hat{\beta}_s$ from a normal model, $\hat{\beta}_s \sim \mathcal{N}\left(\beta, \hat{\Sigma}\right)$.[22] In simulation s, we construct nominal 95% confidence sets for the parameter of interest θ using the pair $(\hat{\beta}_s, \hat{\Sigma})$ for each proposed procedure. The parameter of interest is the causal effect in the first post-treatment period $(\theta = \tau_1)$.[23]

For a given choice of Δ, we compute the identified set $\mathcal{S}(\beta, \Delta)$ and calculate the expected excess length for each of the proposed confidence sets. We benchmark the expected excess length of our proposed confidence sets relative to an efficiency bound for confidence sets that satisfy the uniform coverage requirement.[24] We report the efficiency ratio of each procedure, which is defined as the ratio of the optimal benchmark relative to the average excess length for the procedure. All results are calculated over 1000 simulations per paper.

We consider four choices of Δ to highlight the performance of our proposed confidence sets across a range of conditions: $\Delta^{SD}(M)$, $\Delta^{SDPB}(M)$, $\Delta^{RM}(\bar{M})$, and $\Delta^{SDRM}(\bar{M})$. We consider simulations under the assumption of zero treatment effects, so that $\tau = 0$ and thus $\beta = \delta$. We consider two forms for δ. First, we consider the baseline case of parallel trends $(\delta = 0)$. Second, we consider a "pulse" pre-trend in which δ_{-1} is non-zero and the remaining elements of δ are zero. Such a pre-trend might arise in practice if there are confounding policy changes or other events close to the time of treatment. These different choices of δ

---

[22]We focus on the normal simulations in the main text since it allows for a tractable computation of the optimal excess length of procedures that control size. In the supplementary material, we show that our procedures perform similarly in simulations based on the empirical distribution in the original paper.

[23]In the supplementary material, we provide simulation results in which the parameter of interest is the average causal effect in the post-treatment periods $(\theta = \bar{\tau}_{post})$, with qualitatively similar results.

[24]For choices of Δ that are convex (e.g., $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$), we benchmark the expected excess length of our proposed confidence sets against a sharp optimal bound over confidence sets that satisfy the uniform coverage requirement. This optimal bound is provided in the supplementary materials, and follows as a corollary from results in Armstrong and Kolesar (2018a) on the optimal expected length of a confidence set satisfying the uniform coverage requirement (10). For choices of Δ that can be written as the union of convex sets (e.g., $\Delta^{RM}(\bar{M})$ and $\Delta^{SDRM}(\bar{M})$), we compare the expected excess length of our proposed confidence sets against the maximal optimal bound over each set in the union, which is a potentially non-sharp bound for any confidence set with correct coverage.

24

| | **Parallel Trends** | | **Pulse Pre-Trend** | |
	$\Delta^{SD}(M)$	$\Delta^{SDPB}(M)$	$\Delta^{SDRM}(\bar{M})$	$\Delta^{RM}(\bar{M})$
Conditional (and Hybrids)				
Consistent	✓	✓	✓	✓
Asymptotically (near-)optimal	✓	✓	✓	✗
FLCI				
Consistent	✓	✗	✗	✗
Asymptotically optimal	✓	✗	✗	✗
Finite-sample near-optimal	✓	✗	✗	✗

Table 1: Summary of expected properties for each simulation design

allow us to highlight the relative strengths of the proposed inference procedures, since FLCIs have near-optimal expected length when $\delta = 0$ and $\Delta = \Delta^{SD}(M)$, whereas the conditional test has optimal local asymptotic power under the pulse design when $\Delta = \Delta^{SDPB}(M)$.

In practice, we find that for $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$, the results depend on M but are qualitatively similar across values of δ. By contrast, for $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$, the choice of δ is more important than the choice of $\bar{M}$. Therefore, to highlight the most important dimensions for each of the simulation designs, in the main text of the paper we report results for $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$ under different values of M and $\delta = 0$ (parallel trends), whereas for $\Delta^{RM}(\bar{M})$ and $\Delta^{SDRM}(\bar{M})$ we vary the magnitude of the pre-treatment pulse δ_{-1}, holding $\bar{M} = 1$ constant. In the supplementary materials, we report results for additional choices of the parameters.

We report results for four methods for constructing confidence sets: FLCIs, conditional confidence sets, and two hybrid approaches that combine the conditional test with either a least-favorable moment inequality test or FLCIs (see Remark 15).[25] For $\Delta^{RM}(\bar{M})$ and $\Delta^{SDRM}(\bar{M})$, we omit results for the FLCI and conditional-FLCI hybrid since the FLCIs have infinite length. Table 1 summarizes which of our theoretical results hold for each of the simulation designs.

5.2 Simulation Results

To compare results easily across the 12 papers in the simulation study, we normalize the units of δ_{-1} and M by the standard deviation of $\hat{\beta}_1$ (denoted σ_1). Comparative statics as the normalized values of M or δ_{-1} grow large thus mimic the "small-Σ" asymptotics considered above. In the graphs below, we report the median value of excess length efficiency across the papers in the survey.

---

[25]We use a first-stage test of size $\kappa = \alpha/10$, following ARP and Romano et al. (2014).

Figure 1: Simulation results for $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$: Median efficiency ratios for proposed procedures.
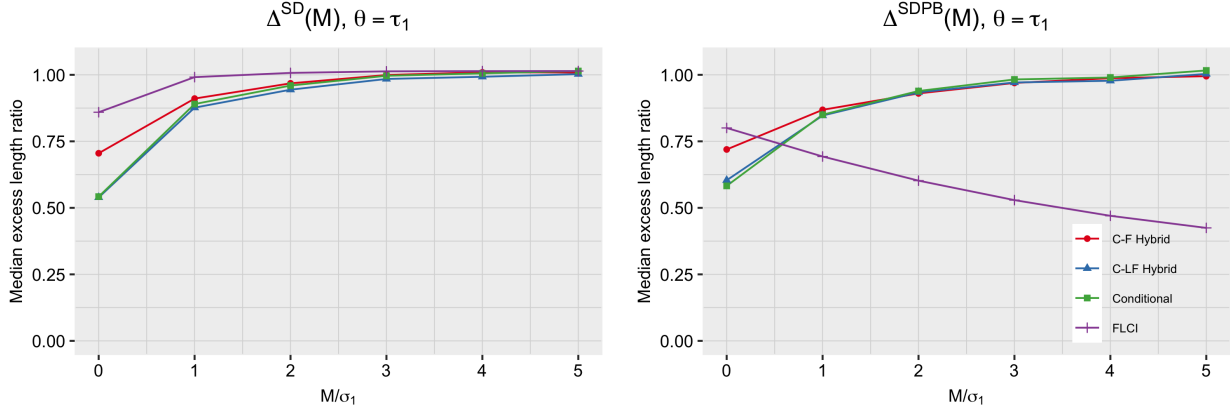
Note: Median efficiency ratios for our proposed confidence sets over $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$ under the assumption of parallel trends and zero treatment effects (i.e., $\beta = 0$). The efficiency ratio for a procedure is defined as the efficiency bound divided by the procedure's expected excess length. The results for the FLCI are plotted in purple, the results for the conditional-FLCI ("C-F Hybrid") confidence interval in red, the results for the conditional-LF ("C-LF Hybrid") hybrid in blue, and the results for the conditional confidence interval in green. Results are averaged over 1000 simulations for each of the 12 papers surveyed, and the median across papers is reported here.

Figure 2: Simulation results for $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$: Median efficiency ratios for proposed procedures.
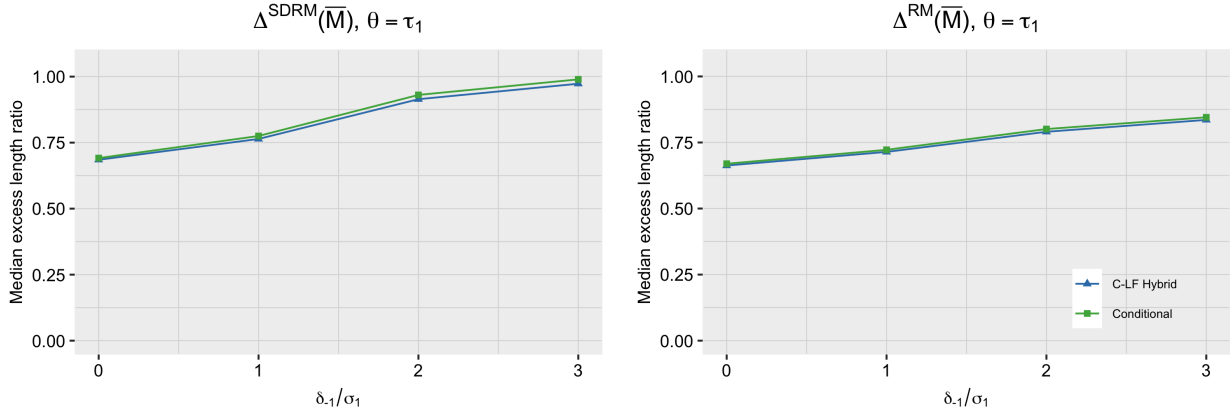
Note: Median efficiency ratios for our proposed confidence sets over $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$ with $\bar{M} = 1$ under the assumption of zero treatment effects and a "pulse" pre-trend (i.e., $\beta_{-1} = \delta_{-1}$ and $\beta_t = 0$ for all $t \neq -1$). The efficiency ratio for a procedure is defined as the efficiency bound divided by the procedure's expected excess length. The results for the conditional-least favorable ("C-LF") hybrid are plotted in blue, and the results for the conditional confidence interval in green. Results are averaged over 1000 simulations for each of the 12 papers surveyed, and the median across papers is reported here.

Results for $\Delta^{SD}(M)$: The left panel of Figure 1 plots the efficiency ratio for each procedure as a function of M/σ_1 when $\Delta = \Delta^{SD}(M)$. All procedures perform well as M/σ_1 grows large

with efficiency ratios approaching 1, illustrating our asymptotic (near-)optimality results for this design. However, the FLCIs perform best for smaller values of M/σ_1, including the point-identified case where $M = 0$, illustrating the finite-sample near-optimality results for the FLCIs when Assumption 2 holds. Although the conditional confidence sets have efficiency approaching the optimal bound for M/σ_1 large, their efficiency is only about 50% when $M/\sigma_1 = 0$ (and thus Assumption 4 fails). The conditional-FLCI hybrid substantially improves efficiency relative to the conditional for small values of M/σ_1, while still retaining near-optimal performance as M/σ_1 grows large.

Results for $\Delta^{SDPB}(M)$: The right panel of Figure 1 plots the efficiency ratio for each procedure as a function of M/σ_1 when $\Delta = \Delta^{SDPB}(M)$. The efficiency ratios for the conditional and hybrid confidence sets are again (near-)optimal as M/σ_1 grows large, highlighting our asymptotic (near-)optimality results for these procedures in this simulation design. By contrast, the efficiency ratios for the FLCIs steadily decrease as M/σ_1 increases, reflecting that the FLCIs are not consistent in this simulation design when $M > 0$. The conditional-FLCI hybrid again improves efficiency relative to the conditional when M/σ_1 is small and retains near-optimal performance as M/σ_1 grows large.

Results for $\Delta^{SDRM}(\bar{M})$: The left panel of Figure 2 plots the efficiency ratios for the conditional and conditional-least favorable hybrid confidence sets as a function of δ_{-1}/σ_1 when $\Delta = \Delta^{SDRM}(\bar{M})$. We omit results for the FLCI and conditional-FLCI hybrid since the optimal FLCI has infinite length for this design. Both procedures perform well as δ_{-1}/σ_1 grows large with efficiency ratios approaching 1, illustrating our asymptotic (near-) optimality result for this design. Both procedures also have similar power curves.

Results for $\Delta^{RM}(\bar{M})$: The right panel of Figure 2 plots the efficiency ratio for the conditional and conditional-least favorable hybrid confidence sets as a function of δ_{-1}/σ_1 when $\Delta = \Delta^{RM}(\bar{M})$. We again omit results for the FLCI and conditional-FLCI hybrid since the optimal FLCI has infinite length for this design. The conditions for our asymptotic (near-) optimality result for unions of convex sets do not hold in this simulation design (as the maximum pre-period violation is not unique). Nonetheless, we find that the conditional-least favorable hybrid confidence set and the conditional confidence set perform quite well for large values of δ_{-1}/σ_1, with efficiency ratios approaching about 83%. This is encouraging as it shows that these procedures may perform well even in cases where the conditions of Corollary 4.1 fail. Once again, we also find that the conditional and conditional-least favorable hybrid have similar power.

5.3 Practical Recommendations

Two clear patterns emerge from our simulations. First, the FLCIs have the best performance for $\Delta^{SD}(M)$, particularly when M is small, which aligns with the finite-sample near-optimality results in Section 3. Second, the conditional approach and its hybrid variants outperform the FLCIs for other choices of Δ where the consistency of the FLCIs is not guaranteed.

We thus recommend the use of FLCIs for the case of $\Delta^{SD}(M)$, where FLCIs are consistent and have good finite-sample properties, and recommend a moment-inequality approach for more general forms of Δ where consistency of the FLCIs is not guaranteed. The choice between the conditional and hybrid approaches is somewhat more nuanced, as their performance is quite similar in our simulation designs. We do find somewhat better performance for the conditional-FLCI hybrid for $\Delta^{SDPB}(M)$, and thus recommend this approach for this choice of Δ. For $\Delta^{RM}(\bar{M})$ and $\Delta^{SDRM}(\bar{M})$, the FLCIs have infinite length, and we find nearly identical performance for the conditional and conditional-least favorable hybrid. We therefore tentatively recommend the least favorable hybrid approach following the recommendations of ARP. We implement these recommendations in our applications in the next section.

6   Empirical Applications

We recommend that applied researchers use our methods to conduct a sensitivity analysis in which they construct robust confidence intervals for different choices of Δ. For example, many of the Δs described above have a parameter M (or $\bar{M}$) that determines the informativeness of the pre-trends about the post-treatment differences in trends. It is natural to report sensitivity to the parameter M, as well as the "breakdown" value at which particular hypotheses of interest can no longer be rejected.[26] This analysis makes transparent what assumptions need to be placed on the relationship between the pre-trends and the counterfactual post-treatment differences in trends in order to draw particular conclusions of interest. We illustrate such a sensitivity analysis in applications to two recently published papers.

---

[26]Similar "breakdown" concepts have been proposed in other settings with partial identification (Horowitz and Manski, 1995; Kline and Santos, 2013; Masten and Poirier, 2020).

6.1 Estimating the incidence of a value-added tax cut

Benzarti and Carloni (2019, henceforth, BC) study the incidence of a decrease in the value-added tax (VAT) on restaurants in France. France reduced its VAT on sit-down restaurants from 19.6 to 5.5 percent in July of 2009. BC analyze the impact of this change using a dynamic difference-in-differences design that compares restaurants to a control group of other market services firms that were not affected by the VAT change, estimating

$$Y_{it} = \sum_{s \neq 2008} \beta_s \times 1[t = s] \times D_i + \phi_i + \lambda_t + \epsilon_{it}, \tag{17}$$

where Y_{it} is the log of (before-tax) profits for firm i in in year t; D_i is an indicator for whether firm i is a restaurant; ϕ_i and λ_t are firm and year fixed effects; and standard errors are clustered at the regional level. BC's main finding is that the VAT reduction had a large effect on restaurant profits. Figure 3 shows the estimated event-study coefficients $\{\hat{\beta}_s\}$ from specification (17). We can formally reject the hypothesis that $\beta_{pre} = 0$ ($p < 0.01$), although visually there appears to be a jump in the coefficients following treatment.

Figure 3: Event-study coefficients $\{\beta_s\}$ for log profits, estimated using the event-study specification in (17).
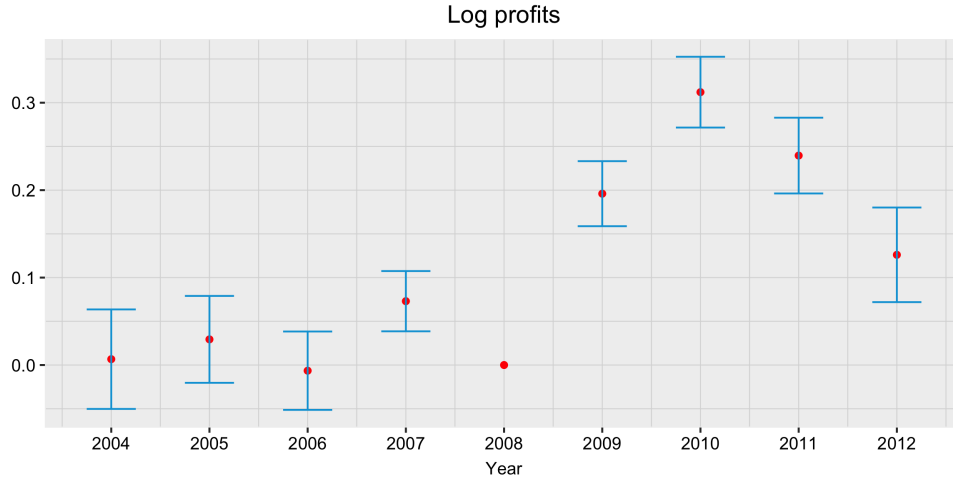

The top left panel of Figure 4 shows a sensitivity analysis that plots robust confidence sets for the treatment effect in 2009 using $\Delta^{SD}(M)$ for different values of M. The confidence sets contain only positive values unless M exceeds 0.22. Thus, we can reject a null effect on profits in 2009 if we are willing to restrict the slope of the differential trend to change by no more than 22 log points between periods. To further contextualize these results, the top right panel of the figure shows a sensitivity analysis using $\Delta^{SDRM}(\bar{M})$, from which we see that the breakdown value of $\bar{M}$ is about 1.5. This indicates that the significant effect

is robust to allowing a non-linearity in the differential trend that is about 1.5 times the maximum observed in the pre-treatment period. Similarly, the bottom left panel shows a sensitivity analysis for $\Delta^{RM}(\bar{M})$, with a breakdown value of $\bar{M} \approx 2$, indicating that we would have to allow the violation of parallel trends between 2008 and 2009 to be roughly twice the magnitude of the maximal pre-treatment violation to include a null effect in our confidence sets.

Figure 4: Sensitivity analysis for $\theta = \tau_{2009}$ for Benzarti and Carloni (2019)
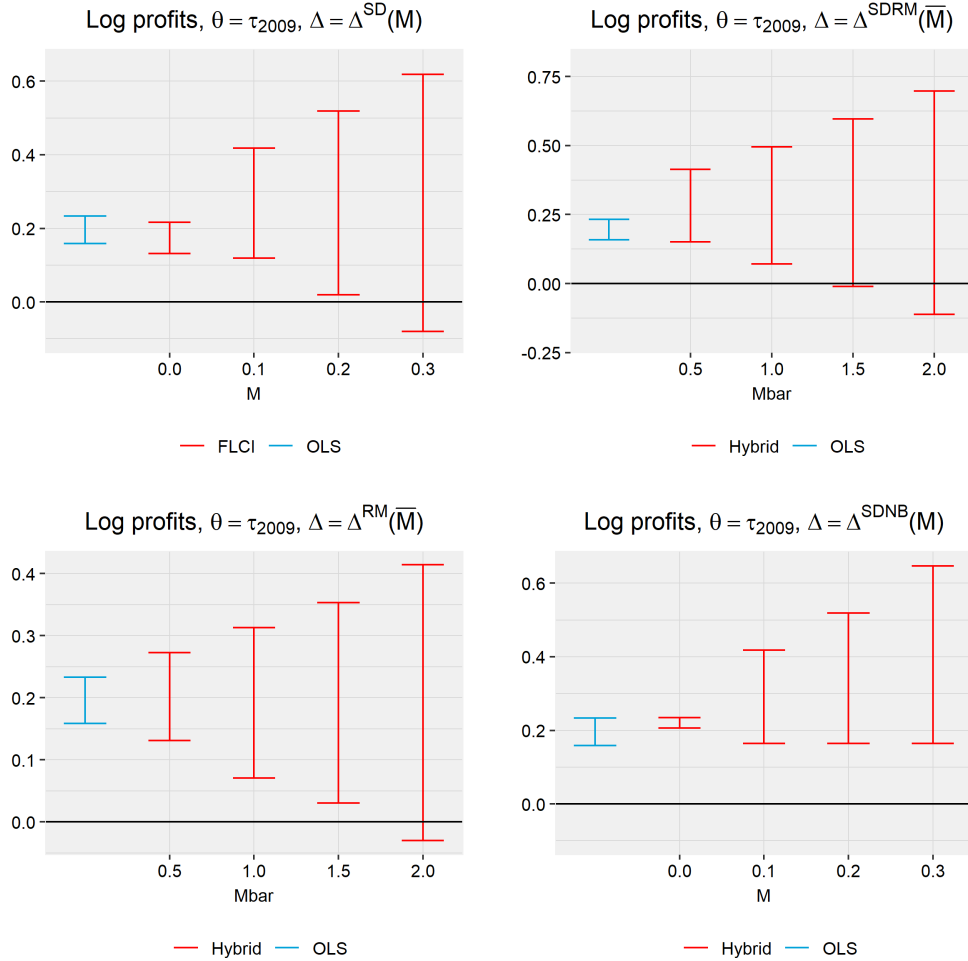
We can further tighten our bounds by including context-specific information. Since the VAT cut occurred at the same time that a payroll subsidy for restaurants was terminated, BC write, "a conservative interpretation of our results is that we are estimating a lower bound on the effect of the VAT cut on profits" (pg. 40). This argument may be made precise by further imposing that the bias of the post-period event-study coefficients is negative. The bottom right panel of Figure 4 imposes the additional constraint that the sign of the bias be negative — that is, we set $\Delta = \Delta^{SDNB}(M) := \Delta^{SD}(M) \cap \{\delta : \delta_{post} \leqslant 0\}$. With this added

constraint, the robust confidence sets now rule out effects on profits smaller than 15 log points for all values of M, highlighting how our approach allows researchers to incorporate informative context-specific knowledge to obtain more precise inference.

6.2 The effect of duty-to-bargain laws on long-run student outcomes

Lovenheim and Willen (2019, henceforth LW) study the impact of state-level public sector duty-to-bargain (DTB) laws, which mandated that school districts bargain in good faith with teachers' unions. LW examine the impacts of these laws on the adult labor market outcomes of people who were students around the time that these laws were passed, comparing individuals across different states and different birth cohorts to exploit the differential timing of the passage of DTB laws across states. The authors estimate the following regression specification separately for men and women, using data from the American Community Survey (ACS),

$$Y_{sct} = \sum_{r=-11}^{21} D_{scr}\beta_r + X'_{sct}\gamma + \lambda_{ct} + \phi_s + \epsilon_{sct}. \tag{18}$$

Y_{sct} is an average outcome for the cohort of students born in state s in cohort c in ACS calendar year t. D_{scr} is an indicator for whether state s passed a DTB law r years before cohort c turned age 18.[27] The event-study coefficients $\{\hat{\beta}_r\}$ estimate the dynamic treatment effects (or placebo effects) r years after DTB passage.[28] The remaining terms include time-varying controls, birth-cohort-by-ACS-year fixed effects, and state fixed effects. We normalize the event-study coefficient β_{-2} to 0.[29] We focus on the results where the outcome is employment.

Figure 5 plots the estimated event-study coefficients $\{\hat{\beta}_r\}$ from specification (18). In the event-study for men (left panel), the pre-period coefficients are relatively close to zero, whereas the longer-run post-period coefficients are negative. By contrast, the results for women (right panel) suggest a downward-sloping pre-existing trend.

---

[27]$D_{sc,-11}$ is set to 1 if state s passed a law 11 years or more after cohort c turned 18. Likewise, $D_{sc,21}$ is set to 1 if state s passed a law 21 or more years before cohort c turned 18.

[28]Treatment timing in LW is staggered, and therefore the results in Sun and Abraham (2020) imply that β_r can be interpreted as a sensible weighted average of causal effects under parallel trends only if treatment effects are homogeneous across adoption cohorts. For simplicity, we focus on the robustness of the results to violations of parallel trends using the original specification in LW, which is valid under the assumption of homogeneous treatment effects. As discussed in Remark 1, our sensitivity analysis can also be applied to estimators that are robust to treatment effect heterogeneity.

[29]LW normalize event time -1 to 0, but discuss how cohorts at event time -1 may have been partially treated, since LW impute the year that a student starts school with error. Since our robust confidence sets assume that there is no causal effect in the pre-period ($\tau_{pre} = 0$), we instead treat event-time -2 as the reference period in our analysis.

Figure 5: Event-study coefficient $\{\beta_r\}$ for employment, estimated using the event-study specification in (18).
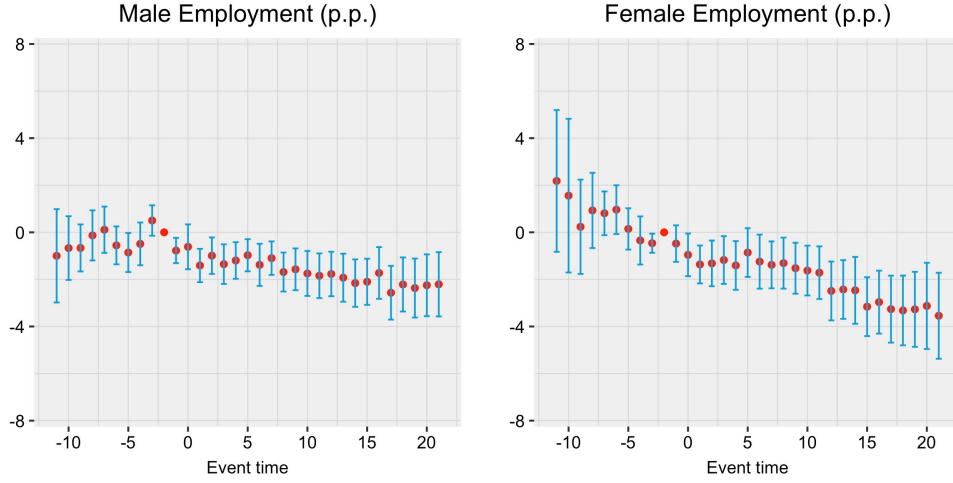

Figure 6 reports results of a sensitivity analysis for the treatment effect on employment for the cohort 15 years after the passage of a DTB law (as in Table 2 of LW), constructing robust confidence sets under varying assumptions on the class of possible violations of parallel trends. In blue, we plot the original OLS confidence intervals for $\hat{\beta}_{15}$ from specification (18). In red, we plot FLCIs when $\Delta = \Delta^{SD}(M)$ for different values of M; recall that $M = 0$ corresponds with allowing only for linear violations of parallel trends, and larger values of M allow for larger deviations from linearity. In the sensitivity analysis for men (left panel), the FLCIs are similar to those from OLS when allowing for violations of parallel trends that are approximately linear ($M \approx 0$), but become wider as we allow for more non-linearity; the

Figure 6: Sensitivity analysis for $\theta = \tau_{15}$ using $\Delta = \Delta^{SD}(M)$
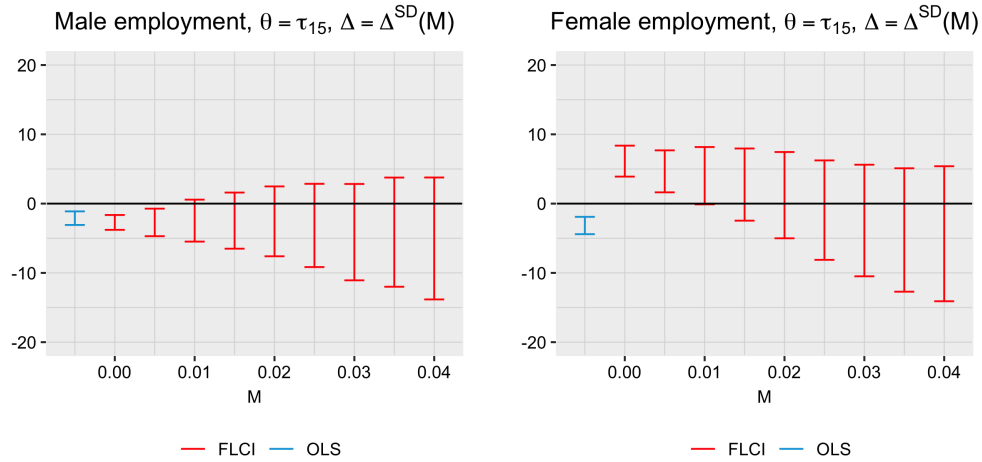
breakdown value for a significant effect is $M \approx 0.01$. For women (right panel), the original OLS estimates are negative and the confidence interval rules out 0. When we allow for linear violations of parallel trends ($M = 0$), however, the picture changes substantially owing to the pre-existing downward trend that is visible in Figure 5. Indeed, for $M < 0.01$ the robust confidence set contains only positive values. Intuitively, this is because the point estimate for $t = 15$ lies above a linear extrapolation of the negative pre-trend. Thus, if we were to impose the same smoothness restrictions for men as for women, we would either have to reconcile significant effects of opposite signs by gender (if $M < 0.01$) or we would not be able to rule out null effects for both genders ($M \geqslant 0.1$).

Sensitivity analyses using $\Delta^{RM}(\bar{M})$ or $\Delta^{SDMB}(\bar{M})$ do not allow us to obtain informative inference unless we are willing to impose that the post-treatment violations of parallel trends (or changes in their slope) are substantially smaller than those in the pre-treatment periods, with breakdown values of $\bar{M}$ of 0.1 or less for both genders and both Δs (see Appendix Figure D.1). We therefore consider a calibration exercise based on the magnitudes of possible possible confounds: if violations of parallel trends were driven by confounding changes in education quality, what would a given value of M imply about the evolution of those confounds? Chetty, Friedman and Rockoff (2014) estimate that a 1 standard deviation increase in teacher value-added (VA) corresponds with a 0.4 percentage point increase in adult employment. Hence, a value of $M = 0.01$ would correspond with allowing the slope of the differential trend to change by the equivalent of a one-fourtieth of a standard deviation of teacher VA across consecutive periods. Since the robust confidence sets for both men and women begin to include zero around this value of M, the strength with which we can rule out a null effect depends on our assessment of the economic plausibility of such non-linearities.

7   Conclusion

This paper considers the problem of conducting inference in difference-in-differences and related designs that is robust to violations of the parallel trends assumption. We introduce a variety of restrictions on the class of possible differences in trends that formalize commonly made arguments in empirical work. We provide inference procedures that are uniformly valid so long as the difference in trends satisfies these restrictions, and derive novel results on the power of these procedures. We recommend that applied researchers use our methods to conduct formal sensitivity analyses, in which they report confidence sets for the causal effect of interest under a variety of possible restrictions on the underlying trends. Such sensitivity analyses make transparent what assumptions are needed in order to obtain informative inference and help researchers assess whether those assumptions are plausible in a given

setting.

References

Abadie, Alberto, "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 2005, *72* (1), 1–19.

__ **and Guido W. Imbens**, "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 2006, *74* (1), 235–267.

__ , **Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge**, "Sampling-Based versus Design-Based Uncertainty in Regression Analysis," *Econometrica*, 2020, *88* (1), 265–296. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA12675.

Andrews, Donald W. K. and Panle Jia Barwick, "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 2012, *80* (6), 2805–2826.

Andrews, Isaiah, Jonathan Roth, and Ariel Pakes, "Inference for Linear Conditional Moment Inequalities," Technical Report, NBER Working Paper No. 26374 2019.

Armstrong, Timothy and Michal Kolesar, "Optimal Inference in a Class of Regression Models," *Econometrica*, 2018, *86*, 655–683.

__ **and** __ , "Simple and Honest Confidence Intervals in Nonparametric Regression," Technical Report 2018.

__ **and** __ , "Sensitivity Analysis using Approximate Moment Condition Models," *Quantitative Economics*, 2020. Forthcoming.

Ashenfelter, Orley, "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, 1978, *60* (1), 47–57. Publisher: The MIT Press.

Athey, Susan and Guido Imbens, "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption," *arXiv:1808.05293 [cs, econ, math, stat]*, August 2018.

Benzarti, Youssef and Dorian Carloni, "Who Really Benefits from Consumption Tax Cuts? Evidence from a Large VAT Reform in France," *American Economic Journal: Economic Policy*, February 2019, *11* (1), 38–63.

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, and Magne Mogstad, "Broadband Internet: An Information Superhighway to Sex Crime?," *The Review of Economics and Statistics*, 2013, *80*, 1237—-1266.

Bilinski, Alyssa and Laura A. Hatfield, "Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions," *arXiv:1805.03273 [stat.ME]*, 2020.

Borusyak, Kirill and Xavier Jaravel, "Revisiting Event Study Designs," SSRN Scholarly Paper ID 2826228, Social Science Research Network, Rochester, NY August 2016.

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi, "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 2017, *8* (1), 1–38.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *arXiv:1803.09015 [econ.EM]*, 2020.

Canay, Ivan A and Azeem M Shaikh, "Practical and theoretical advances in inference for partially identified models," *Advances in Economics and Econometrics*, 2017, *2*, 271–306.

Chen, Xiaohong, Timothy M. Christensen, and Elie Tamer, "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 2018, *86* (6), 1965–2018.

Chernozhukov, Victor, Whitney K. Newey, and Andres Santos, "Constrained Conditional Moment Restriction Models," *arXiv:1509.06311 [math, stat]*, September 2015. arXiv: 1509.06311.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, September 2014, *104* (9), 2633–2679.

Cho, JoonHwan and Thomas M. Russell, "Simple Inference on Functionals of Set-Identified Parameters Defined by Linear Moments," *arXiv:1810.03180 [econ.EM]*, 2019.

Cox, Gregory and Xiaoxia Shi, "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," *arXiv:1907.06317 [econ, math, stat]*, August 2020.

de Chaisemartin, Clément and Xavier D'Haultfoeuille, "Difference-in-Differences Estimators of Intertemporal Treatment Effects," *arXiv:2007.04267 [econ]*, May 2021. arXiv: 2007.04267.

___ **and Xavier D'Haultfœuille**, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, September 2020, *110* (9), 2964–2996.

Dette, Holger and Martin Schumann, "Difference-in-Differences Estimation Under Non-Parallel Trends," *Working paper*, 2020.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo, "The Economic Consequences of Hospital Admissions," *American Economic Review*, 2018, *108* (2), 308–352.

Donoho, David L., "Statistical Estimation and Optimal Recovery," *The Annals of Statistics*, 1994, *22* (1), 238–270.

Flynn, Zach, "Inference Based on Continuous Linear Inequalities via Semi-Infinite Programming," SSRN Scholarly Paper ID 3390788, Social Science Research Network, Rochester, NY May 2019.

Frandsen, Brigham R, "The Effects of Collective Bargaining Rights on Public Employee Compensation: Evidence from Teachers, Firefighters, and Police," *ILR Review*, 2016, *69* (1), 84–112.

Freyaldenhoven, Simon, Christian Hansen, and Jesse Shapiro, "Pre-event Trends in the Panel Event-study Design," *American Economic Review*, 2019, *109* (9), 3307–3338.

Gafarov, Bulat, "Inference in high-dimensional set-identified affine models," *arXiv:1904.00111 [econ.EM]*, 2019.

Goodman-Bacon, Andrew, "Public Insurance and Mortality: Evidence from Medicaid Implementation," *Journal of Public Economics*, 2018, *126* (1), 216–262.

___ , "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, June 2021.

Greenstone, Michael and Rema Hanna, "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India," *American Economic Review*, October 2014, *104* (10), 3038–3072.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 1998, *66* (5), 1017–1098.

Horowitz, Joel L. and Charles F. Manski, "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 1995, *63* (2), 281–302.

Hudson, Sally, Peter Hull, and Jack Liebersohn, "Interpreting Instrumented Difference-in-Differences," Working Paper 2017.

Kadane, Joseph B., "Comparison of k-Class Estimators When the Disturbances Are Small," *Econometrica*, 1971, *39* (5), 723–737.

Kahn-Lang, Ariella and Kevin Lang, "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications," *Journal of Business and Economic Statistics*, 2020, *38* (3), 613–620.

Kaido, Hiroaki and Andres Santos, "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities," *Econometrica*, 2014, *82* (1), 387–413.

_ , **Francesca Molinari, and Jörg Stoye**, "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 2019, *87*, 1397–1432.

_ , _ , **and** _ , "Constraint Qualifications in Partial Identification," *arXiv:1908.09103 [econ.EM]*, 2020.

Keele, Luke J., Dylan S. Small, Jesse Y. Hsu, and Colin B. Fogarty, "Patterns of Effects and Sensitivity Analysis for Differences-in-Differences," *arXiv:1901.01869 [stat.AP]*, 2019.

Kim, Wooyoung, Koohyun Kwon, Soonwoo Kwon, and Sokbae Lee, "The identification power of smoothness assumptions in models with counterfactual outcomes," *Quantitative Economics*, 2018, *9* (2), 617–642.

Kline, Patrick and Andres Santos, "Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure," *Quantitative Economics*, 2013, *4* (2), 231–267.

Kolesar, Michal and Christoph Rothe, "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 2018, *108* (8), 2277–2304.

Leavitt, Thomas, "Beyond Parallel Trends: Improvements on Estimation and Inference in the Difference-in-Differences Design," *Working paper*, 2020.

Lee, Jin Young and Gary Solon, "The Fragility of Estimated Effects of Unilateral Divorce Laws on Divorce Rates," *The B.E. Journal of Economic Analysis & Policy*, 2011, *11* (1).

Li, Ker-Chau, "Honest Confidence Regions for Nonparametric Regression," *The Annals of Statistics*, 1989, *17* (3), 1001–1008.

Lovenheim, Michael F. and Alexander Willen, "The Long-Run Effects of Teacher Collective Bargaining," *American Economic Journal: Economic Policy*, 2019, *11* (3), 292–324.

Low, Mark G., "Bias-Variance Tradeoffs in Functional Estimation Problems," *The Annals of Statistics*, 1995, *23* (3), 824–835.

Malani, Anup and Julian Reif, "Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform," *Journal of Public Economics*, April 2015, *124*, 1–17.

Manski, Charles F. and John V. Pepper, "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," *Review of Economics and Statistics*, 2018, *100* (2), 232–244.

Masten, Matthew A. and Alexandre Poirier, "Inference on Breakdown Frontiers," *Quantitative Economics*, 2020, *11* (1), 41–111.

Molinari, Francesca, "Chapter 5 - Microeconometrics with partial identification," in Steven N. Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa L. Matzkin, eds., *Handbook of Econometrics*, Vol. 7 of *Handbook of Econometrics, Volume 7A*, Elsevier, January 2020, pp. 355–486.

Moreira, Marcelo J. and Geert Ridder, "Efficiency Loss of Asymptotically Efficient Tests in an Instrumental Variables Regression," SSRN Scholarly Paper ID 3348716, Social Science Research Network, Rochester, NY March 2019.

Noack, Claudia and Christoph Rothe, "Bias-Aware Inference in Fuzzy Regression Discontinuity Designs," *arXiv:1906.04631 [econ.EM]*, 2020.

Rambachan, Ashesh and Jonathan Roth, "Design-Based Uncertainty for Quasi-Experiments," *arXiv:2008.00602 [econ, stat]*, August 2020.

Romano, Joseph P. and Azeem M. Shaikh, "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, September 2008, *138* (9), 2786–2807.

_ , _ , and Michael Wolf, "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 2014, *82* (5), 1979–2002.

Roth, Jonathan, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *Working paper*, 2019.

Sant'Anna, Pedro H. C. and Jun B. Zhao, "Doubly Robust Difference-in-Differences Estimators," *Journal of Econometrics*, 2020, *219* (1), 101–122.

Schrijver, Alexander, *Theory of Linear and Integer Programming*, Wiley-Interscience, 1986.

Sun, Liyan and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2020. Forthcoming.

Wolfers, Justin, "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results," *American Economic Review*, 2006, *96*, 1802–1820.

Ye, Ting, Luke Keele, Raiden Hasegawa, and Dylan S. Small, "A Negative Correlation Strategy for Bracketing in Difference-in-Differences with Application to the Effect of Voter Identification Laws on Voter Turnout," *arXiv:2006.02423 [stat.ME]*, 2020.

An Honest Approach to Parallel Trends

Online Appendix

Ashesh Rambachan   Jonathan Roth

July 18, 2021

This online appendix contains proofs and additional results for the paper "An Honest Approach to Parallel Trends" by Ashesh Rambachan and Jonathan Roth. Section A collects together additional results that are referenced in the main text. Section B discusses a hybrid approach using FLCIs and the conditional test. Section C contains proofs and auxiliary lemmas for the results in the main text. The supplementary materials provide statements and proofs of uniform asymptotic results along with additional simulation results.

A   Additional Results

A.1   Optimal local asymptotic power of FLCIs

As discussed in Remark 9, the FLCIs have local asymptotic power converging to the power envelope provided that Assumptions 3-4 are satisfied. We now formally state this result; the proof is given in Section C.

Proposition A.1. *Fix $\delta_A \in \Delta, \tau_A \in \mathbb{R}^{\bar{T}}$ and suppose Σ^* is positive definite. Let $\theta_A^{ub} = \sup_\theta \mathcal{S}(\Delta, \delta_A + \tau_A)$ be the upper bound of the identified set. Suppose that Assumption 4 holds and $\delta_{A,pre}$ satisfies Assumption 3. Then, for any $x > 0$ and $\alpha \in (0, 0.5]$,*

$$\lim_{n \to \infty} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}^{FLCI} \right) = \lim_{n \to \infty} \sup_{\mathcal{C}_{\alpha,n} \in \mathcal{I}_\alpha(\Delta, \Sigma_n)} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n} \right).$$

The analogous result holds replacing $\theta_A^{ub} + \frac{1}{\sqrt{n}}x$ with $\theta_A^{lb} - \frac{1}{\sqrt{n}}x$, for θ_A^{lb} the lower bound of the identified set.

Thus, $\mathcal{C}_{\alpha,n}^{FLCI}$ behaves similarly to $\mathcal{C}_{\alpha,n}^{C}$ as $n \to \infty$ when both Assumptions 3-4 hold.

A.2   Connection to linear independence constraint qualification (LICQ)

We now draw connections between linear independence constraint qualification (LICQ) and Assumption 4, under which the power of the conditional test converges to the power envelope

asymptotically. We show that LICQ implies Assumption 4. Our discussion follows the notation of Kaido et al. (2020).

Suppose $\Delta = \{\delta \,:\, A\delta \leqslant d\}$. Let $m(\tau_{post}; \beta) = A(\beta - M_{post}\tau_{post}) - d$, and let $\mathrm{T}(\Delta, \beta) := \{\tau_{post} \,:\, m(\tau_{post}; \beta) \leqslant 0\}$ be the identified set for the full parameter vector τ_{post}. Define the set of support points in direction p to be $S(p, \mathrm{T}) := \{\tau_{post} \,:\, p'\tau_{post} = \sup_{\tau_{post} \in \mathrm{T}} p'\tau_{post}\}$.

Definition 2. The **linear constraint qualification** (LICQ) is satisfied in the direction p if, for all support points in the direction p, the gradients of the binding constraints are linearly independent. That is, for all $\tau_{post} \in S(p, \mathrm{T})$, the set $\{D_{\tau_{post}} m_j(\tau_{post}, \beta) \,:\, m_j(\tau_{post}, \beta) = 0\}$ is linearly independent, where $D_{\tau_{post}}$ denotes the gradient with respect to τ_{post}.

Our next result shows that LICQ in the directions l and $-l$ is equivalent to a slightly stronger version of Assumption 4.

Lemma A.1. *Suppose $\beta_A = \delta_A + M_{post}\tau_{A,post}$ for some $\delta_A \in \Delta = \{\delta \,:\, A\delta \leqslant d\}$ and $\tau_{A,post} \in \mathbb{R}^{\bar{T}}$. Then the following are equivalent: (i) LICQ is satisfied in the direction l; (ii) For any solution δ^{**} to the linear program*

$$b^{min}(\delta_{A,pre}) = \min_{\delta} l'\delta_{post} \ s.t. \ A\delta \leqslant d, \delta_{pre} = \delta_{A,pre},$$

the matrix $A_{(B(\delta^{**}),post)}$ with rows corresponding with the binding inequality constraints at δ^{**} has rank $|B(\delta^{**})|$. Analogous results hold replacing l with $-l$ in (i) and $\min$ with $\max$ in (ii).

Proof. We first show (i) implies (ii). Let δ^{**} be a solution to the minimization problem for b^{min}. Let $\tau_{post}^{**} = \beta_{A,post} - \delta_{post}^{**}$. Observe that $l'\tau_{post}^{**} = l'\beta_{A,post} - b^{min}(\delta_{A,post})$. From (7), we then see that $l'\tau_{post}^{**} = \theta^{ub}$ and hence $\tau_{post}^{**} \in S(l, \mathrm{T})$. Now, note that by construction, $m(\beta_A, \tau_{post}^{**}) = A(\beta_A - M_{post}\tau_{post}^{**} - d) = A\delta^{**} - d$, so the binding constraints in $m(\beta_A, \tau_{post}^{**})$ correspond with the binding constraints in the minimization for b^{min}. Finally, observe that $D_{\tau_{post}} m(\beta_A, \tau_{post}^{**}) = A_{(\cdot, post)}$. It then follows from (i) that the rows of $A_{(B(\delta^{**}),post)}$ are linearly independent, which gives the desired result.

Conversely, suppose $\tau_{post}^{**} \in S(l, \tau)$. By definition, there exists some $\delta^{**} \in \Delta$ such that $\delta^{**} = \beta_A - M_{post}\tau_{post}^{**}$, and $l'\tau_{post}^{**} = \theta^{ub}$. Thus, $\theta^{ub} = l'\beta_{A,post} - l'\delta^{**}$. It then follows from (7) that $l'\delta^{**} = b^{min}(\delta_{A,pre})$, so δ^{**} is a solution to the optimization b^{min}. (ii) then implies that $A_{(B(\delta^{**},post)}$ has linearly independent rows. By the same argument as earlier in the proof, $A_{(B(\delta^{**},post)}$ corresponds with the matrix of gradients for the binding constraints in $m(\beta_A, \tau_{post}^{**})$, from which we see that LICQ is satisfied. $\square$

Therefore, if LICQ holds in the directions l and $-l$, then Assumption 4 is satisfied. Indeed, if LICQ holds, then Lemma 4 implies that the rank condition in Assumption 4 holds

for *any* solutions δ^* and δ^{**} to the problems b^{max} and b^{min}. By contrast, Assumption 4 only requires the rank condition to hold for *at least one* solution to b^{max} and b^{min}. It is possible for a linear program to have multiple solutions, and for the rows of the binding constraints to be linearly independent (non-degenerate) for some solutions but not for others (e.g., see Example 1 on p. 146 of Sierksma (2001)). Assumption 4 is thus potentially weaker than LICQ if there are multiple solutions to the optimizations for b^{max} or b^{min}, but equivalent when the solutions are unique.

B Conditional-FLCI Hybrid Confidence Sets

As discussed in Remark 15, ARP recommend a Bonferroni-type hybrid approach that uses tests based on least-favorable critical values in combination with the conditional test. In this section, we show that a similar hybrid approach can be applied using FLCIs in place of the least-favorable tests.

The conditional-FLCI hybrid confidence set is constructed by first testing whether a candidate parameter value lies within the level-$(1 - \kappa)$ optimal FLCI, and then applying a conditional test to all parameter values that lie within the optimal FLCI. The second stage uses a modified version of the conditional test that i) adjusts size to account for the first-stage test, and ii) conditions on the event that the first-stage test fails to reject.

Formally, suppose that $0 < \kappa < \alpha$.[30] Consider the level $(1 - \kappa)$ optimal FLCI, $\mathcal{C}_{\kappa,n}^{FLCI} = a_n + v_n' \hat{\beta}_n \pm \chi_n$. Lemma B.2 below shows that the distribution of the test statistic $\hat{\eta}$ defined in (15) follows a truncated normal distribution conditional on the parameter value $\bar{\theta}$ falling within the level $(1 - \kappa)$ optimal FLCI. With this result, the construction of the second-stage of the conditional-FLCI hybrid test is analogous to the construction of the conditional test, except it uses the modified size $\tilde{\alpha} = \frac{\alpha - \kappa}{1 - \kappa}$ to account for the first-stage test. The conditional-FLCI hybrid test $\psi_{\kappa,\alpha}^{C\text{-}FLCI}(\hat{\beta}_n, \bar{\theta}, \tilde{\Sigma}_n)$ rejects if and only if either $\bar{\theta} \notin \mathcal{C}_{\kappa,n}^{FLCI}$ or $F_{\xi \,|\, \xi \in [v_{C\text{-}FLCI}^{lo}, v_{C\text{-}FLCI}^{up}]}(\hat{\eta}) > 1 - \tilde{\alpha}$, where $F_{\xi \,|\, \xi \in [v_{C\text{-}FLCI}^{lo}, v_{C\text{-}FLCI}^{up}]}(\cdot)$ denotes the CDF of the truncated normal distribution derived in Lemma B.2 below.

Since the FLCI controls size, the first stage test rejects with probability at most κ under the null that $\theta = \bar{\theta}$. The second-stage test rejects with probability at most $\tilde{\alpha} = \frac{\alpha - \kappa}{1 - \kappa}$ conditional on $\theta \in \mathcal{C}_{\kappa,n}^{FLCI}$. Together, this implies that the conditional-FLCI hybrid test controls size,

$$\sup_{\delta \in \Delta, \tau} \sup_{\bar{\theta} \in \mathcal{S}(\Delta, \delta + \tau)} \mathbb{E}_{(\delta, \tau, \Sigma_n)} \left[\psi_{\kappa,\alpha}^{C\text{-}FLCI}(\hat{\beta}_n, \bar{\theta}, \tilde{\Sigma}_n) \right] \leqslant \alpha. \tag{19}$$

---

[30]In practice, we set $\kappa = \alpha/10$ following Romano et al. (2014) and ARP, although the optimal choice of κ is an interesting question for future research.

We therefore construct a conditional-FLCI hybrid confidence set for the parameter θ that satisfies (10) by inverting the conditional-FLCI test, denoting this confidence set as $\mathcal{C}_{\kappa,\alpha,n}^{C\text{-}FLCI} :=$ $\{\theta : \psi_{\kappa,\alpha}^{C\text{-}FLCI}(\hat{\beta}_n, \bar{\theta}, \tilde{\Sigma}_n) = 0\}$.

The following two results show that the conditional-FLCI hybrid confidence set inherits some desirable asymptotic properties from the conditional approach: it is asymptotically consistent, and under the same conditions as Proposition 4.2, the conditional-FLCI hybrid test has local asymptotic power at least as good as the optimal $\frac{\alpha-\kappa}{1-\kappa}$ test. (The proofs of these results are provided in Section C.)

Proposition B.1 (Consistency)**.** *The conditional-FLCI hybrid test is consistent. For any* $\delta_A \in \Delta, \tau_A \in \mathbb{R}^{\bar{T}}, \theta^{out} \notin \mathcal{S}(\Delta, \delta_A + \tau_A), \alpha \in (0, 0.5],$ *and* $\kappa \in (0, \alpha),$

$$\lim_{n \to \infty} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left(\theta^{out} \notin \mathcal{C}_{\kappa,\alpha,n}^{C\text{-}FLCI} \right) = 1.$$

Proposition B.2 (Local asymptotic power)**.** *Fix* $\delta_A \in \Delta, \tau_A,$ *and* Σ^* *positive definite. Suppose Assumption 4 holds. Suppose* $\alpha \in (0, 0.5], \kappa \in (0, \alpha),$ *and let* $\tilde{\alpha} = \frac{\alpha-\kappa}{1-\kappa}.$ *Then,*

$$\liminf_{n \to \infty} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\kappa,\alpha,n}^{C\text{-}FLCI} \right) \geqslant \lim_{n \to \infty} \sup_{\mathcal{C}_{\alpha,n} \in \mathcal{I}_{\tilde{\alpha}}(\Delta, \Sigma_n)} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n} \right).$$

The analogous result holds replacing $\theta_A^{ub} + \frac{1}{\sqrt{n}}x$ *with* $\theta_A^{lb} - \frac{1}{\sqrt{n}}x,$ *for* θ_A^{lb} *the lower bound of the identified set (although the constant* c^* *may differ).*

ARP show that a hybrid that uses least-favorable critical values in the first stage always rejects when the size $\tilde{\alpha}$ conditional test rejects. It is thus immediate that analogs of the previous two propositions hold for the LF hybrid as well.

B.1 Auxiliary Lemmas for the Conditional-FLCI Hybrid Confidence Sets

We now derive the truncated normal distribution used to construct the conditional-FLCI hybrid confidence sets. We first provide a lemma that implies that the affine estimator at which the optimal FLCI is centered can be written as an affine function of $A\hat{\beta}$, where recall that A is the matrix defining the polyhedral set, $\Delta = \{\delta : A\delta \leqslant d\}$.

Lemma B.1. *Suppose* $\Delta = \{\delta : A\delta \leqslant d\} \neq \varnothing,$ *and* (a, v) *are such that* $\bar{b}(a, v) < \infty.$ *Then, there exists* $\tilde{v}$ *such that* $v' = \tilde{v}'A.$

Proof. Note that

$$\bar{b}(a,v) = \max_{\delta,\tau_{post}} |v'(\delta + M_{post}\tau_{post}) - l'\tau_{post}| \text{ s.t. } A\delta - d \leqslant 0.$$

We will show that if $\bar{b}(a,v)$ is finite, then for all $\tilde{\delta} \in \Delta$, $A\tilde{\delta} = 0$ implies $v'\tilde{\delta} = 0$. This implies that v is in the rowspace of A, from which the result follows. To prove this, suppose towards contradiction that $\tilde{\delta}$ is such that $A\tilde{\delta} = 0$ and $v'\tilde{\delta} \neq 0$. Since $A\tilde{\delta} = 0$, it follows that $\delta_c := (\delta_0 + c \cdot \tilde{\delta})$ is in Δ for any $\delta_0 \in \Delta$ and $c \in \mathbb{R}$. However, it then follows that for any fixed τ_{post} and δ_0, the objective in the previous display can be made arbitrarily large at (δ_c, τ_{post}) by taking $c \to \infty$. $\qquad\square$

Consider the level $1 - \kappa$ optimal FLCI, $\mathcal{C}_{\kappa,n}^{FLCI} = a_n + v_n'\hat{\beta}_n \pm \chi_n$. By Lemma B.1, there exists some vector $\tilde{v}_n$ such that the level $1 - \kappa$ optimal FLCI can be written as $a_n + \tilde{v}_n'A\hat{\beta}_n \pm \chi_n$. Since $\tilde{Y}_n(\theta) = A\hat{\beta}_n - \tilde{A}_{(\cdot,1)}\theta - d$, it follows that $\bar{\theta} \in \mathcal{C}_{\kappa,n}^{FLCI}$ if and only if

$$\tilde{v}_n'\tilde{Y}_n(\bar{\theta}) \leqslant \chi_n - a_n - \tilde{v}_n'd + (1 - \tilde{v}_n'\tilde{A}_{(\cdot,1)})\bar{\theta}),$$
$$-\tilde{v}_n'\tilde{Y}_n(\bar{\theta}) \leqslant \chi_n + a_n + \tilde{v}_n'd - (1 - \tilde{v}_n'\tilde{A}_{(\cdot,1)})\bar{\theta}).$$

One can further show that $(1 - \tilde{v}_n'\tilde{A}_{(\cdot,1)})\bar{\theta}) = 0$, which simplifies the upper bounds above.[31] Defining the matrix $\tilde{V}_n = (\tilde{v}_n', -\tilde{v}_n')'$ and the vector $d_n(\bar{\theta})$ which stacks the upper-bounds of the inequalities in the previous display, we see that the optimal level $1 - \kappa$ FLCI contains the parameter value $\bar{\theta}$ if and only if $\tilde{V}_n\tilde{Y}_n(\bar{\theta}) \leqslant \bar{d}_n(\bar{\theta})$.

With this equivalent representation of the optimal FLCI, we can now characterize the distribution of the test statistic $\hat{\eta}$ (15) conditional on the parameter θ falling within the optimal FLCI.

Lemma B.2. $\hat{\eta} \mid \left\{\gamma_* \in \hat{V}_n, S_n = s, \bar{\theta} \in \mathcal{C}_{\kappa,n}^{FLCI}\right\} \sim \xi \mid \xi \in [v_{C\text{-}FLCI}^{lo}, v_{C\text{-}FLCI}^{up}], \text{ where } \xi \sim \mathcal{N}\left(\gamma_*'\tilde{\mu}, \gamma_*'\tilde{\Sigma}_n\gamma_*\right),$ $v_{C\text{-}FLCI}^{lo} := \max\{v^{lo}, v_{FLCI}^{lo}\}, v_{C\text{-}FLCI}^{up} := \min\{v^{up}, v_{FLCI}^{up}\}; v^{lo} \text{ and } v^{up} \text{ are as defined in Section 4.2}, v_{FLCI}^{lo} := \max_{\{j:(\tilde{V}_n c_{n,\gamma*})_j < 0\}} \frac{\bar{d}_n(\theta)_j - (\tilde{V}_n S_n)_j}{(\tilde{V}_n c_{n,\gamma*})_j}, v_{FLCI}^{up} := \min_{\{j:(\tilde{V}_n c_{n,\gamma*})_j > 0\}} \frac{\bar{d}_n(\theta)_j - (\tilde{V}_n S_n)_j}{(\tilde{V}_n c_{n,\gamma*})_j},$ $c_{n,\gamma*} = \frac{\tilde{\Sigma}_n\gamma_*}{\gamma_*'\tilde{\Sigma}_n\gamma_*} \text{ and } S_n = (I - \frac{\tilde{\Sigma}_n\gamma_*}{\gamma_*'\tilde{\Sigma}_n\gamma_*}\gamma_*')\tilde{Y}_n(\bar{\theta}), \text{ and } \gamma_* \text{ is the vector of Lagrange multipliers for the primal problem (16).}$

Proof. The proof follows an analogous argument to Lemma 9 in ARP. Recall that conditional on $\gamma_* \in \hat{V}_n$, $\hat{\eta} = \gamma_*'\tilde{Y}_n$. Recall also that $\theta \in \mathcal{C}_{\kappa,n}^{FLCI}$ if and only if $\tilde{V}_n\tilde{Y}_n \leqslant \bar{d}_n$. Observe that the set of values of $\tilde{Y}_n$ such that

$$\gamma_*'\tilde{Y}_n = \left(\max \gamma'\tilde{Y}_n \text{ s.t. } \gamma \geqslant 0, \gamma'\tilde{A}_{(\cdot,-1)} = 0, \gamma'\tilde{\sigma}_n = 1\right) \text{ and } \tilde{V}_n\tilde{Y}_n \leqslant \bar{d}_n$$

---

[31] Applying the definitions of $\tilde{A}$ and $\tilde{v}_n$, we obtain that $\tilde{v}_n'\tilde{A}_{(\cdot,1)} = v_{n,post}'\Gamma^{-1}e_1$. However, we show in the proof to Lemma C.19 that $v_{n,post} = l$, so $\tilde{v}_n'\tilde{A}_{(\cdot,1)} = l'\Gamma^{-1}e_1$. The result then follows from the fact that $e_1'\Gamma = l'$ by construction.

is convex. This follows from the fact that if the expression above holds for both $\tilde{Y}_n$ and $\tilde{Y}_n^*$, then $\gamma_*'\tilde{Y}_n \geqslant \gamma'\tilde{Y}_n$ and $\gamma_*'\tilde{Y}_n^* \geqslant \gamma'\tilde{Y}_n^*$ for all γ feasible in the maximization. It then follows that $\gamma_*'(\alpha\tilde{Y}_n^* + (1-\alpha)\tilde{Y}_n) \geqslant \gamma'(\alpha\tilde{Y}_n^* + (1-\alpha)\tilde{Y}_n)$ for any $\alpha \in (0,1)$. Thus, $(\alpha\tilde{Y}_n^* + (1-\alpha)\tilde{Y}_n)$ is also equal to the maximum. It is likewise clear that the second constraint holds for a convex combination of $\tilde{Y}_n$ and $\tilde{Y}_n^*$.

Thus, once we condition on S_n, the set of values of $\gamma_*'\tilde{Y}_n$ such that $\gamma_* \in \hat{V}_n$ and $\tilde{V}_n\tilde{Y}_n \leqslant \bar{d}_n$ is an interval. It thus suffices to find the endpoints. Without loss of generality, we focus on the lower bound. For ease of notation, let $F := \{\gamma \geqslant 0, \gamma'\tilde{A}_{(\cdot,-1)} = 0, \gamma'\tilde{\sigma}_n = 1\}$ denote the feasible region for the maximization. Then we are interested in

$$\min_{\{\tilde{Y}_n : S_n = s\}} \left\{ \gamma_*'\tilde{Y}_n \,:\, \gamma_*'\tilde{Y}_n = \max_{\gamma \in F} \gamma'\tilde{Y}_n \,, \tilde{V}_n\tilde{Y}_n \leqslant \bar{d}_n \right\}.$$

Recalling that $S_n = (I - c_{n,\gamma_*}\gamma_*')\tilde{Y}_n$ for $c_{n,\gamma_*} = \frac{\tilde{\Sigma}_n\gamma_*}{\gamma_*'\tilde{\Sigma}_n\gamma_*}$, the expression becomes

$$\min_{\{\tilde{Y}_n : S_n = s\}} \left\{ \gamma_*'\tilde{Y}_n \,:\, \gamma_*'\tilde{Y}_n = \max_{\gamma \in F} \gamma'\left(s + c_{n,\gamma_*}\gamma_*'\tilde{Y}_n\right) \,, \tilde{V}_n\tilde{Y}_n \leqslant \bar{d}_n \right\},$$

which is equivalent to

$$\min\left\{ \left\{ x \,:\, x = \max_{\gamma \in F} \gamma'\left(s + c_{n,\gamma_*}x\right) \right\} \cap \left\{ \gamma_*'\tilde{Y}_n \,:\, \tilde{Y}_n \text{ s.t. } S_n = s, \tilde{V}_n\tilde{Y}_n \leqslant \bar{d}_n \right\} \right\}.$$

However, the first set in the minimization above is the interval $[v^{lo}, v^{up}]$, and the polyhedral lemma in Lee, Sun, Sun and Taylor (2016) (Lemma 5.1) implies that second set is the interval $[v_{FLCI}^{lo}, v_{FLCI}^{up}]$. Thus, the expression above is $\max\{v^{lo}, v_{FLCI}^{lo}\}$, as desired. The argument for the lower bound of the interval is analogous. Finally, the independence of $\gamma_*'\tilde{Y}_n$ and S_n implies that the distribution of $\gamma_*'\tilde{Y}_n$ conditional on $\left\{ \gamma_* \in \hat{V}_n, S_n = s, \bar{\theta} \in \mathcal{C}_{n,\kappa}^{FLCI} \right\}$ is truncated normal. $\qquad\square$

C   Proofs of Finite Sample Normal Results

C.1   Proofs of Main Finite Sample Normal Results

Throughout the proofs, we will use the following notation. Let $\tilde{Y}(\hat{\beta}_n, A, d, \bar{\theta}) := A\hat{\beta}_n - d - \tilde{A}_{(\cdot,1)}\bar{\theta}$. Define $\psi_\alpha^C(\hat{\beta}_n; A, d, \bar{\theta}, \Sigma) := \psi_\alpha^C(\tilde{Y}(\hat{\beta}_n; A, d, \bar{\theta}), A\Sigma A')$ to be an indicator for whether the conditional test constructed using $(\hat{\beta}_n, A, d, \bar{\theta}, \Sigma)$ rejects. For a matrix A, we use $A_{(B,\cdot)}$ to denote the sub-matrix of A defined by the index B.

Proof of Lemma 2.2

Proof. By equation (8), we can write the coverage requirement as

$$\inf_{\delta \in \Delta, \tau} \ \inf_k \ \inf_{\theta \in \mathcal{S}(\Delta_k, \delta + \tau)} \mathbb{P}_{(\delta, \tau, \Sigma_n)} \left(\theta \in \bigcup_{k'} \mathcal{C}_{n, k'} \right),$$

which is bounded below by

$$\inf_{\delta \in \Delta, \tau} \ \inf_k \ \inf_{\theta \in \mathcal{S}(\Delta_k, \delta + \tau)} \mathbb{P}_{(\delta, \tau, \Sigma_n)} \left(\theta \in \mathcal{C}_{n, k} \right),$$

which is at least $1 - \alpha$ since $\mathcal{C}_{n,k}$ satisfies (10) for $\Delta = \Delta_k$. $\qquad\square$

Proof of Proposition 3.2

Proof. First, suppose Assumption 3 holds. Without loss of generality, we show $\mathbb{P}\left((\theta^{ub} + x) \in \mathcal{C}_{\alpha, n}^{FLCI} \right) \to 0$ for any $x > 0$. By Lemma C.20 there exists $(\bar{a}, \bar{v})$ such that $\bar{b}(\bar{a}, \bar{v}) = \frac{1}{2} LID(\Delta, \delta_{pre}) =: \bar{b}_{min}$ and $\mathbb{E}_{(\delta_A, \tau_A, \Sigma_n)} \left[\bar{a} + \bar{v}' \hat{\beta}_n \right] = \frac{1}{2} (\theta^{ub} + \theta^{lb}) =: \theta^{mid}$. Let $\bar{\mathcal{C}}_n := \bar{a} + \bar{v}' \hat{\beta}_n \pm \chi_n(\bar{a}, \bar{v})$ denote the fixed length confidence interval based on $(\bar{a}, \bar{v})$.

By construction, $\bar{\chi}_n := \chi_n(\bar{a}, \bar{v})$ is the $1 - \alpha$ quantile of the $| \mathcal{N} \left(\bar{b}_{min}, \sigma_{\bar{v}, n}^2 \right) |$ distribution. Since $\sigma_{\bar{v}, n}^2 = \frac{1}{n} \sigma_{\bar{v}, 1}^2 \to 0$, the $| \mathcal{N} \left(\bar{b}_{min}, \sigma_{\bar{v}, n}^2 \right) |$ distribution collapses to a point mass at $\bar{b}_{min}$, and thus $\bar{\chi}_n \to \bar{b}_{min}$. By construction the length of the shortest FLCI $\chi_n := \chi_n(a_n, v_n)$ must be less than or equal to $\bar{\chi}_n$, and so $\limsup_{n \to \infty} \chi_n \leqslant \bar{b}_{min}$. Let $b_n := \bar{b}(a_n, v_n)$ be the worst-case bias of the optimal FLCI. Since $\alpha \in (0, .5]$, Lemma C.21 implies that $\chi_n \geqslant b_n$. Additionally, Lemma C.19 implies that $b_n \geqslant \frac{1}{2} LID(\Delta, \delta_{pre}) = \bar{b}_{min}$, and thus $\chi_n \geqslant \bar{b}_{min}$. Hence, $\chi_n \to \bar{b}_{min}$ implies $b_n \to \bar{b}_{min}$. Additionally, note that for $\alpha \in (0, 0.5]$, $\chi_n(a, v)$ is increasing in both $\bar{b}(a, v)$ and $\sigma_{v, n}$. Since $\bar{b}_{min} \leqslant b_n$ and $\chi_n \leqslant \bar{\chi}_n$, it must be that $\sigma_{v_n, n} \leqslant \sigma_{\bar{v}, n}$, from which it follows that $\sigma_{v_n, n} \to 0$.

Now, we claim that $\mu_n := \mathbb{E}_{(\delta_A, \tau_A, \Sigma_n)} \left[a_n + v_n' \hat{\beta}_n \right]$ converges to $\theta^{mid} := \frac{1}{2} (\theta^{ub} + \theta^{lb})$. To show this, note that $\mu_n = a_n + v_n' \beta_A$ for $\beta_A = \delta_A + \tau_A$. Since $\theta^{ub}, \theta^{lb} \in \mathcal{S}(\Delta, \beta_A)$, by the definition of the identified set there exist $\delta^{ub}, \delta^{lb} \in \Delta$ and τ^{ub}, τ^{lb} such that $\beta_A = \delta^{ub} + \tau^{ub} = \delta^{lb} + \tau^{lb}$, $\theta^{ub} = l' \tau_{post}^{ub}$, and $\theta^{lb} = l' \tau_{post}^{lb}$. Thus, $\theta^{ub} - \mathbb{E}_{(\delta^{ub}, \tau^{ub}, \Sigma_n)} \left[a_n + v_n' \hat{\beta}_n \right] = \theta^{ub} - \mu_n$ and $\mathbb{E}_{(\delta^{lb}, \tau^{lb}, \Sigma_n)} \left[a_n + v_n' \hat{\beta}_n \right] - \theta^{lb} = \mu_n - \theta^{lb}$. This implies that $b_n \geqslant \max\{\theta^{ub} - \mu_n, \mu_n - \theta^{lb}\} = \bar{b}_{min} + |\mu_n - \theta^{mid}|$, where the equality uses the fact that $\theta^{ub} - \theta^{lb} = LID(\Delta, \delta_{A, pre}) = 2 \bar{b}_{min}$. Since we've shown that $b_n \to \bar{b}_{min}$, it follows that $\mu_n \to \theta^{mid}$, as desired.

Next, note that if $\hat{\beta}_n \sim \mathcal{N}(\delta_A + \tau_A, \Sigma_n)$, then $a_n + v_n' \hat{\beta}_n \sim \mathcal{N}\left(\mu_n, \sigma_{v_n, n}^2 \right)$. Observe that $\bar{\theta} \in \mathcal{C}_{\alpha, n}^{FLCI}$ if and only if $a_n + v_n' \hat{\beta}_n \in [\bar{\theta} - \chi_n, \bar{\theta} + \chi_n]$. Thus,

$$\mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left(\bar{\theta} \in \mathcal{C}_{\alpha,n}^{FLCI} \right) = \Phi \left(\frac{\bar{\theta} + \chi_n - \mu_n}{\sigma_{v_n,n}} \right) - \Phi \left(\frac{\bar{\theta} - \chi_n - \mu_n}{\sigma_{v_n,n}} \right).$$

Now, recalling that $\theta^{ub} = \theta^{mid} + \bar{b}_{min}$ by construction, we have $\mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left((\theta^{ub} + x) \in \mathcal{C}_{\alpha,n}^{FLCI} \right)$
equals

$$\Phi \left(\frac{\theta^{mid} + \bar{b}_{min} + x + \chi_n - \mu_n}{\sigma_{v_n,n}} \right) - \Phi \left(\frac{\theta^{mid} + \bar{b}_{min} + x - \chi_n - \mu_n}{\sigma_{v_n,n}} \right). \qquad (20)$$

Note that the term inside the second normal CDF in the previous display equals

$$-\frac{\chi_n - b_n}{\sigma_{v_n,n}} + \frac{x + \theta^{mid} - \mu_n + \bar{b}_{min} - b_n}{\sigma_{v_n,n}}.$$

However, the first summand above is bounded between $-z_{1-\alpha/2}$ and $-z_{1-\alpha}$ by Lemma C.21. Additionally, we've shown that $\theta^{mid} - \mu_n \to 0$ and $\bar{b}_{min} - b_n \to 0$, so the numerator of the second summand converges to $x > 0$. Since the denominator $\sigma_{v_n,n} \to 0$, the expression in the previous display diverges to ∞, and hence the second normal CDF term in (20) converges to 1, which implies that $\mathbb{P} \left((\theta^{ub} + x) \in \mathcal{C}_{\alpha,n}^{FLCI} \right) \to 0$, as needed.

Conversely, suppose Assumption 3 fails. Let $L_A := LID(\Delta, \delta_{A,pre})$ and $\bar{L} := \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\Delta, \tilde{\delta}_{pre})$. By Lemma C.19, $b_n := \bar{b}(a_n, v_n) \geqslant \frac{1}{2} \bar{L} =: \bar{b}_{min}$. As argued earlier in the proof, since $\alpha \in (0, .5]$, $\chi_n \geqslant b_n = \frac{1}{2} \bar{L}$. If $\bar{L} = \infty$, then $\mathcal{C}_{\alpha,n}^{FLCI}$ is the real line, and thus never rejects, so $\mathcal{C}_{\alpha,n}^{FLCI}$ is trivially inconsistent under the assumption that $\mathcal{S}_\theta(\Delta, \delta_A + \tau_A) \neq \mathbb{R}$. For the remainder of the proof, we assume $L_A < \bar{L} < \infty$. From Lemma 2.1, $\mathcal{S}(\delta_A + \tau_A, \Delta) = [\theta^{lb}, \theta^{ub}]$, where $\theta^{ub} - \theta^{lb} = LID(\Delta, \delta_{A,pre}) = L_A$. Let $\epsilon = \frac{1}{4}(\bar{L} - L_A)$, and set $\theta_1^{out} := \theta^{ub} + \epsilon$ and $\theta_2^{out} := \theta^{lb} - \epsilon$. Let $\theta^{mid} = \frac{1}{2}(\theta^{ub} + \theta^{lb})$ be the midpoint of the identified set. By construction, $\theta_1^{out} - \theta^{mid} = \theta^{mid} - \theta_2^{out} = \frac{1}{2} L_A + \epsilon < \frac{1}{2} \bar{L}$. Since $\mathcal{C}_{\alpha,n}^{FLCI}$ is an interval with half-length at least $\frac{1}{2} \bar{L}$, it follows that if $\theta^{mid} \in \mathcal{C}_{\alpha,n}^{FLCI}$ then at least one of $\theta_1^{out}, \theta_2^{out}$ is also in $\mathcal{C}_{\alpha,n}^{FLCI}$. Hence, $\mathbb{P} \left(\theta_1^{out} \in \mathcal{C}_{\alpha,n}^{FLCI} \right) + \mathbb{P} \left(\theta_2^{out} \in \mathcal{C}_{\alpha,n}^{FLCI} \right) \geqslant \mathbb{P} \left(\theta^{mid} \in \mathcal{C}_{\alpha,n}^{FLCI} \right) \geqslant 1 - \alpha$, where the final bound follows since $\mathcal{C}_{\alpha,n}^{FLCI}$ has correct coverage. It follows that $\limsup_{n \to \infty} \mathbb{P} \left(\theta_j^{out} \in \mathcal{C}_{\alpha,n}^{FLCI} \right) \geqslant \frac{1}{2}(1 - \alpha) > 0$ for at least one $j \in \{1, 2\}$. $\qquad \square$

Proof of Proposition 4.1

Proof. Lemma 2.1 showed that the identified set is an interval, $\mathcal{S}(\delta_A + \tau_A, \Delta) = [\theta^{lb}, \theta^{ub}]$, and so if $\theta^{out} \notin \mathcal{S}(\delta_A + \tau_A, \Delta)$, then we must have either $\theta^{out} = \theta^{ub} + x$ or $\theta^{out} = \theta^{lb} - x$ for some $x > 0$. Without loss of generality, consider the case $\theta^{out} = \theta^{ub} + x$, so

$$\liminf_{n \to \infty} \mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left(\theta^{out} \notin \mathcal{C}_{\alpha,n}^C \right) = \liminf_{n \to \infty} \mathbb{E}_{(\delta_A, \tau_A, \Sigma_n)} \left[\psi_\alpha^C(\hat{\beta}_n; A, d, \theta^{ub} + x, \Sigma_n) \right].$$

Lemma C.2 along with $\Sigma_n = \frac{1}{n}\Sigma^*$ imply $\psi_\alpha^C(\hat{\beta}_n; A, d, \theta^{ub}+x, \Sigma_n) = \psi_\alpha^C(\sqrt{n}\hat{\beta}_n; A, \sqrt{n}d, \sqrt{n}\theta^{ub}+\sqrt{n}x, \Sigma^*)$. Thus,

$$\liminf_{n\to\infty} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left(\theta^{out} \notin \mathcal{C}_{\alpha,n}^C\right) = \liminf_{n\to\infty} \mathbb{E}_{(\sqrt{n}\delta_A,\sqrt{n}\tau_A,\Sigma^*)}\left[\psi_\alpha^C(\hat{\beta}_n; A, \sqrt{n}d, \sqrt{n}\theta^{ub} + \sqrt{n}x, \Sigma^*)\right],$$

where we further used that $\hat{\beta}_n \sim \mathcal{N}(\beta_A, \Sigma_n)$ implies $\sqrt{n}\hat{\beta}_n \sim \mathcal{N}(\sqrt{n}\beta_A, \Sigma^*)$. Lemma C.1 implies that $\sqrt{n}\theta^{ub} = \theta_n^{ub}$, for $\theta_n^{ub} = \sup \mathcal{S}(\sqrt{n}\delta_A + \sqrt{n}\tau_A, \Delta_n)$ and $\Delta_n = \{\delta : A\delta \leqslant \sqrt{n}d\}$. It follows from Lemma C.18 that

$$\mathbb{E}_{(\sqrt{n}\delta_A,\sqrt{n}\tau_A,\Sigma^*)}\left[\psi_\alpha^C(\hat{\beta}_n; A, \sqrt{n}d, \sqrt{n}\theta^{ub} + \sqrt{n}x, \Sigma^*)\right] \geqslant \rho_{LB}(\sqrt{n}x, \Sigma^*),$$

for ρ_{LB} a function with $\lim_{\tilde{x}\to\infty} \rho_{LB}(\tilde{x}, \Sigma^*) = 1$. It is immediate from the previous two displays that $\lim_{n\to\infty} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left(\theta^{out} \notin \mathcal{C}_{\alpha,n}^C\right) = 1$ as desired.

$\square$

Proof of Proposition 4.2

Proof. We show that each of the limits in the proposition equals $\Phi(c^*x - z_{1-\alpha})$. Following the same steps as proof of Proposition 4.1, the first limit of interest can be written as

$$\lim_{n\to\infty} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}^C\right) = \lim_{n\to\infty} \mathbb{E}_{(\sqrt{n}\delta_A,\sqrt{n}\tau_A,\Sigma^*)}\left[\psi_\alpha^C(\hat{\beta}_n; A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*)\right].$$

The term on the right-hand side converges to $\Phi(c^*x - z_{1-\alpha})$ by Lemma C.8.

We next turn to the second limit. Consider testing $H_0 : \delta \in \Delta = \{\delta : A\delta \leqslant d\}, \theta = \bar{\theta}$ against $H_1 : (\delta, \tau) = (\delta_A, \tau_A)$. Observe that the null is equivalent to $H_0 : \beta \in \mathcal{B}_0(\bar{\theta}) := \{\beta : \exists \tau_{post} \text{ s.t. } l'\tau_{post} = \bar{\theta}, A\beta - d - AM_{post}\tau_{post} \leqslant 0\}$ and the alternative is equivalent to $H_1 : \beta = \delta_A + \tau_A =: \beta_A$. It is clear from the definition of $\mathcal{B}_0$ that it is convex. From Lemma C.5, the most powerful test that controls size is a one-sided t-test (Neyman-Pearson test) that rejects for large values of $(\beta_A - \tilde{\beta}_n)'\Sigma_n^{-1}\hat{\beta}_n$, where $\tilde{\beta} := \arg\min_{\beta \in \mathcal{B}_0} ||\beta_A - \beta||_{\Sigma_n}$. Define $\psi_\alpha^{MP}(\hat{\beta}_n; A, d, \bar{\theta}, \Sigma_n, \delta_A, \tau_A)$ to be an indicator for whether the Neyman-Pearson test rejects H_0 in favor H_1 given a draw $\hat{\beta}_n$ that is assumed to be normally distributed with covariance Σ_n. The second limit can thus be written as

$$\lim_{n\to\infty} \sup_{\mathcal{C}_{\alpha,n}\in\mathcal{I}_\alpha(\Delta,\frac{1}{n}\Sigma^*)} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}\right) =$$

$$\lim_{n\to\infty} \mathbb{E}_{(\delta_A,\tau_A,\Sigma_n)}\left[\psi_\alpha^{MP}(\hat{\beta}_n; A, d, \theta^{ub} + \frac{1}{\sqrt{n}}x, \Sigma_n, \delta_A, \tau_A)\right]$$

A-9

From Lemma C.6 and the fact that $\Sigma_n = \frac{1}{n}\Sigma^*$, $\psi_\alpha^{MP}(\hat{\beta}_n; A, d, \theta^{ub} + \frac{1}{\sqrt{n}}x, \Sigma_n, \delta_A, \tau_A) = \psi_\alpha^{MP}(\sqrt{n}\hat{\beta}_n; A, \sqrt{n}d, \sqrt{n}\theta^{ub} + x, \Sigma^*, \sqrt{n}\delta_A, \sqrt{n}\tau_A)$. It follows that

$$\lim_{n\to\infty} \sup_{\mathcal{C}_{\alpha,n}\in\mathcal{I}_\alpha(\Delta,\Sigma_n)} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}\right) \overset{(1)}{=}$$

$$\mathbb{E}_{(\sqrt{n}\delta_A,\sqrt{n}\tau_A,\Sigma^*)}\left[\psi_\alpha^{MP}(\hat{\beta}_n; A, \sqrt{n}d, \sqrt{n}\theta^{ub} + x, \Sigma^*, \sqrt{n}\delta_A, \sqrt{n}\tau_A)\right] \overset{(2)}{=}$$

$$\mathbb{E}_{(\sqrt{n}\delta_A,\sqrt{n}\tau_A,\Sigma^*)}\left[\psi_\alpha^{MP}(\hat{\beta}_n; A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*, \sqrt{n}\delta_A, \sqrt{n}\tau_A)\right] \overset{(3)}{\to} \Phi(c^*x - z_{1-\alpha}),$$

where (1) used that if $\hat{\beta}_n \sim \mathcal{N}\left(\beta_A, \frac{1}{n}\Sigma^*\right)$, then $\sqrt{n}\hat{\beta}_n \sim \mathcal{N}\left(\sqrt{n}\beta_A, \Sigma^*\right)$, (2) used that $\theta_n^{ub} = \sqrt{n}\theta^{ub}$ and (3) follows by Lemma C.12. $\qquad\square$

Proof of Corollary 4.1

Proof. Let $\mathcal{C}_{\alpha,n}^{C,k}$ be the conditional confidence set for Δ_k. Proposition 4.2 implies that

$$\lim_{n\to\infty} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}^{C,1}\right) = \lim_{n\to\infty} \sup_{\mathcal{C}_{\alpha,n}\in\mathcal{I}_\alpha(\Delta_1,\Sigma_n)} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}\right)$$

$$\geqslant \lim_{n\to\infty} \sup_{\mathcal{C}_{\alpha,n}\in\mathcal{I}_\alpha(\Delta,\Sigma_n)} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}\right)$$

where the inequality follows from the fact that $\mathcal{I}_\alpha(\Delta,\Sigma_n) \subseteq \mathcal{I}_\alpha(\Delta_1,\Sigma_n)$ since $\mathcal{S}(\beta,\Delta_1) \subseteq \mathcal{S}(\beta,\Delta)$. To complete the proof, we will show that

$$\lim_{n\to\infty} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}^{C,1}\right) = \lim_{n\to\infty} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \bigcup_k \mathcal{C}_{\alpha,n}^{C,k}\right),$$

for which it suffices to show that for all $k > 1$,

$$\lim_{n\to\infty} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)}\left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\alpha,n}^{C,k}\right) = 1.$$

Since $\mathcal{S}(\delta_A + \tau_A, \Delta_k) \subsetneq \mathcal{S}(\delta_A + \tau_A, \Delta_1)$, we have from equation (8) that θ_A^{ub} is strictly larger than the upper bound of $\mathcal{S}(\delta_A + \tau_A, \Delta_k)$. The convergence in the previous display then follows almost immediately from Proposition 4.1. The one technical complication is that rather than a fixed point outside the identified set, we are considering power against a sequence converging to a fixed point outside the identified set. It is, however, straightforward to verify that the argument in Proposition 4.1 applies if we replace $\theta^{ub} + x$ with $\theta^{ub} + x + O(n^{-1/2})$. $\qquad\square$

Proof of Proposition A.1

Proof. Following the same argument as in the proof to Proposition 3.2, we can show that $\mathbb{P}_{(\delta_A, \tau_A, \Sigma_n)} \left((\theta^{ub} + \frac{x}{\sqrt{n}}) \in \mathcal{C}_{\alpha,n}^{FLCI} \right)$ equals

$$\Phi \left(\frac{\theta^{mid} + \bar{b}_{min} + \frac{x}{\sqrt{n}} + \chi_n - \mu_n}{\sigma_{v_n,n}} \right) - \Phi \left(\frac{\theta^{mid} + \bar{b}_{min} + \frac{x}{\sqrt{n}} - \chi_n - \mu_n}{\sigma_{v_n,n}} \right), \qquad (21)$$

where $\bar{b}_{min} = \frac{1}{2} LID(\Delta, \delta_{A,pre})$, $\mu_n = a_n + v_n'(\delta_A + \tau_A)$, and θ^{mid} is the midpoint of $\mathcal{S}(\Delta, \delta_A + \tau_A)$.

The term inside the second normal CDF in the previous display equals

$$-\frac{\chi_n - b_n}{\sigma_{v_n,n}} + \frac{x}{\sqrt{n}\sigma_{v_n,n}} + \frac{\theta^{mid} - \mu_n + \bar{b}_{min} - b_n}{\sigma_{v_n,n}}. \qquad (22)$$

We first show the first term in (22) converges to $-z_{1-\alpha}$. Since χ_n is the $1 - \alpha$ quantile of the $| \mathcal{N} \left(b_n, \sigma_{v_n,n}^2 \right) |$ distribution, $\Phi \left(\frac{\chi_n - b_n}{\sigma_{v_n,n}} \right) - \Phi \left(\frac{-\chi_n - b_n}{\sigma_{v_n,n}} \right) = 1 - \alpha$. Lemma C.25 implies that $\bar{b}_{min} = \frac{1}{2} \sup_{\tilde{\delta}_{pre}} LID(\Delta, \tilde{\delta}_{pre}) > 0$. We argued in the proof to Proposition 3.2 that $b_n \geqslant \bar{b}_{min} > 0$, $\chi_n \geqslant 0$, and $\sigma_{v_n,n} \to 0$, from which we see that $\frac{-\chi_n - b_n}{\sigma_{v_n,n}} \to -\infty$. It follows that $\Phi \left(\frac{\chi_n - b_n}{\sigma_{v_n,n}} \right) \to 1 - \alpha$, and hence $\frac{\chi_n - b_n}{\sigma_{v_n,n}} \to z_{1-\alpha}$.

Next, we show the second term in (22) converges to c^*x, for the same constant c^* as in Proposition 4.2. Lemma C.23 implies that $\lim_{n\to\infty} \frac{x}{\sqrt{n}\sigma_{v_n,n}} = \lim_{n\to\infty} \frac{x}{\sqrt{n}\sigma_{\bar{v},n}} = \frac{x}{\sigma_{\bar{v},1}}$, where $\bar{v}$ is the unique value such that there exists $(\bar{a}, \bar{v})$ with $\bar{b}(\bar{a}, \bar{v}) = \bar{b}_{min}$. Moreover, Lemma C.22 implies that $1/\sigma_{\bar{v},1} = c^*$, from which we see that the limit of the second term is c^*x, as desired.

Now, we show the third term in (22) converges to 0. We argued in the proof to Proposition 3.2 that $|\mu_n - \theta^{mid}| \leqslant b_n - \bar{b}_{min}$. It thus suffices to show that $\frac{b_n - \bar{b}_{min}}{\sigma_{v_n,n}} \to 0$. Lemma C.22 implies that there is a unique pair $(\bar{a}, \bar{v})$ such that $\bar{b}(\bar{a}, \bar{v}) = \bar{b}_{min}$. Let $\bar{\chi}_n = \chi_n(\bar{a}, \bar{v})$ and $\chi_n = \chi_n(a_n, v_n)$. Note that $\chi_n \leqslant \bar{\chi}_n$ by construction, and $b_n \geqslant \bar{b}_{min}$ by Lemma C.19. Hence, using the bounds from Lemma C.21, we have that $b_n + \sigma_{v_n,n} z_{1-\alpha} \leqslant \chi_n \leqslant \bar{\chi}_n = \sigma_{\bar{v},n} cv_\alpha \left(\frac{\bar{b}_{min}}{\sigma_{\bar{v},n}} \right)$,

which, along with the inequality $b_n \geqslant \bar{b}_{min}$, implies that

$$
\begin{aligned}
0 \leqslant \frac{b_n - \bar{b}_{min}}{\sigma_{v_n,n}} \leqslant & \frac{\sigma_{\bar{v},n}}{\sigma_{v_n,n}} cv_\alpha \left(\frac{\bar{b}_{min}}{\sigma_{\bar{v},n}} \right) - \left(z_{1-\alpha} + \frac{\bar{b}_{min}}{\sigma_{v_n,n}} \right) \\
= & \left[cv_\alpha \left(\frac{\bar{b}_{min}}{\sigma_{\bar{v},n}} \right) - \left(z_{1-\alpha} + \frac{\bar{b}_{min}}{\sigma_{\bar{v},n}} \right) \right] + \\
& \left[\left(\frac{\sigma_{\bar{v},n}}{\sigma_{v_n,n}} - 1 \right) cv_\alpha \left(\frac{\bar{b}_{min}}{\sigma_{\bar{v},n}} \right) - \left(\frac{\bar{b}_{min}}{\sigma_{v_n,n}} - \frac{\bar{b}_{min}}{\sigma_{\bar{v},n}} \right) \right].
\end{aligned}
$$

The first bracketed expression in the upper bound above converges to 0 by Lemma C.24. Applying the upper bound from Lemma C.21 to the cv_α term in the second bracketed expression, we obtain that the second bracketed expression is bounded above by $\left(\frac{\sigma_{\bar{v},n}}{\sigma_{v_n,n}} - 1 \right) z_{1-\alpha/2}$, which converges to 0 by Lemma C.23.

Combining the results above, we see that the expression in (22) converges to $c^*x - z_{1-\alpha}$. It follows that $\limsup_{n\to\infty} \mathbb{P}\left((\theta^{ub} + \frac{x}{\sqrt{n}}) \in \mathcal{C}_{\alpha,n}^{FLCI} \right) \leqslant 1 - \Phi(c^*x - z_{1-\alpha})$, and hence $\liminf_{n\to\infty} \mathbb{P}\left((\theta^{ub} + \frac{x}{\sqrt{n}}) \notin \mathcal{C}_{\alpha,n}^{FLCI} \right) \geqslant \Phi(c^*x - z_{1-\alpha})$. Proposition 4.2 gives that $\Phi(c^*x - z_{1-\alpha})$ is the optimal local asymptotic power over procedures that control size, from which the result follows. $\qquad \square$

Proof of Proposition B.1

Proof. The proof follows from the same argument as for Proposition 4.1, replacing Lemma C.2 with Lemma C.4, and Lemma C.16 with Lemma C.17. $\qquad \square$

Proof of Proposition B.2

Proof. By an invariance to scale argument analogous to that in the proof of Proposition 4.2, $\liminf_{n\to\infty} \mathbb{P}_{(\delta_A,\tau_A,\Sigma_n)} \left((\theta_A^{ub} + \frac{1}{\sqrt{n}}x) \notin \mathcal{C}_{\kappa,\alpha,n}^{C\text{-}FLCI} \right)$ equals

$$
\liminf_{n\to\infty} \mathbb{E}_{(\sqrt{n}\delta_A,\sqrt{n}\tau_A,\Sigma^*)} \left[\psi_{\kappa,\alpha}^{C\text{-}FLCI}(\hat{\beta}_n, A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) \right].
$$

From Proposition 4.2, it thus suffices to show

$$
\begin{aligned}
& \liminf_{n\to\infty} \mathbb{E}_{(\sqrt{n}\delta_A,\sqrt{n}\tau_A,\Sigma^*)} \left[\psi_{\kappa,\alpha}^{C\text{-}FLCI}(\hat{\beta}_n, A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) \right] \geqslant \\
& \liminf_{n\to\infty} \mathbb{E}_{(\sqrt{n}\delta_A,\sqrt{n}\tau_A,\Sigma^*)} \left[\psi_{\tilde{\alpha}}^{C}(\hat{\beta}_n, A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) \right].
\end{aligned}
$$

Note that the second stage of the test $\psi_{\kappa,\alpha}^{C\text{-}FLCI}$ is nearly identical to $\psi_{\tilde{\alpha}}^{C}$ except it uses $v_{C\text{-}FLCI}^{lo} := \max\{v^{lo}, v_{FLCI}^{lo}\}$ and $v_{C\text{-}FLCI}^{up} := \min\{v^{up}, v_{FLCI}^{up}\}$ instead of v^{lo} and v^{up}. Since

A-12

$F_{\xi:\xi\in[v^{lo},v^{up}]}(\hat\eta)$ is decreasing in v^{lo} and v^{up}, it suffices to show that $v^{lo}_{FLCI} \xrightarrow{P_n}_p -\infty$, where P_n denotes the sequence of distributions under which $(\delta,\tau,\Sigma) = (\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)$.

Let $\Delta_n = \{\delta \ : \ A\delta \leqslant \sqrt{n}d\}$. Let $v_n = v_n(\Delta_n, \Sigma^*)$ and $\breve{v}_n = v_n(\Delta, \Sigma_n)$. Define a_n and $\breve{a}_n$, and $\breve\chi_n$ and χ_n analogously. By Lemma C.3, $v_n = \breve{v}_n$, $a_n = \sqrt{n}\breve{a}_n$, $\chi_n = \sqrt{n}\breve\chi_n$, and $\bar{b}(a,v;\Delta_n) = \sqrt{n}\bar{b}(\breve{a},\breve{v},\Delta)$. We argued in the proof to Lemma C.23 that $\breve{v}_n \to \bar{v}$. Further, we showed in the proof to Lemma C.22 that $\bar{v} = -\bar\gamma A$, where $\bar\gamma_{-B} = 0$ and $\bar\gamma_B$ is the unique vector such that $\bar\gamma'_B \tilde{A}_{(B,1)} = 1, \bar\gamma'_B \tilde{A}_{(B,-1)} = 0$. Likewise, we argued in the proof to Lemma C.23 that $\breve{a}_n \to \bar{a}$, for $\bar{a}$ the unique value such that $\bar{b}(\bar{a},\bar{v};\Delta) = b_{min}(\Delta)$. We also showed in the proof to Proposition 3.2 that $\breve\chi_n \to b_{min}(\Delta)$.

Let $\tilde{v}_n$ be a vector such that $\tilde{v}'_n A = v'$ (which exists by Lemma B.1). Observe that

$$\tilde{v}'_n \tilde\Sigma \bar\gamma = \tilde{v}'_n A\Sigma^* A'\bar\gamma \to \bar{v}'\Sigma^* A'\bar\gamma = -\bar\gamma' A\Sigma^* A'\bar\gamma = -\bar\gamma'\tilde\Sigma\bar\gamma,$$

where we use $\tilde{v}'_n A = v_n$, $v_n \to \bar{v} = -\bar\gamma'A$ as shown above, and the identity $\tilde\Sigma = A\Sigma^* A'$.

Now, Lemma C.27 implies that there is a constant $c > 0$ such that, with probability approaching one under P_n, $c\cdot\bar\gamma$ is an optimal vertex of the dual problem for $\psi^{C\text{-}FLCI}_{\kappa,\alpha}(\hat\beta_n, A, \sqrt{n}d, \theta^{ub}_n + x, \Sigma^*)$. Observe from Lemma B.2 that if γ_* is an optimal vertex, and $\frac{v'_n\tilde\Sigma\gamma_*}{\gamma'_*\tilde\Sigma\gamma_*} < 0$, then the value of v^{lo}_{FLCI} used in $\psi^{C\text{-}FLCI}_{\kappa,\alpha}(\hat\beta_n, A, \sqrt{n}d, \theta^{ub}_n + x, \Sigma^*)$ is

$$v^{lo}_{FLCI} = \frac{d_{n,1} - \tilde{v}'_n\left(I - \frac{\tilde\Sigma\gamma_*}{\gamma'_*\tilde\Sigma\gamma_*}\gamma'_*\right)\tilde{Y}_n}{\tilde{v}'_n\tilde\Sigma\gamma_*/\gamma'_*\tilde\Sigma\gamma_*} = \frac{-(\tilde{v}'_n\tilde{Y}_n - d_{n,1})}{v'_n\tilde\Sigma\gamma_*/\gamma'_*\tilde\Sigma\gamma_*} + \gamma'_*\tilde{Y}_n, \tag{23}$$

where $\tilde{Y}_n = A\hat\beta_n - \sqrt{n}d - \tilde{A}_{(\cdot,1)}(\theta^{ub}_n + x)$. Since $c\bar\gamma$ is optimal in the dual with probability approaching one under P_n, and

$$\frac{v'_n\tilde\Sigma c\bar\gamma}{c^2\bar\gamma'\tilde\Sigma\bar\gamma} \to -\frac{1}{c} < 0 \tag{24}$$

by the argument above, we have that with probability approaching 1 under P_n,

$$v^{lo}_{FLCI} = \frac{-(\tilde{v}'_n\tilde{Y}_n - d_{n,1})}{v'_n\tilde\Sigma c\bar\gamma/c^2\bar\gamma'\tilde\Sigma\bar\gamma} + c\bar\gamma'\tilde{Y}_n. \tag{25}$$

Now, we showed in the proof to Lemma C.8 that $\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left[\breve{Y}_{n,B}\right] = -\tilde{A}_{(B,1)}x$ regardless of n, where $\breve{Y}_n = \tilde{Y}_n - \tilde{A}_{(\cdot,-1)}(\sqrt{n}\tilde\tau^{ub}_1)$ for a vector $\tilde\tau^{ub}_1$. Since $\bar\gamma_{-B} = 0$ and $\bar\gamma'_B \tilde{A}_{(B,-1)} = 0$, it follows that $\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left[\bar\gamma'\tilde{Y}_n\right] = -\bar\gamma'_B\tilde{A}_{(B,1)}x$ regardless of n. Thus,

$$c\bar\gamma'\tilde{Y}_n \overset{P_n}{\sim} \mathcal{N}\left(-c\bar\gamma'_B\tilde{A}_{(B,1)}x, \ c^2\bar\gamma'\tilde\Sigma\bar\gamma\right). \tag{26}$$

Now, note that by construction, $\tilde{v}_n' \tilde{Y}_n - d_{n,1} = a_n + v_n' \hat{\beta} - (\theta_n^{ub} + x) - \chi_n$. Further, we have that $\hat{\beta}_n \overset{P_n}{\sim} \mathcal{N}(\sqrt{n}\beta_A, \Sigma^*)$, where $\beta_A = \delta_A + \tau_A$. It follows that under P_n, $\tilde{v}_n' \tilde{Y}_n - d_{n,1} = a_n + v_n' \sqrt{n}\beta_A - (\theta_n^{ub} + x) - \chi_n + v_n'\xi$, where $\xi \sim \mathcal{N}(0, \Sigma^*)$. Applying the equalities $v_n = \breve{v}_n$, $a = \sqrt{n}\breve{a}_n$, $\chi_n = \sqrt{n}\breve{\chi}$ derived above, along with the fact that $\theta_n^{ub} = \sqrt{n}\theta_1^{ub}$ by Lemma C.1, we see that under P_n,

$$\tilde{v}_n' \tilde{Y}_n - \tilde{d}_1 = \sqrt{n}\left(\breve{a}_n + \breve{v}_n'\beta_A - \theta_1^{ub} - \breve{\chi}_n\right) - x + \breve{v}_n'\xi, \tag{27}$$

Since $\breve{v}_n' \to \bar{v}$, it follows that $\breve{v}_n'\xi \to_d \bar{v}'\xi$ by Slutsky's lemma.

Additionally, the results above imply that $\breve{a}_n + \breve{v}_n'\beta_A - \theta_1^{ub} - \breve{\chi}_n \to \bar{a} + \bar{v}'\beta_A - \theta_1^{ub} - \bar{b}_{min}(\Delta)$. We claim that this limit is strictly negative. Since Assumption 4 holds, Lemma C.25 implies that $LID(\Delta, \delta_A) > 0$. Hence, for $\epsilon > 0$ sufficiently small, we have that $\theta_1^{ub} - \epsilon \in \mathcal{S}(\Delta, \beta_A)$. If the limit above were weakly positive, then we would have $\bar{a} + \bar{v}'\beta_A - (\theta_1^{ub} - \epsilon) - \bar{b}_{min}(\Delta) > 0$. However, this implies that $\bar{b}(\bar{a}, \bar{v}) > \bar{b}_{min}(\Delta)$, which is a contradiction. The limit must thus be strictly negative, as desired. We then see from (27) that

$$\tilde{v}_n' \tilde{Y}_n(\theta_n^{ub} + x) - d_{n,1} \overset{P_n}{\to}_p -\infty. \tag{28}$$

Displays (24), (25), (26), and (28) together give that $v_{FLCI}^{lo} \overset{P_n}{\to}_p -\infty$, as desired.

$\square$

C.2   Auxiliary Lemmas for Finite Sample Normal Results

Lemma C.1 (Scale Invariance of Identified Set). *For any $n > 0$, let $\Delta_n = \{\delta : A\delta \leq \sqrt{n}d\}$. Fix $\delta_A \in \Delta_1$ and τ_A. Then, $\mathcal{S}(\sqrt{n}\delta_A + \sqrt{n}\tau_A, \Delta_n) = \sqrt{n}\mathcal{S}(\delta_A + \tau_A, \Delta_1)$. This implies $\theta_n^{ub} = \sqrt{n}\theta_1^{ub}$, where $\theta_n^{ub} := \sup_\theta \mathcal{S}(\sqrt{n}\delta_A + \sqrt{n}\tau_A, \Delta_n)$, and $\theta_n^{lb} = \sqrt{n}\theta_1^{lb}$, for $\theta_n^{lb} := \inf_\theta \mathcal{S}(\sqrt{n}\delta_A + \sqrt{n}\tau_A, \Delta_n)$*

Proof. Let $\mathcal{S}_n = \mathcal{S}(\sqrt{n}\delta_A + \sqrt{n}\tau_A, \Delta_n)$ and $\beta_A = \delta_A + \tau_A$. By definition, $\bar{\theta}_n \in \mathcal{S}_n$ iff there exists a vector $\tau_{post} \in \mathbb{R}^{\bar{T}}$ such that $l'\tau_{post} = \bar{\theta}_n$ and $A(\sqrt{n}\beta_A - M_{post}\tau_{post} - \sqrt{n}d) \leq 0$. Using the change of basis described in Section 4.1, it follows that $\bar{\theta}_n \in \mathcal{S}_n$ iff there exists $\tilde{\tau}_n \in \mathbb{R}^{\bar{T}-1}$ such that

$$A\sqrt{n}\beta_A - \sqrt{n}d - \tilde{A}_{(\cdot,1)}\bar{\theta}_n - \tilde{A}_{(\cdot,-1)}\tilde{\tau}_n \leq 0. \tag{29}$$

Thus, $\bar{\theta}_1 \in \mathcal{S}_1$ iff there exists $\tilde{\tau}_1$ such that

$$A\beta_A - d - \tilde{A}_{(\cdot,1)}\bar{\theta}_1 - \tilde{A}_{(\cdot,-1)}\tilde{\tau}_1 \leq 0. \tag{30}$$

If there exists a $\tilde{\tau}_1$ such that (30) holds for $\bar{\theta}_1$, then multiplying both sides of (30) by $\sqrt{n}$

implies that (29) holds with $\bar{\theta}_n = \sqrt{n}\bar{\theta}_1$ and $\tilde{\tau}_n = \sqrt{n}\tilde{\tau}_1$. Likewise, if there exists a $\tilde{\tau}_n$ such that (29) holds for $\bar{\theta}_n$, then multiplying both sides of (29) by $\frac{1}{\sqrt{n}}$ implies that (30) holds with $\bar{\theta}_1 = \frac{1}{\sqrt{n}}\bar{\theta}_n$ and $\tilde{\tau}_1 = \frac{1}{\sqrt{n}}\tilde{\tau}_n$. The desired result follows immediately.

$\square$

Lemma C.2 (Scale Invariance of Conditional Test). *For any $n > 0$ and $(\hat{\beta}; A, d, \bar{\theta}, \Sigma)$,*
$$\psi_\alpha^C(\hat{\beta}; A, d, \bar{\theta}, \Sigma) = \psi_\alpha^C(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, n\Sigma).$$

Proof. Using the change of basis described in Section 4.1, the test statistic used to calculate $\psi_\alpha^C(\hat{\beta}; A, d, \bar{\theta}, \Sigma)$ is

$$\min_{\eta,\tilde{\tau}} \eta \quad \text{s.t} \quad A\hat{\beta} - d - \tilde{A}_{(\cdot,1)}\bar{\theta} - \tilde{A}_{(\cdot,-1)}\tilde{\tau} \leqslant \eta\tilde{\sigma},$$

where $\tilde{\sigma}$ is the vector containing the square roots of the diagonal elements of $\tilde{\Sigma} = A\Sigma A'$. Since multiplying the constraints by $\sqrt{n}$ does not change the feasible set, this optimization is equivalent to

$$\min_{\eta,\tilde{\tau}} \eta \quad \text{s.t.} \quad A\sqrt{n}\hat{\beta} - \sqrt{n}d - \tilde{A}_{(\cdot,1)}\sqrt{n}\bar{\theta} - \tilde{A}_{(\cdot,-1)}\sqrt{n}\tilde{\tau} \leqslant \eta\sqrt{n}\tilde{\sigma}.$$

However, since $\tilde{\tau}$ enters only in the constraint, and $\{\sqrt{n}\tilde{\tau} : \tilde{\tau} \in \mathbb{R}^{\bar{T}-1}\} = \{\tilde{\tau} \in \mathbb{R}^{\bar{T}-1}\}$, this linear program is equivalent to

$$\min_{\eta,\tilde{\tau}} \eta \quad \text{s.t} \quad A\sqrt{n}\hat{\beta} - \sqrt{n}d - \tilde{A}_{(\cdot,1)}\sqrt{n}\bar{\theta} - \tilde{A}_{(\cdot,-1)}\tilde{\tau} \leqslant \eta\sqrt{n}\tilde{\sigma},$$

which is the test statistic used to calculate $\psi_\alpha^C(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, n\Sigma)$. Thus, the test statistics used for the two problems are the same. Additionally, the feasible set for the dual for the unscaled problem is $F_1 = \{\gamma : \gamma'\tilde{A}_{(\cdot,-1)} = 0, \gamma'\tilde{\sigma} = 1, \gamma \geqslant 0\}$, whereas for the problem scaled by $\sqrt{n}$ it is $F_n = \{\gamma : \gamma'\tilde{A}_{(\cdot,-1)} = 0, \gamma'\sqrt{n}\tilde{\sigma} = 1, \gamma \geqslant 0\} = \frac{1}{\sqrt{n}}F_1$. It follows that $V_n = \frac{1}{\sqrt{n}}V_1$, for V_1 and V_n respectively the vertices of F_1 and F_n. Moreover, it is immediate that if γ_1 is an optimal vertex of the unscaled problem, then $\gamma_n = \frac{1}{\sqrt{n}}\gamma_1$ will be an optimal vertex of the problem scaled by $\sqrt{n}$.

Recall that the critical value for the conditional test depends on $\gamma_*'\tilde{\Sigma}\gamma_*$, where γ_* is an optimal vertex, and the values v^{lo} and v^{up} which are functions of γ_*, $\tilde{\Sigma}$, and a sufficient statistic S. However, for $\gamma_n = \frac{1}{\sqrt{n}}\gamma_1$, we have that $\gamma_n'(n\tilde{\Sigma})\gamma_n = \frac{1}{\sqrt{n}}\gamma_1'(n\tilde{\Sigma})\frac{1}{\sqrt{n}}\gamma_1 = \gamma_1'\tilde{\Sigma}\gamma_1$, and so the variances are the same. Let $\tilde{Y}_1 = A\hat{\beta} - d - \tilde{A}_{(\cdot,1)}\bar{\theta}$ and $\tilde{Y}_n = A\sqrt{n}\hat{\beta} - \sqrt{n}d - \tilde{A}_{(\cdot,1)}\sqrt{n}\bar{\theta} = \sqrt{n}\tilde{Y}_1$. The sufficient statistic used to construct v^{lo} and v^{up} in the first problem is $S_1 = (I - \frac{\tilde{\Sigma}\gamma_1\gamma_1'}{\gamma_1'\tilde{\Sigma}\gamma_1})\tilde{Y}_1$, whereas for the second problem it is $S_n = (I - \frac{n\tilde{\Sigma}\gamma_n\gamma_n'}{\gamma_n'n\tilde{\Sigma}\gamma_n})\tilde{Y}_n$. The identities $\tilde{Y}_n = \sqrt{n}\tilde{Y}_1$ and $\gamma_1 = \sqrt{n}\gamma_n$ immediately imply that $S_n = \sqrt{n}S_1$. The values v^{lo} and v^{up} for the first

problem are then the minimum and maximum of $C_1 = \{c : c = \max_{\tilde\gamma_1 \in V_1} \tilde\gamma'_1(S_1 + \frac{\tilde\Sigma\gamma_1}{\gamma'_1\tilde\Sigma\gamma_1}c)\}$.
Likewise, the values v^{lo} and v^{up} for the second problem are the the minimum and maximum
of $C_n = \{c : c = \max_{\tilde\gamma_n \in V_n} \tilde\gamma'_n(S_n + \frac{n\tilde\Sigma\gamma_n}{\gamma'_n n\tilde\Sigma\gamma_n}c)\}$, However, since $V_n = \sqrt{n}V_1$, $S_n = \sqrt{n}S_1$, and
$\gamma_n = \frac{1}{\sqrt{n}}\gamma_1$, we have that for any c,

$$\max_{\tilde\gamma_1 \in V_1} \tilde\gamma'_1(S_1 + \frac{\tilde\Sigma\gamma_1}{\gamma'_1\tilde\Sigma\gamma_1}c) = \max_{\tilde\gamma_n \in V_n} \sqrt{n}\tilde\gamma'_n(\frac{1}{\sqrt{n}}S_n + \frac{\tilde\Sigma\sqrt{n}\gamma_n}{\gamma'_n n\tilde\Sigma\gamma_n}c) = \max_{\tilde\gamma_n \in V_n} \tilde\gamma'_n(S_n + \frac{n\tilde\Sigma\gamma_n}{\gamma'_n n\tilde\Sigma\gamma_n}c)\},$$

from which it is immediate that $C_1 = C_n$, and hence the values of v^{lo} and v^{up} are the
same across the two problems as well. Since the test statistics and critical values of the two
problems are the same, they are equivalent. $\qquad\square$

Lemma C.3 (Scale Invariance for FLCIs). *Let $v_n(\Delta, \Sigma_n)$ be the vector v_n used in the optimal
FLCI as defined in Section 3, making the dependence on (Δ, Σ_n) explicit. Define, $\chi_n(\Delta, \Sigma_n)$,
$a_n(\Delta, \Sigma_n)$, and $\bar{b}_n(\Delta, \Sigma_n)$ analogously. Let $\Sigma_n = \frac{1}{n}\Sigma^*$ and $\Delta_n = \sqrt{n}\Delta$. Then*

1. $v_n(\Delta_n, \Sigma^*) = v_n(\Delta, \frac{1}{n}\Sigma_n)$

2. $a_n(\Delta_n, \Sigma^*) = \sqrt{n}a_n(\Delta, \frac{1}{n}\Sigma_n)$

3. $\chi_n(\Delta_n, \Sigma^*) = \sqrt{n}\chi_n(\Delta, \frac{1}{n}\Sigma_n)$

4. $\bar{b}(a_n(\Delta_n, \Sigma^*), v_n(\Delta_n, \Sigma^*); \Delta_n) = \sqrt{n}\bar{b}(a_n(\Delta, \Sigma_n), v_n(\Delta, \Sigma_n); \Delta)$.

Proof. We show in the proof to Lemma C.19 that $\bar{b}(a, v; \Delta_n)$ is finite only if $v_{post} = l$, in
which case $\bar{b}(a, v; \Delta) = \max_{\delta \in \Delta} |a + v'\delta|$. Likewise, $\bar{b}(\sqrt{n}a, v; \Delta_n)$ is finite only if $v_{post} = l$, in
which case

$$\bar{b}(\sqrt{n}a, v; \Delta) = \max_{\delta_n \in \Delta_n} |\sqrt{n}a + v'\delta_n| = \max_{\delta \in \Delta} |\sqrt{n}a + v'\sqrt{n}\delta| = \sqrt{n}\bar{b}(a, v, \Delta).$$

Next, observe that using the invariance above and $\Sigma_n = \frac{1}{n}\Sigma^*$,

$$\chi(\sqrt{n}a, v; \Sigma^*, \Delta_n) = \sqrt{v'\Sigma^*v} \cdot cv_\alpha\left(\frac{\bar{b}(\sqrt{n}a, v; \Delta_n)}{\sqrt{v'\Sigma^*v}}\right)$$

$$= \sqrt{n} \cdot \sqrt{v'\Sigma_n v} \cdot cv_\alpha\left(\frac{\bar{b}(a, v; \Delta)}{\sqrt{v'\Sigma_n v}}\right) = \sqrt{n}\chi(a, v; \Sigma^*, \Delta).$$

It is then immediate that if $(a^*, v^*) = \arg\min_{(a,v)} \chi(a, v; \Delta, \Sigma_n)$, then $(\sqrt{n}a^*, v^*) = \arg\min_{(a,v)}$
$\chi(a, v; \Delta_n, \Sigma^*)$, from which the first two results follow. The second two results then follow
from the two invariances derived above. $\qquad\square$

Lemma C.4 (Scale Invariance for FLCI Hybrid). *For any $n > 0$ and $(\hat{\beta}; A, d, \bar{\theta}, \Sigma)$,*

$$\psi_{\kappa,\alpha}^{C\text{-}FLCI}(\hat{\beta}; A, d, \bar{\theta}, \Sigma) = \psi_{\kappa,\alpha}^{C\text{-}FLCI}(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, n\Sigma).$$

Proof. From Lemma C.3, if $\mathcal{C}_{\kappa}^{FLCI}(\hat{\beta}; A, d, \bar{\theta}, \Sigma) = a_1 + v_1'\hat{\beta} \pm \chi_1$, then $\mathcal{C}_{\kappa}^{FLCI}(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, \sqrt{n}\Sigma) = \sqrt{n}\left(a_1 + v_1'\hat{\beta} \pm \chi_1\right)$. Thus, $\bar{\theta} \in \mathcal{C}_{\kappa}^{FLCI}(\hat{\beta}; A, d, \bar{\theta}, \Sigma)$ iff $\sqrt{n}\bar{\theta} \in \mathcal{C}_{\kappa}^{FLCI}(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, \sqrt{n}\Sigma)$, so the first stage tests are equivalent. The second stage test is almost identical to ψ_{α}^{C}, which is invariant to scale by Lemma C.2, except it replaces v^{lo} with $\max\{v^{lo}, v_{FLCI}^{lo}\}$ and v^{up} with $\min\{v^{up}, v_{FLCI}^{up}\}$. It thus suffices to show that v_{FLCI}^{lo} and v_{FLCI}^{up} are invariant to scale. We show in the proof to Lemma C.2 that $S_n = \sqrt{n}S_1$, and if $\gamma_n \in \hat{V}_n$ then $\gamma_1 = \sqrt{n}\gamma_n \in \hat{V}_n$; where objects subscripted by 1 indicate values based on $(\hat{\beta}; A, d, \bar{\theta}, \Sigma)$ and values subscripted by n indicate those based on $(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, \sqrt{n}\Sigma)$. Additionally, Lemma C.3 implies that $\tilde{V}_1 = \tilde{V}_n$, and $\bar{d}_n = \sqrt{n}\bar{d}_n$. The desired invariance is then immediate from the formulas in Lemma B.2. $\qquad\square$

Lemma C.5. *Suppose $\hat{\beta} \sim \mathcal{N}(\beta, \Sigma)$ for Σ known. Let B_0 be a closed, convex set. Then the most-powerful size α test of $H_0 : \beta \in B_0$ against the point alternative $H_A : \beta = \beta_A$ is equivalent to the most powerful test of $H_0 : \beta = \tilde{\beta}$ against $H_A : \beta = \beta_A$, where $\tilde{\beta} = \arg\min_{\beta} ||\beta - \beta_A||_{\Sigma}$ and $||\cdot||_{\Sigma}$ is the Mahalanobis norm in Σ, $||x||_{\Sigma} = \sqrt{x'\Sigma^{-1}x}$. The most powerful test rejects for values of $(\beta_A - \tilde{\beta})'\Sigma^{-1}\hat{\beta}$ greater than $(\beta_A - \tilde{\beta})'\Sigma^{-1}\tilde{\beta} + z_{1-\alpha}||\beta_A - \tilde{\beta}||_{\Sigma}$, and has power against the alternative of $\Phi(||\beta_A - \tilde{\beta}||_{\Sigma} - z_{1-\alpha})$, for $z_{1-\alpha}$ the $1-\alpha$ quantile of the standard normal.*

Proof. Define $<\cdot, \cdot>_{\Sigma}$ by $<x, y>_{\Sigma} = x'\Sigma^{-1}y$, and observe that $<\cdot, \cdot>_{\Sigma}$ is an inner product. The result then follows immediately from the discussion in Section 2.4.3 of Ingster and Suslina (2003), replacing all instances of the usual euclidean inner product with $<\cdot, \cdot>_{\Sigma}$. $\qquad\square$

Lemma C.6 (Scale Invariance of Optimal Test). *Suppose $\Delta = \{\delta : A\delta \leqslant d\}$. As in the proof to Proposition 4.2, let $\psi_{\alpha}^{MP}(\hat{\beta}; A, d, \bar{\theta}, \Sigma, \delta_A, \tau_A)$ be an indicator for whether the most powerful (Neyman-Pearson) test between the null hypothesis $H_0 : \delta \in \Delta, \theta = \bar{\theta}$ and the alternative $H_A : (\delta, \tau) = (\delta_A, \tau_A)$ rejects the null, given the realization $\hat{\beta}$ which is assumed to come from a normal distribution with variance Σ. Then for any $n > 0$,*

$$\psi_{\alpha}^{MP}(\hat{\beta}; A, d, \bar{\theta}, \Sigma, \delta_A, \tau_A) = \psi_{\alpha}^{MP}(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, n\Sigma, \sqrt{n}\delta_A, \sqrt{n}\tau_A)$$

Proof. As argued in the proof to Proposition 4.2, the null $H_0 : \delta \in \Delta, \theta = \bar{\theta}$ is equivalent to the null $H_0 : \beta \in \mathcal{B}_0(\bar{\theta}, d) := \{\beta : \exists \tau_{post} \text{ s.t. } l'\tau_{post} = \bar{\theta}, A\beta - d - AM_{post}\tau_{post} \leqslant 0\}$. Likewise, the alternative that $(\delta, \tau) = (\delta_A, \tau_A)$ is equivalent to $H_A : \beta = \delta_A + \tau_A =: \beta_A$. It is clear from the definition that $\mathcal{B}_0$ is convex. Thus, by Lemma C.5, the most powerful test of H_0 against

A-17

H_A when the covariance of $\hat{\beta}$ is Σ is a t-test that rejects for large values of $(\beta_A - \tilde{\beta}_1)'\Sigma^{-1}\hat{\beta}$, where $\tilde{\beta}_1 = \arg\min_{\beta \in \mathcal{B}_0(\bar{\theta},d)} ||\beta_A - \beta||_\Sigma$. Its critical value is $(\beta_A - \tilde{\beta})'\Sigma^{-1}\tilde{\beta}_1 + z_{1-\alpha}||\beta_A - \tilde{\beta}_1||_\Sigma$, for $z_{1-\alpha}$ the $1 - \alpha$ quantile of the standard normal distribution.

Similarly, the null hypothesis $\delta \in \{\delta : A\delta \leqslant \sqrt{n}d\}, \theta = \sqrt{n}\bar{\theta}$ is equivalent to $H_0 : \beta \in \mathcal{B}_0(\sqrt{n}\bar{\theta}, \sqrt{n}d) = \{\beta : \exists \tau_{post} \text{ s.t. } l'\tau_{post} = \sqrt{n}\bar{\theta}, A\beta - \sqrt{n}d - AM_{post}\tau_{post} \leqslant 0\}$. Likewise, the alternative that $(\delta, \tau) = (\sqrt{n}\delta_A, \sqrt{n}\tau_A)$ is equivalent to $H_A : \beta = \sqrt{n}\delta_A + \sqrt{n}\tau_A = \sqrt{n}\beta_A$. It is clear from the definition that $\mathcal{B}_0(\sqrt{n}\bar{\theta}, \sqrt{n}d)$ is convex. Thus, by Lemma C.5, the most powerful test of H_0 against H_A when the covariance of $\hat{\beta}$ is $n\Sigma$ is a t-test that rejects for large values of $(\sqrt{n}\beta_A - \tilde{\beta}_2)'(n\Sigma)^{-1}\hat{\beta}$, where $\tilde{\beta}_2 = \arg\min_{\beta \in \mathcal{B}_0(\sqrt{n}\bar{\theta}, \sqrt{n}d)} ||\sqrt{n}\beta_A - \beta||_{(n\Sigma)}$. Its critical value is $(\sqrt{n}\beta_A - \tilde{\beta}_2)'(n\Sigma)^{-1}\tilde{\beta}_2 + z_{1-\alpha}||\beta_A - \tilde{\beta}_2||_{(n\Sigma)}$.

Now, define

$$\eta(\hat{\beta}, A, d, \bar{\theta}, \Sigma) := \min_{\eta, \tilde{\tau}} \eta \text{ s.t. } A\hat{\beta} - d - \tilde{A}_{(\cdot,1)}\bar{\theta} - \tilde{A}_{(\cdot,-1)}\tilde{\tau} \leqslant \eta\tilde{\sigma}, \tag{31}$$

where $\tilde{\sigma}$ is the square root of the diagonal elements of $A\Sigma A'$. It follows immediately from the definition of $\mathcal{B}_0$ and the function η that we can write

$$\mathcal{B}_0(\bar{\theta}, d) = \{\beta : \eta(\beta, A, d, \bar{\theta}, \Sigma) \leqslant 0\}$$
$$\mathcal{B}_0(\sqrt{n}\bar{\theta}, \sqrt{n}d) = \{\beta : \eta(\beta, A, \sqrt{n}d, \sqrt{n}\bar{\theta}, n\Sigma) \leqslant 0\}$$

As argued in the proof to Lemma C.2 above, for any $n > 0$, $\eta(\beta, A, d, \bar{\theta}, \Sigma) = \eta(\sqrt{n}\beta, A, \sqrt{n}d, \sqrt{n}\bar{\theta}, n\Sigma)$, from which it follows that $\sqrt{n}\mathcal{B}_0(\bar{\theta}, d) = \mathcal{B}_0(\sqrt{n}\bar{\theta}, \sqrt{n}d)$. Thus,

$$\tilde{\beta}_2 = \sqrt{n} \arg\min_{\beta \in \mathcal{B}_0(\bar{\theta},d)} ||\sqrt{n}\beta_A - \sqrt{n}\beta||_{(n\Sigma)}$$
$$= \sqrt{n} \arg\min_{\beta \in \mathcal{B}_0(\bar{\theta},d)} ||\beta_A - \beta||_\Sigma = \sqrt{n}\tilde{\beta}_1,$$

where the second equality uses the fact that $||\sqrt{n}x||_{(n\Sigma)} = ||x||_\Sigma$. Thus, the test statistic used for $\psi_\alpha^{MP}(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, \sqrt{n}\delta_A, \sqrt{n}\tau_A)$ is

$$(\sqrt{n}\beta_A - \tilde{\beta}_2)'(n\Sigma)^{-1}(\sqrt{n}\hat{\beta}) = (\sqrt{n}\beta_A - \sqrt{n}\tilde{\beta}_1)'(n\Sigma)^{-1}(\sqrt{n}\hat{\beta}) = (\beta_A - \tilde{\beta}_1)'\Sigma^{-1}\hat{\beta},$$

which is the test statistic used for $\psi_\alpha^{MP}(\hat{\beta}; A, d, \bar{\theta}, \delta_A, \tau_A)$.

Likewise, the critical value used for $\psi_\alpha^{MP}(\sqrt{n}\hat{\beta}; A, \sqrt{n}d, \sqrt{n}\bar{\theta}, \sqrt{n}\delta_A, \sqrt{n}\tau_A)$ is

$$
\begin{aligned}
(\sqrt{n}\beta_A - \tilde{\beta}_2)'(n\Sigma)^{-1}\tilde{\beta}_2 + z_{1-\alpha}||\sqrt{n}\beta_A - \tilde{\beta}_2||_{(n\Sigma)} &= \\
(\sqrt{n}\beta_A - \sqrt{n}\tilde{\beta}_1)'(n\Sigma)^{-1}\sqrt{n}\tilde{\beta}_1 + z_{1-\alpha}||\sqrt{n}\beta_A - \sqrt{n}\tilde{\beta}_1||_{(n\Sigma)} &= \\
(\beta_A - \tilde{\beta}_1)'\Sigma^{-1}\tilde{\beta}_1 + z_{1-\alpha}||\beta_A - \tilde{\beta}_1||_\Sigma, &
\end{aligned}
$$

which is the critical value used for $\psi_\alpha^{MP}(\hat{\beta}; A, d, \bar{\theta}, \delta_A, \tau_A)$. We have thus shown that the test statistics and critical values for the two tests align, which gives the desired result.

$\square$

Lemma C.7 (Rank of binding moments). *Suppose Assumption 4 holds. Let $\theta^{ub} := \sup_\theta \mathcal{S}(\Delta, \delta_A + \tau_A)$ and $\beta_A = \delta_A + \tau_A$. Then there exists a vector $\tilde{\tau}^{ub} \in \mathbb{R}^{\bar{T}-1}$ such that for $B = B(\delta^{**})$ as defined in Assumption 4,*

$$
\begin{aligned}
A_{(B,\cdot)}\beta_A - d_B - \tilde{A}_{(B,1)}\theta^{ub} - \tilde{A}_{(B,-1)}\tilde{\tau}^{ub} &= 0 \\
A_{(-B,\cdot)}\beta_A - d_{-B} - \tilde{A}_{(-B,1)}\theta^{ub} - \tilde{A}_{(-B,1)}\tilde{\tau}^{ub} &= -\epsilon < 0,
\end{aligned}
$$

for ϵ a vector with strictly positive entries. Additionally, the matrix $\tilde{A}_{(B,-1)}$ has rank equal to $|B| - 1$, and $\{\gamma_B : \gamma_B'\tilde{A}_{(B,-1)} = 0\} = \{c\bar{\gamma}_B : c \in \mathbb{R}\}$ for a non-zero vector $\bar{\gamma}_B \geqslant 0$.

Proof. From (7), we have that $\theta^{ub} = l'\beta_{A,post} - l'\delta_{post}^{**}$, for δ^{**} a solution to

$$
\min_\delta l'\delta_{post} \text{ s.t. } A\delta \leqslant d, \delta_{pre} = \delta_{A,pre}. \tag{32}
$$

Let $B = B(\delta^{**})$ index the binding inequalities of the optimization above at δ^{**}, so that

$$
A_{(B,\cdot)}\delta^{**} - d_B = 0 \tag{33}
$$
$$
A_{(-B,\cdot)}\delta^{**} - d_{-B} = -\epsilon < 0. \tag{34}
$$

By Assumption 4, $A_{(B,post)}$ has rank $|B|$.

Now, let $\tau^{**} = (\delta_{A,post} + \tau_{A,post}) - \delta_{post}^{**}$. Since by construction $\theta^{ub} \in \mathcal{S}_\theta(\Delta, \beta_A)$, we have

$$
\begin{pmatrix} \delta_{pre}^{**} \\ \delta_{post}^{**} + \tau^{**} \end{pmatrix} = \begin{pmatrix} \delta_{A,pre} \\ \delta_{A,post} + \tau_{A,post} \end{pmatrix} = \beta_A.
$$

It follows that

$$
\begin{aligned}
A\delta^{**} &= A\beta_A - AM_{post}\tau^{**} \\
&= A\beta_A - AM_{post}\Gamma^{-1}\Gamma\tau^{**} \\
&= A\beta_A - \tilde{A}_{(\cdot,1)}(l'\tau^{**}) - \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}\tau^{**} \\
&= A\beta_A - \tilde{A}_{(\cdot,1)}(\theta^{ub}) - \tilde{A}_{(\cdot,-1)}\tilde{\tau}^{ub},
\end{aligned}
$$

where the third equality uses the definition of $\tilde{A}$ and the fact that the first row of Γ is l'; and the fourth equality uses the fact that $\theta^{ub} = l'((\delta_{A,post} + \tau_{A,post}) - \delta^{**}_{post}) = l'\tau^{**}$ and defines $\tilde{\tau}^{ub} := -\Gamma_{(-1,\cdot)}\tau^{**}$. The first result then follows immediately from the previous display along with (33) and (34).

To show the second set of results, note that $\tilde{A}_{(B,\cdot)} = A_{(B,post)}\Gamma^{-1}$. Since $A_{(B,post)}$ has rank $|B|$ by assumption and Γ^{-1} is full rank, $\tilde{A}_{(B,\cdot)}$ also has rank $|B|$. This implies that $\tilde{A}_{(B,-1)}$ has rank of either $|B| - 1$ or $|B|$. To show that the rank must be $|B| - 1$, note that the optimization (32) can be re-written as

$$
\min_{\delta_{post}} l'\delta_{post} \quad \text{s.t.} \quad A_{(\cdot,post)}\delta_{post} \leq d - A_{(\cdot,pre)}\delta_{A,pre}. \tag{35}
$$

Since the optimization is assumed to have a finite solution, it is equivalent to its dual formulation,

$$
\max_{\gamma} \gamma'(A_{(\cdot,pre)}\delta_{A,pre} - d) \text{ s.t. } -\gamma'A_{(\cdot,post)} = l', \gamma \geq 0.
$$

Let $\bar{\gamma}$ be a solution to the dual problem. Since $\bar{\gamma}$ is feasible in the dual, $-\bar{\gamma}'A_{(\cdot,post)} = l'$ and $\bar{\gamma} \geq 0$. Additionally, by the complementary slackness conditions, it must be that $\bar{\gamma}_{-B} = 0$. Hence, we have $-\bar{\gamma}'_B A_{(B,post)} = l'$. Multiplying on the right by Γ^{-1}, we obtain $-\bar{\gamma}'_B \tilde{A}_{(B,\cdot)} = l'\Gamma^{-1}$. Recall that by construction the first row of Γ is l', so $l' = e'_1\Gamma$, and thus $-\bar{\gamma}'_B \tilde{A}_{(B,\cdot)} = l'\Gamma^{-1} = e'_1$. This shows, however, that $\bar{\gamma}_B$ is in the nullspace of $\tilde{A}'_{(B,-1)}$ but not in the null space of $\tilde{A}'_{(B,\cdot)}$. It follows that the rank of $\tilde{A}_{(B,-1)}$ is strictly less than that of $\tilde{A}_{(B,\cdot)}$, and thus must be equal to $|B| - 1$. Since $\tilde{A}_{(B,-1)}$ has $|B|$ rows and rank $|B| - 1$, by the rank nullity theorem the set $\{\gamma_B : \gamma'_B \tilde{A}_{(B,-1)} = 0\}$ must be one dimensional. We've shown that $\bar{\gamma}'_B \tilde{A}_{(B,-1)} = 0$, and $\bar{\gamma}_B \neq 0$ since $\bar{\gamma}'_B A_{(B,post)} = -l' \neq 0$, which implies that $\{\gamma_B : \gamma'_B \tilde{A}_{(B,-1)} = 0\} = \{c\bar{\gamma}_B : c \in \mathbb{R}\}$, as needed.

$\square$

Lemma C.8 (Limiting power of conditional test)**.** *Let* $\Delta = \{\delta : A\delta \leq d\}$, *and fix* $\delta_A \in \Delta$, τ_A, *and* Σ^* *positive definite. If Assumption 4 holds, then for any* $x > 0$,

$$\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)} \left[\psi_\alpha^C(\hat{\beta}_n; A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) \right] \to 1 - \Phi(z_{1-\alpha} - c^* x),$$

where $\theta_n^{ub} := \sup \mathcal{S}(\Delta_n, \sqrt{n}\delta_A + \sqrt{n}\tau_A)$, $\Delta_n = \{\delta : A\delta \leqslant \sqrt{n}d\}$, *and* c^* *is a positive constant (not depending on* x *or* α*). In particular,* $c^* = -\bar{\gamma}_B' \tilde{A}_{(B,1)}/\sigma_B$, *where* $\sigma_B = \sqrt{\bar{\gamma}_B' A_{(B,\cdot)} \Sigma^* A_{(B,\cdot)}' \bar{\gamma}_B}$ *and* $\bar{\gamma}_B$ *is the unique vector such that* $\bar{\gamma}_B' \tilde{A}_{(B,-1)} = 0, \bar{\gamma}_B \geqslant 0, ||\bar{\gamma}_B|| = 1$.

Proof. From Lemma C.7, there exists a vector $\tilde{\tau}_1^{ub}$ and a set of indices B such that

$$A_{(B,\cdot)}\beta_A - d_B - \tilde{A}_{(B,1)}\theta_1^{ub} - \tilde{A}_{(B,-1)}\tilde{\tau}_1^{ub} = 0 \tag{36}$$

$$A_{(-B,\cdot)}\beta_A - d_{-B} - \tilde{A}_{(-B,1)}\theta_1^{ub} - \tilde{A}_{(B,-1)}\tilde{\tau}_1^{ub} = -\epsilon < 0, \tag{37}$$

and the set $\{\gamma_B \in \mathbb{R}^{|B|} : \gamma_B' \tilde{A}_{(B,-1)} = 0\} = \{c\bar{\gamma}_B : c \in \mathbb{R}\}$ for some non-zero vector $\bar{\gamma}_B \geqslant 0$, which without loss of generality we can normalize so that $||\bar{\gamma}_B|| = 1$. Let $\tilde{\sigma}$ be the vector containing the square roots of the diagonal elements of $A\Sigma^* A'$. It follows that the set $\{\gamma_B \in \mathbb{R}^{|B|} : \gamma_B' \tilde{A}_{(B,-1)} = 0, \gamma_B' \tilde{\sigma}_B = 1, \gamma_B \geqslant 0\}$ is a singleton. In particular, its lone element is $\gamma_B^* := (\tilde{\sigma}_B' \bar{\gamma}_B)^{-1} \bar{\gamma}_B$. Note that $(\tilde{\sigma}_B' \bar{\gamma}_B)^{-1}$ is well-defined since $\bar{\gamma}_B \geqslant 0$ and has at least one strictly positive element, and $\tilde{\sigma} > 0$ since by assumption A has no all-zero rows and Σ^* is positive definite.

Now, consider $\psi_\alpha^C(\hat{\beta}_n, A_{(B,\cdot)}, \sqrt{n}d_B, \theta_n^{ub} + x, \Sigma^*)$, the conditional test that uses only the moments in B. The test statistic for the conditional test that uses only the moments in B is

$$\eta(\hat{\beta}_n, A_{(B,\cdot)}, \sqrt{n}d_B, \theta_n^{ub} + x, \Sigma^*) = \min_{\eta, \tilde{\tau}} \eta$$

$$\text{s.t. } A_{(B,\cdot)}\hat{\beta}_n - \sqrt{n}d_B - \tilde{A}_{(B,1)}(\theta_n^{ub} + x) - \tilde{A}_{(B,-1)}\tilde{\tau} \leqslant \eta\tilde{\sigma}_B.$$

The equivalent dual problem is

$$\max_{\gamma_B} \gamma_B' \tilde{Y}_{B,n} \text{ s.t. } \gamma_B' \tilde{A}_{(B,-1)} = 0, \gamma_B' \tilde{\sigma}_B = 1, \gamma_B \geqslant 0,$$

where $\tilde{Y}_{B,n} = A_{(B,\cdot)}\hat{\beta}_n - \sqrt{n}d_B - \tilde{A}_{(B,1)}(\theta_n^{ub} + x)$. We have shown, however, that there is a single value, γ_B^*, that satisfies the constraints of the dual problem, and so the solution to the problem in the previous display is $\gamma_B^{*\prime} \tilde{Y}_{B,n}$. Additionally, since the set of dual vertices is a singleton, the conditioning event that γ_B^* is optimal is trivial, so $v^{lo} = -\infty$ and $v^{up} = \infty$. It follows that the conditional test using only the moments B is a one-sided t-test that rejects for large values of $\gamma_B^{*\prime} \tilde{Y}_{B,n}$. Specifically, the critical value is $z_{1-\alpha}\sigma_B^*$, for $\sigma_B^* = \sqrt{\gamma_B^{*\prime} A_{(B,\cdot)} \Sigma^* A_{(B,\cdot)}' \gamma_B^*}$ the standard deviation of the test statistic $\gamma_B^{*\prime} \tilde{Y}_{B,n}$. We claim that $\sigma_B^* > 0$. To see why this is the case, observe that Assumption 4 implies that $A_{(B,\cdot)}$ has full

row rank, and by construction $\gamma_B^* \neq 0$, so $\gamma_B^{*\prime} A_{(B,\cdot)} \neq 0$. That $\sigma_B^* > 0$ then follows from the fact that Σ^* is positive definite.

Additionally, observe that

$$
\begin{aligned}
\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)} \left[\gamma_B^{*\prime} \tilde{Y}_{B,n} \right] &= \gamma_B^{*\prime} \left[A_{(B,\cdot)} \sqrt{n}\beta_A - \sqrt{n}d_B - \tilde{A}_{(B,1)}(\theta_n^{ub} + x) \right] \\
&= \gamma_B^{*\prime} \sqrt{n} \left[A_{(B,\cdot)}\beta_A - d_B - \tilde{A}_{(B,1)}\theta_1^{ub} \right] - \gamma_B^{*\prime} \tilde{A}_{(B,1)} x \\
&= \gamma_B^{*\prime} [\sqrt{n}\tilde{A}_{(B,-1)}\tilde{\tau}_1^{ub}] - \gamma_B^{*\prime} \tilde{A}_{(B,1)} x = -\gamma_B^{*\prime} \tilde{A}_{(B,1)} x
\end{aligned}
$$

where the second equality uses $\theta_n^{ub} = \sqrt{n}\theta_1^{ub}$ from Lemma C.1, the third equality uses (36) to substitute for the term in brackets, and the final equality follows from the fact that $\gamma_B^{*\prime} \tilde{A}_{(B,-1)} = 0$ by construction. Thus, regardless of n, the conditional test using only the moments in B rejects with probability

$$
\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)} \left[\psi_\alpha^C(\hat{\beta}_n, A_{(B,\cdot)}, \sqrt{n}d_B, \theta_n^{ub} + x, \Sigma^*) \right] = 1 - \Phi(z_{1-\alpha} - (-\gamma_B^{*\prime} \tilde{A}_{(B,1)}/\sigma_B^*) \cdot x). \quad (38)
$$

Note also that we showed in the proof of Lemma C.7 that $-\bar{\gamma}_B' \tilde{A}_{(B,\cdot)} = e_1'$, which implies that $-\bar{\gamma}_B' \tilde{A}_{(B,1)} = 1$, and hence $c^* := -\bar{\gamma}_B^{*\prime} \tilde{A}_{(B,1)}/\sigma_B^* > 0$ since γ_B^* is a positive multiple of $\bar{\gamma}_B$. Moreover, observe that if we define $\sigma_B = \sqrt{\bar{\gamma}_B' A_{(B,\cdot)} \Sigma^* A_{(B,\cdot)}' \bar{\gamma}_B}$, then $\bar{\gamma}_B^*/\sigma_B^* = \bar{\gamma}_B/\sigma_B$, so $c^* = -\bar{\gamma}_B' \tilde{A}_{(B,1)}/\sigma_B$.

Recall that $\psi_\alpha^C(\hat{\beta}_n; A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) = \psi_\alpha^C(\tilde{Y}_n, A\Sigma^* A')$ for $\tilde{Y}_n = A\hat{\beta}_n - \sqrt{n}d - \tilde{A}_{(\cdot,1)}(\theta_n^{ub} + x)$. Since the conditional test optimizes over $\tilde{\tau}$, and $\tilde{\tau}$ appears in this optimization only in the term $\tilde{A}_{(\cdot,-1)}\tilde{\tau}$, the result of the conditional test using $\tilde{Y}_n$ is equivalent to the result of the conditional test that replaces $\tilde{Y}_n$ with $\breve{Y}_n = \tilde{Y}_n - \tilde{A}_{(\cdot,-1)}(\sqrt{n}\tilde{\tau}_1^{ub})$ (see Lemma 16 in ARP for a formal justification). That is, $\psi_\alpha^C(\tilde{Y}_n, A\Sigma^* A') = \psi_\alpha^C(\breve{Y}_n, A\Sigma^* A')$. The expectation of the elements of $\breve{Y}_n$ corresponding with the rows B is

$$
\begin{aligned}
\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)} \left[\breve{Y}_{n,B} \right] &= A_{(B,\cdot)} \sqrt{n}\beta_A - \sqrt{n}d_B - \tilde{A}_{(B,1)}(\theta_n^{ub} + x) - \tilde{A}_{(B,-1)}(\sqrt{n}\tilde{\tau}_1^{ub}) \\
&= \sqrt{n} \left(A_{(B,\cdot)}\beta_A - d_B - \tilde{A}_{(B,1)}\theta_1^{ub} - \tilde{A}_{(B,-1)}\tilde{\tau}_1^{ub} \right) - \tilde{A}_{(B,1)} x \\
&= -\tilde{A}_{(B,1)} x,
\end{aligned}
$$

where the second line uses the fact that $\theta_n^{ub} = \sqrt{n}\theta_1^{ub}$ from Lemma C.1, and the third uses (36). Similarly, the expectation of the elements of $\breve{Y}_n$ corresponding to the rows other than B is

$$\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left[\breve{Y}_{n,-B}\right] = A_{(-B,\cdot)}\sqrt{n}\beta_A - \sqrt{n}d_{-B} - \tilde{A}_{(-B,1)}(\theta_n^{ub} + x) - \tilde{A}_{(-B,1)}(\sqrt{n}\tilde{\tau}_1^{ub})$$

$$= \sqrt{n}\left(A_{(-B,\cdot)}\beta_A - d_{-B} - \tilde{A}_{(-B,1)}\theta_1^{ub} - \tilde{A}_{(-B,1)}\tilde{\tau}_1^{ub}\right) - \tilde{A}_{(-B,1)}x$$

$$= \sqrt{n}(-\epsilon) - \tilde{A}_{(-B,1)}x,$$

where the last line uses (37). Since $-\epsilon < 0$, all of the elements of $\mathbb{E}\left[\breve{Y}_{n,-B}\right]$ converge to $-\infty$ as $n \to \infty$, whereas $\mathbb{E}\left[\breve{Y}_{n,B}\right]$ does not depend on n. It follows from Proposition 3 in ARP that the conditional test based on the full set of moments is equal to the conditional test that only uses the moments B with probability approaching one,

$$\lim_{n\to\infty} \mathbb{P}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left(\psi_\alpha^C(\hat{\beta}_n; A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) = \psi_\alpha^C(\hat{\beta}_n, A_{(B,\cdot)}, \sqrt{n}d_B, \theta_n^{ub} + x, \Sigma^*)\right) = 1.$$

This, combined with (38), gives the desired result. $\qquad\square$

Lemma C.9. *Let $\mathcal{B}$ be a closed, convex subset of $\mathbb{R}^K$, and $\beta_A \notin \mathcal{B}$. Let $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} ||\beta - \beta_A||_\Sigma$, where $||x||_\Sigma^2 = x'\Sigma^{-1}x$ for some positive definite matrix Σ. Then for any $\beta \in \mathcal{B}$, $(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta - \tilde{\beta}) \geqslant 0$.*

Proof. Consider any $\beta \in \mathcal{B}$. Define $\beta_\theta = \theta(\beta - \tilde{\beta}) + \tilde{\beta}$, and note that since $\mathcal{B}$ is convex $\beta_\theta \in \mathcal{B}$ for any $\theta \in [0,1]$. Further,

$$||\beta_\theta - \beta_A||_\Sigma^2 = \theta^2||\beta - \tilde{\beta}||_\Sigma^2 + 2\theta(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta - \tilde{\beta}) + ||\tilde{\beta} - \beta_A||_\Sigma^2.$$

Differentiating with respect to θ, we have

$$\frac{\partial}{\partial\theta}||\beta_\theta - \beta_A||_\Sigma^2 = 2\theta||\beta - \tilde{\beta}||_\Sigma^2 + 2(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta - \tilde{\beta}),$$

from which we see that the derivative evaluated at $\theta = 0$ is $2(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta_A - \tilde{\beta})$. Since $\tilde{\beta}$ minimizes the norm, it follows that we must have $2(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta_A - \tilde{\beta}) \geqslant 0$, else we could achieve a lower value of the norm at β_θ by choosing θ sufficiently small. $\qquad\square$

Lemma C.10. *Let $\mathcal{B} = \{\beta \in \mathbb{R}^K : v'\beta \leqslant d\}$ for some $v \in \mathbb{R}^K\backslash\{0\}$ and $d \in \mathbb{R}$. Let $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} ||\beta - \beta_A||_\Sigma$ for some $\beta_A \notin \mathcal{B}$, where $||x||_\Sigma^2 = x'\Sigma^{-1}x$ and Σ is positive definite. Then $(\beta_A - \tilde{\beta})'\Sigma^{-1} = c \cdot v'$ for the positive constant $c = \frac{v'\beta_A - d}{\sqrt{v'\Sigma v}}$.*

Proof. Note that we can form a basis $v, \tilde{v}_2, ..., \tilde{v}_K$ such that $v'\tilde{v}_j = 0$ for $j = 2, ..., K$. It follows by construction that for any $j = 2, .., K$ and any $t \in \mathbb{R}$, $\tilde{\beta} + t \cdot \tilde{v}_j \in \mathcal{B}$. Hence, from

A-23

Lemma C.9, $-(\beta_A - \tilde{\beta})'\Sigma^{-1}(t\tilde{v}_j) \geqslant 0$. Since we can choose t both positive and negative, it follows that $(\beta_A - \tilde{\beta})'\Sigma^{-1}\tilde{v}_j = 0$ for all j. Since $(\beta_A - \tilde{\beta})'\Sigma^{-1}$ is orthogonal to $\{\tilde{v}_2, ..., \tilde{v}_K\}$, and $\{v, \tilde{v}_2, ..., \tilde{v}_K\}$ form a basis, we have that $(\beta_A - \tilde{\beta})'\Sigma^{-1} = c \cdot v'$, for some $c \in \mathbb{R}$. Multiplying both sides of the equation on the right by Σv, we obtain that $(\beta_A - \tilde{\beta})'v = c \cdot v'\Sigma v$. However, since $\tilde{\beta}$ is the closest point to β_A in Mahalanobis distance, it must be on the boundary of $\mathcal{B}$, and so $v'\tilde{\beta} = d$. It follows that $c = (v'\beta_A - d)/(v'\Sigma v)$, which is clearly positive since $\beta_A \notin \mathcal{B}$ and thus $v'\beta_A > d$. $\qquad\square$

Lemma C.11 (Power of optimal test for linear subspace). *Let $\mathcal{B} = \{\beta \in \mathbb{R}^K \,:\, v'\beta \leqslant d\}$ for some $v \in \mathbb{R}^K \backslash \{0\}$ and $d \in \mathbb{R}$. Suppose $\hat{\beta} \sim \mathcal{N}(\beta, \Sigma)$ for Σ positive definite known, and consider the problem of testing $H_0 : \beta \in \mathcal{B}$ against $H_A : \beta = \beta_A$ for some $\beta_A \notin \mathcal{B}$. Then the most powerful size-α test of H_0 against H_A is a one-sided t-test that rejects for large values of $v'\hat{\beta}$, and has power equal to $\Phi((v'\beta_A - d)/\sqrt{v'\Sigma v} - z_{1-\alpha})$.*

Proof. From Lemma C.5, the most powerful test rejects for large values of $(\beta_A - \tilde{\beta})'\Sigma^{-1}\hat{\beta}$, where $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} ||\beta - \beta_A||_\Sigma$, and has power $\Phi(||\beta_A - \tilde{\beta}||_\Sigma - z_{1-\alpha})$. By Lemma C.10, $(\beta_A - \tilde{\beta})'\Sigma^{-1} = cv'$, for $c = (v'\beta_A - d)/(v'\Sigma v)$. It follows that

$$
\begin{aligned}
||\beta_A - \tilde{\beta}||_\Sigma^2 &= (\beta_A - \tilde{\beta})'\Sigma^{-1}(\beta_A - \tilde{\beta}) \\
&= cv'(\beta_A - \tilde{\beta}) \\
&= c(v'\beta_A - d) = (v'\beta_A - d)^2/(v'\Sigma v),
\end{aligned}
$$

where we use the fact that $v'\tilde{\beta} = d$, since $\tilde{\beta}$ must be on the boundary of $\mathcal{B}$, as argued in the proof to Lemma C.10. The result then follows immediately. $\qquad\square$

Lemma C.12 (Asymptotic Power Envelope). *Let $\Delta = \{\delta \,:\, A\delta \leqslant d\}$, and fix $\delta_A \in \Delta$, τ_A, and Σ^* positive definite. If Assumption 4 holds, then for any $x > 0$,*

$$
\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left[\psi_\alpha^{MP}(\hat{\beta}_n; A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*)\right] \to 1 - \Phi(z_{1-\alpha} - c^*x),
$$

where ψ_α^{MP} is as defined in the proof to Proposition 4.2, $\theta_n^{ub} := \sup \mathcal{S}(\Delta_n, \sqrt{n}\delta_A + \sqrt{n}\tau_A)$, $\Delta_n = \{\delta \,:\, A\delta \leqslant \sqrt{n}d\}$, and c^* is the same positive constant as in Lemma C.8.

Proof. As argued in the proof to Lemma C.6, the null hypothesis $H_0 : \theta = \bar{\theta}, \delta \in \{A\delta \leqslant d\}$ is equivalent to the null $H_0 : \beta \in \mathcal{B}_0(\bar{\theta}, d) = \{\beta \,:\, \exists \tau_{post} \text{ s.t. } l'\tau_{post} = \bar{\theta}, A\beta - d - AM_{post}\tau_{post} \leqslant 0\}$, which we showed in Lemma C.6 to be equivalent to $\mathcal{B}_0(\bar{\theta}, d) = \{\beta \,:\, \eta(\beta, A, d, \bar{\theta}, \Sigma^*) \leqslant 0\}$ for the function η as defined in (31). Thus, the null hypothesis for the test associated with $\psi_\alpha^{MP}(\hat{\beta}_n; A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*)$ can be written as $H_0 : \beta_n \in \mathcal{B}_{n,0} := \{\beta \,:\, \eta(\beta, A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) \leqslant 0\}$. Under the alternative for this test, $\beta_n = \sqrt{n}\beta_A$, so by Lemma C.5 the most

powerful test uses the test statistic $(\sqrt{n}\beta_A - \tilde{\beta}_n)'\Sigma^{*-1}\hat{\beta}_n$, where $\tilde{\beta}_n = \arg\min_{\beta \in \mathcal{B}_{n,0}} ||\beta - \sqrt{n}\beta_A||_{\Sigma^*}$.

Now, from Lemma C.7, there exists a vector $\tilde{\tau}_1^{ub}$ and a set of indices B such that

$$A_{(B,\cdot)}\beta_A - d_B - \tilde{A}_{(B,1)}\theta_1^{ub} - \tilde{A}_{(B,-1)}\tilde{\tau}_1^{ub} = 0 \tag{39}$$

$$A_{(-B,\cdot)}\beta_A - d_{-B} - \tilde{A}_{(-B,1)}\theta_1^{ub} - \tilde{A}_{(B,-1)}\tilde{\tau}_1^{ub} = -\epsilon < 0, \tag{40}$$

where $\{\gamma_B \in \mathbb{R}^{|B|} : \gamma_B'\tilde{A}_{(B,-1)} = 0\} = \{c\bar{\gamma}_B : c \in \mathbb{R}\}$ for some non-zero vector $\bar{\gamma}_B \geqslant 0$. Define $\mathcal{B}_n^B := \{\beta : \eta(\beta, A_{(B,\cdot)}, \sqrt{n}d_B, \theta_n^{ub} + x, \Sigma^*) \leqslant 0\}$, the analog to $\mathcal{B}_{n,0}$ that restricts attention only to the set of moments B. By an argument analogous to that in the proof to Lemma C.8 (replacing $\hat{Y}$ with μ), we can show that $\eta(\beta, A_{(B,\cdot)}, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) = \gamma_B^{*'}\mu_{B,n}(\beta)$, where $\mu_{B,n}(\beta) = A_{(B,\cdot)}\beta - \sqrt{n}d_B - \tilde{A}_{(B,1)}(\theta_n^{ub} + x)$ and $\gamma_B^* = (\bar{\gamma}_B'\tilde{\sigma})^{-1}\bar{\gamma}_B$. Note also that (39) implies that $\tilde{A}_{(B,1)}\theta_1^{ub} = A_{(B,\cdot)}\beta_A - d_B - \tilde{A}_{(B,-1)}\tilde{\tau}^{ub}$. Substituting into the expression for $\mu_{B,n}(\beta)$ and using the fact that $\theta_n^{ub} = \sqrt{n}\theta_1^{ub}$ by Lemma C.1, we obtain $\mu_{B,n}(\beta) = A_{(B,\cdot)}(\beta - \sqrt{n}\beta_A) - \tilde{A}_{(B,1)}x + \sqrt{n}\tilde{A}_{(B,-1)}\tilde{\tau}^{ub}$. Since $\gamma_B^{*'}\tilde{A}_{(B,-1)} = 0$ by construction, this implies that $\gamma_B^{*'}\mu_{B,n}(\beta) = \gamma_B^{*'}(A_{(B,\cdot)}(\beta - \sqrt{n}\beta_A) - \tilde{A}_{(B,1)}x)$. Hence,

$$\begin{aligned}
\mathcal{B}_n^B &= \{\beta : \eta(\beta, A_{(B,\cdot)}, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*) \leqslant 0\} \\
&= \{\beta : \gamma_B^{*'}(A_{(B,\cdot)}(\beta - \sqrt{n}\beta_A) - \tilde{A}_{(B,1)}x) \leqslant 0\} \\
&= \{\beta : \gamma_B^{*'}(A_{(B,\cdot)}((\beta - (\sqrt{n} - 1)\beta_A) - \beta_A) - \tilde{A}_{(B,1)}x) \leqslant 0\} \\
&= \{\beta : (\beta - (\sqrt{n} - 1)\beta_A) \in \mathcal{B}_1^B\} = (\sqrt{n} - 1)\beta_A + \mathcal{B}_1^B.
\end{aligned}$$

Now, define $\beta_n^* = \arg\min_{\beta \in \mathcal{B}_n^B} ||\beta - \sqrt{n}\beta_A||_{\Sigma^*}$. The results above imply that

$$\begin{aligned}
\beta_n^* &= \arg\min_{\beta \in \mathcal{B}_n^B} ||\beta - \sqrt{n}\beta_A||_{\Sigma^*} \\
&= \arg\min_{\beta \in (\sqrt{n} - 1)\beta_A + \mathcal{B}_1^B} ||\beta - \sqrt{n}\beta_A||_{\Sigma^*} \\
&= (\sqrt{n} - 1)\beta_A + \arg\min_{\beta \in \mathcal{B}_1^B} ||\underbrace{\beta + (\sqrt{n} - 1)\beta_A - \sqrt{n}\beta_A}_{= (\beta - \beta_A)}||_{\Sigma^*} \\
&= (\sqrt{n} - 1)\beta_A + \beta_1^*.
\end{aligned}$$

Observe that $\mathcal{B}_n^B \supseteq \mathcal{B}_{n,0}$ since $\mathcal{B}_n^B$ is the set of values β that are consistent with a subset of the moments used in $\mathcal{B}_{n,0}$ (formally, $\eta(\beta, A_{(B,\cdot)}, \sqrt{n}d_B\theta_n^{ub} + x, \Sigma^*) \leqslant \eta(\beta, A, \sqrt{n}d\theta_n^{ub} + x, \Sigma^*)$ since the RHS minimizes the same objective function subject to additional constraints). Thus, $\beta_n^* = \tilde{\beta}_n$ iff $\beta_n^* \in \mathcal{B}_{n,0}$. From the definition of $\mathcal{B}_{n,0}$, this occurs iff there exists a value $\tilde{\tau}_n$ such

that

$$A_{(B,\cdot)}\beta_n^* - \sqrt{n}d_B - \tilde{A}_{(B,1)}(\theta_n^{ub} + x) - \tilde{A}_{(B,-1)}\tilde{\tau}_n \leqslant 0$$
$$A_{(-B,\cdot)}\beta_n^* - \sqrt{n}d_{-B} - \tilde{A}_{(-B,1)}\sqrt{n}(\theta_n^{ub} + x) - \tilde{A}_{(B,-1)}\tilde{\tau}_n \leqslant 0$$

Now, since $\beta_1^* \in \mathcal{B}_1^B$, there exists a value $\tilde{\tau}_1^*$ such that

$$A_{(B,\cdot)}\beta_1^* - d_B - \tilde{A}_{(B,1)}(\theta_1^{ub} + x) - \tilde{A}_{(B,-1)}\tilde{\tau}_1^* \leqslant 0.$$

It follows that

$$A_{(B,\cdot)}\beta_n^* - \sqrt{n}d_B - \tilde{A}_{(B,1)}(\theta_n^{ub} + x) - \tilde{A}_{(B,-1)}(\tilde{\tau}_1^* + (\sqrt{n} - 1)\tilde{\tau}_1^{ub})$$
$$= A_{(B,\cdot)}\beta_1^* - d_B - \tilde{A}_{(B,1)}(\theta_1^{ub} + x) - \tilde{A}_{(B,-1)}\tilde{\tau}_1^* + (\sqrt{n} - 1)\left[A_{(B,\cdot)}\beta_A - d_B - \tilde{A}_{(B,1)}\theta_1^{ub} - \tilde{A}_{(B,-1)}\tilde{\tau}_1^{ub}\right]$$
$$= A_{(B,\cdot)}\beta_1^* - d_B - \tilde{A}_{(B,1)}(\theta_1^{ub} + x) - \tilde{A}_{(B,-1)}\tilde{\tau}_1^* \leqslant 0,$$

where the first equality uses the fact that $\theta_n^{ub} = \sqrt{n}\theta_1^{ub}$ by Lemma C.1 and $\beta_n^* = \beta_1^* + (\sqrt{n} - 1)\beta_A$ as shown above, and the second equality uses (39).

Similarly, we have

$$A_{(-B,\cdot)}\beta_n^* - \sqrt{n}d_{-B} - \tilde{A}_{(-B,1)}\sqrt{n}(\theta_n^{ub} + x) - \tilde{A}_{(B,-1)}(\tilde{\tau}_1^* + (\sqrt{n} - 1)\tilde{\tau}_1^{ub}) =$$
$$= A_{(-B,\cdot)}\beta_1^* - d_{-B} - \tilde{A}_{(-B,1)}(\theta_1^{ub} + x) - \tilde{A}_{(-B,1)}\tilde{\tau}_1^* +$$
$$(\sqrt{n} - 1)\left[A_{(-B,\cdot)}\beta_A - d_{-B} - \tilde{A}_{(-B,1)}\theta_1^{ub} - \tilde{A}_{(-B,1)}\tilde{\tau}_1^{ub}\right]$$
$$= \left[A_{(-B,\cdot)}\beta_1^* - d_{-B} - \tilde{A}_{(-B,1)}(\theta_1^{ub} + x) - \tilde{A}_{(-B,1)}\tilde{\tau}_1^*\right] - (\sqrt{n} - 1)\epsilon,$$

for ϵ a vector with strictly positive elements, where the first equality again uses that $\theta_n^{ub} = \sqrt{n}\theta_1^{ub}$ and $\beta_n^* = \beta_1^* + (\sqrt{n} - 1)\beta_A$, and the second equality uses (40). Since the term in brackets in the final expression in the previous display does not depend on n and all elements of the final term go to $-\infty$, for n sufficiently large the expression in the previous display will be less than or equal to 0. Thus, for n sufficiently large, $\beta_n^* = \tilde{\beta}_n$, and hence the MP test of $H_0 : \beta \in \mathcal{B}_{n,0}$ against $H_A : \beta = \sqrt{n}\beta_A$ is equivalent to the most powerful test of $H_0 : \beta \in \mathcal{B}_n^B$ against $H_A : \beta = \sqrt{n}\beta_A$.

We showed earlier in the proof that $\mathcal{B}_n^B = \{\beta : v'\beta \leqslant \bar{d}_n\}$, for $v = \gamma_B^{*\prime}A_{(B,\cdot)}$ and $\bar{d}_n = \gamma_B^{*\prime}\tilde{A}_{(B,1)}x + v'\sqrt{n}\beta_A$. From Lemma C.11, the MP test of $H_0 : \beta \in \mathcal{B}_n^B$ against $H_A : \beta = \beta_A$ has power equal to $\Phi((v'\sqrt{n}\beta_A - \bar{d}_n)/(v'\Sigma^*v) - z_{1-\alpha})$. Plugging in the definitions of v and $\bar{d}$ and cancelling like terms, we obtain that the power of the test is $\Phi(-\gamma_B^{*\prime}\tilde{A}_{(B,1)}x/\sigma_B^* - z_{1-\alpha})$, for $\sigma_B^* = \sqrt{\gamma_B^{*\prime}A_{(B,\cdot)}\Sigma^*A_{(B,\cdot)}'\gamma_B^*}$, which coincides with the expression for the limiting power

of the conditional test in Lemma C.8, as needed. $\qquad\square$

Lemma C.13 (Lower bound for η). *Let $\eta(\beta, A, d, \bar{\theta}, \Sigma)$ be as defined in (31). Fix Σ^* positive definite. For any δ_A, τ_A, and d, let $\beta_A(\delta_A, \tau_A) = \delta_A + \tau_A$ and $\theta^{ub}(\delta_A, \tau_A, d) = \sup \mathcal{S}(\Delta, \delta_A + \tau_A)$ for $\Delta = \{\delta : A\delta \leqslant d\}$. Let $\eta^*(x; \delta_A, \tau_A, d) := \eta(\beta_A(\delta_A, \tau_A), A, d, \theta^{ub}(\delta_A, \tau_A, d) + x, \Sigma^*)$. Then there exists a scalar $c(\Sigma^*, A) > 0$ such that $\eta^*(x; \delta_A, \tau_A, d) \geqslant c(\Sigma^*, A) \cdot x$ for all δ_A, τ_A, and d.*

Proof. Observe that $\eta(\beta_A, A, d, \bar{\theta}, x)$ is equivalent to the linear program

$$\min_{\eta, \tau} \eta \ \text{s.t.} \ A\beta_A - d - A\tau \leqslant \eta\tilde{\sigma}, \ l'\tau = \bar{\theta}.$$

The dual formulation for this problem is

$$\max_{\gamma} \begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \bar{\theta} \end{pmatrix} \ \text{s.t.} \ \gamma_A'A + \gamma_\theta l' = 0, \ \gamma_A'\tilde{\sigma} = 1, \ \gamma_A \geqslant 0,$$

where γ_A is a vector with length equal to the number of rows of A, and γ_θ is a scalar. Note that the feasible set for the dual depends on A and Σ^* but not on d, δ_A, or τ_A. Let V_D denote the set of vertices of the dual, which is finite, and recall that maximizing over the feasible set is equivalent to maximizing over the set of vertices.

Now, we first claim that $\eta(\beta_A, A, d, \theta^{ub}, \Sigma^*) = 0$. Note that since θ^{ub} is in the identified set, it must be that $\eta(\beta_A, A, d, \theta^{ub}, \Sigma^*) \leqslant 0$. Towards contradiction, suppose that $\eta(\beta_A, A, d, \theta^{ub}, \Sigma^*) = -\epsilon_1 < 0$. Then for all $\gamma = \begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix} \in V_D$,

$$\begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} \end{pmatrix} \leqslant -\epsilon_1.$$

Since V_D is finite, $\bar{\gamma}_\theta := \max_{\gamma \in V_D} \gamma_\theta$ is finite. But then for $\epsilon_2 > 0$,

$$\begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} + \epsilon_2 \end{pmatrix} \leqslant -\epsilon_1 + \bar{\gamma}_\theta \epsilon_2.$$

By choosing ϵ_2 sufficiently small, we can make the upper bound in the previous display less than or equal to 0. However, this implies that $\eta(\beta_A, A, d, \theta^{ub} + \epsilon_2, \Sigma^*) \leqslant 0$. But this in turn implies $\theta^{ub} + \epsilon_2$ is in the identified set, which contradicts θ^{ub} being maximal. Therefore, $\eta(\beta_A, A, d, \theta^{ub}, \Sigma^*) = 0$.

Additionally, we claim that for $\bar{\theta} = \theta^{ub}$, there must be an optimal dual vertex with $\gamma_\theta > 0$. Towards contradiction, suppose not. Then there exists $\epsilon_3 > 0$ such that for all

A-27

$$\gamma = \begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix} \in V_{D,+} := \{\gamma \in V_D : \gamma_\theta > 0\}, \quad \begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} \end{pmatrix} < -\epsilon_3. \text{ Letting } \epsilon_4 =$$

$\epsilon_3 / \max_{\gamma \in V_{D,+}} \gamma_\theta$, it follows that for all $\gamma \in V_{D,+}$, $\begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} + \epsilon_4 \end{pmatrix} < 0$. Additionally,

for $\gamma = \begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix} \in V_D \backslash V_{D,+}$, we have $\gamma_\theta \leqslant 0$, and so

$$\begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} + \epsilon_4 \end{pmatrix} \leqslant \begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} \end{pmatrix} \leqslant 0.$$

Thus, $\begin{pmatrix} \gamma_A \\ \gamma_\theta \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} + \epsilon_4 \end{pmatrix} \leqslant 0$ for all $\gamma \in V_D$, and so $\eta(\beta_A, A, d, \theta^{ub} + \epsilon_4, \Sigma^*) \leqslant 0$. However, this implies that $\theta^{ub} + \epsilon_4$ is in the identified set, which contradicts θ^{ub} being maximal. Thus, there must be at least one $\gamma^* \in V_{D,+}$ such that

$$\begin{pmatrix} \gamma_A^* \\ \gamma_\theta^* \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} \end{pmatrix} = 0.$$

Since γ^* remains feasible in the dual with $\bar{\theta} = \theta^{ub} + x$, it follows that $\eta(\beta_A, A, d, \theta^{ub} + x, \Sigma^*)$ is lower bounded by

$$\begin{pmatrix} \gamma_A^* \\ \gamma_\theta^* \end{pmatrix}' \begin{pmatrix} A\beta_A - d \\ \theta^{ub} + x \end{pmatrix} = \gamma_\theta^* \cdot x.$$

Note that the choice of $\gamma^* \in V_{D,+}$ depended on d, δ_A, and τ_A. However, as noted earlier in the proof, the set $V_{D,+}$ depends on A and Σ^* but does not on d, δ_A, τ_A. Since $V_{D,+}$ is finite and $\gamma_\theta > 0$ for all $\gamma \in V_{D,+}$, there is a value $c > 0$ such that $\gamma_\theta \geqslant c$ for all $\gamma \in V_{D,+}$. Hence, $\eta^*(x; \delta_A, \tau_A, d) \geqslant c \cdot x$ for all δ_A, τ_A, d, as needed. $\qquad \square$

Lemma C.14. *Let $\alpha \in (0,1)$ and $c > z_{1-\alpha}$. Then there exists a unique constant $\zeta(c) > 0$ such that*

$$\frac{\Phi(c) - \Phi(c - \zeta(c))}{1 - \Phi(c - \zeta(c))} = 1 - \alpha.$$

Additionally, for any values $z_{lo} < z_{up}$, with z_{lo} and z_{up} potentially infinite-valued, and $\eta > \max\{c, z_{lo} + \zeta(c)\}$,

$$F_{\xi | \xi \in [z^{lo}, \infty)}(\eta) > 1 - \alpha,$$

where $F_{\xi | \xi \in [z^{lo}, z^{up})}(\cdot)$ is the CDF of $\xi \sim \mathcal{N}(0,1)$ truncated to $[z^{lo}, z^{up})$.

Proof. First, we show that $F_{\xi | \xi \in [z^{lo}, z^{up})}(t)$ is increasing in t and decreasing in z_{lo} and z_{up}, and

these comparative statics are strict for $t \in (z_{lo}, z_{up})$. To see this, note that

$$F_{\xi \,|\, \xi \in [z^{lo}, z^{up})}(t) = \begin{cases} 0 & \text{for } t \leqslant z_{lo} \\[2mm] \dfrac{\Phi(t) - \Phi(z_{lo})}{\Phi(z_{up}) - \Phi(z_{lo})} & \text{for } t \in (z_{lo}, z_{up}) \\[2mm] 1 & \text{for } t \geqslant z_{up} \end{cases}$$

It is immediate that $F_{\xi \,|\, \xi \in [z^{lo}, z^{up})}(t)$ is increasing in t and decreasing in z_{up}, and strictly so when $t \in (z_{lo}, z_{up})$. Additionally, we have

$$\frac{\partial}{\partial z_{lo}} \frac{\Phi(t) - \Phi(z_{lo})}{\Phi(z_{up}) - \Phi(z_{lo})} = \frac{-\phi(z_{lo})\,(\Phi(z_{up}) - \Phi(t))}{(\Phi(z_{up}) - \Phi(z_{lo}))^2},$$

which is clearly negative for $t \in (z_{lo}, z_{up})$, which gives the desired result for z_{lo}.

Next, consider the function

$$f(\zeta) = \frac{\Phi(c) - \Phi(c - \zeta)}{1 - \Phi(c - \zeta)}.$$

Observe that $f(0) = 0$ and $\lim_{\zeta \to \infty} f(\zeta) = \Phi(c) > 1 - \alpha$. Additionally, the derivative in the previous paragraph (with $z_{up} = \infty$) implies that $\dfrac{d}{d\zeta} f(\zeta) > 0$ for $\zeta > 0$. It follows that there is a unique value $\zeta(c) > 0$ such that

$$f(c, \zeta(c)) = \frac{\Phi(c) - \Phi(c - \zeta(c))}{1 - \Phi(c - \zeta(c))} = 1 - \alpha,$$

which gives the first result.

Next, we claim that for $z_{lo} \in (-\infty, \infty)$ and $\zeta > 0$, $F_{\xi \,|\, \xi \in [z^{lo}, \infty)}(z_{lo} + \zeta)$ is increasing in z_{lo}. To see why this is the case, note that

$$F_{\xi \,|\, \xi \in [z^{lo}, \infty)}(z_{lo} + \zeta) = \frac{\Phi(z_{lo} + \zeta) - \Phi(z_{lo})}{1 - \Phi(z_{lo})}.$$

Differentiating with respect to z_{lo}, we obtain

$$\frac{\phi(z_{lo} + \zeta)(1 - \Phi(z_{lo})) - \phi(z_{lo})(1 - \Phi(z_{lo} + \zeta))}{[1 - \Phi(z_{lo})]^2},$$

which is greater than zero iff

$$\frac{\phi(z_{lo} + \zeta)}{1 - \Phi(z_{lo} + \zeta)} > \frac{\phi(z_{lo})}{1 - \Phi(z_{lo})},$$

which holds since the normal hazard function is strictly increasing.

Now, suppose that $\eta > \max\{c, z_{lo} + \zeta(c)\}$. Then

$$
\begin{aligned}
F_{\xi \mid \xi \in [z^{lo}, z^{up})}(\eta) &\geqslant F_{\xi \mid \xi \in [z^{lo}, \infty)}(\eta) \\
&> F_{\xi \mid \xi \in [z^{lo}, \infty)}(\max\{c, z_{lo} + \zeta(c)\}) \\
&= \max\{F_{\xi \mid \xi \in [z^{lo}, \infty)}(c), F_{\xi \mid \xi \in [z^{lo}, \infty)}(z_{lo} + \zeta(c))\}, \qquad (41)
\end{aligned}
$$

where the first inequality uses the fact that $F_{\xi \mid \xi \in [z^{lo}, z^{up})}(t)$ is decreasing in z_{up} and the second inequality uses the fact that $F_{\xi \mid \xi \in [z^{lo}, z^{up})}(t)$ is strictly increasing in t when $t \in (z_{lo}, \infty)$, and that $\max\{c, z_{lo} + \zeta(c)\} > z_{lo}$ since $\zeta(c) > 0$. The final equality again uses the fact that $F_{\xi \mid \xi \in [z^{lo}, z^{up})}(t)$ is increasing in t.

However, if $z_{lo} \leqslant c - \zeta(c)$, then

$$
F_{\xi \mid \xi \in [z^{lo}, \infty)}(c) \geqslant F_{\xi \mid \xi \in [c - \zeta(c), \infty)}(c) = 1 - \alpha,
$$

since we've shown that the expression on the left hand side is decreasing in z_{lo}. On the other hand, if $z_{lo} \geqslant c - \zeta(c)$, then

$$
F_{\xi \mid \xi \in [z^{lo}, \infty)}(z_{lo} + \zeta(c)) \geqslant F_{\xi \mid \xi \in [c - \zeta(c), \infty)}(c) = 1 - \alpha,
$$

since we've shown that $F_{\xi \mid \xi \in [z^{lo}, \infty)}(z_{lo} + \zeta)$ is increasing in z_{lo}. We have thus shown that the max on the right-hand side of (41) is at least $1 - \alpha$, which gives the desired result. □

Lemma C.15. *For any $t \in \mathbb{R}$, $\int_{-\infty}^{t} \Phi(x) dx$ is finite. In particular, $\int_{-\infty}^{t} \Phi(x) dx = t\Phi(t) + \phi(t)$.*

Proof. We have

$$
\begin{aligned}
\int_{-\infty}^{t} \Phi(x) dx &= \int_{-\infty}^{\infty} \mathbb{1}[x \leqslant t] \int_{-\infty}^{\infty} \mathbb{1}[s \leqslant x] \phi(s) \, ds \, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}[s \leqslant x \leqslant t] \phi(s) \, ds \, dx \\
&= \int_{-\infty}^{\infty} (t - s) \mathbb{1}[s \leqslant t] \phi(s) \, ds \\
&= t \int_{-\infty}^{\infty} \mathbb{1}[s \leqslant t] \phi(s) \, ds - \int_{-\infty}^{\infty} s \mathbb{1}[s \leqslant t] \phi(s) \, ds \\
&= t\Phi(t) - \Phi(t) \mathbb{E}\left[\xi \mid \xi \leqslant t, \xi \sim \mathcal{N}(0, 1)\right] = t\Phi(t) - \Phi(t) \frac{-\phi(t)}{\Phi(t)},
\end{aligned}
$$

where the last line uses the formula for the mean of a truncated normal distribution. Note that we exchange the order of integration via Fubini's theorem, which is valid since the integrand is weakly positive everywhere and thus equal to its absolute value, and we've

shown that the integral after switching the order is finite. □

Lemma C.16 (Lower bound on conditional test power). *Suppose $\tilde{Y}(x) \sim \mathcal{N}\left(\tilde{\mu}(x), \tilde{\Sigma}\right)$ for some $\tilde{\mu}(x)$ such that $\max_{\gamma \in V(\tilde{\Sigma})} \gamma' \tilde{\mu}(x) \geqslant x > 0$, where $V(\tilde{\Sigma})$ is the set of vertices of the dual feasible set, $F = \{\gamma : \gamma \geqslant 0, \gamma' \tilde{A}_{(\cdot, -1)} = 0, \gamma' \tilde{\sigma} = 1\}$, and $\tilde{\sigma}$ contains the square root of the diagonal elements of $\tilde{\Sigma}$. Then there exists a function $\underline{\rho}(x, \tilde{\Sigma})$, not depending on $\tilde{\mu}(x)$, such that $\mathbb{E}\left[\psi_\alpha^C(\tilde{Y}(x), \tilde{\Sigma})\right] \geqslant \underline{\rho}(x, \tilde{\Sigma})$ and $\lim_{n \to \infty} \underline{\rho}(x, \tilde{\Sigma}) = 1$.*

Proof. Recall that ψ_α^C is based on the solution to the dual problem, $\hat{\eta} = \max_{\gamma \in V(\tilde{\Sigma})} \gamma' \tilde{Y}$. Specifically, $\psi_\alpha^C(\tilde{Y}, \tilde{\Sigma}) = 1$ iff

$$F_{\xi \mid \xi \in [v_{\gamma*}^{lo}, v_{\gamma*}^{up}]}(\hat{\eta}; \gamma_*' \tilde{\Sigma} \gamma_*)) > 1 - \alpha,$$

where γ_* is an optimal solution to the dual ($\gamma_* \in \hat{V}$), and v^{lo}, v^{up} are functions of γ_*, $\tilde{\Sigma}$, and a sufficient statistic $S_{\gamma*}(\tilde{Y})$ that by construction is independent of $\gamma_*' \tilde{Y}$. (In this proof only, we make the dependence of v^{lo} and v^{up} on γ_* explicit in the notation.) If $\gamma_*' \Sigma \gamma_* \neq 0$, then using the standard formula for the CDF of a truncated normal distribution, we have that the conditional test rejects iff

$$\frac{\Phi(\hat{\eta}/\sigma_{\gamma*}) - \Phi(z_{\gamma*}^{lo})}{\Phi(z_{\gamma*}^{up}) - \Phi(z_{\gamma*}^{lo})} > 1 - \alpha, \tag{42}$$

where $\sigma_{\gamma*} = \sqrt{\gamma_*' \tilde{\Sigma} \gamma_*}$ and $z_{\gamma*}^{lo} = \hat{v}_{\gamma*}^{lo}/\sigma_{\gamma*}$, $z_{\gamma*}^{up} = \hat{v}_{\gamma*}^{up}/\sigma_{\gamma*}$. By Lemma C.14, for any $c > z_{1-\alpha}$, (42) holds whenever $\hat{\eta}/\sigma_{\gamma*} > \max\{c, z_{\gamma*}^{lo} + \zeta(c)\}$, where $\zeta(c)$ is the unique value that solves

$$\frac{\Phi(c) - \Phi(c - \zeta(c))}{1 - \Phi(c - \zeta(c))} = 1 - \alpha.$$

Thus, when $\sigma_{\gamma*} \neq 0$, $\psi_\alpha^C = 1$ whenever $\eta/\sigma_{\gamma*} > \max\{c, z_{\gamma*}^{lo} + \zeta(c)\}$, or equivalently, whenever $\hat{\eta} > \sigma_{\gamma*} c$ and $\hat{\eta}/\sigma_{\gamma*} - z_{\gamma*}^{lo} > \zeta(c)$. Additionally, if $\sigma_{\gamma*} = 0$, then $\psi_\alpha^C = 1$ whenever $\hat{\eta} > 0$.

Let $\bar{\sigma} = \max_{\gamma \in V(\tilde{\Sigma})} \sigma_\gamma$, which is finite since $V(\tilde{\Sigma})$ is finite. Then the preceding discussion implies that for any $c > \max\{z_{1-\alpha}, 0\}$, $\psi_\alpha^C = 1$ whenever

1) $\hat{\eta} > \bar{\sigma} c$, AND

2) $\exists \gamma_* \in \hat{V}$ such that either i) $\sigma_{\gamma*} = 0$, OR ii) $\sigma_{\gamma*} > 0$ and $\gamma_*' \tilde{Y}/\sigma_{\gamma*} - z_{\gamma*}^{lo} > \zeta(c)$,

where for the second part of condition 2) we use the fact that $\hat{\eta} = \gamma_*' \tilde{Y}$ when $\gamma_* \in \hat{V}$. Hence, $\psi_\alpha^C = 0$ only if either

A) $\hat{\eta} \leqslant \bar{\sigma} c$, OR

A-31

B) $\exists \gamma_* \in \hat{V}$ such that $\sigma_{\gamma_*} > 0$ and $\gamma_*' \tilde{Y} / \sigma_{\gamma_*} - z_{\gamma_*}^{lo} \leqslant \zeta(c)$.

Now, by assumption there exists some $\bar{\gamma} \in V(\tilde{\Sigma})$ such that $\bar{\gamma}' \tilde{\mu}(x) \geqslant x$. Since $\bar{\gamma}$ is feasible in the dual problem for $\hat{\eta}$, we see that $\hat{\eta}$ is lower bounded by $\bar{\gamma}' \tilde{Y}$, which is distributed $\mathcal{N}\left(\bar{\gamma}' \tilde{\mu}(x), \sigma_{\bar{\gamma}}^2\right)$. Thus, the probability that condition A) holds is bounded above by the probability that $\bar{\gamma}' \tilde{Y} \leqslant \bar{\sigma} c$. If $\sigma_{\bar{\gamma}} = 0$, then the probability condition A) holds is 0 so long as $c < \frac{x}{\bar{\sigma}}$. If $\sigma_{\bar{\gamma}} > 0$, then $\mathbb{P}\left(\bar{\gamma}' \tilde{Y} \leqslant \bar{\sigma} c\right) = \Phi\left(\dfrac{\bar{\sigma} c - \bar{\gamma}' \tilde{\mu}(x)}{\sigma_{\bar{\gamma}}}\right)$. If $\sigma_{\bar{\gamma}} > 0$, then the set $V^+(\tilde{\Sigma}) := \{\gamma \in V(\tilde{\Sigma}) \,;\, \sigma_\gamma > 0\}$ is non-empty. In this case, let $\underline{\sigma} = \min_{\gamma \in V^+} \sigma_\gamma$ and note that $\underline{\sigma} > 0$ since V^+ is finite. Then

$$\Phi\left(\frac{\bar{\sigma} c - \bar{\gamma}' \tilde{\mu}(x)}{\sigma_{\bar{\gamma}}}\right) \leqslant \Phi\left(\frac{\bar{\sigma} c - x}{\sigma_{\bar{\gamma}}}\right) \leqslant \Phi\left(\frac{\bar{\sigma}}{\underline{\sigma}} c - \frac{x}{\bar{\sigma}}\right),$$

where we use the fact that $\Phi(\cdot)$ is increasing, $c \geqslant 0$ and $\bar{\gamma}' \tilde{\mu}(x) \geqslant x > 0$. Thus, if $c < \frac{x}{\bar{\sigma}}$, we have that condition A) holds with probability bounded above by $\Phi\left(\frac{\bar{\sigma}}{\underline{\sigma}} c - \frac{x}{\bar{\sigma}}\right)$.

Now, the probability that condition B) holds is equal to

$$\mathbb{P}\left(\exists \gamma_* \in \hat{V} \text{ s.t. } \sigma_{\gamma_*} > 0 \text{ and } \gamma_*' \tilde{Y} / \sigma_{\gamma_*} - z_{\gamma_*}^{lo} \leqslant \zeta(c)\right) =$$
$$\mathbb{P}\left(\exists \gamma_* \in \hat{V} \text{ s.t. } \sigma_{\gamma_*} > 0 \text{ and } \left|\gamma_*' \tilde{Y} / \sigma_{\gamma_*} - z_{\gamma_*}^{lo}\right| \leqslant \zeta(c)\right) \leqslant$$
$$\mathbb{P}\left(\exists \gamma_+ \in V^+ \text{ s.t. } \left|\gamma_+' \tilde{Y} / \sigma_{\gamma_+} - z_{\gamma_+}^{lo}\right| \leqslant \zeta(c)\right) \leqslant$$
$$\sum_{\gamma_+ \in V^+} \mathbb{P}\left(|\gamma_+' \tilde{Y} / \sigma_{\gamma_+} - z_{\gamma_+}^{lo}| \leqslant \zeta(c)\right).$$

The equality above uses the fact that $\gamma_* \in \hat{V}$ implies that $\gamma_*' \tilde{Y} / \sigma_{\gamma_*} - z_{\gamma_*}^{lo} \geqslant 0$ since $\hat{\eta} \geqslant v^{lo}$ by construction; and the remaining inequalities follow from standard properties of probability. Next, observe that $\gamma_+' \tilde{Y} / \sigma_{\gamma_+}$ is normally distributed with variance 1 for every $\gamma_+ \in V^+(\tilde{\Sigma})$. Additionally, the random variable $z_{\gamma_+}^{lo}$ is by construction independent of $\gamma_+' \tilde{Y} / \sigma_{\gamma_+}$. However, for any variable ξ that is normally distributed with variance 1 and any variable Z independent of ξ,

$$\mathbb{P}_{(\xi, Z)}\left(|\xi - Z| \leqslant \zeta\right) = \mathbb{E}_Z\left[\mathbb{P}_{\xi|Z}\left(\xi \in [z - \zeta, z + \zeta] \mid Z = z\right)\right]$$
$$\leqslant \max_{v \in \mathbb{R}} \mathbb{P}_\xi\left(\xi \in [v - \zeta, v + \zeta]\right) = \Phi(\zeta) - \Phi(-\zeta),$$

where the first equality follows from iterated expectations, the inequality uses the fact that the distribution of ξ is independent of Z, and the final equality uses the fact that the normal distribution is single-peaked at its mean, so the maximal probability that a normal distribution with variance 1 falls in an interval of length 2ζ is $\Phi(\zeta) - \Phi(-\zeta)$. Additionally,

observe that $\Phi(\zeta) - \Phi(-\zeta) = \int_{-\zeta}^{\zeta} \phi(t)dt \leqslant 2\phi(0)\zeta$. It follows that for any constant $c > \max\{z_{1-\alpha}, 0\}$, the probability condition B) holds is bounded above by $\kappa\zeta(c)$, where we define the constant $\kappa = 2|V^+|\phi(0)$.

Since $\psi_\alpha^C = 0$ only if either condition A) or condition B) holds, the probability that $\psi_\alpha^C = 0$ is bounded above by $\Phi\left(\frac{\bar{\sigma}}{\sigma}c - \frac{x}{\bar{\sigma}}\right) + \kappa\zeta(c)$, for any $c \in [\max\{z_{1-\alpha}, 0\}, \frac{x}{\bar{\sigma}})$. Let $c(x) = c_0 \cdot x$ for $c_0 = \frac{1}{2}\frac{\sigma}{\bar{\sigma}^2}$. Note that $c(x) > \max\{z_{1-\alpha}, 0\}$ for $x > \max\{z_{1-\alpha}/c_0, 0\} =: x_{min}$. Note also that $c_0 = \frac{1}{2}\frac{\sigma}{\bar{\sigma}}\frac{1}{\bar{\sigma}} < \frac{1}{\bar{\sigma}}$, so $c(x) < \frac{1}{\bar{\sigma}}x$. For $x > x_{min}$, we then have that the probability $\psi_\alpha^C = 0$ is bounded above by $\Phi\left(-\frac{1}{2\bar{\sigma}}x\right) + \kappa\zeta(c_0 x)$.

Define $\underline{\rho}(x, \tilde{\Sigma}) = 1 - \Phi\left(-\frac{1}{2\bar{\sigma}}x\right) - \kappa\zeta(c_0 x)$ for $x > x_{min}$ and $\underline{\rho}(x, \tilde{\Sigma}) = 0$ otherwise. By construction, $\mathbb{E}\left[\psi_\alpha^C(\tilde{Y}, \tilde{\Sigma})\right] \geqslant \underline{\rho}(x, \tilde{\Sigma})$. Note that as $x \to \infty$, $\Phi\left(-\frac{1}{2\bar{\sigma}}x\right) \to 0$. To complete the proof that $\underline{\rho} \to 1$, we show that $\kappa\zeta(c) \to 0$ as $c \to \infty$. To show this, observe that for any $\epsilon > 0$, by L'Hospitale's rule,

$$
\begin{aligned}
\lim_{c \to \infty} \frac{\Phi(c) - \Phi(c - \epsilon)}{1 - \Phi(c - \epsilon)} &= \lim_{c \to \infty} \frac{\phi(c) - \phi(c - \epsilon)}{-\phi(c - \epsilon)} \\
&= 1 - \lim_{c \to \infty} \frac{\phi(c)}{\phi(c - \epsilon)} \\
&= 1 - \lim_{c \to \infty} \exp\left(-\frac{1}{2}(2c\epsilon - \epsilon^2)\right) = 1.
\end{aligned}
$$

Additionally, as shown in the proof to Lemma C.14, $\frac{\Phi(c) - \Phi(c - \zeta)}{1 - \Phi(c - \zeta)}$ is increasing in ζ. It is then immediate that $\limsup_{c \to \infty} \zeta(c) < \epsilon$ for all $\epsilon > 0$, and hence $\lim_{c \to \infty} \zeta(c) = 0$. $\qquad \square$

Lemma C.17 (Lower bound on hybrid power). *Suppose $\tilde{Y}(x) \sim \mathcal{N}\left(\tilde{\mu}, \tilde{\Sigma}\right)$ for some $\tilde{\mu}(x)$ such that $\max_{\gamma \in V(\tilde{\Sigma})} \gamma'\tilde{\mu}(x) \geqslant x > 0$, where $V(\tilde{\Sigma})$ is the set of vertices of the dual feasible set, $F = \{\gamma : \gamma \geqslant 0, \gamma'\tilde{A}_{(\cdot, -1)} = 0, \gamma'\tilde{\sigma} = 1\}$, and $\tilde{\sigma}$ contains the square root of the diagonal elements of $\tilde{\Sigma}$. Then there exists a function $\underline{\rho}(x, \tilde{\Sigma})$, not depending on $\tilde{\mu}(x)$, such that $\mathbb{E}\left[\psi_{\kappa,\alpha}^{C\text{-}FLCI}(\tilde{Y}(x), \tilde{\Sigma})\right] \geqslant \underline{\rho}(x, \tilde{\Sigma})$ and $\lim_{n \to \infty} \underline{\rho}(x, \tilde{\Sigma}) = 1$.*

Proof. The proof is nearly identical to that of Lemma C.16. In particular, by analogous argument we can show that the test $\psi_{\kappa,\alpha}^{C\text{-}FLCI} = 0$ only if A) $\hat{\eta} \leqslant \bar{\sigma}c_{\tilde{\alpha}}$, or B) $\exists \gamma_* \in \hat{V}$ such that $\sigma_{\gamma_*} > 0$ and $0 \leqslant \gamma_*'\tilde{Y}/\sigma_{\gamma_*} - z_{C\text{-}FLCI,\gamma_*}^{lo} \leqslant \zeta(c_{\tilde{\alpha}})$, where $z_{C\text{-}FLCI,\gamma_*}^{lo} = v_{C\text{-}FLCI,\gamma_*}^{lo}/\sigma_{\gamma_*}$, and $c_{\tilde{\alpha}}$ solves $\frac{\Phi(c_{\tilde{\alpha}}) - \Phi(c_{\tilde{\alpha}} - \zeta(c_{\tilde{\alpha}}))}{1 - \Phi(c_{\tilde{\alpha}} - \zeta(c_{\tilde{\alpha}}))} = 1 - \tilde{\alpha}$. Noting that $z_{C\text{-}FLCI,\gamma_*}^{lo}$ is independent of $\gamma_*'\tilde{Y}$, we can then obtain upper bounds on the probability that conditions A) or B) hold by an analogous argument to that in the proof to Lemma C.16. $\qquad \square$

Lemma C.18 (Lower bounds on power). *Let $\Delta = \{\delta : A\delta \leqslant d\}$, and let $\theta^{ub}(\Delta, \beta) := \sup \mathcal{S}(\Delta, \beta)$. Then there exists a function $\rho_{LB}(\cdot, \cdot)$ such that for any $\delta \in \Delta$, τ, Σ^*, and $x > 0$,*

$$\mathbb{E}_{(\delta,\tau,\Sigma^*)}\left[\psi_\alpha^C(\hat{\beta}_n; A, d, \theta^{ub}(\Delta, \delta + \tau) + x, \Sigma^*)\right] \geqslant \rho_{LB}(x, \Sigma^*),$$

and for any Σ^* fixed, $\rho_{LB}(x, \Sigma^*) \to 1$ as $x \to \infty$. Analogously, there exists a function $\tilde{\rho}_{LB}(\cdot, \cdot)$ such that for any $\delta \in \Delta$, τ, Σ^*, and $x > 0$,

$$\mathbb{E}_{(\delta,\tau,\Sigma^*)}\left[\psi_{\kappa,\alpha}^{C\text{-}FLCI}(\hat{\beta}_n; A, d, \theta^{ub}(\Delta, \delta + \tau) + x, \Sigma^*)\right] \geqslant \tilde{\rho}_{LB}(x, \Sigma^*),$$

and for any Σ^* fixed, $\tilde{\rho}_{LB}(x, \Sigma^*) \to 1$ as $x \to \infty$.

Proof. Lemma C.13 implies that there exists a scalar $c(\Sigma^*, A) > 0$ such that

$$c(\Sigma^*, A) \cdot x \leqslant \eta^*(x; \delta, \tau, \Sigma^*) := \min_{\eta, \tilde{\tau}} \eta \quad \text{s.t.} \quad \underbrace{A\beta - d - \tilde{A}_{(\cdot,1)}(\theta^{ub} + x)}_{:=\tilde{\mu}} - \tilde{A}_{(\cdot,-1)}\tilde{\tau} \leqslant \eta\tilde{\sigma},$$

where $\beta = \delta + \tau$. Reformulating the minimization above in terms of its dual, we have that $c(\Sigma^*, A) \cdot x \leqslant \max_{\gamma \in V(\tilde{\Sigma}^*)} \gamma'\tilde{\mu}$, where $V(\tilde{\Sigma}^*)$ is the set of vertices of $F = \{\gamma'\tilde{A}_{(\cdot,-1)} = 0, \gamma'\tilde{\sigma} = 0, \gamma \geqslant 0\}$. Next, recall that by definition, $\psi_\alpha^C(\hat{\beta}; A, d, \bar{\theta}, \Sigma^*) = \psi_\alpha^C(\tilde{Y}(\hat{\beta}, A, d, \bar{\theta}), A\Sigma^*A')$, where $\tilde{Y}(\hat{\beta}, A, d, \bar{\theta}) = A\hat{\beta} - d - \tilde{A}_{(\cdot,-1)}\bar{\theta}$. Observe that $\mathbb{E}_{(\delta,\tau,\Sigma^*)}\left[\tilde{Y}(\hat{\beta}, A, d, \bar{\theta})\right] = \tilde{\mu}$. Lemma C.16 then implies that there exists a function $\rho(\cdot, \cdot)$ such that

$$\mathbb{E}_{(\delta,\tau,\Sigma^*)}\left[\psi_\alpha^C(\hat{\beta}_n; A, d, \theta^{ub}(\Delta, \delta + \tau) + x, \Sigma^*)\right] \geqslant \rho(c(\Sigma^*, A) \cdot x, A\Sigma^*A'),$$

and $\rho(\tilde{x}, A\Sigma^*A') \to 1$ as $\tilde{x} \to \infty$. The first desired result then follows by defining $\rho_{LB}(x, \Sigma^*) := \rho(c(\Sigma^*, A) \cdot x, A\Sigma^*A')$. The second desired result follows from an analogous argument, appealing to Lemma C.17 instead of Lemma C.16. $\qquad\square$

Lemma C.19 (Bounds for worst-case bias). *For any (a, v), $\bar{b}(a, v) \geqslant \frac{1}{2}\sup_{\delta_{pre} \in \Delta_{pre}} LID(\Delta, \delta_{pre})$.*

Proof. Since $\beta = \delta + \tau$, we can write the bias of the affine estimator $a + v'\hat{\beta}$ as $b = a + v'\delta + (v_{post} - l)'\tau_{post}$. Since τ_{post} is unrestricted in the maximization in (11), we see that the worst-case bias will be infinite if $v_{post} \neq l$ and the lemma holds trivially. We can thus restrict attention to affine estimators with $v_{post} = l$, in which case the worst-case bias reduces to

$$\bar{b}(a, v) = \sup_{\delta \in \Delta} |a + v'\delta| = \sup_{\delta \in \Delta} |a + v_{pre}'\delta_{pre} + l'\delta_{post}|. \tag{43}$$

Now, pick any $\delta_{pre}^* \in \Delta_{pre}$. First, suppose that the minimum $\left(\min_\delta l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta_{pre}^*\right)$

and the equivalent maximum $\left(\max_\delta l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta^*_{pre}\right)$ are finite. Let δ^{min} and δ^{max} be the associated solutions. By construction, $\delta^{max}_{pre} = \delta^{min}_{pre} = \delta^*_{pre}$. For any v_{pre}, we apply the triangle inequality to show that

$$\left|a + v'_{pre}\delta^{max}_{pre} + l'\delta^{max}_{post}\right| + \left|a + v'_{pre}\delta^{min}_{pre} + l'\delta^{min}_{post}\right| \geq \left|\left(a + v'_{pre}\delta^{max}_{pre} + l'\delta^{max}_{post}\right) - \left(a + v'_{pre}\delta^{min}_{pre} + l'\delta^{min}_{post}\right)\right|$$
$$= \left|l'\delta^{max}_{post} - l'\delta^{min}_{post}\right| = LID(\Delta, \delta^*_{pre}).$$

Note that for any $x_1, x_2 \geq 0$, $\max\{x_1, x_2\} \geq \frac{1}{2}(x_1 + x_2)$. It then follows from the previous display that

$$\max\{\left|a + v'_{pre}\delta^{max}_{pre} + l'\delta^{max}_{post}\right|, \left|a + v'_{pre}\delta^{min}_{pre} + l'\delta^{min}_{post}\right|\} \geq \frac{1}{2}LID(\Delta, \delta^*_{pre}).$$

Since δ^{max}_{pre} and δ^{min}_{pre} are feasible in the maximization (43), we see that $\bar{b} \geq \frac{1}{2}LID(\Delta, \delta^*_{pre})$, as needed. To complete the proof, now suppose without loss of generality that

$$\left(\max_\delta l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta^*_{pre}\right) = \infty.$$

Then, we can replay the argument above replacing δ^{max} with a sequence of values $\{\delta_j\}$ such that $l'\delta_j$ diverges, which gives that $\bar{b}$ is infinite and the result follows. $\square$

Lemma C.20. *Suppose Δ is convex. Suppose there exists $\delta \in \Delta$ such that $LID(\Delta, \delta_{pre}) = \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\Delta, \tilde{\delta}_{pre}) < \infty$. Then there exists (a, v) such that $\bar{b}(a, v) = \frac{1}{2} \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\Delta, \tilde{\delta}_{pre})$. Additionally, for any τ and Σ_n, $\mathbb{E}_{(\delta, \tau, \Sigma_n)}\left[a + v'\hat{\beta}_n\right] = \frac{1}{2}(\theta^{ub} + \theta^{lb})$, where θ^{ub} and θ^{lb} are the upper and lower bounds of the identified set $\mathcal{S}(\Delta, \delta + \tau)$.*

Proof. Let $b^{max}(\delta^*_{pre}) := \left(\max_\delta l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta^*_{pre}\right)$, where we define $b^{max} = -\infty$ if $\delta^*_{pre} \notin \Delta_{pre}$. Likewise, define $b^{min}(\delta^*_{pre}) := \left(\min_\delta l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta^*_{pre}\right)$, where we define $b^{min} = \infty$ if $\delta^*_{pre} \notin \Delta_{pre}$. Note that Δ convex implies that b^{max} is concave and b^{min} is convex. Thus, $-LID(\delta^*_{pre}) = b^{min}(\delta^*_{pre}) - b^{max}(\delta^*_{pre})$ is convex (where we define $LID(\delta^*_{pre}) = -\infty$ if $\delta^*_{pre} \notin \Delta_{pre}$). The domain of $-LID(\delta^*_{pre})$ (i.e. the set of values for which it is finite) is Δ_{pre}, since it is infinite for $\delta^*_{pre} \notin \Delta_{pre}$ by construction, and by assumption, $LID(\delta^*_{pre})$ is finite for all $\delta^*_{pre} \in \Delta_{pre}$. Since Δ is assumed to be convex, it is easy to verify that Δ_{pre} is a non-empty convex set, and thus has non-empty relative interior, so the relative interior of the domain of $-LID$ is non-empty.[32] It follows from Theorem 8.2 in Mau Nam (2019) that $\partial(-LID) = \partial(-b^{max}) + \partial(b^{min})$ where for a convex function f, ∂f is the subdifferential $\partial f(\bar{x}) := \{v : f(\bar{x}) + v'(x - \bar{x}) \leq f(x), \forall x\}$ and $\partial(-b^{max}) + \partial(b^{min})$ is the

---

[32]The relative interior of a set is the interior of the set relative to its affine hull. See, e.g., Mau Nam (2019), Chapter 5.

Minkowski sum of the two subdifferentials.

Additionally, if $LID(\delta_{pre}) = \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\tilde{\delta}_{pre})$, then $-LID(\delta_{pre}) = \inf_{\tilde{\delta}_{pre} \in \Delta_{pre}} -LID(\tilde{\delta}_{pre})$. Thus, standard results in convex analysis (see, e.g., Theorem 16.2 in Mau Nam (2019)) give that $0 \in \partial(-LID)(\delta_{pre}) + N(\Delta; \delta_{pre})$, where $N(\Delta; \delta_{pre}) = \{v_{pre} \ : \ v'_{pre}(\tilde{\delta}_{pre} - \delta_{pre}) \leqslant 0, \forall \tilde{\delta}_{pre} \in \Delta_{pre}\}$ is the normal cone to Δ_{pre} at δ_{pre}. Hence, there exist vectors $\bar{v}_{min}, \bar{v}_{max}$ such that for all $\tilde{\delta}_{pre} \in \Delta_{pre}$,

$$b^{min}(\delta_{pre}) + \bar{v}'_{min}(\tilde{\delta}_{pre} - \delta_{pre}) \leqslant b^{min}(\tilde{\delta}_{pre}) \tag{44}$$

$$- b^{max}(\delta_{pre}) + \bar{v}'_{max}(\tilde{\delta}_{pre} - \delta_{pre}) \leqslant -b^{max}(\tilde{\delta}_{pre}) \tag{45}$$

$$- (\bar{v}_{min} + \bar{v}_{max})'(\tilde{\delta}_{pre} - \delta_{pre}) \leqslant 0. \tag{46}$$

The inequalities (45) and (46) together imply that for all $\tilde{\delta}_{pre} \in \Delta_{pre}$,

$$b^{max}(\delta_{pre}) + \bar{v}'_{min}(\tilde{\delta}_{pre} - \delta_{pre}) \geqslant b^{max}(\tilde{\delta}_{pre}). \tag{47}$$

Now, let v be the vector such that $v_{post} = l$ and $v_{pre} = -\bar{v}_{min}$. Observe that

$$
\begin{aligned}
\max_{\tilde{\delta} \in \Delta} a + v'_{pre}\tilde{\delta}_{pre} + l'\tilde{\delta}_{post} &= \max_{\tilde{\delta}_{pre} \in \Delta_{pre}} \left(a + v'_{pre}\tilde{\delta}_{pre} + \max_{\tilde{\delta} \in \Delta, \tilde{\delta}_{pre} = \tilde{\delta}_{pre}} l'\tilde{\delta}_{post} \right) \\
&= \max_{\tilde{\delta}_{pre} \in \Delta_{pre}} a + v'_{pre}\tilde{\delta}_{pre} + b^{max}(\tilde{\delta}_{pre}) \\
&\leqslant a + v'_{pre}\delta_{pre} + b^{max}(\delta_{pre}), \tag{48}
\end{aligned}
$$

where the first equality nests the maximization, the second equality uses the definition of b^{max}, and the inequality follows from (47). An analogous argument using (44) yields that

$$
\begin{aligned}
\min_{\tilde{\delta} \in \Delta} a + v'_{pre}\tilde{\delta}_{pre} + l'\tilde{\delta}_{post} &= \min_{\tilde{\delta}_{pre} \in \Delta_{pre}} a + v'_{pre}\tilde{\delta}_{pre} + b^{min}(\tilde{\delta}_{pre}) \\
&\geqslant a + v'_{pre}\delta_{pre} + b^{min}(\delta_{pre}). \tag{49}
\end{aligned}
$$

Now, it is apparent from equation (43) that

$$\bar{b}(a, v) = \max \left\{ \left| \max_{\tilde{\delta} \in \Delta} a + v'_{pre}\tilde{\delta}_{pre} + l'\tilde{\delta}_{post} \right|, \left| \min_{\tilde{\delta} \in \Delta} a + v'_{pre}\tilde{\delta}_{pre} + l'\tilde{\delta}_{post} \right| \right\}, \tag{50}$$

which is bounded above by $\max\left\{ \left| a + v'_{pre}\delta_{pre} + b^{max}(\delta_{pre}) \right|, \left| a + v'_{pre}\delta_{pre} + b^{min}(\delta_{pre}) \right| \right\}$ from the results above. Setting $a = -v'_{pre}\delta_{pre} - \frac{1}{2}(b^{max}(\delta_{pre}) + b^{min}(\delta_{pre}))$, the upper bound in the previous display reduces to $\frac{1}{2}(b^{max}(\delta_{pre}) - b^{min}(\delta_{pre}))$. Since $LID(\Delta, \delta_{pre}) = b^{max}(\delta_{pre}) - b^{min}(\delta_{pre})$ and $LID(\Delta, \delta_{pre}) = \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\Delta, \tilde{\delta}_{pre})$ by assumption, it is then immediate

that $\bar{b} \leqslant \frac{1}{2}\sup_{\tilde{\delta}_{pre}\in\Delta_{pre}} LID(\Delta,\tilde{\delta}_{pre})$. The inequality in the opposite direction follows from Lemma C.19.

Finally, substituting in the definition of a and v above and simplifying, we see that $\mathbb{E}_{(\delta,\tau,\Sigma_n)}\left[a + v'\hat{\beta}_n\right] = l'\beta_{post} - \frac{1}{2}(b^{max}(\delta_{pre}) + b^{min}(\delta_{pre}))$, which from (6) and (7) we see is the midpoint of the identified set. □

Lemma C.21. *Let χ_α be the $1-\alpha$ quantile of the $|\mathcal{N}(b,\sigma^2)|$ distribution for $b \geqslant 0$. Then $b + \sigma z_{1-\alpha} \leqslant \chi_\alpha \leqslant b + \sigma z_{1-\alpha/2}$.*

Proof. Since $|\xi| \geqslant \xi$, we have that $q_{1-\alpha}(|\xi| \,|\, \xi \sim \mathcal{N}(b,\sigma^2)) \geqslant q_{1-\alpha}(\xi \,|\, \xi \sim \mathcal{N}(b,\sigma^2)) = b + \sigma z_{1-\alpha}$, which yields the first inequality. For the second inequality, observe that

$$
\begin{aligned}
q_{1-\alpha}(|\xi| \,|\, \xi \sim \mathcal{N}(b,\sigma^2)) &= q_{1-\alpha}(|\xi + b| \,|\, \xi \sim \mathcal{N}(0,\sigma^2)) \\
&\leqslant b + q_{1-\alpha}(|\xi| \,|\, \xi \sim \mathcal{N}(0,\sigma^2)) = b + \sigma z_{1-\alpha/2}
\end{aligned}
$$

where the first inequality uses the triangle inequality, and the final equality uses the fact that a mean-zero normal distribution is symmetric about 0. □

Lemma C.22. *Suppose the conditions of Proposition A.1 hold. Then there is a unique pair $(\bar{a},\bar{v})$ such that $\bar{b}(\bar{a},\bar{v}) = \frac{1}{2}\sup_{\tilde{\delta}_{pre}\in\Delta_{pre}} LID(\Delta,\tilde{\delta}_{pre}) =: \bar{b}_{min}$. Additionally, $\sqrt{\bar{v}'A\Sigma^*A'\bar{v}} = 1/c^*$, for the same constant c^* as in Proposition 4.2.*

Proof. Existence of an $(\bar{a},\bar{v})$ satisfying $\bar{b}(\bar{a},\bar{v}) = \bar{b}_{min}$ follows from Lemma C.20, so to establish the existence of a unique solution it suffices to establish uniqueness. In the proof to Lemma C.7, we showed that $b^{min}(\delta_{A,pre})$ is equivalent to the problem (35). Assumption 4 implies that there is a solution δ_{post}^{**} to the optimization (35) such that $A_{(B,post)}$ has rank $|B|$, where B indexes the binding moments. The solution δ_{post}^{**} to the problem (35) is thus non-degenerate. It follows that in a neighborhood of $\delta_{pre,A}$, $b^{min}(\delta_{pre}) = b^{min}(\delta_{A,pre}) + \bar{\gamma}'A_{(\cdot,pre)}(\delta_{pre} - \delta_{A,pre})$, where $\bar{\gamma}$ is a solution to the dual problem (see, e.g., Section 10.4 of Schrijver (1986)). By the complementary slackness conditions, $\bar{\gamma}_{-B} = 0$. Moreover, we showed in the proof to Lemma C.7 that $\bar{\gamma}_B$ is the unique vector that satisfies $\bar{\gamma}_B'\tilde{A}_{(B,-1)} = 0, \bar{\gamma}_B'\tilde{A}_{(B,1)} = 1$.

Next, combining the expression for $\bar{b}$ in (50) along with the equalities in (48) and (49) in the proof to Lemma C.20, we see that for any (a,v),

$$
\bar{b}(a,v) = \max\left\{\left|\max_{\tilde{\delta}_{pre}\in\Delta_{pre}} a + v'_{pre}\tilde{\delta}_{pre} + b^{max}(\tilde{\delta}_{pre})\right|, \left|\min_{\tilde{\delta}_{pre}\in\Delta_{pre}} a + v'_{pre}\tilde{\delta}_{pre} + b^{min}(\tilde{\delta}_{pre})\right|\right\}.
\tag{51}
$$

This implies that if $(\bar{a}, \bar{v})$ are such that $\bar{b}(\bar{a}, \bar{v}) = \bar{b}_{min}$, then for all $\tilde{\delta}_{pre} \in \Delta_{pre}$,

$$\bar{b}_{min} \geqslant \max\left\{\left|\bar{a} + \bar{v}'_{pre}\tilde{\delta}_{pre} + b^{max}(\tilde{\delta}_{pre})\right|, \left|\bar{a} + \bar{v}'_{pre}\tilde{\delta}_{pre} + b^{min}(\tilde{\delta}_{pre})\right|\right\}. \tag{52}$$

Now, note that by the triangle inequality, for any scalars x_1, x_2, x_3 with $x_2 \geqslant x_3$, $\max\{|x_1 + x_2|, |x_1 + x_3|\} \geqslant \frac{1}{2}|x_2 - x_3|$, with equality if and only if $x_1 + x_3 = -(x_1 + x_2)$. Further, recall that $b^{max}(\delta_{A,pre}) - b^{min}(\delta_{A,pre}) = LID(\Delta, \delta_{A,pre}) = 2\bar{b}_{min}$. It follows from these two facts along with the expression in the previous display that

$$\bar{b}_{min} = \bar{a} + \bar{v}'_{pre}\delta_{A,pre} + b^{max}(\delta_{A,pre}) = -\left(\bar{a} + \bar{v}'_{pre}\delta_{A,pre} + b^{min}(\delta_{A,pre})\right). \tag{53}$$

Displays (52) and (53) imply that for all $\tilde{\delta}_{pre} \in \Delta_{pre}$,

$$\bar{v}'_{pre}(\tilde{\delta}_{pre} - \delta_{A,pre}) + b^{min}(\tilde{\delta}_{pre}) - b^{min}(\delta_{A,pre}) \geqslant 0$$

which using the local linearization derived above implies that

$$(\bar{v}'_{pre} + \bar{\gamma}'A_{(\cdot,pre)}))(\tilde{\delta}_{pre} - \delta_{A,pre}) \geqslant 0$$

for all $\tilde{\delta}_{pre} \in \Delta_{pre}$ in a sufficiently small neighborhood of $\delta_{A,pre}$. However, Assumption 4 implies that $\delta_{A,pre}$ is in the interior of Δ_{pre}, and so the equality in the previous display can hold for all such $\tilde{\delta}_{pre}$ only if $\bar{v}'_{pre} = -\bar{\gamma}'A_{(\cdot,pre)}$. We argued in the proof to Lemma C.19 that $\bar{v}_{post}$ must equal l, so we have shown that there is a unique value of $\bar{v}$. Further, (53) uniquely pins downs $\bar{a}$ in terms of $\bar{v}$, and so the pair $(\bar{a}, \bar{v})$ is unique, as claimed.

Finally, recall from the proof to Lemma C.7 that $-\bar{\gamma}'A_{(\cdot,post)} = l'$. Hence $\bar{v}' = (-\bar{\gamma}'A_{(\cdot,pre)}, -\bar{\gamma}'A_{(\cdot,post)}) = -\bar{\gamma}'A$ and thus $\bar{v}'\Sigma^*\bar{v} = \bar{\gamma}'A\Sigma^*A'\bar{\gamma}$. Since $\bar{\gamma}_{-B} = 0$ and $\bar{\gamma}'_B\tilde{A}_{(B,1)} = 1$, we see that $1/\sqrt{\bar{v}'\Sigma^*\bar{v}}$ corresponds with the formula for c^* given in Lemma C.8.

$\square$

Lemma C.23. *Suppose the conditions of Proposition A.1 hold. Then $\dfrac{\sigma_{v_n,n}}{\sigma_{\bar{v},n}} \to 1$, where the optimal FLCI is based on the affine estimator $a_n + v'_n\hat{\beta}_n$ and $\bar{v}$ is the unique value such that $\bar{b}(\bar{a}, \bar{v}) = \bar{b}_{min}$.*

Proof. It suffices to show that $\dfrac{\sqrt{n}\sigma_{v_n,n}}{\sqrt{n}\sigma_{\bar{v},n}} \to 1$. Note that $\sqrt{n}\sigma_{\bar{v},n} = \sigma_{\bar{v},1} = \sqrt{\bar{v}'\Sigma^*\bar{v}}$. By assumption, Σ^* is positive definite, and we showed in the proof to Lemma C.19 that $\bar{v}_{post} = l$, so $\bar{v} \neq 0$. Hence $\sigma_{\bar{v},1} > 0$. Next, observe that $\sqrt{n}\sigma_{v_n,n} = \sqrt{v'_n\Sigma^*v_n}$. It thus suffices to show that $v_n \to \bar{v}$, since then both the numerator and denominator converge to the same non-zero limit. To do this, we will show that every subsequence of v_n has a convergent subsequence. Consider

A-38

a subsequence v_{n_m}. We argued in the proof to Proposition 3.2 that $\sigma_{v_n,n} \leqslant \sigma_{\bar{v},n}$, which implies that $\sqrt{v_n' \Sigma^* v_n} \leqslant \sqrt{\bar{v}' \Sigma^* \bar{v}}$. Thus, v_n is bounded in the Mahalanobis norm using Σ^*, which implies that v_n is bounded in the standard euclidean norm since Σ^* positive definite. It follows that v_{n_m} has a convergent subsequence, $v_{n_m,1} \to v^*$. We argued in the proof to Proposition 3.2 that $\bar{b}(a_n, v_n) \to \bar{b}_{min}$. This implies, however, that $a_{n_m,1}$ is bounded. To see why this is the case, note that if there is a divergent subsequence $a_{n_m,2}$, then for any fixed $\tilde{\delta}_{pre} \in \Delta_{pre}$, $|a_{n_m,2} + v_{n_m,2,pre}' \tilde{\delta}_{pre} + b^{max}(\tilde{\delta}_{pre})|$ diverges since $v_{n_m,2,pre} \to v_{pre}^*$. Equation (51) then implies that $\bar{b}(a_{n_m,2}, v_{n_m,2})$ diverges, which is a contradiction. Thus $a_{n_m,1}$ is bounded, and so we can extract a further subsequence such that $(a_{n_m,2}, v_{n_m,2}) \to (a^*, v^*)$. For ease of notation, suppose without loss of generality that these convergences hold for the original subsequence n_m. To complete the proof, we will show that $\bar{b}(a^*, v^*) = \bar{b}_{min}$, which then implies that $v^* = \bar{v}$ by Lemma C.22. To show this, note that (43) together with the identity $|x| = \max\{x, -x\}$ imply that $\bar{b}(a, v) = \max\{(\max_\delta a + v'\delta \text{ s.t. } A\delta \leqslant d), (\max_\delta -a - v'\delta \text{ s.t. } A\delta \leqslant d)\}$. Consider the first inner maximization, and let δ_{n_m} denote the optimal value using $v = v_{n_m}$, and δ^* the optimal value using $v = v^*$. Since δ^* is feasible in the optimization using v_{n_m}, we have $a_{n_m} + v_{n_m}' \delta^* \leqslant a_{n_m} + v_{n_m}' \delta_{n_m}$. Taking limits on both sides of this inequality implies that

$$a^* + (v^*)'\delta^* = \left(\max_\delta a^* + (v^*)'\delta \text{ s.t. } A\delta \leqslant d\right) \leqslant \liminf_{m \to \infty} \left(\max_\delta a_{n_m} + v_{n_m}' \delta \text{ s.t. } A\delta \leqslant d\right).$$

Applying a similar argument to the second inner maximization, it follows that

$$\bar{b}(a^*, v^*) \leqslant \lim_{m \to \infty} \bar{b}(a_{n_m}, v_{n_m}) = \bar{b}_{min}.$$

But $\bar{b}(a^*, v^*) \geqslant \bar{b}_{min}$ by Lemma C.19, which gives the desired equality. $\square$

Lemma C.24. $\lim_{x \to \infty}(cv_\alpha(x) - (z_{1-\alpha} + x)) = 0$.

Proof. $cv_\alpha(x)$ solves $\Phi(cv_\alpha(x) - x) - \Phi(-cv_\alpha(x) - x) = 1 - \alpha$. By Lemma C.21, $cv_\alpha(x) \geqslant x + z_{1-\alpha}$, which diverges as $x \to \infty$. Thus, $\Phi(-cv_\alpha(x) - x)$ converges to 0 and $\Phi(cv_\alpha(x) - x) \to 1 - \alpha$, together implying $cv_\alpha(x) - x \to z_{1-\alpha}$. $\square$

Lemma C.25. *Suppose that Assumption 4 holds at $\delta_{A,pre}$. Then $LID(\Delta, \delta_{A,pre}) > 0$.*

Proof. From (6) and (7), we see that that $LID(\Delta, \delta_{A,pre}) = 0$ if and only if $b^{max}(\delta_{A,pre}) = b^{min}(\delta_{A,pre})$, where $b^{min}(\delta_{pre,A}) := (\min_\delta l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta_{pre,A})$, and b^{max} is defined analogously. In the proof to Lemma C.7, we showed that b^{min} is equivalent to the problem (35). Assumption 4 implies that there is a solution δ_{post}^{**} such that

$$A_{(B,post)}\delta_{post}^{**} = d_B - A_{(B,pre)}\delta_{A,pre} \text{ and } A_{(-B,post)}\delta_{post}^{**} < d_{-B} - A_{(-B,pre)}\delta_{A,pre},$$

where $A_{B,post}$ has rank $|B|$. Observe that if $b^{min}(\delta_{A,pre}) = b^{max}(\delta_{A,pre})$, then it must be that $l'\delta_{pre}^{**} = l'\tilde{\delta}_{pre}$ for any $\tilde{\delta}_{pre}$ that is feasible in the problem (35). It thus suffices to construct a feasible value $\tilde{\delta}_{pre}$ such that $l'\tilde{\delta}_{pre} \neq l'\delta_{pre}^{**}$. Since $A_{(B,post)}$ has rank $|B|$, its image is $\mathbb{R}^{|B|}$, so there exists $\bar{\delta}_{post}$ such that $A_{(B,post)}\bar{\delta}_{post} = -\iota$, for ι the vector of ones. Thus, for any $\epsilon_1 > 0$, we have that $A_{(B,post)}(\delta_{post}^{**} + \epsilon_1\bar{\delta}_{post}) < d_B - A_{(B,pre)}\delta_{A,pre}$. However, since the moments $-B$ are slack at $\delta_{A,pre}$, for ϵ_1 sufficiently small, we also have $A_{(-B,post)}(\delta_{post}^{**} + \epsilon_1\bar{\delta}_{post}) < d_{-B} - A_{(-B,pre)}\delta_{A,pre}$. If $l'\bar{\delta}_{post} \neq 0$, then we are done. If $l'\bar{\delta}_{post} = 0$, then since all of the moments are slack at $\delta_{post}^{**} + \epsilon_1\bar{\delta}_{post}$, for $\epsilon_2 > 0$ sufficiently small, $\tilde{\delta}_{post} = \delta_{post}^{**} + \epsilon_1\bar{\delta}_{post} + \epsilon_2 l$ is also feasible, and by construction $l'(\tilde{\delta}_{post} - \delta_{post}^{**}) = \epsilon_2 l'l > 0$. $\qquad\square$

Lemma C.26. *Suppose Δ is convex and centrosymmetric, and δ_A is such that $\delta \in \Delta$ implies $\delta - \delta_A \in \Delta$. Then δ_A satisfies Assumption 3.*

Proof. Recall from the proof to Lemma C.20 that for any $\delta_{pre}^* \in \Delta_{pre}$, $LID(\Delta, \delta_{pre}^*) = b^{max}(\delta_{pre}^*) - b^{min}(\delta_{pre}^*)$, where the functions b^{min} and $-b^{max}$ are convex. Observe that

$$
\begin{aligned}
b^{min}(\delta_{pre}^*) &= \left(\min_{\delta} l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta_{pre}^* \right) \\
&= -\left(\max_{\delta} l'(-\delta_{post}), \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta_{pre}^* \right) \\
&= -\left(\max_{\delta} l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = -\delta_{pre}^* \right) = -b^{max}(-\delta_{pre}^*),
\end{aligned}
$$

where the third equality uses the fact that Δ is centrosymmetric. Hence, $-LID(\Delta, \delta_{pre}^*) = -b^{max}(\delta_{pre}^*) - b^{max}(-\delta_{pre}^*)$. It follows from the subdifferential sum and chain rules for convex functions (e.g., Theorems 8.2 and 9.3 in Mau Nam (2019)) that $\partial - LID(\Delta, \delta_{pre}^*) = \partial(-b^{max})(\delta_{pre}^*) + (-\partial(-b^{max})(-\delta_{pre}^*))$, for $+$ the Minkowski sum. It is then immediate that $0 \in \partial(-LID(\Delta, 0))$, and hence $0 \in \arg\min_{\tilde{\delta}_{pre}\in\Delta_{pre}} -LID(\Delta, \tilde{\delta}_{pre})$. This implies that $LID(\Delta, 0) = \sup_{\tilde{\delta}_{pre}\in\Delta_{pre}} LID(\Delta, \tilde{\delta}_{pre})$.

To complete the proof, we show that $LID(\Delta, \delta_{A,pre}) \geqslant LID(\Delta, 0)$. We first claim that for any $\delta \in \Delta$, we also have $\delta + \delta_A \in \Delta$. Indeed, by centrosymmetry, $-\delta \in \Delta$. By assumption, this implies that $-\delta - \delta_A \in \Delta$. Applying centrosymmetry again, we see that $\delta + \delta_A \in \Delta$, as desired. Next, suppose that δ^{max} is optimal in the maximization $b^{max}(0) = (\max_\delta l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = 0)$. Then $\delta^{max} + \delta_A$ is feasible in the optimization $(\max_\delta l'\delta_{post}, \text{ s.t. } \delta \in \Delta, \delta_{pre} = \delta_{A,pre})$, and thus $b^{max}(\delta_{A,pre}) \geqslant b^{max}(0) + l'\delta_{A,post}$. By analogous argument, we can obtain that $b^{min}(\delta_{A,pre}) \leqslant b^{min}(0) + l'\delta_{A,post}$. It follows that $LID(\Delta, \delta_{A,pre}) = b^{max}(\delta_{A,pre}) - b^{min}(\delta_{A,pre}) \geqslant b^{max}(0) - b^{min}(0) = LID(\Delta, 0)$, as needed. $\qquad\square$

Lemma C.27. *Fix Σ^* positive definite, $\delta_A \in \Delta$, and τ_A. Suppose Assumption 4 holds at δ_A, and let $B = B(\delta^{**})$. Let $\hat{V}_n$ denote the set of optimal vertices used in $\psi_\alpha^C(\hat{\beta}_n; A, \sqrt{n}d, \theta_n^{ub} + x, \Sigma^*)$, where $\theta_n^{ub} = \sup S(\Delta_n, \sqrt{n}(\delta_A + \tau_A))$, $\Delta_n = \sqrt{n}\Delta$. Then*

A-40

$$\lim_{n\to\infty} \mathbb{P}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left(\hat{V}_n = \{c\bar{\gamma}\}\right) = 1,$$

where $c > 0$ and $\bar{\gamma}$ is the vector such that $\bar{\gamma}_{-B} = 0$ and $\bar{\gamma}_B$ is the unique vector such that $\bar{\gamma}'\tilde{A}_{(\cdot,-1)} = 0$, $\bar{\gamma} \geqslant 0$, $||\bar{\gamma}|| = 1$.

Proof. Observe that $\hat{V}_n = \arg\min_{\gamma \in V(\Sigma^*)} \gamma'\tilde{Y}_n$, where $\tilde{Y}_n = A\hat{\beta}_n - \sqrt{n}d - \tilde{A}_{(\cdot,1)}(\theta_n^{ub} + x)$. Since all vertices $\gamma \in V(\Sigma^*)$ satisfy $\gamma'\tilde{A}_{(\cdot,-1)} = 0$ by definition, we have that $\hat{V}_n = \arg\min_{\gamma \in V(\Sigma^*)} \gamma'\breve{Y}_n$, for $\breve{Y}_n = \tilde{Y}_n - \tilde{A}_{(\cdot,-1)}(\sqrt{n}\tilde{\tau}_1^{ub})$ and $\tilde{\tau}_1^{ub}$ the vector constructed in the proof to Lemma C.8. However, we showed in the proof to Lemma C.8 that $\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left[\breve{Y}_{n,B}\right] = -\tilde{A}_{(B,1)}x$ and $\mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left[\breve{Y}_{n,-B}\right] \to -\infty$ as $n \to \infty$. Lemmas E.1 and F.7 in the supplementary material together imply that there is a unique vector $\gamma^* \in V(\Sigma^*)$ such that $\gamma^*_{-B} = 0$, which satisfies $\gamma^*_B = c\bar{\gamma}$ for $c > 0$. By definition, $\gamma \geqslant 0$ for all $\gamma \in V(\Sigma^*)$, and thus γ_{-B} has at least one strictly positive element for all $\gamma \in V(\Sigma^*)\backslash\{\gamma^*\}$. It follows that

$$\lim_{n\to\infty} \mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left[\gamma^{*\prime}\breve{Y}_n\right] = -\gamma_B^{*\prime}\tilde{A}_{(B,1)}x \text{ and } \lim_{n\to\infty} \mathbb{E}_{(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)}\left[\gamma'\breve{Y}_n\right] = -\infty, \;\; \forall \gamma \in V(\Sigma^*)\backslash\{\gamma^*\}.$$

Let P_n denote the sequence of data-generating processes characterized by $(\sqrt{n}\delta_A, \sqrt{n}\tau_A, \Sigma^*)$. Note that that for all n, $(\gamma^* - \gamma)'\tilde{Y}_n$ is normally distributed with variance $(\gamma^* - \gamma)'\Sigma^*(\gamma^* - \gamma)$ under P_n. This combined with the results in the previous display imply that $\gamma^{*\prime}\breve{Y}_n - \gamma'\breve{Y}_n \xrightarrow{P_n}_p \infty$ for all $\gamma \in V(\Sigma^*)\backslash\{\gamma^*\}$. Since $V(\Sigma^*)\backslash\{\gamma^*\}$ is finite, this implies that $\min_{\gamma \in V(\Sigma^*)\backslash\{\gamma^*\}}(\gamma^{*\prime}\breve{Y}_n - \gamma'\breve{Y}_n) \xrightarrow{P_n}_p \infty$, from which we see that $\gamma^{*\prime}\breve{Y}_n = \max_{\gamma \in V(\Sigma^*)} \gamma'\breve{Y}_n$ with probability approaching 1 under P_n, which gives the desired result. $\qquad\square$

D   Additional Tables and Figures

Figure D.1: Sensitivity analysis for $\theta = \tau_{15}$ for Lovenheim and Willen (2019) using $\Delta = \Delta^{RM}(\bar{M})$ and $\Delta = \Delta^{SDRM}(\bar{M})$
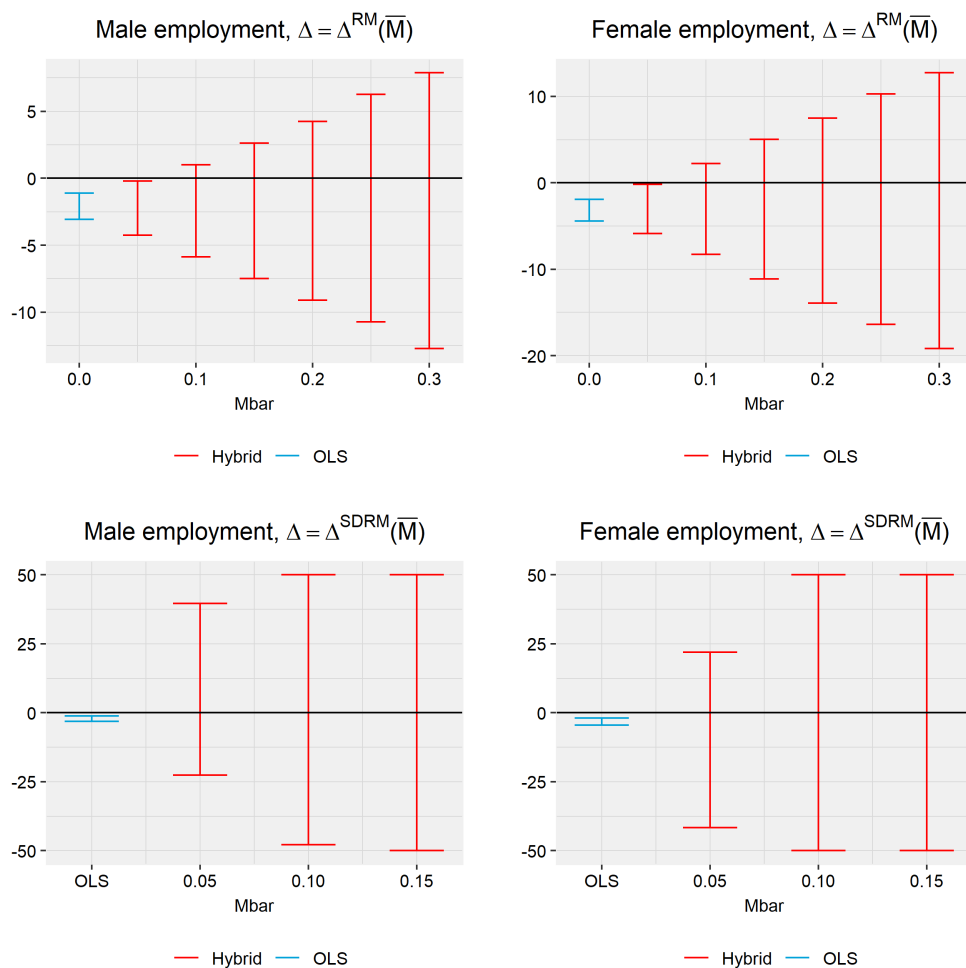
Note: Confidence sets for $\Delta^{SDMB}(\bar{M})$ are truncated at ± 50 to preserve readability.

Appendix References

Ingster, Yuri and I. A. Suslina, *Nonparametric Goodness-of-Fit Testing Under Gaussian Models* Lecture Notes in Statistics, New York: Springer-Verlag, 2003.

Kaido, Hiroaki, Francesca Molinari, and Jörg Stoye, "Constraint Qualifications in Partial Identification," *arXiv:1908.09103 [econ.EM]*, 2020.

Lee, Jason D., Dennis L. Sun, Yuekai Sun, and Jonathan E. Taylor, "Exact Post-Selection Inference, with Application to the Lasso," *The Annals of Statistics*, June 2016, *44* (3), 907–927.

Lovenheim, Michael F. and Alexander Willen, "The Long-Run Effects of Teacher Collective Bargaining," *American Economic Journal: Economic Policy*, 2019, *11* (3), 292–324.

Nam, Nguyen Mau, "Convex Analysis: An introduction to convexity and nonsmooth analysis," https://maunamn.wordpress.com/ 2019.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf, "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 2014, *82* (5), 1979–2002.

Schrijver, Alexander, *Theory of Linear and Integer Programming*, Wiley-Interscience, 1986.

Sierksma, Gerard, *Linear and Integer Programming: Theory and Practice, Second Edition*, CRC Press, November 2001. Google-Books-ID: jP6H4xrTBkUC.