Numerical Analysis of Test Optimality[*]

Philipp Ketz[†]   Adam McCloskey[‡]   Jan Scherer[§]

December 22, 2025

Abstract

In nonstandard testing environments, researchers often derive *ad hoc* tests with correct (asymptotic) size, but their optimality properties are typically unknown a priori and difficult to assess. This paper develops a numerical framework for determining whether an ad hoc test is *effectively* optimal—approximately maximizing a weighted average power criterion for some weights over the alternative and attaining a power envelope generated by a single weighted average power–maximizing test. Our approach uses nested optimization algorithms to approximate the weight function that makes an ad hoc test's weighted average power as close as possible to that of a true weighted average power–maximizing test, and we show the surprising result that the rejection probabilities corresponding to the latter form an approximate power envelope for the former. We provide convergence guarantees, discuss practical implementation and apply the method to the weak-instrument–robust conditional likelihood ratio test and a recently-proposed test for when a nuisance parameter may be on or near its boundary.

KEYWORDS: Hypothesis Testing, Weighted Average Power, Power Envelope, Weak Instrumental Variables, Boundary-Robust Inference

JEL CODES: C12, C15, C26

1 Introduction

In nonstandard testing environments involving composite null and/or alternative hypotheses, a uniformly most powerful test often does not exist. It is therefore common for researchers to devise implementable *ad hoc* tests that are designed to have provably correct (asymptotic) size. Although researchers may go to great lengths to demonstrate that their devised ad hoc tests have "good" power properties via Monte Carlo simulation, the theoretical power optimality properties of these tests are frequently unknown and elusive. It is therefore difficult for both those who design hypothesis tests and those who are meant to use them to assess whether the power of a given test can be improved upon by a different test, a central question for its usefulness.

Instances for which researchers have sought to show that an ad hoc test is theoretically optimal in some sense have been limited. Such approaches can be roughly divided into two types of power comparisons. The first compares the power function of the ad hoc test to a point-wise power envelope. This point-wise power envelope is constructed as the power of a *collection* of point-optimal tests computed at each point in the corresponding *collection* of point alternative hypotheses for points lying in the composite alternative space (e.g., Andrews et al., 2006; Andrews et al., 2019 and Guggenberger et al., 2019). Although it is possible for the power function of an ad hoc test to coincide with a power envelope constructed in this fashion, this is rarely the case because this form of power envelope may be unattainable by the power of any *single* test. To avoid or at least mitigate such unfair comparisons, researchers typically impose additional constraints on the point-optimal tests that are used to produce the point-wise power envelope such as unbiasedness (e.g., the two-sided *t*-test and Moreira and Moreira, 2013, 2019), rotational invariance (e.g., Andrews et al., 2006 and Moreira et al., 2021) and similarity (e.g., Andrews et al., 2006 and Montiel Olea, 2020). However, the imposition of constraints on the underlying collection of (constrained) point-optimal tests still does not generally produce a power envelope that is necessarily attainable by a single test and leads to an optimality analysis that is limited to tests satisfying the constraints. So when the power function of an ad hoc test falls short of a power envelope constructed in this way, what is a researcher to conclude? Is the ad hoc test suboptimal? Or is the power envelope simply an unfair and unattainable comparison because it does not correspond to the power function of any single test of the composite hypothesis?

The second approach to test optimality assessment in the literature compares either (i) the power function of the ad hoc test to the power function of a test that is known to

be weighted average power (WAP) maximizing for a particular weight function over the parameter values in the composite alternative or (ii) the WAPs of these two tests, or both (e.g., Andrews and Ploberger, 1994; Andrews et al., 2008; Elliott and Müller, 2014; Elliott et al., 2015 and Cox, 2024). This approach is also incomplete, leading to its own set of questions. How should the researcher choose the weight function when constructing the WAP-maximizing test under comparison? If the power function or WAP of the ad hoc test is not dominated by that of the WAP-maximizing test, should we conclude that it is optimal or could it be dominated by a WAP-maximizing test for a different set of weights?

In this paper, we propose to numerically determine whether there exists a weight function over the composite alternative that justifies the ad hoc test. More precisely, we propose a numerical procedure for automatically finding the weight function that minimizes the WAP difference between the WAP-maximizing test corresponding to that weight function and the ad hoc test over a discretization of the composite alternative space. This eliminates the ambiguity of the second approach to test optimality implicit in the questions above. When this WAP-difference is (not) approximately equal to zero, this approach immediately allows us to conclude that the ad hoc test under study is (not) effectively optimal in the sense of (not) being an approximately WAP-maximizing test itself. A side benefit to our approach is that in the case for which the ad hoc test is determined to be effectively optimal, we know which weights over the alternative space make it approximately WAP-maximizing, allowing us to interpret how the test "directs power".

In addition to determining whether an ad hoc test is approximately WAP-maximizing for some weight function, we show the surprising result that our procedure immediately produces an attainable power envelope for the ad hoc test. Over a discretization of the composite alternative space, we provide a set of sufficient conditions under which the power function of the WAP-maximizing test using the weights that minimize the WAP difference with the ad hoc test is a power envelope for the ad hoc test. Since this power envelope is constructed from a single test, it is indeed attainable. In addition, this power envelope is "most favorable" to the ad hoc test in the sense that it is constructed from a single WAP-maximizing test with WAP as close as possible to that of the ad hoc test.

Our numerical procedure is composed of an inner and an outer loop algorithm. The inner loop algorithm computes an approximate WAP-maximizing test for a given set of weights over the discretized alternative space, in analogy with Moreira and Moreira (2013), Elliott et al. (2015) and Fernández et al. (2025). Our outer loop algorithm, which has no counterparts that we are aware of in the literature, approximates the weight function that

2

minimizes the WAP difference between the corresponding WAP-maximizing test and the ad hoc test over the discretized alternative space.[1] Using the results mentioned above, the power function of this corresponding WAP-maximizing test can then be used as a power envelope for the ad hoc test. We provide new convergence results for both the inner and outer loop algorithms, formally guaranteeing that they approximate the WAP-maximizing test and weights mentioned above. Our algorithms and their convergence guarantees are general and not case-specific.

One may wonder why it would be more desirable to construct an ad hoc test with correct size and subsequently analyze its optimality properties rather than simply constructing a WAP-maximizing test directly using, e.g., Moreira and Moreira (2013) or Elliott et al. (2015). There are two main reasons: one conceptual and one computational. The first reason for constructing an ad hoc test is that it is often unclear which weight function one should use when constructing a WAP-maximizing test. Our procedure instead simply tells the researcher whether there exists a weight function that rationalizes the ad hoc test they have devised (as well as providing an approximation to it). The second reason is that even with a clear idea of a desirable weight function, computing the corresponding WAP-maximizing test can become computationally prohibitive when the dimension of the alternative parameter space is too high. Since our inner loop algorithm also computes WAP-maximizing tests iteratively, it is also of course subject to this criticism. However, our numerical procedure can be used to provide partial evidence on the optimality of a test if the researcher applies it to lower-dimensional special cases and is able to show that the ad hoc test is optimal in those special cases. On the other hand, a WAP-maximizing test for a lower-dimensional special case does not readily generalize up to higher dimensions.

Finally, we note that discovering an ad hoc test to be effectively optimal for a particular testing problem can be used to motivate the development of tests for related problems. For example, the theoretical optimality results established by Andrews et al. (2006, 2008) in the specific setting of a homoskedastic linear IV model with randomly sampled data for Moreira's (2003) conditional likelihood ratio (CLR) test, apparently helped motivate Andrews and Mikusheva (2016) and Andrews and Guggenberger (2019) to produce generalizations of this test to settings that allow for dependent data, nonlinear models and

---

[1]Ketz and McCloskey (2025) made the first attempt we are aware of at numerically finding weights that minimize the WAP difference between an ad hoc test and a WAP-maximizing test. However, their approach was confined to the particular problem of testing with inequality restrictions on nuisance parameters, carried no formal convergence guarantees and did not contain any results on producing power envelopes. Nevertheless, it helped motivate us to write the current paper.

singular variance matrices.

We apply our results and algorithms to two different testing problems, the first of which has been previously analyzed using the first approach to optimality assessment described above and the second of which has been analyzed using the second. For the first problem, we substantively contribute to the ongoing debate over the optimality properties of the CLR test. Specifically, we revisit classic optimality results from Andrews et al. (2006, 2008) and more recent critiques by Andrews et al. (2019) and Van de Sijpe and Windmeijer (2023) in the context of the Gaussian homoskedastic linear IV model. Using the first existing approach to test optimality assessment described above, Andrews et al. (2006, 2008) show that the CLR test effectively attains the point-wise power envelope for invariant (similar) tests under a design that holds the reduced-form error variance matrix constant. On the other hand, Andrews et al. (2019) argue that the CLR test falls short of the point-wise power envelope obtained by fixing the true value of the parameter of interest and instead varying its hypothesized value, which Van de Sijpe and Windmeijer (2023) show is equivalent to producing the more standard point-wise power envelope that fixes the hypothesized value and varies the true value under a design that instead holds the variance matrix of the structural errors constant. Using our new algorithms, we show that the CLR test's power function is virtually indistinguishable from that of a *single* WAP-maximzing test *for both designs*, leading us to conclude that the CLR test is in fact effectively optimal in the class of invariant asymptotically efficient tests for this problem.

For our second application, we analyze the optimality properties of the new inequality-imposed confidence interval (IICI) of Cox (2024), designed to improve inference when a scalar nuisance parameter may lie on or near the boundary of the parameter space. Using the second existing approach to test optimality assessment described above, Cox (2024) compares the WAP of the test implied by the IICI to that of the WAP-maximizing test proposed for this problem by Elliott et al. (2015), finding them to be very close. We plot the power functions of both tests and find that they intersect, making the optimality properties of the IICI-implied test unclear. Instead, we show that while the power function of the IICI-implied test is very close to the power function of a *single* most favorable WAP-maximizing test, it does fall slightly short with a maximum gap of about 0.3 percentage points. Nevertheless, its WAP is extremely close to that of the WAP-maximizing test, implying that it is highly-competitive but slightly short of optimal for problems with a scalar nuisance parameter that may lie on or near its boundary.

The scope for other applications of our results and numerical procedure is very wide

and we just briefly mention some additional examples here. In addition to analyzing the CLR test, one may seek to analyze the (unknown) optimality properties of other weak IV-robust tests in the literature such as the Lagrange multiplier test of Kleibergen (2002), which could be particularly interesting due to its non-monotonic power function. Campbell and Yogo's (2006) test of stock return predictability was not found to be dominated by the WAP-maximizing test with weights chosen by Elliott et al. (2015). Using our approach, it would be interesting to learn if there is a WAP-maximizing test that indeed dominates it. We could use our approach to determine whether the apparent power deficiencies of the test of Guggenberger et al. (2019) when compared to the point-wise power envelope disappear when using our approach that makes the attainable comparison to a most favorable single WAP-maximizing test. More broadly, our approach could be used to analyze tests in the large literatures on inference for moment inequality models, inference robust to identification failure, inference in structural change models and inference with highly-persistent (e.g., local-to-unit root) processes. We intend to analyze some of these examples in follow-up work. And we of course hope that our results and numerical procedures will be useful for the optimality analysis of new ad hoc tests that have yet to be developed.

The remainder of the paper proceeds as follows. In Section 2, we provide the intuition underlying our numerical approach by illustrating how to find the weight function that justifies the two-sided t-test as a WAP-maximizing test and contrasting it with the point-wise power envelope that is not attained by the two-sided t-test. Section 3 formalizes the general hypothesis testing framework we study, defines the numerical problem of finding weights that make an ad hoc test approximately WAP-maximizing and shows that the resulting WAP-maximizing test yields a "most favorable" approximate power envelope under a set of sufficient conditions. Section 4 describes our numerical procedure, detailing the inner loop algorithm that computes approximate WAP-maximizing tests and the outer loop algorithm that adjusts weights, along with practical considerations such as switching tests, thresholding, adding support points, and Monte Carlo smoothing. We provide the theoretical convergence guarantees for these algorithms in Section 5, establishing that our numerical method yields valid approximate WAP-maximizing tests and power envelopes. In Section 6, we apply our results and numerical procedures to the two testing applications described above. Mathematical proofs and implementation details for the applications are contained in the Appendix.

5

2   Intuition and Motivation

We begin by analyzing a simple canonical hypothesis testing example to impart intuition and motivate our approach to assessing the optimality of a hypothesis test. Consider a two-sided test of the mean of a Gaussian random variable with known unit variance: $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$ for $Y \sim \mathcal{N}(\beta, 1)$.[2] The standard level-$\alpha$ two-sided t-test that rejects when $|Y|$ exceeds $z_{1-\alpha/2}$ is well-known to be the uniformly most powerful test amongst all unbiased tests. However, suppose that we would not like to confine our analysis to unbiased tests. Since there is no uniformly most powerful test, we take the common approach of comparing the t-test to the point-wise power envelope for this testing problem. Specifically, the point-wise power envelope is equal to the rejection probabilities of the collection of the most powerful point-wise tests of H_0 vs. $H_{\beta'} : \beta = \beta'$, as a function of a collection of β' values. By the Neyman-Pearson lemma, we know that each of these point-wise tests is the likelihood ratio test of H_0 vs. $H_{\beta'}$.

Figure 1 plots the power function of the two-sided t-test and the point-wise power envelope. Notably, the power function of the two-sided t-test lies substantially below the power envelope. Using the point-wise power envelope could thus lead us to believe that the two-sided t-test is suboptimal. However, note that the point-wise power envelope does not produce a power comparison that is necessarily attainable because it corresponds to the power of a *collection* of point-optimal tests that are each optimal against a point alternative $H_{\beta'}$ with $\beta' \in \mathbb{R} \setminus \{0\}$, none of which correspond to the composite alternative hypothesis of interest H_1. In other words, it is still possible for the two-sided t-test of the composite alternative H_1 to be optimal even though its power function lies below the power function of a collection of point-wise optimal tests of $H_{\beta'}$—we may simply be comparing its power function to an unattainable upper bound.

In the context of this testing problem, our proposed approach is to instead determine whether there exists a set of weights over the composite alternative space $\Theta_1 = \mathbb{R} \setminus \{0\}$ such that the WAP of the two-sided t-test is well-approximated by the WAP of a WAP-maximizing test with this set of weights. If so, we may conclude that the two-sided t-test is "nearly optimal" for this set of weights. To address this task, we discretize the composite alternative space Θ_1 into support points $\bar{\Theta}_1 \subset \Theta_1$ and introduce an algorithm that successively adjusts the weights on each support point to produce a WAP-maximizing test with WAP as close as possible to that of the two-sided t-test. Surprisingly, we show below that the rejection probabilities of this WAP-maximizing test produce a power envelope for

---

[2]The analysis of this section applies without loss of generality to all two-sided t-tests of the mean of a Gaussian random variable with known variance via a simple scale transformation.
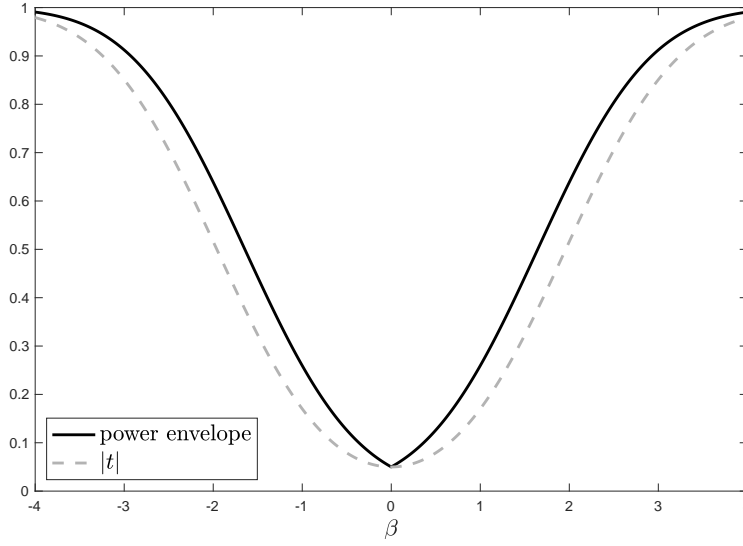
6

Figure 1: Power function of two-sided t-test and point-wise power envelope

the two-sided t-test. An important feature of this power envelope is that it corresponds to the rejection probabilities of a *single* test and is therefore necessarily attainable, in contrast to the point-wise power envelope. This power envelope is not only attainable but it is a "most favorable" power envelope in the sense that it is based upon a WAP-maximizing test with WAP as close as possible to that of the two-sided t-test itself.

For the sake of illustration, let us work with the coarse discretization $\bar{\Theta}_1 = \{-1, 1\}$. The WAP-maximizing test of H_0 vs. the discretized composite alternative $\bar{H}_1 : \beta \in \bar{\Theta}_1$ that weights $\beta = 1$ by ω_1 (and $\beta = -1$ by $1 - \omega_1$) rejects when

$$LR(Y) = \frac{(1 - \omega_1) f(Y; -1) + \omega_1 f(Y; 1)}{f(Y; 0)} > \mathrm{cv}$$

for cv satisfying $\mathbb{P}_{H_0}(LR(Y) > \mathrm{cv}) = \alpha$, where $f(\cdot; \mu)$ denotes the density function of a $\mathcal{N}(\mu, 1)$ random variable and $\mathbb{P}_{H_0}$ denotes the probability under H_0. Our approach begins with some initial weight ω_1, computes the power function of the WAP-maximizing test using this weight and adjusts the value of ω_1 according to the difference between the power function of the t-test and the WAP-function at $\beta = 1$: if the power of the t-test lies below that of the WAP-maximizing test at $\beta = 1$, we adjust ω_1 downwards, putting more weight on $\beta = -1$, and vice versa. We then repeat this procedure iteratively until the power function of the t-test lies weakly below that of the WAP-maximizing test at all points $\beta \in \bar{\Theta}_1$. If the power

7

function of the t-test lies weakly below that of the WAP-maximizing test at all points $\beta \in \Theta_1$, the power function of this final WAP-maximizing test produces a power envelope that is based upon a *single* test. If the power function of the t-test lies within a small value of the power function of this final WAP-maximizing test, we deem the t-test to be "effectively optimal" since its rejection probabilities nearly coincide with those of a test with known optimality. The weights given in the final iteration tell us how the nearly optimal t-test weights points in the alternative space since they are the weights for which the corresponding WAP-maximizing test has WAP nearly equal to that of the t-test. Otherwise, we deem the t-test to be suboptimal since its power function is dominated by that of another test.

We provide the specifics of how the weight ω_1 is adjusted at each iteration in the outer loop algorithm in Section 4.2, along with a theoretical justification for the algorithm in Section 5, but the intuition is as follows. Since any WAP-maximizing test is optimal for its set of weights by definition, we know (i) its power function cannot be dominated by that of the t-test on $\bar{\Theta}_1$ and (ii) if the t-test is itself WAP-optimal, there exists a set of weights for which its power function must match that of a WAP-maximizing test. These two facts justify our iterative weighting adjustment: at each iteration, we want to increase the weight at support points for which the WAP-maximizing test has lower power than the t-test and therefore decrease the weights at other support points. Naturally, these decreased weights will be at points for which the WAP-maximizing test has greater power than the t-test. The five panels of Figure 2 illustrate this re-weighting principle in five iterative steps, starting at $\omega_1 = 0.1$, for which ω_1 is successively increased until the power function of the WAP-maximizing test coincides with that of the t-test. The upward-pointing (downward-pointing) arrows indicate that the weight on the support point for the WAP-maximizing test needs to be adjusted downward (upward) to bring the power function of the WAP-maximizing test closer to lying (weakly) above that of the t-test. Since the power function of the WAP-maximizing test with weight $\omega_1 = 0.5$ in the final panel coincides with that of the two-sided t-test, we can say that the two-sided t-test is optimal in the class of all tests of H_0 vs. H_1 while no longer needing to constrain ourselves to the class of unbiased tests.

3   General Framework

Having illustrated the intuition behind our approach in a simple problem, we now move to the general hypothesis testing framework of interest.
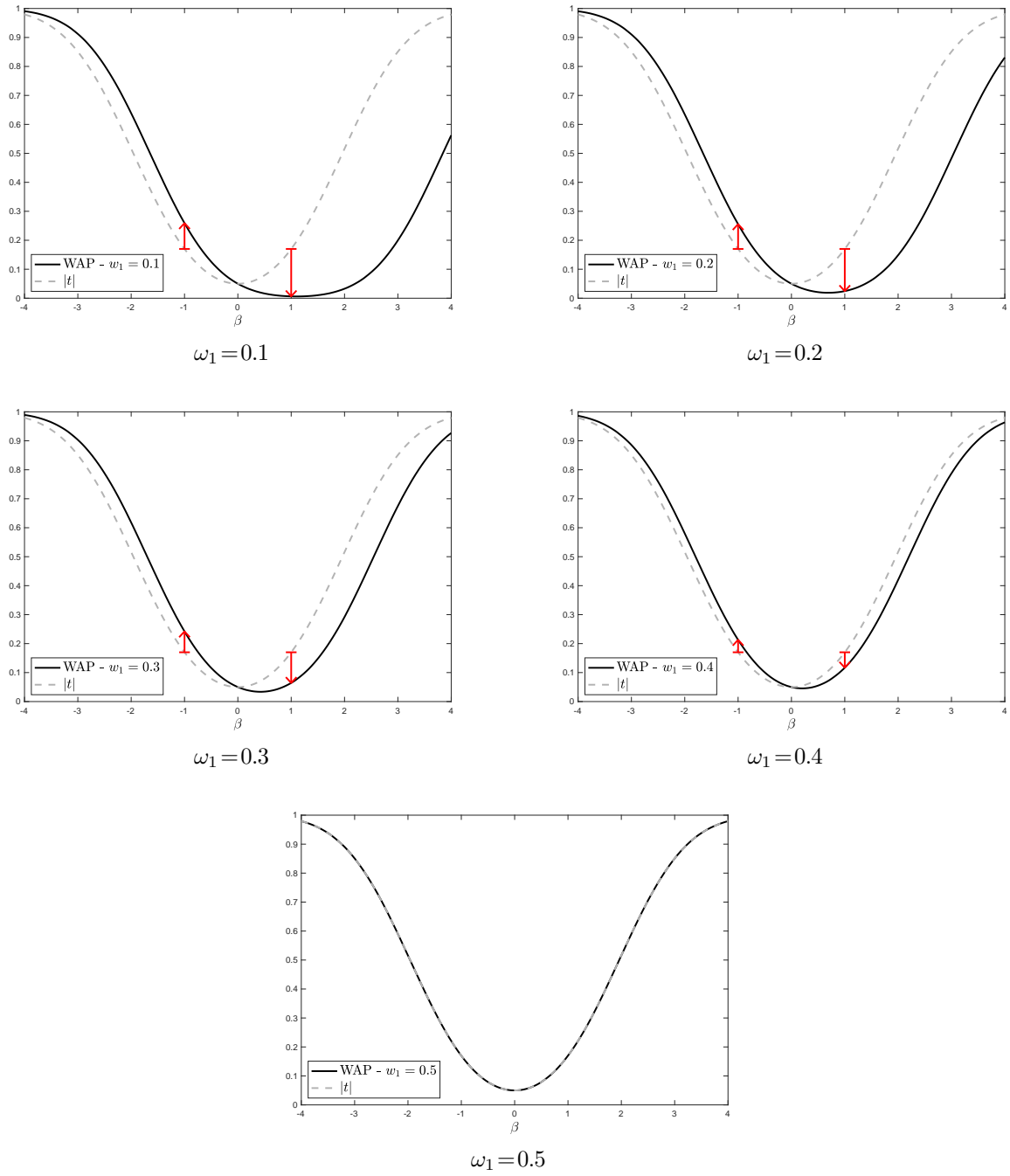
$\omega_1 = 0.1$

$\omega_1 = 0.2$

$\omega_1 = 0.3$

$\omega_1 = 0.4$

$\omega_1 = 0.5$

Figure 2: Power functions for two-sided t-tests and WAP-maximizing tests with different weights

3.1 Testing Problem

Suppose that we observe a random element Y taking values in the metric space $\mathcal{Y}$ and that Y has a probability density function $f_\theta(\cdot)$ relative to some sigma-finite measure ν, where the parameter governing its distribution $\theta \in \Theta \subset \mathbb{R}^k$ is finite-dimensional. We are interested in testing

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1, \tag{1}$$

where $\Theta_0, \Theta_1 \subset \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$ and Θ_0 is not a singleton. This testing problem with a single observation Y typically arises as the limiting problem under an asymptotic approximation to a finite-sample problem with many observations via a local asymptotic embedding corresponding to a limit experiment or by using the asymptotic equivalence approach of Müller (2011). See Elliott et al. (2015) for a more detailed discussion.

A generic test of (1) is a measurable function $\varphi : \mathcal{Y} \mapsto [0,1]$ for which $\varphi(y)$ is the probability of rejecting H_0 upon observing the realization $Y = y$. For the parameter value $\theta \in \Theta$, $\int \varphi f_\theta d\nu$ is thus equal to the rejection probability of the test when the true value of the parameter is θ. The starting point of our analysis is to suppose that we have an ad hoc test with correct size, a measurable function $\varphi_{ah} : \mathcal{Y} \mapsto [0,1]$ that is known to satisfy the (uniform) size constraint $\sup_{\theta \in \Theta_0} \int \varphi_{ah} f_\theta d\nu \leq \alpha$. We would like to assess whether φ_{ah} is "nearly optimal" among tests that control size. For this problem to be nontrivial, we focus on tests of the hypothesis (1) for which no uniformly most powerful test is known to exist.

3.2 Numerical Assessment of Near Optimality

As illustrated above in the simple context of a two-sided t-test for the mean of a Gaussian random variable with known variance, one could attempt to assess the optimality of φ_{ah} by comparing it to the point-wise power envelope obtained by computing the rejection probabilities of a collection of level-α Neyman-Pearson tests indexed by $\theta' \in \Theta_1$ of $H_{0,\Lambda_{\theta'}}$: the density of Y is $\int f_\theta d\Lambda_{\theta'}(\theta)$ vs. $H_{\theta'} : \theta = \theta'$ given by

$$\varphi_{\Lambda_{\theta'},\theta'}(y) = \begin{cases} 1 & \text{if } f_{\theta'}(y) > \mathrm{cv} \int f_\theta(y) d\Lambda_{\theta'}(\theta) \\ \varkappa & \text{if } f_{\theta'}(y) = \mathrm{cv} \int f_\theta(y) d\Lambda_{\theta'}(\theta), \\ 0 & \text{if } f_{\theta'}(y) < \mathrm{cv} \int f_\theta(y) d\Lambda_{\theta'}(\theta) \end{cases}$$

for some $\mathrm{cv} \geq 0$ and $0 \leq \varkappa \leq 1$ satisfying $\int \varphi_{\Lambda_{\theta'},\theta'}(\int f_\theta d\Lambda_{\theta'}(\theta)) d\nu = \alpha$, where $\Lambda_{\theta'}$ is the least-favorable probability distribution over Θ_0 corresponding to the point alternative

$H_{\theta'}$.[3] Indeed, this approach has been applied in the literature, oftentimes after imposing additional side constraints such as similarity or invariance (e.g., Andrews et al., 2006, 2008; Moreira and Moreira, 2013, 2019; Guggenberger et al., 2019; Andrews et al., 2019; Montiel Olea, 2020). However, as in the simple example of Section 2, comparison to this collection of Neyman-Pearson tests may produce an unattainable power envelope.

Instead, we propose to determine whether the WAP of φ_{ah} is nearly equal to that of a *single* WAP-maximizing test corresponding to some set of weights over Θ_1 and is thus "nearly optimal". More formally, we seek to determine whether there exists a weight function Ω, a probability distribution with support on Θ_1, such that φ_{ah} nearly maximizes the WAP criterion (Wald, 1943)

$$\text{WAP}(\varphi) = \int \left(\int \varphi f_\theta d\nu \right) d\Omega(\theta) \tag{2}$$

within the set of level-α tests

$$\Phi_\alpha \equiv \left\{ \varphi : \mathcal{Y} \mapsto [0,1] : \varphi \text{ is measurable}, \sup_{\theta \in \Theta_0} \int \varphi f_\theta d\nu \leq \alpha \right\}.$$

If a least-favorable probability distribution Λ_Ω over Θ_0 corresponding to the simple alternative $H_{1,\Omega}$: the density of Y is $\int f_\theta d\Omega(\theta)$ exists, this WAP-maximizing test takes the familiar Neyman-Pearson form of a test of H_{0,Λ_Ω}: the density of Y is $\int f_\theta d\Lambda_\Omega(\theta)$ vs. $H_{1,\Omega}$ given by

$$\varphi_{\Lambda_\Omega,\Omega}(y) = \begin{cases} 1 & \text{if } \int f_\theta(y)d\Omega(\theta) > \text{cv}_\Omega \int f_\theta(y)d\Lambda_\Omega(\theta) \\ \varkappa_\Omega & \text{if } \int f_\theta(y)d\Omega(\theta) = \text{cv}_\Omega \int f_\theta(y)d\Lambda_\Omega(\theta) \ , \\ 0 & \text{if } \int f_\theta(y)d\Omega(\theta) < \text{cv}_\Omega \int f_\theta(y)d\Lambda_\Omega(\theta). \end{cases} \tag{3}$$

Here, $\text{cv}_\Omega \geq 0$ and $0 \leq \varkappa_\Omega \leq 1$ are defined to satisfy $\int \varphi_{\Lambda_\Omega,\Omega}(\int f_\theta d\Lambda_\Omega(\theta))d\nu = \alpha$. As noted by Elliott et al. (2015), even if this least-favorable distribution does not exist, we can find an "approximate least-favorable distribution" to approximate the test the maximizes (2).

It is not typically possible to determine the existence of a weight function Ω for which an ad hoc test φ_{ah} maximizes the WAP criterion $\text{WAP}(\varphi)$ analytically. We thus propose a numerical approach that computes the weights that make the WAP of a WAP-maximizing test as close as possible to the WAP of the ad hoc test *for those weights* over a discretization

---

[3]See Lehmann and Romano (2005) for details on least-favorable distributions and Andrews et al. (2008), Elliott et al. (2015) and Fernández et al. (2025) for details on computing approximations to them.

of the alternative space. Specifically, let $\bar{\Theta}_1 = (\theta_1,...,\theta_{M_1}) \subset \Theta_1$ denote a finite set of support points for the alternative space with M_1 elements and $\bar{\Omega} = (\omega_1,...,\omega_{M_1}) \in \Delta_{M_1}$ denote a probability distribution over those support points, where Δ_{M_1} is the M_1-dimensional unit simplex. Formally, the numerical problem we wish to solve is given by

$$\inf_{\bar{\Omega} \in \Delta_{M_1}} \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu. \tag{4}$$

A test $\varphi^* \in \Phi_\alpha$ that solves (4) is a WAP-maximizing test with corresponding WAP as close as possible to that of φ_{ah} since

$$\sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu = \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int \varphi f_{\theta_j} d\nu - \sum_{j=1}^{M_1} \omega_j \int \varphi_{ah} f_{\theta_j} d\nu. \tag{5}$$

A natural byproduct of our approach is that a weight function $\bar{\Omega}$ solving (4) provides valuable information about the power properties of the ad hoc test: higher weight over a set of points in $\bar{\Theta}_1$ indicates that the test prioritizes rejection in that set relative to other points in $\bar{\Theta}_1$.

3.3 Using WAP-Maximizing Tests to Produce an Approximate Power Envelope

In addition to determining whether φ_{ah} nearly maximizes a WAP criterion for some set of weights via (4), our numerical approach has the added benefit of being able to produce an approximate power envelope (APE) for φ_{ah}, whether or not it is WAP-maximizing. This power envelope is again obtained from a *single* test and is therefore necessarily attainable, in contrast to power envelopes derived from a collection of tests. To see this, first note that the rejection probabilities over Θ_1 of the test that solves

$$\sup_{\varphi \in \Phi_\alpha} \inf_{\theta \in \Theta_1} \int (\varphi - \varphi_{ah}) f_\theta d\nu$$

constitute a power envelope for $\varphi_{ah} \in \Phi_\alpha$ since

$$\sup_{\varphi \in \Phi_\alpha} \inf_{\theta \in \Theta_1} \int (\varphi - \varphi_{ah}) f_\theta d\nu \geq \inf_{\theta \in \Theta_1} \int (\varphi_{ah} - \varphi_{ah}) f_\theta d\nu = 0.$$

For an appropriately chosen $\bar{\Theta}_1 = (\theta_1, ..., \theta_{M_1}) \subset \Theta_1$—see Section 4.3.3 for details, we can then view the rejection probabilities of the test that solves

$$\sup_{\varphi \in \Phi_\alpha} \min_{j=1,...,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu \qquad (6)$$

as an APE for φ_{ah} since it constitutes a power envelope for φ_{ah} over $\bar{\Theta}_1$ by definition. In Theorem 1 below, we show that the value of the maximin problem in (6) is equal to the value of the minimax problem in (4) and any test that solves (6) also solves (4). Since we show how to approximate the solution to (4) in Sections 4 and 5 below, this latter result provides a practical strategy for obtaining an APE for φ_{ah}: solve (4) and numerically check whether the solution yields an APE for φ_{ah}. Absent further structure on the problem, this strategy is not guaranteed to produce an APE but it can still provide a useful guide for obtaining an APE in practice. In Theorem 2 below, we impose additional structure on the problem that is sufficient for guaranteeing that this strategy indeed produces an APE.

Let

$$\Phi^* = \underset{\varphi \in \Phi_\alpha}{\mathrm{argmax}} \min_{j=1,...,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu$$

and

$$\Delta^*_{M_1} = \underset{\bar{\Omega} \in \Delta_{M_1}}{\mathrm{argmin}} \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu.$$

We can now state our first result.

Theorem 1
For $\bar{\Omega} = (\omega_1, ..., \omega_{M_1}) \in \Delta_{M_1}$ and any $\alpha \in (0,1)$,

$$\sup_{\varphi \in \Phi_\alpha} \min_{j=1,...,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu = \inf_{\bar{\Omega} \in \Delta_{M_1}} \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu,$$

$\Phi^* \neq \emptyset$ and $\Delta^*_{M_1} \neq \emptyset$. *Furthermore, for any $\varphi^* \in \Phi^*$ and any $\bar{\Omega}^* = (\omega_1^*, ..., \omega_{M_1}^*) \in \Delta^*_{M_1}$, we have*

$$\sum_{j=1}^{M_1} \omega_j^* \int (\varphi^* - \varphi_{ah}) f_{\theta_j} d\nu = \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j^* \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu. \qquad (7)$$

In words, equation (7) states that any $\varphi^* \in \Phi^*$, i.e., any test that solves the maximin problem in (6), is WAP-maximizing with respect to any $\bar{\Omega}^* \in \Delta^*_{M_1}$. Therefore, if one follows the strategy mentioned above and finds that a test that solves (4) yields an APE for φ_{ah},

this strategy comes with the added benefit that the APE is "most-favorable" for φ_{ah} in the sense that it corresponds to the rejection probabilities of a WAP-maximizing test with WAP as close as possible to that of φ_{ah} over $\bar{\Theta}_1$.

It can be shown that Theorem 1 further implies that if the least-favorable distribution $\Lambda_{\bar{\Omega}^*}$ exists for all $\bar{\Omega}^* \in \Delta_{M_1}^*$, then any test $\varphi^* \in \Phi^*$—which serves as a power envelope for φ_{ah} on $\bar{\Theta}_1$—is equal to a Neyman-Pearson test $\varphi_{\Lambda_{\bar{\Omega}^*},\bar{\Omega}^*}$ for some $\bar{\Omega}^* \in \Delta_{M_1}^*$, except possibly on the event:[4]

$$\left\{ y \in \mathcal{Y} : \sum_{j=1}^{M_1} \omega_j^* f_{\theta_j} = \mathrm{cv}_{\bar{\Omega}^*} \int f_\theta(y) d\Lambda_{\bar{\Omega}^*}(\theta) \right\}. \tag{8}$$

We impose the commonly-satisfied sufficient condition that this event has ν-measure zero for all $\bar{\Omega}^*$ in Theorem 2 below and formally show that any $\varphi^* \in \Phi^*$ is equal to a Neyman-Pearson test $\varphi_{\Lambda_{\bar{\Omega}^*},\bar{\Omega}^*}$ ν-almost everywhere (a.e.), and this Neyman-Pearson test is unique ν-a.e. This uniqueness result trivially implies that the test that solves (4) also solves (6) so that the test obtained by solving the minimax problem according to our algorithms in Section 4 readily provides a most-favorable APE for φ_{ah}.

Theorem 2

Let $\alpha \in (0,1)$. For all $\bar{\Omega}^* \in \Delta_{M_1}^*$, suppose that the least favorable distribution $\Lambda_{\bar{\Omega}^*}$ over Θ_0 exists and that (8) has ν-measure zero. Then, for any $\bar{\Omega}^* \in \Delta_{M_1}^*$, $\varphi_{\Lambda_{\bar{\Omega}^*},\bar{\Omega}^*}$ is the ν-a.e. unique maximizer of

$$\sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j^* \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu.$$

Furthermore, for any $\varphi_1^*, \varphi_2^* \in \Phi^*$ and any $\bar{\Omega}_1^*, \bar{\Omega}_2^* \in \Delta_{M_1}^*$, we have

$$\varphi_1^* = \varphi_2^* = \varphi_{\Lambda_{\bar{\Omega}_1^*},\bar{\Omega}_1^*} = \varphi_{\Lambda_{\bar{\Omega}_2^*},\bar{\Omega}_2^*} \ \nu\text{-}a.e.$$

The condition that (8) has ν-measure zero typically holds when Y is an absolutely continuous random vector since (8) is typically a lower-dimensional submanifold of $\mathcal{Y}$. The existence of a least favorable distribution has been established under weak conditions for a Euclidean sample space $\mathcal{Y}$ when f_θ is continuous in θ and Θ_0 is a closed Borel set in a finite-dimensional Euclidean space. See Lehmann and Romano (2005) and references therein.

---

[4]This follows from Lemma 3 under the assumption that $\Lambda^*(\Theta_0) > 0$ for all $\Lambda^* \in \mathcal{M}_0^*$.

4   Numerical Implementation

To numerically approximate the weight function that justifies an ad hoc test φ_{ah} in terms of WAP or produce a power function that dominates it, we aim to solve (4). Our numerical algorithm for solving (4) is composed of an inner loop and an outer loop. The inner loop computes an approximation to a WAP-maximizing test $\phi_{\Lambda_\Omega,\Omega}$ *for a given weight function* Ω in the spirit of Moreira and Moreira (2013), Elliott et al. (2015) and Fernández et al. (2025).[5] The outer loop approximates a weight function $\bar{\Omega}$ that solves (4) using the inner loop as input at each step since it involves searching over WAP-maximizing tests via the relation (5). In Section 5, we provide theoretical convergence results justifying the use of our algorithm for solving (4).

4.1   Inner Loop for Computing Approximate WAP-Maximizing Test

Discretize the null parameter space Θ_0 into a finite set of support points $\bar{\Theta}_0 = \{\tilde{\theta}_1,...,\tilde{\theta}_{M_0}\} \subset \Theta_0$ with M_0 elements. In light of (2) (and Fubini's Theorem), we aim to solve the following discretized optimization problem:

$$\max_{\varphi \in \Phi} \int \varphi g d\nu \quad \text{s.t.} \quad \int \varphi f_{\tilde{\theta}_i} d\nu \leq \alpha, \quad \text{for } i=1,...,M_0, \tag{9}$$

where $\Phi = \{\varphi : \mathcal{Y} \to [0,1] : \varphi \text{ is measurable}\}$ and $g(y) = \int f_\theta(y) d\Omega(\theta)$ for a given weight function Ω. Its dual problem is given by

$$\min_{\widetilde{\Lambda} \geq 0} \max_{\varphi \in \Phi} \int \varphi g d\nu - \sum_{i=1}^{M_0} \tilde{\lambda}_i \left(\int \varphi f_{\tilde{\theta}_i} d\nu - \alpha \right),$$

where $\widetilde{\Lambda} = (\tilde{\lambda}_1,...,\tilde{\lambda}_{M_0})$. It is not hard to see that Slater's condition is satisfied in this setting and therefore strong duality holds between these two problems.[6] This implies that for any solution $\widetilde{\Lambda}^*$ of the dual problem, any solution $\tilde{\varphi}_{\widetilde{\Lambda}^*}$ of

$$\tilde{\phi}(\widetilde{\Lambda}^*) := \max_{\varphi \in \Phi} \int \varphi g d\nu - \sum_{i=1}^{M_0} \tilde{\lambda}_i^* \left(\int \varphi f_{\tilde{\theta}_i} d\nu - \alpha \right)$$

---

[5]Guggenberger and Huang (2024) also briefly note that their numerical algorithm for approximating minimax regret treatment rules could also potentially be modified to numerically approximate a WAP-maximizing test for a given weight function.

[6]See Theorem 1 in Chapter 8.3 in Luenberger (1997).

is also a solution of the primal problem (9). We can operationalize this observation via the following simple result.

Proposition 1

For any $\widetilde{\Lambda} \geq 0$,

$$\tilde{\varphi}_{\widetilde{\Lambda}} \in \underset{\varphi \in \Phi}{\operatorname{argmax}} \int \varphi g d\nu - \sum_{i=1}^{M_0} \tilde{\lambda}_i \left(\int \varphi f_{\tilde{\theta}_i} d\nu - \alpha \right)$$

for

$$\tilde{\varphi}_{\widetilde{\Lambda}}(y) = \begin{cases} 1 & , \text{ if } g(y) \geq \sum_{i=1}^{M_0} \tilde{\lambda}_i f_{\tilde{\theta}_i}(y) \\ 0 & , \text{ if } g(y) < \sum_{i=1}^{M_0} \tilde{\lambda}_i f_{\tilde{\theta}_i}(y). \end{cases}$$

Since $\tilde{\varphi}_{\widetilde{\Lambda}^*}$ is given in closed form as soon as we know $\widetilde{\Lambda}^*$, it is sufficient to solve the dual problem in order to find a tractable solution to the primal problem (9). Also, note that given the above, we have

$$\tilde{\phi}(\widetilde{\Lambda}) = \int \tilde{\varphi}_{\widetilde{\Lambda}} g d\nu - \sum_{i=1}^{M_0} \tilde{\lambda}_i \left(\int \tilde{\varphi}_{\widetilde{\Lambda}} f_{\tilde{\theta}_i} d\nu - \alpha \right). \tag{10}$$

This observation motivates the following algorithm for approximating $\tilde{\varphi}_{\widetilde{\Lambda}^*}$.

1. Initialization:

   (a) Choose $\widetilde{\Lambda}^{(0)} = (\tilde{\lambda}_1^{(0)}, ..., \tilde{\lambda}_{M_0}^{(0)}) \geq 0$ and $\{\tilde{h}_k\}_{k=0}^{\infty}$ satisfying

   $$\tilde{h}_k > 0 \quad \text{for } k = 0, 1, ..., \quad \tilde{h}_k \to 0 \quad \text{as } k \to \infty \quad \text{and} \quad \sum_{k=0}^{\infty} \tilde{h}_k = \infty.$$

   (b) Compute $\tilde{\varphi}_{\widetilde{\Lambda}^{(0)}}$ and $\tilde{\phi}(\widetilde{\Lambda}^{(0)})$.

2. kth iteration:

   (a) Compute $\tau = (\tau_1, ..., \tau_{M_0})$, where

   $$\tau_i = \int \tilde{\varphi}_{\widetilde{\Lambda}^{(k)}} f_{\tilde{\theta}_i} d\nu - \alpha.$$

16

(b) Update $\widetilde{\Lambda}^{(k+1)} = (\tilde{\lambda}_1^{(k+1)},...,\tilde{\lambda}_{M_0}^{(k+1)})$:

$$\tilde{\lambda}_i^{(k+1)} = \left(\tilde{\lambda}_i^{(k)} + \tilde{h}_k \frac{\tau_i}{\|\tau_i\|_2}\right)_+ \quad \text{for } i=1,...,M_0,$$

where $(a)_+ = \max\{a,0\}$ for $a \in \mathbb{R}$.

(c) Compute $\tilde{\varphi}_{\widetilde{\Lambda}^{(k+1)}}$ and $\tilde{\phi}(\widetilde{\Lambda}^{(k+1)})$.

The intuition for this algorithm is essentially the same as that of Elliott et al. (2015). However, in contrast to the algorithm in Elliott et al. (2015), our algorithm carries proven convergence guarantees—see Section 5.[7]

For k large enough, a $\tilde{\varphi}_{\widetilde{\Lambda}^{(k)}}$ approximately solves the original optimization problem (9) in a sense we make precise in Section 5. By Lemma 1 of Elliott et al. (2015), if $\sup_{\theta \in \Theta_0} \int \tilde{\varphi}_{\widetilde{\Lambda}^{(k)}} f_\theta d\nu \leq \alpha$, then $\tilde{\varphi}_{\widetilde{\Lambda}^{(k)}}$ attains an upper bound on the WAP for the weight function Ω for any test $\varphi \in \Phi_\alpha$, no matter how large k is. This size constraint needs to be checked numerically but approximately holds by design if $\bar{\Theta}_0 \subset \Theta_0$ is rich enough. Finally, to map any test $\tilde{\varphi}_{\widetilde{\Lambda}}$ given in Lagrange multiplier form back to the Neyman-Pearson form (3), simply set $\text{cv}_\Omega = \sum_{i=1}^{M_0} \tilde{\lambda}_i$, $\varkappa_\Omega = 1$ and $\Lambda = (\lambda_1,...,\lambda_{M_0}) \in \Delta_{M_0}$ with $\lambda_i = \tilde{\lambda}_i/\text{cv}_\Omega$ for $i=1,...,M_0$.

4.2 Outer Loop Algorithm for Computing Weight Function

Let

$$\phi(\bar{\Omega}) = \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu$$

and

$$\varphi_{\bar{\Omega}}^* \in \arg\max_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu.$$

The function ϕ may fail to be fully differentiable but is convex, therefore enabling us to work with the following projected subgradient descent algorithm to solve (4).

1. Initialization:

(a) Choose $\bar{\Omega}^{(0)} = (\omega_1^{(0)}, ..., \omega_{M_1}^{(0)}) \in \Delta_{M_1}$ and $\{h_k\}_{k=0}^{\infty}$ satisfying

$$h_k > 0 \quad \text{for } k = 0, 1, ..., \quad h_k \to 0 \quad \text{as } k \to \infty \quad \text{and} \quad \sum_{k=0}^{\infty} h_k = \infty.$$

(b) Compute $\phi(\bar{\Omega}^{(0)})$ and

$$\varphi^*_{\bar{\Omega}^{(0)}} \in \operatorname*{argmax}_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j^{(0)} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu$$

using the inner loop algorithm.

2. $\underline{k\text{th iteration:}}$

(a) Compute $\gamma = (\gamma_1, ..., \gamma_{M_1})$, where

$$\gamma_i = \int (\varphi^*_{\bar{\Omega}^{(k)}} - \varphi_{ah}) f_{\theta_i} d\nu, \quad \text{for } i = 1, ..., M_1.$$

(b) Update $\bar{\Omega}^{(k+1)} = (\omega_1^{(k+1)}, ..., \omega_{M_1}^{(k+1)})$:

$$\omega_i^{(k+1)} = \pi_\Delta \left(\omega_i^{(k)} - h_k \frac{\gamma_i}{\|\gamma_i\|_2} \right) \quad \text{for } i = 1, ..., M_1,$$

where π_Δ denotes the Euclidean projection onto the unit simplex Δ_{M_1}.

(c) Compute $\phi(\bar{\Omega}^{(k+1)})$ and

$$\varphi^*_{\bar{\Omega}^{(k+1)}} \in \operatorname*{argmax}_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j^{(k+1)} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu$$

using the inner loop algorithm.

The intuition underlying the algorithm is simple. Since our goal is to obtain a power envelope for φ_{ah} from a WAP-maximizing test, we aim to find weights $\bar{\Omega}$ such that the power of $\varphi^*_{\bar{\Omega}}$ is (weakly) greater than that of φ_{ah} at all support points in $\bar{\Theta}_1$. So at each iteration k, we aim to increase the weight given to support points for which $\varphi^*_{\bar{\Omega}^{(k)}}$ currently has lower power than φ_{ah}. Since the weights must add to one, we therefore must decrease the weight given to other support points. Naturally, we do so at points for which $\varphi^*_{\bar{\Omega}^{(k)}}$ currently has

18

higher power than φ_{ah}. Step 2(a) computes the support points for which we aim to increase (decrease) the corresponding weights while Step 2(b) computes the weight adjustment.

For k large enough, a $\bar{\Omega}^{(k)}$ approximately solves the original optimization problem (4) in a sense we make precise in Section 5. Step 2(b) of the algorithm relies on a Euclidean projection onto the unit simplex Δ_{M_1}. This is a quadratic programming problem and there are efficient computational algorithms in the literature designed for solving this problem (e.g., Wang and Carreira-Perpinan, 2013).

4.3 Implementation Details

In this section, we provide precise recommendations on how to implement our numerical procedure. We follow these recommendations ourselves when analyzing the applications in Section 6.

4.3.1 Switching Tests

We follow Elliott et al. (2015) and implement switching tests when the parameter space for the nuisance parameter is unbounded, albeit with an additional related objective. Tests that do not (approximately) reduce to a standard test in the "standard" part of the parameter space, typically characterized by large values of nuisance parameters, tend to significantly sacrifice power in the "standard" part of the parameter space, a conclusively undesirable property (see Section 4 of Elliott et al., 2015). Therefore, we do not wish to limit the scope of our analysis on test optimality to WAP-maximizing tests that are not able to "switch" to standard tests in the "standard" parts of the parameter space. Indeed, ad hoc tests are typically purposefully designed to reduce to standard tests that are known to be optimal in some sense in the "standard" part of the parameter space for this very reason. For example, in the linear instrumental variables (IV) model, the CLR test reduces to the two-sided t-test when the concentration parameter is large. Since our optimality-assessment procedure relies upon approximating WAP-maximizing tests on a finite set of support points $\bar{\Theta}_1 \subset \Theta_1$, comparing to WAP-maximizing tests that do no allow for "switching" would inherently disadvantage an ad hoc test that reduces to a standard test in the "standard" part of the parameter space because the WAP-maximizing test would place zero weight over any region in Θ_1 outside of $\bar{\Theta}_1$. Due to the typical noncompactness of the parameter spaces Θ_0 and Θ_1, there are also numerical benefits to focusing on switching tests, for which we also refer the interested reader to Section 4 of Elliott et al. (2015).

Since an ad hoc test can immediately be seen as suboptimal when it does not reduce to a known "standard best test" in the "standard" part of the parameter space, we focus here

on ad hoc tests that do. Section 4.1 of Elliott et al. (2015) formalizes how the "standard" part of the parameter space can be characterized in terms of large values of a parameter δ. Let δ_S be the point where the "standard" part of the parameter space begins in the sense that for $\delta > \delta_S$ the ad hoc test and the standard test have essentially the same rejection profile. Let $D(Y)$ and δ_{SP} denote a statistic and a "switching point" such that with probability very close to zero, $D(Y) > \delta_{SP}$ whenever $\delta \leq \delta_S$. The motivation for this choice is that we only want the test to which we compare φ_{ah} to switch to a standard test φ_S that has the best rejection profile in the standard part of the parameter space when we know that the ad hoc test has essentially the same rejection profile, enabling us to analyze its optimality properties outside of this region. In practice, all this amounts to changing $\tilde{\varphi}_{\tilde{\Lambda}}(y)$ in the inner loop algorithm from the expression in Proposition 1 to the switching form

$$\tilde{\varphi}_{\tilde{\Lambda}}(y) = \mathbb{1}(D(y) > \delta_{SP})\varphi_S(y) + \mathbb{1}(D(y) \leq \delta_{SP})\mathbb{1}\left(g(y) \geq \sum_{i=1}^{M_0} \tilde{\lambda}_i f_{\tilde{\theta}_i}(y)\right).$$

See Elliott et al. (2015) for a formal definition of the standard best test φ_S and further details on switching tests.

4.3.2 Numerical Approximation Thresholds

Acknowledging that we cannot perfectly compute $\bar{\Omega}^*$ (due to the numerical approximation of the outer loop) or φ_Ω^* for any $\Omega \in \Delta_{M_1}$ (due to the numerical approximation of the inner loop), let $\hat{\varphi}_{\bar{\Omega}}^*$ denote the test obtained from the outer loop algorithm. In addition to potential numerical error arising from the use of our algorithms, we must rely on approximations of rejection probabilities within step 2.(a) of the inner loop algorithm and step 2.(a) of the outer loop algorithm. Nevertheless, the theoretical results in the following section and standard laws of large numbers imply that if we apply our algorithms with enough iterations and simulate rejection probabilities from enough Monte Carlo replications, these approximation errors will be small so that (i) $\int \hat{\varphi}_{\bar{\Omega}}^* f_{\tilde{\theta}} d\nu$ cannot lie substantially above α for any $\tilde{\theta} \in \bar{\Theta}_0$ and (ii) $\int (\hat{\varphi}_{\bar{\Omega}}^* - \varphi_{ah}) f_\theta d\nu$ cannot lie substantially below zero for any $\theta \in \bar{\Theta}_1$. We therefore use the following numerical approximation threshold rules for some small threshold $\epsilon > 0$:

- For $\theta \in \Theta_0$, we conclude that $\int \hat{\varphi}_{\bar{\Omega}}^* f_\theta d\nu \lesssim \alpha$ if $\int \hat{\varphi}_{\bar{\Omega}}^* f_\theta d\nu \leq \alpha + \epsilon$.

- For $\theta \in \Theta_1$, we conclude that

    - $\int (\hat{\varphi}_{\bar{\Omega}}^* - \varphi_{ah}) f_\theta d\nu \lesssim 0$ if $\int (\hat{\varphi}_{\bar{\Omega}}^* - \varphi_{ah}) f_\theta d\nu < \varepsilon$,
    - $\int (\hat{\varphi}_{\bar{\Omega}}^* - \varphi_{ah}) f_\theta d\nu \gtrsim 0$ if $\int (\hat{\varphi}_{\bar{\Omega}}^* - \varphi_{ah}) f_\theta d\nu > -\varepsilon$,

$$- \int(\hat{\varphi}_{\hat{\Omega}}^* - \varphi_{ah})f_\theta d\nu \approx 0 \text{ if } |\int(\hat{\varphi}_{\hat{\Omega}}^* - \varphi_{ah})f_\theta d\nu| \leq \varepsilon.$$

4.3.3 Constructing the Approximate Power Envelope

Using the numerical approximation threshold rules above, we implement the algorithm by potentially adding support points according to the following steps:

1. For a given $\bar{\Theta}_0$ and $\bar{\Theta}_1$, run the inner and outer loop algorithms until $\int\hat{\varphi}_{\hat{\Omega}}^* f_{\tilde{\theta}} d\nu \lesssim \alpha$ for all $\tilde{\theta} \in \bar{\Theta}_0$ and $\int(\hat{\varphi}_{\hat{\Omega}}^* - \varphi_{ah})f_\theta d\nu \gtrsim 0$ for all $\theta \in \bar{\Theta}_1$. Proceed to step 2.

2. For a fine grid $\Theta_0^f \subset \Theta_0$, compute $\int\hat{\varphi}_{\hat{\Omega}}^* f_{\tilde{\theta}} d\nu$ for each $\tilde{\theta} \in \Theta_0^f$.

   - If $\int\hat{\varphi}_{\hat{\Omega}}^* f_{\tilde{\theta}} d\nu \lesssim \alpha$ for all $\tilde{\theta} \in \Theta_0^f$, proceed to step 3.
   - Otherwise, add (some of) the values in Θ_0^f for which $\int\hat{\varphi}_{\hat{\Omega}}^* f_{\tilde{\theta}} d\nu \not\lesssim \alpha$ to $\bar{\Theta}_0$ and return to step 1.

3. For a fine grid $\Theta_1^f \subset \Theta_1$, compute $\int(\hat{\varphi}_{\hat{\Omega}}^* - \varphi_{ah})f_\theta d\nu$ for each $\theta \in \Theta_1^f$.

   - If $\int(\hat{\varphi}_{\hat{\Omega}}^* - \varphi_{ah})f_\theta d\nu \gtrsim 0$ for all $\theta \in \Theta_1^f$, then $\hat{\varphi}_{\hat{\Omega}}^*$ constitutes an *approximate power envelope* and either

     - $\int(\hat{\varphi}_{\hat{\Omega}}^* - \varphi_{ah})f_\theta d\nu \approx 0$ for all $\theta \in \Theta_1^f \Rightarrow$ Conclude that φ_{ah} is *effectively optimal*.
     - $\int(\hat{\varphi}_{\hat{\Omega}}^* - \varphi_{ah})f_\theta d\nu \not\lesssim 0$ for some $\theta \in \Theta_1^f \Rightarrow$ Conclude that φ_{ah} is *effectively dominated*.

   - Otherwise, add (some of) the values in Θ_1^f for which $\int(\hat{\varphi}_{\hat{\Omega}}^* - \varphi_{ah})f_\theta d\nu \not\gtrsim 0$ to $\bar{\Theta}_1$ and return to step 1.

The motivation for step 2. is to ensure $\hat{\varphi}_{\hat{\Omega}}^*$ has approximately correct size, in analogy with step 8. of Elliott et al.'s (2015) algorithm. Similarly, step 3. is used to ensure that the power function of $\hat{\varphi}_{\hat{\Omega}}^*$ provides a good approximation to that of a WAP-maximizing test that yields a power envelope for φ_{ah}.

4.3.4 Approximating Rejection Probabilities

To approximate rejection probabilities, we rely on Monte Carlo simulation but incorporate several modifications designed to improve the numerical stability and convergence of our algorithms. These adjustments impose qualitative features that the true rejection probabilities are known to satisfy, thereby reducing simulation noise that would otherwise distort the optimization.

First, in all examples we consider, the power function of the ad hoc test is continuous in θ. Independent simulation draws across different θ values produce artificially jagged power functions. To avoid this, we employ common random numbers by generating a single set of baseline draws and obtaining simulated values of Y for each parameter value by transforming these baseline draws (i.e., by translating them by θ). This induces smoothness in the simulated rejection probabilities.

Second, we ensure that key moments of the distribution of Y are matched exactly in the simulations. For example, when Y is standard bivariate normal, we standardize the baseline draws to have mean zero and identity covariance. In some instances we also impose symmetry by symmetrizing the baseline draws. For the bivariate normal case, this guarantees that the marginal distributions are symmetric around zero. In the boundary-robust testing application of Section 6.2, this ensures that the estimated power function of the two-sided t-test is symmetric, as implied by theory.

Third, we choose the random seed so that the estimated null rejection probabilities are close to the nominal level whenever the true null rejection probabilities are known to equal the nominal level. For example, in the linear IV model the CLR test is similar by construction. If the simulation draws happen to make the CLR test appear to overreject at some points in $\bar{\Theta}_0$, then the algorithm—which searches for a WAP-maximizing test that weakly dominates the CLR test while satisfying the size constraint—may struggle to converge. Ensuring that simulated null rejection rates are close to their theoretical values avoids these problems and yields more reliable numerical results.

5 Theoretical Justification of Numerical Implementation

In this section, we present the theoretical justification for both the inner and outer loop algorithms for computing approximate WAP-maximizing tests that can produce a power envelope for a given ad hoc test.

5.1 Inner Loop

We now present the theoretical result guaranteeing the convergence of the inner loop algorithm for minimizing $\tilde{\phi}$ in (10), which as discussed in Section 4.1, also solves the primal problem (9). This algorithm is a dual (projected) subgradient descent algorithm and therefore its convergence properties readily follow from known results in the literature. Let

$$\tilde{\phi}^{(k)} = \min_{i=0,1,\ldots,k} \tilde{\phi}(\widetilde{\Lambda}^{(i)}).$$

Theorem 3

The inner loop algorithm described in Section 4.1 satisfies for any $k \geq 1$,[8]

$$\tilde{\phi}^{(k)} - \min_{\widetilde{\Lambda} \geq 0} \tilde{\phi}(\widetilde{\Lambda}) \leq \sqrt{M_0 \max\{1-\alpha, \alpha\}} \frac{\|\widetilde{\Lambda}^{(0)} - \widetilde{\Lambda}^*\|_2^2 + \sum_{i=0}^{k} \tilde{h}_i^2}{2 \sum_{i=0}^{k} \tilde{h}_i} =: \widetilde{\Gamma}_k.$$

The tuning parameter choice

$$\tilde{h}_i = \frac{\varepsilon}{\sqrt{M_0 \max\{1-\alpha, \alpha\}}}, \qquad for \ i = 0, 1, \dots$$

with $\varepsilon > 0$ implies

$$\tilde{\phi}^{(N)} - \min_{\widetilde{\Lambda} \geq 0} \tilde{\phi}(\widetilde{\Lambda}) \leq \varepsilon$$

for all $N \geq \max\{1-\alpha, \alpha\} M_0 \|\widetilde{\Lambda}^{(0)} - \widetilde{\Lambda}^*\|_2^2 / \varepsilon^2$.

The convergence rate is controlled by $\widetilde{\Gamma}_k$, which depends on the starting value $\widetilde{\Lambda}^{(0)}$ through $\|\widetilde{\Lambda}^{(0)} - \widetilde{\Lambda}^*\|_2^2$ and on the tuning parameter sequence $\{\tilde{h}_k\}$. It holds that $\widetilde{\Gamma}_k \to 0$ as $k \to \infty$ if $\sum \tilde{h}_k = \infty$ and $\sum \tilde{h}_k^2 < \infty$. The particular proposed choice of the tuning parameters $\{\tilde{h}_k\}$ is taken from Section 3.2.3 of Nesterov et al. (2018). This choice is convenient as it only depends on known quantities and guarantees that the algorithm finds an ε-solution if the number of iterations is sufficiently large.

Elliott et al. (2015) also propose an algorithm for approximating $\tilde{\varphi}_{\widetilde{\Lambda}^*}$ but they do not provide any convergence guarantees analogous to our Theorem 3 for our inner loop algorithm. However, while we were writing this paper, Fernández et al. (2025) proposed an algorithm that slightly modifies that of Elliott et al. (2015) and produces a test that provably approximates $\tilde{\varphi}_{\widetilde{\Lambda}^*}$ with convergence guarantees analogous to our Theorem 3 as well as bounds the Monte Carlo error. Their results also enable them to bound the size-distortions of their approximately optimal test and provide specific theoretically-grounded tuning parameter recommendations for an updating parameter, number of iterations and initial weights. Either the original algorithm of Elliott et al. (2015), the new modified algorithm of Fernández et al. (2025) or the linear-programming algorithm of Moreira and Moreira (2013) can be used in place of our inner loop algorithm if desired.

---

[8]As there are potentially multiple solutions $\widetilde{\Lambda}^*$ of the dual problem, the distance on the right hand side can be interpreted as the smallest distance of $\Lambda^{(0)}$ to the set of solutions.

5.2 Outer Loop

Moving to the theoretical result guaranteeing the convergence of the outer loop algorithm for approximating the weight function that solves (4), we again note that since this algorithm is a (projected) subgradient descent algorithm, its convergence properties readily follow from known results in the literature. Let

$$\phi^{(k)} = \min_{i=0,1,\ldots,k} \phi(\bar{\Omega}^{(i)}).$$

Theorem 4

The outer loop algorithm described in Section 4.2 satisfies for any $k \geq 1$,[9]

$$\phi^{(k)} - \min_{\bar{\Omega} \in \Delta_{M_1}} \phi(\bar{\Omega}) \leq \sqrt{M_1} \frac{\|\Omega^{(0)} - \bar{\Omega}^*\|_2^2 + \sum_{i=0}^{k} h_i^2}{2\sum_{i=0}^{k} h_i} =: \Gamma_k.$$

The tuning parameter choice

$$h_i = \frac{\varepsilon}{\sqrt{M_1}}, \qquad \text{for } i = 0,1,\ldots$$

with $\varepsilon > 0$ implies

$$\phi^{(N)} - \min_{\bar{\Omega} \in \Delta_{M_1}} \phi(\Omega) \leq \varepsilon$$

for all $N \geq M_1 \|\Omega^{(0)} - \bar{\Omega}^*\|_2^2 / \varepsilon^2$.

The convergence rate is controlled by Γ_k, which depends on the starting value $\Omega^{(0)}$ through $\|\Omega^{(0)} - \bar{\Omega}^*\|_2^2$ and on the tuning parameter sequence $\{h_k\}$. It holds $\Gamma_k \to 0$ as $k \to \infty$ if $\sum h_k = \infty$ and $\sum h_k^2 < \infty$, which as with the inner loop is a convenient choice taken from Section 3.2.3 of Nesterov et al. (2018).

6 Applications

In this section, we apply our results and algorithms to shed new light on the optimality properties of a test whose optimality properties have already been thoroughly analyzed in the literature as well as a brand-new test that has not yet. Specifically, we analyze the optimality of the CLR test of Moreira (2003) and the test implied by the inequality-imposed confidence

---

[9]As there are potentially multiple solutions $\bar{\Omega}^*$ of (4), the distance on the right hand side can be interpreted as the smallest distance of $\bar{\Omega}^{(0)}$ to the set of solutions.

interval of Cox (2024). Throughout this section, the nominal level α is taken equal to 5%.

6.1 CLR Test in the Homoskedastic Linear IV Model

As alluded to in the introduction, Andrews et al. (2006, 2008) (AMS06 and AMS08, henceforth) investigate the optimality of the CLR test in the homoskedastic linear IV model, holding the variance matrix for the reduced-form errors fixed in their analysis (coined as the "fixed-Ω design" by Van de Sijpe and Windmeijer, 2023). In this setting, AMS06 construct an asymptotically efficient two-sided point-wise power envelope for invariant similar tests as the rejection probabilities arising from a collection of point-optimal invariant similar tests, finding the striking result that the CLR test numerically attains this power envelope. AMS08 further strengthen this result by showing that one essentially obtains the same power envelope without imposing similarity. However, Andrews et al. (2019) (AMY, henceforth) provide a counterpoint to these optimality results by finding that the CLR test falls short of a point-wise power envelope that is constructed by varying the value of the parameter of interest and keeping its hypothesized value fixed, rather than the more standard analysis that keeps the value of the parameter of interest fixed and varies its hypothesized value. In recent work, Van de Sijpe and Windmeijer (2023) show that this latter analysis is essentially the same as the standard analysis that varies the parameter of interest while keeping its hypothesized value fixed but instead of holding the variance matrix for the reduced-form errors fixed, fixes the variance matrix of the structural and first-stage errors (coined as the "fixed-Σ design" by Van de Sijpe and Windmeijer, 2023). Van de Sijpe and Windmeijer (2023) further argue that the fixed-Ω design implicitly favors the power function of the CLR test and that the fixed-Σ design is better suited for analyzing power in cases of low to moderate endogeneity as well as differing signs of the parameter of interest and the correlation between the structural errors in the IV model.

These recent results of AMY and Van de Sijpe and Windmeijer (2023) motivate us to revisit the optimality analysis of the CLR test, looking at both the fixed-Ω and fixed-Σ designs. Analyzing the CLR test under the fixed-Σ design using our new numerical optimality assessment enables us to assess whether the previously examined point-wise power envelopes of AMY may simply be setting an unattainable power bound. Before proceeding, we formally introduce the linear IV model and the CLR test, and briefly describe the power envelope of AMS06. We closely follow the notation of AMS06.

The linear IV model is given by

$$y_1 = y_2\beta + u,$$

$$y_2 = Z\pi + v_2,$$

where $y_1, y_2, u, v_2 \in \mathbb{R}^n$, $Z \in \mathbb{R}^{n \times k}$, $\beta \in \mathbb{R}$, and $\pi \in \mathbb{R}^k$. Here, y_1, y_2 and Z are observed, where y_1 denotes the outcome of interest, y_2 the (potentially) endogenous regressor of interest, and Z the instruments.[10] The random vectors u and v_2 are unobserved structural error terms. We assume that Z is fixed. Plugging the reduced-form equation for y_2 into the structural equation for y_1, we obtain the reduced-form equation for y_1, i.e.,

$$y_1 = Z\pi\beta + v_1,$$

where $v_1 = u + v_2\beta$. The resulting set of reduced-form equations can be written as[11]

$$Y = Z\pi a' + V,$$

where

$$Y = [y_1, y_2], \ V = [v_1, v_2] \text{ and } a = (\beta, 1)'.$$

Let V_i denote the ith row of V and assume that V_i is iid across i with

$$V_i \sim \mathcal{N}(0, \Omega).$$

Having defined the linear IV model, we can now formally state the testing problem of interest, which is given by

$$H_0: \ \beta = \beta_0, \ \pi \in \mathbb{R}^k \text{ vs. } H_1: \ \beta \in \mathbb{R} \backslash \{\beta_0\}, \ \pi \in \mathbb{R}^k.$$

We follow AMS06 and further simplify this testing problem. To that end, we define the following transformations of Y:

$$S = (Z'Z)^{-1/2} Z'Y b_0 \cdot (b_0' \Omega b_0)^{-1/2} \text{ and } T = (Z'Z)^{-1/2} Z'Y \Omega^{-1} a_0 \cdot (a_0' \Omega a_0)^{-1/2},$$

where $b_0 = (1, -\beta_0)'$ and $a = (\beta_0, 1)'$. Lemma 2 of AMS06 shows that S and T are jointly normally distributed and independent. AMS06 argue that the coordinate system used to

---

[10]We omit additional exogenous regressors without loss of generality, as the above model can always be obtained by partialling them out.

[11]Here, we use Y to be consistent with the notation in AMS06. This Y should, however, not be confused with the random element Y that enters the general testing problem in Section 3.1.

specify S and T should not affect inference and, therefore, only consider statistics that are invariant to rotation of the coordinate system. This is achieved by considering test statistics that only depend on S and T through

$$Q = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix} = \begin{bmatrix} S'S & S'T \\ T'S & T'T \end{bmatrix};$$

see Theorem 1 in AMS06 and the surrounding discussion. The distribution of Q is noncentral Wishart and, importantly, depends on π only through

$$\lambda = \pi' Z' Z \pi,$$

which implies that if we restrict our attention to test statistics based on Q, the testing problem of interest simplifies to

$$H_0: \ \beta = \beta_0, \ \lambda \in \mathbb{R}_+ \ \text{ vs. } \ H_1: \ \beta \in \mathbb{R} \backslash \{\beta_0\}, \ \lambda \in \mathbb{R}_+. \tag{11}$$

The LR statistic for testing (11) can be written as

$$\mathrm{LR} = \frac{1}{2} \left(Q_S - Q_T + \sqrt{(Q_S - Q_T)^2 + 4 Q_{ST}^2} \right).$$

Moreira (2003) observes that Q_T is a sufficient statistic for λ under H_0. This, in turn, allows the construction of a size α test using *conditional* critical values. In particular, the CLR test rejects when $\mathrm{LR} > \mathrm{cv}_{1-\alpha}(Q_T)$, where $\mathrm{cv}_{1-\alpha}(Q_T)$ is such that $P_{\beta_0}(\mathrm{LR} > \mathrm{cv}_{1-\alpha}(Q_T)|Q_T) = \alpha$.

The power envelope proposed by AMS06 is constructed from a collection of point-optimal invariant similar two-sided (POIS2) tests. Invariance is imposed by restricting attention to tests that only depend on the data through Q. In the context at hand, imposing similarity is tantamount to relying on conditional critical values, conditional on the observed value of Q_T. This avoids the need for approximating the least-favorable distribution, as done in AMS08. The two-sidedness that AMS06 consider is such that the resulting test is asymptotically efficient, i.e., the test has the same power as the two-sided t-test when instruments are strong ($\lambda = \infty$). The corresponding point-optimal test, POIS2, puts equal weight on (β^*, λ^*) and (β_2^*, λ_2^*), where the latter is a function of the former, ensuring asymptotic efficiency. The corresponding power envelope is then mapped out by the POIS2 test as (β^*, λ^*) varies.

6.1.1 Fixed-Ω Design

AMS06 consider a fixed-Ω design in their simulations, setting $\Omega_{11} = \Omega_{22} = 1$ and $\beta_0 = 0$. They numerically compare the power of the CLR test to their point-wise power envelope as functions of β and λ for various values of k and Ω_{12} (ρ in their paper). For given values of k and Ω_{12}, they consider several λ-slices of the CLR power function and the power envelope, for which β is varied over a fine grid for a given value of λ. An example of such a λ-slice is given in Figure 3.

(a) Screenshot of Figure 1(c) in AMS06   **(b)** Power of CLR test together with our APE
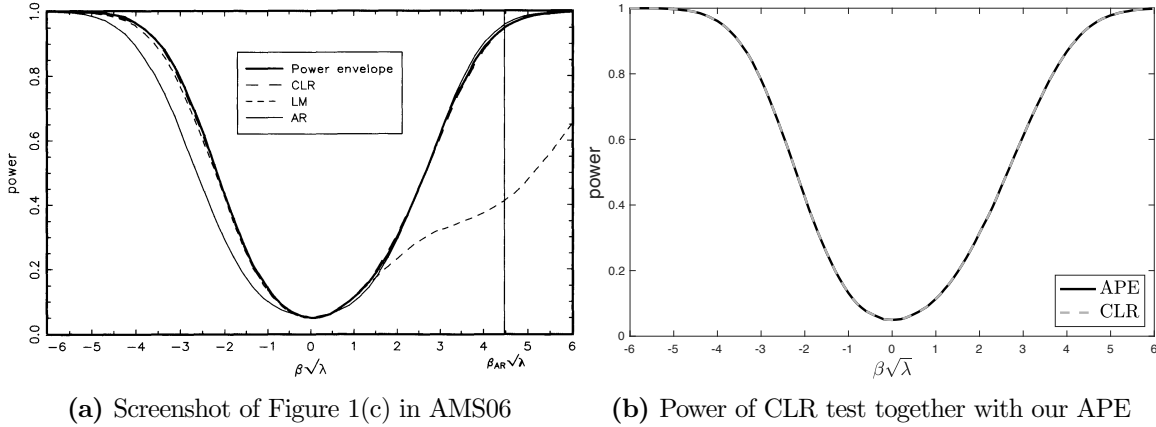
Figure 3: Power of CLR test as a function of β for $\lambda = 5$ together with AMS06's power envelope and our approximate power envelope for $k = 5$ and $\Omega_{12} = 0.5$.

Panel (a) of Figure 3 is a screenshot of Figure 1(c) in AMS06, which shows the power of the CLR test (and the Anderson-Rubin and Lagrange multiplier tests of Anderson and Rubin, 1949 and Kleibergen, 2002) together with their proposed power envelope as β varies and $\lambda = 5$ for $k = 5$ and $\Omega_{12} = 0.5$. The power of the CLR test and the power envelope are virtually indistinguishable (or "very close" in this and many other figures in AMS06 where k, Ω_{12}, and λ are varied), which leads AMS06 to conclude that the CLR test attains the power envelope "in a numerical sense".

We reconsider the above testing problem for $k = 5$ and $\Omega_{12} = 0.5$. In the case at hand, Q takes the role of Y and λ replaces δ.[12] We set $\lambda_S = 75$ and $\lambda_{SP} = 160$, where Q_T takes the role of $D(Y)$, since the probability that $Q_T > \lambda_{SP}$ is less than 0.01 whenever $\lambda \leq \lambda_S$. The standard test is implemented as the Lagrange multiplier test, which rejects

---

[12]In principle, we could use (S,T) as Y, replace δ by π and investigate the optimality of the CLR test in the class of asymptotically efficient tests. However, the computational cost of approximating of Λ increases rapidly in the dimension of the nuisance parameter.

H_0 when $Q^2_{ST}/Q_T > \chi^2_{1-\alpha}(1)$.[13] Note that switching to the standard test (for large λ) is equivalent to imposing asymptotic efficiency as defined by AMS06. We obtain our APE using $\bar{\Theta}_0 = \{(\beta,\lambda) : \beta = 0, \ \lambda \in \{1,5,10,...,30,40,...,170\}\}$ and $\bar{\Theta}_1 = \{(b/\sqrt{\lambda},\lambda) : b \in \{-4,-3,-2,2,3,4\}, \ \lambda \in \{1,5,10,...,30,40,...,170\}\}$. Implementation details, including the number of simulation draws and the choices of $\{h_k\}$, Θ_0^f and Θ_1^f are given in Appendix A.3.

Panel (b) of Figure 3 shows the power function of the CLR test (for the same parameter constellation as in panel (a)) together with our APE. As in panel (a), the power of the CLR test is indistinguishable from the power envelope for the λ-slice under consideration. Instead of considering multiple λ-slices, we produce a heatmap showing the difference between our APE and the power of the CLR test over a grid of values for β and λ.[14] This heatmap is shown in Figure 4.
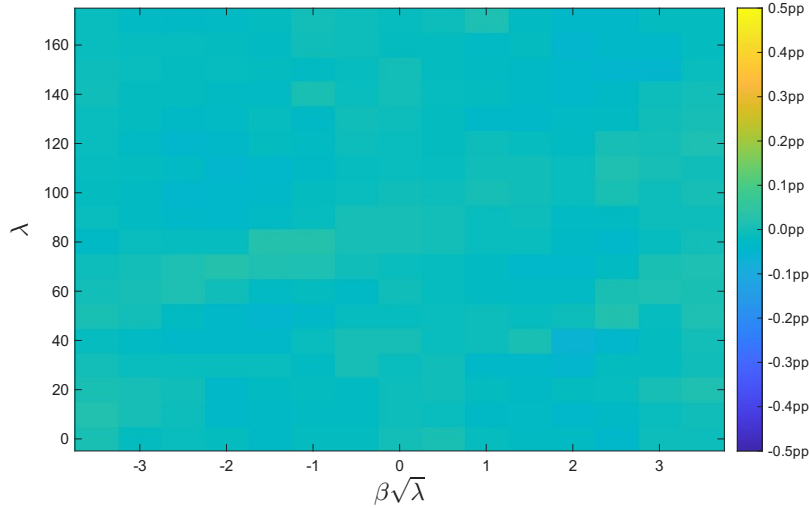

Figure 4: Difference between APE and power of CLR test in percentage points for $k=5$ and $\Omega_{12}=0.5$.

The differences between the APE and the power of the CLR test are very close to zero over the entire grid; note that the scale on the right ranges only from -0.5 percentage points (pp) to 0.5pp. In fact, the largest difference (in absolute value) is below 0.1pp. We therefore conclude that the CLR test is effectively optimal (for $k=5$ and $\Omega_{12}=0.5$). Our finding is complementary to that of AMS06: we find that the CLR test is (effectively) optimal

---

[13]For $\lambda > \lambda_S = 75$, the CLR test, the Lagrange multiplier test and the two-sided t-test all approximately coincide.

[14]The grid is given by $\{(b/\sqrt{\lambda},\lambda) : b \in \{-3.5,-3,...,3.5\}, \ \lambda \in \{0.1,10,20,...,170\}\}$. Note that this grid does not coincide with $\bar{\Theta}_0 \cup \bar{\Theta}_1$ and that the underlying rejection probabilities are evaluated using draws that are independent from those used to obtain the APE. This protects from a potential winner's curse, cf. Andrews et al. (2024).

in the class of invariant asymptotically efficient tests, while AMS06 find the stronger result of point-optimality in the smaller class of tests that are similar and two-sided (in the sense that they impose).[15]

6.1.2 Fixed-Σ Design

As shown by Van de Sijpe and Windmeijer (2023), AMY consider a fixed-Σ design in their simulations, where the variance matrix of the structural errors u and v_2, Σ, is held constant. AMY set $\Sigma_{11} = \Sigma_{22} = 1$ and consider several values for Σ_{12} (ρ_{uv} in their paper). Their Figure 1 shows the power functions of the CLR test (and the Anderson-Rubin test) and the power envelope of AMS06 based on the POIS2 test for different values of Σ_{12}, keeping $k = 10$ and $\lambda = 15$ fixed. In contrast to AMS06, however, AMY set $\beta = 0$ and vary β_0.

(a) Screenshot of Figure 1(b) in AMY   **(b)** Power of CLR test together with our APE
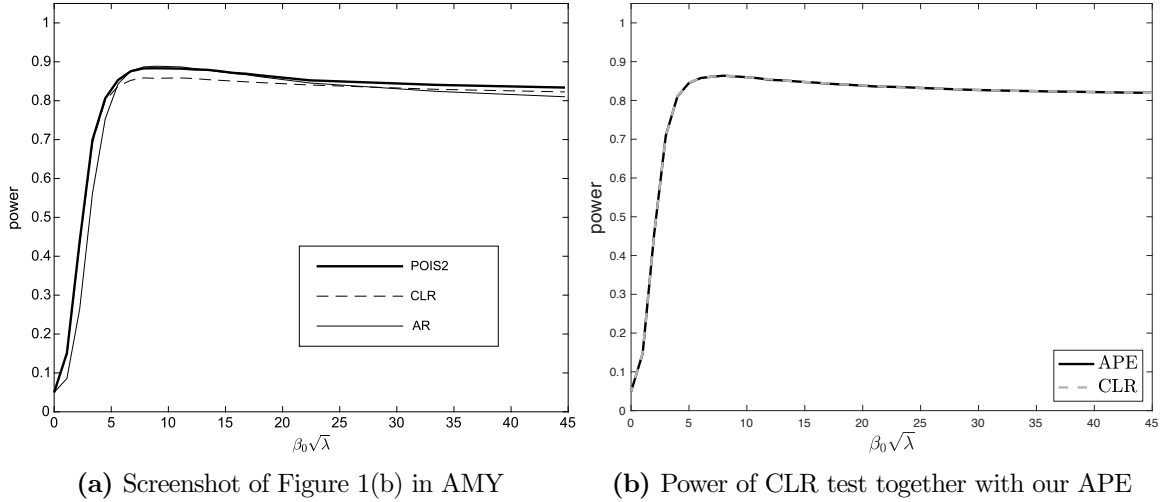
Figure 5: Power of CLR test as a function of β_0 for $\lambda = 15$ together with AMS06's power envelope and our APE for $k = 10$ and $\Sigma_{12} = 0.5$.

Panel (a) of Figure 5 is a screenshot of Figure 1(b) in AMY, where $\Sigma_{12} = 0.5$. The power of the CLR test is on the power envelope for values of $\beta_0\sqrt{\lambda}$ between 0 and 5, but drops below for values further away from 0. The maximal gap of the CLR test's power function with the POIS2 power envelope is around 3–4pp. Based on this finding, AMY conclude that the finding of AMS06 "that the CLR test is essentially on the [...] power envelope does not hold...". However, as discussed extensively above, this does not mean that the CLR test is not optimal since it is unclear if the point-wise power envelope used by AMY is attainable. We therefore reconsider the testing problem considered in AMY, where $k = 10$

---

[15]Our finding is also complementary to that of AMS08: while we find a weaker result (optimality as opposed to point-optimality), we do not impose two-sidedness (in the sense that they impose).

and $\Sigma_{12} = 0.5$, and construct our APE. We set $\lambda_S = 160$ and $\lambda_{SP} = 320$. The underlying $\bar{\Theta}_0$ and $\bar{\Theta}_1$ as well as additional implementation details are provided in Appendix A.3.

Panel (b) of Figure 5 displays the power function of the CLR test (for the same parameter constellation as in panel (a)) together with our APE. We find that the CLR test attains our APE, at least for the λ-slice under consideration. As before, we produce a heatmap showing the difference between our APE and the power of the CLR test over a grid of values for β and λ, while setting $\beta_0 = 0$. While this setting seemingly differs from the setting of AMY (where $\beta = 0$ while β_0 varies), it follows from Corollary 1 of Van de Sijpe and Windmeijer (2023) and the subsequent discussion that the resulting power curves are simply mirror images of one another.[16] The heatmap is shown in Figure 6.
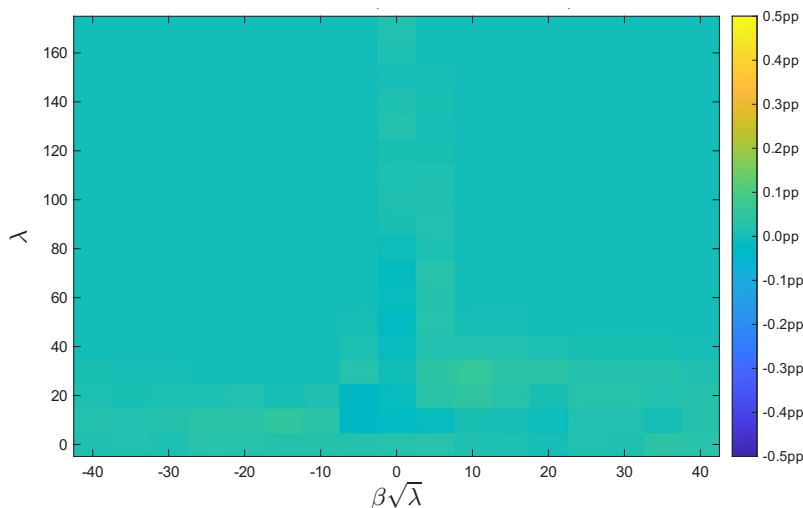

Figure 6: Difference between APE and power of CLR test in percentage points for $k = 10$ and $\Sigma_{12} = 0.5$.

The heatmap in Figure 6 uses the same scale as the heatmap in Figure 4 and leads us to the same conclusion: the difference between our APE and the power function of the CLR test is very close to zero across the entire grid. We conclude that the CLR test is, in fact, effectively optimal *under both* the fixed-Ω and fixed-Σ designs.

6.2 Test Implied by IICI in Boundary Problem

A line of recent work has examined testing problems involving uniformly-valid inference when nuisance parameters may lie on or near the boundary of the parameter space.[17]

---

[16]That is, the power of the CLR test for testing $H_0 : \beta = \beta_0$ when the true value of β, say β^*, is equal to 0, is equal to the power of the CLR test for testing $H_0 : \beta = 0$ when $\beta^* = -\beta_0$.

[17]See, for example, Andrews and Guggenberger (2010), Ketz (2018), Ketz and McCloskey (2025) and Cavaliere et al. (2025).

Focusing on the special case of a scalar nuisance parameter to analyze the properties of a new confidence interval proposed by Cox (2024), we work with the Gaussian experiment

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta \\ \delta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

where β is the scalar parameter of interest and $\delta \geq 0$ is a scalar nuisance parameter. As discussed in Section 3.1, this formulation corresponds to the asymptotic behavior of a broad class of finite-sample models (see, e.g., Elliott et al., 2015 and Ketz and McCloskey, 2025). Formally, the testing problem of interest is given by[18]

$$H_0 : \beta = \beta_0, \ \delta \geq 0 \ \text{vs.} \ H_1 : \beta \neq \beta_0, \ \delta \geq 0. \tag{12}$$

Cox (2024) proposes a new confidence interval for β, called the inequality-imposed confidence interval (IICI). The IICI has several desirable properties: (i) it is easy to compute, (ii) it does not require simulation or tuning parameters, (iii) it is adaptive, and (iv) it has weakly shorter length than the standard two-sided confidence interval that ignores the information contained in Y_2. The IICI is constructed as the union of the standard two-sided CI when Y_2 is large and positive, the two-sided confidence interval that imposes $\delta = 0$ when Y_2 is large and negative, and the intersection of the two when Y_2 is close to zero; see Panel A of Figure 1 in Cox (2024) for a visual illustration. Formally, the lower and upper bounds of the $(1-\alpha)$-nominal IICI are given by

$$\begin{cases} Y_1 - z_{1-\alpha/2} & \text{if } Y_2 > c, \\ Y_1 - \rho Y_2 - \sqrt{1-\rho^2} z_{1-\alpha/2} & \text{otherwise,} \end{cases}$$

and

$$\begin{cases} Y_1 + z_{1-\alpha/2} & \text{if } Y_2 > -c, \\ Y_1 - \rho Y_2 + \sqrt{1-\rho^2} z_{1-\alpha/2} & \text{otherwise,} \end{cases}$$

respectively, where $c = \frac{1-\sqrt{1-\rho^2}}{\rho} z_{1-\alpha/2}$ and $z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the $\mathcal{N}(0,1)$ distribution. The test implied by the IICI rejects H_0 in (12) when β_0 lies outside of it.

To gauge the optimality of the IICI, Cox (2024) compares the WAP of the test im-

---

[18]This limiting Gaussian formulation also applies to problems for which nuisance parameters are restricted to be greater/less than or equal to any known value via a simple affine transformation.
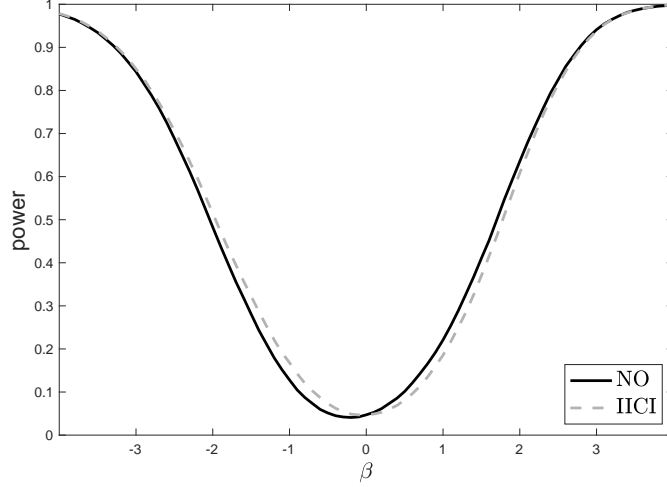
Figure 7: Power of the nearly optimal test of Elliott et al. (2015) and the test implied by the IICI for $\rho = 0.7$ and $\delta = 1$.

plied by the IICI with the power bound on WAP derived in Elliott et al. (2015) (EMW, henceforth), using equal weight on $\beta = -2$ and $\beta = 2$ and uniform weights on $\delta \in [0,9]$. For $\rho = 0.7$, the WAP of the IICI is 53.1% and the WAP of the EMW power bound is 53.5%. Using $\epsilon = 0.005$ (following EMW), Cox (2024) concludes that the test implied by the IICI is nearly optimal in the sense of EMW, as its WAP lies within ϵ of the power bound. However, some ambiguity with respect to the optimality of the test remains:[19] as shown in Figure 7, the power functions of the test implied by the IICI and the nearly optimal test of EMW (under the above weights) cross, which prevents us from concluding whether the test implied by the IICI is optimal or dominated.

To address this ambiguity, we compute our APE for the testing problem with $\rho = 0.7$. We follow EMW and choose $\delta_{\mathrm{SP}} = 6$. The standard test is the two-sided t-test. We use $\bar{\Theta}_1 = \{(b,d) : b \in \{-3,-2,-1,1,2,3\},\ d \in \{0,0.5,...,8\}\}$ and follow EMW in discretizing Θ_0 in terms of "base" distributions. Details on this and other implementation choices are provided in Appendix A.3. As in Section 6.1, we produce a heatmap showing the difference between our APE and the power of the test implied by the IICI over a grid of values for β and δ, fixing $\beta_0 = 0$; see Figure 8.

The scale of the heatmap is the same as for the heatmaps in Section 6.1. The difference between the APE and the power of the test implied by the IICI is close to zero over a large

---

[19]This is related to the fact that ϵ is given as an absolute (rather than a relative) value, making it difficult to interpret what it means for the difference in WAPs to be "small".
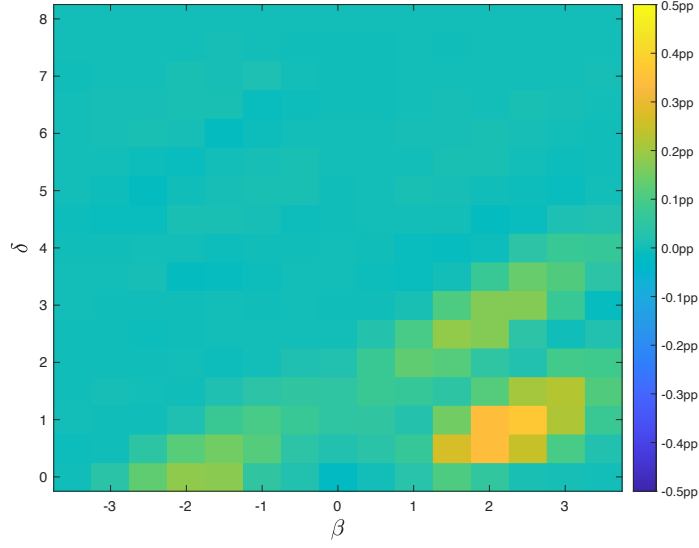
Figure 8: Difference between APE and power of test implied by the IICI in percentage points for $\rho=0.7$.

portion of the grid. However, there are some values for which the test implied by the IICI falls short of the APE—the maximal difference is about 0.3pp and occurs at $(\beta,\delta)=(2,1)$. We, therefore, conclude that the test implied by the IICI is effectively dominated, albeit by a very small margin. In fact, the WAPs of the test underlying the APE and the test implied by the IICI are 52.532% and 52.529%, respectively, with the corresponding weights given in Appendix A.3. Abstracting from numerical approximations, the WAP of the test implied by the IICI is thus within 0.00003 of the lowest possible upper bound (given $\bar{\Theta}_1$), improving on the finding in Cox (2024) and providing a strong argument in favor of using the IICI in practice. Although the test is effectively dominated, the loss in WAP is negligible considering the simplicity of the procedure.

A Appendix

A.1 Proofs of Main Results

Proof of Theorem 1: Starting with the first claim of the theorem, for any $\varphi \in \Phi_\alpha$,

$$\min_{j=1,\ldots,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu = \inf_{\bar{\Omega} \in \Delta_{M_1}} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu. \tag{13}$$

This can be seen as follows: for any $\bar{\Omega} \in \Delta_{M_1}$,

$$\sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu \geq \min_{j=1,\ldots,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu \sum_{i=1}^{M_1} \omega_i = \min_{j=1,\ldots,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu,$$

and therefore

$$\inf_{\bar{\Omega} \in \Delta_{M_1}} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu \geq \min_{j=1,\ldots,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu.$$

Moreover, since the canonical basis vectors of $\mathbb{R}^{M_1}$ are included in Δ_{M_1},

$$\min_{j=1,\ldots,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu \geq \inf_{\bar{\Omega} \in \Delta_{M_1}} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu$$

and (13) follows. Thus,

$$\sup_{\varphi \in \Phi_\alpha} \min_{j=1,\ldots,M_1} \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu = \sup_{\varphi \in \Phi_\alpha} \inf_{\bar{\Omega} \in \Delta_{M_1}} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu.$$

Further, in Lemma 2, we show that the Ky Fan minimax theorem implies

$$\sup_{\varphi \in \Phi_\alpha} \inf_{\bar{\Omega} \in \Delta_{M_1}} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu = \inf_{\bar{\Omega} \in \Delta_{M_1}} \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu. \tag{14}$$

The fact that $\Phi^* \neq \emptyset$ follows directly from Lemma 1.

To see that $\Delta_{M_1}^* \neq \emptyset$, note that since the criterion function on the right hand side of

35

(14) is continuous in $(\bar{\Omega}, \varphi)$, $\phi : \Delta_{M_1} \to \mathbb{R}$ given by

$$\phi(\bar{\Omega}) = \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu,$$

is lower semi-continuous and therefore by compactness of Δ_{M_1}, the infimum is attained at some $\bar{\Omega}^* = (\omega_1^*, ..., \omega_{M_1}^*) \in \Delta_{M_1}$:

$$\sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j^* \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu = \inf_{\bar{\Omega} \in \Delta_{M_1}} \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu. \qquad (15)$$

To show (7), note that for any $\varphi^* \in \Phi^*$

$$\sum_{j=1}^{M_1} \omega_j^* \int (\varphi^* - \varphi_{ah}) f_{\theta_j} d\nu \geq \inf_{\bar{\Omega} \in \Delta_{M_1}} \sum_{j=1}^{M_1} \omega_j \int (\varphi^* - \varphi_{ah}) f_{\theta_j} d\nu$$

$$= \sup_{\varphi \in \Phi_\alpha} \inf_{\bar{\Omega} \in \Delta_{M_1}} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu$$

$$= \inf_{\bar{\Omega} \in \Delta_{M_1}} \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu$$

$$= \sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j^* \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu,$$

where the inequality follows by definition of the infimum, the first equality by the definition of Φ^* and (13), the second by (14) and the last by (15). By the definition of the supremum, we also have

$$\sup_{\varphi \in \Phi_\alpha} \sum_{j=1}^{M_1} \omega_j^* \int (\varphi - \varphi_{ah}) f_{\theta_j} d\nu \geq \sum_{j=1}^{M_1} \omega_j^* \int (\varphi^* - \varphi_{ah}) f_{\theta_j} d\nu.$$

Therefore, the inequality in the preceding display can be replaced by an equality and (7) follows. $\square$

Proof of Theorem 2: Under the given conditions, the first statement follows directly from Lemma 3. The second statement follows from the first, together with equation (7), which implies that φ_1^* and φ_2^* are WAP-maximizing tests with respect to $\bar{\Omega}_1^*$ and $\bar{\Omega}_2^*$. $\square$

Proof of Proposition 1: For any $\varphi \in \Phi$,

$$\int \varphi g d\nu - \sum_{i=1}^{M_0} \tilde{\lambda}_i \left(\int \varphi f_{\tilde{\theta}_i} d\nu - \alpha \right) = \int \varphi \left(g - \sum_{i=1}^{M_0} \tilde{\lambda}_i f_{\tilde{\theta}_i} \right) d\nu + \alpha \sum_{i=0}^{M_0} \tilde{\lambda}_i,$$

which is clearly maximized at $\tilde{\varphi}_{\widetilde{\Lambda}}$. $\square$

Proof of Theorem 3: The proof follows from Theorem 3.2.2 of Nesterov et al. (2018). It only remains to bound the Lipschitz constant of $\tilde{\phi}$ and derive its subdifferential, which is done in Lemma 4. The second claim follows from equations (3.2.16)–(3.2.21) in section 3.2.3 in Nesterov et al. (2018). $\square$

Proof of Theorem 4: The proof follows from Theorem 3.2.2 of Nesterov et al. (2018). It only remains to bound the Lipschitz constant of ϕ and derive its subdifferential, which is done in Lemma 5. The second claim follows from equations (3.2.16)–(3.2.21) in section 3.2.3 in Nesterov et al. (2018). $\square$

A.2 Auxiliary Results

Lemma 1 (Based on Theorem 6.1.5 of Rüschendorf (2014))
For $\mathbb{P}_\theta$ the probability measure associated with density function f_θ, if $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\} \ll \nu$, there exists a test $\varphi^* \in \Phi_\alpha$ for any $\alpha \in [0,1]$ satisfying

$$\inf_{\theta \in \Theta_1} \int (\varphi^* - \varphi_{ah}) f_\theta d\nu = \sup_{\varphi \in \Phi_\alpha} \inf_{\theta \in \Theta_1} \int (\varphi - \varphi_{ah}) f_\theta d\nu.$$

Proof: The existence result relies on properties of the set of test functions Φ and the weak-$*$ topology on the space of measurable and ν-essentially bounded functions $\mathcal{L}_\infty(\nu) := \{\varphi \in \mathcal{L}(\mathcal{Y}) : \exists K \in \mathbb{R} \text{ s.th. } |\varphi| \leq K \nu - \text{a.e.}\}$, where $\mathcal{L}(\mathcal{Y})$ denotes the set of measurable functions mapping $\mathcal{Y}$ into $\mathbb{R}$. As usual, we endow $\mathcal{L}_\infty(\nu)$ with the norm $\|\varphi\|_\infty := \inf\{K : |\varphi| \leq K \ \nu - \text{a.e.}\}$ and treat functions $\varphi, \varphi' \in \mathcal{L}_\infty(\nu)$ as equal when $\varphi = \varphi'$ ν-a.e. Further, a sequence $\{\varphi_n\} \subset \mathcal{L}_\infty(\nu)$ converges with respect to the weak-$*$ topology if

$$\int \varphi_n \delta d\nu \to \int \varphi \delta d\nu \qquad \forall \delta \in \mathcal{L}_1(\nu),$$

as $n \to \infty$, where $\mathcal{L}_1(\nu)$ denotes the space of (equivalence classes) of absolutely ν-integrable functions.

The importance of the weak-$*$ topology lies first in the Banach-Alaoglu theorem, according to which the closed unit ball $B = \{\varphi \in L_\infty(\nu) : \|\varphi\|_\infty \leq 1\}$ is weak-$*$ compact.

Since $\mathcal{P} \ll \nu$, Φ is a weak-$*$ closed subset of B and therefore Φ is weak-$*$ compact. Secondly, the power function $\beta : \Phi \to [0,1]^{\Theta}$, $\varphi \mapsto \beta_{\varphi}$ is weak-$*$ continuous. In order to see this, note that β is continuous if and only if $\varphi \mapsto \beta_{\varphi}(\theta)$ is continuous for all $\theta \in \Theta$. Now, since for all $\theta \in \Theta$,

$$\beta_{\varphi}(\theta) = \int \varphi \frac{d\mathbb{P}_{\theta}}{d\nu} d\nu$$

with $\frac{d\mathbb{P}_{\theta}}{d\nu} \in \mathcal{L}_1(\nu)$, the continuity follows by the definition of weak-$*$ convergence, given above. Thus, not only is Φ weak-$*$ compact, but also the set of power function differences $G := \{\beta_{\varphi} - \beta_{ah} : \Theta \to [-1,1] \mid \forall \theta \in \Theta \text{ s.th. } \beta_{\varphi}(\theta) - \beta_{ah}(\theta) = \int (\varphi - \varphi_{ah}) f_{\theta} d\nu\}$ is weak-$*$ compact as a continuous image of a compact set.[20]

These observations readily imply the claimed existence of such minimax-type tests. Since

$$\Phi_{\alpha} = \bigcap_{\theta \in \Theta_0} \{\varphi \in \Phi : \beta_{\varphi}(\theta) \leq \alpha\},$$

Φ_{α} is a weak-$*$ closed subset of Φ and hence weak-$*$ compact. Now, let $\kappa : \Phi_{\alpha} \to [-1,1]$ be defined as $\kappa(\varphi) := \inf_{\theta \in \Theta_1} \{\beta_{\varphi}(\theta) - \beta_{ah}(\theta)\}$. By continuity of $\varphi \mapsto \beta_{\varphi}(\theta) - \beta_{ah}(\theta)$, κ is upper semicontinuous and hence attains its supremum on the compact set Φ_{α}. Thus, there exists $\varphi^* \in \Phi_{\alpha}$ such that

$$\kappa(\varphi^*) = \inf_{\theta \in \Theta_1} \{\beta_{\varphi^*}(\theta) - \beta_{ah}(\theta)\} = \sup_{\varphi \in \Phi_{\alpha}} \inf_{\theta \in \Theta_1} \{\beta_{\varphi}(\theta) - \beta_{ah}(\theta)\}$$

and φ^* is a minimax test. $\square$

Lemma 2

If $\mathcal{P} = \{\mathbb{P}_{\theta} : \theta \in \Theta\} \ll \nu$, then

$$\sup_{\varphi \in \Phi_{\alpha}} \inf_{\Omega \in \Delta} \iint (\varphi - \varphi_{ah}) f_{\theta} d\nu d\Omega(\theta) = \inf_{\Omega \in \Delta} \sup_{\varphi \in \Phi_{\alpha}} \iint (\varphi - \varphi_{ah}) f_{\theta} d\nu d\Omega(\theta),$$

where Δ denotes the set of probability measures over either Θ_1 or Δ_{M_1}.

Proof: Note that Δ and Φ_{α} are convex and that $r(\Omega, \varphi) := \iint (\varphi - \varphi_{ah}) f_{\theta} d\nu d\Omega(\theta)$ is linear in both of its arguments. Further, by similar arguments to those in the proof of Lemma 1, $\varphi \mapsto r(\Omega, \varphi)$ is weak-$*$ continuous for all $\Omega \in \Delta$ and Φ_{α} is weak-$*$ compact. We can further endow Δ with the weak topology on $\mathcal{L}_1(\nu)$ when we identify Ω with its implied density

---

[20]See Theorem 6.1.4 in Rüschendorf (2014).

$h_\Omega = \int f_\theta d\Omega(\theta) \in \mathcal{L}_1(\nu)$. Hence, the Ky Fan minimax theorem[21] applies and implies

$$\sup_{\varphi \in \Phi_\alpha} \inf_{\Omega \in \Delta} r(\Omega, \varphi) = \inf_{\Omega \in \Delta} \sup_{\varphi \in \Phi_\alpha} r(\Omega, \varphi).$$

This proves the claim. $\square$

Lemma 3

Suppose $\alpha \in (0,1)$ and $\Omega \in \Delta$, where Δ denotes the set of probability measures over either Θ_1 or Δ_{M_1}, and consider the problem

$$\sup_{\varphi \in \Phi_\alpha} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu. \tag{16}$$

Then, the following statements hold:

1. *There exists a solution to (16), i.e.,*

$$\Phi_\Omega^* := \underset{\varphi \in \Phi_\alpha}{\mathrm{argmax}} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu \neq \emptyset.$$

2. *Let $\mathcal{M}_0$ denote the set of finite measures over Θ_0. For any $\Lambda \in \mathcal{M}_0$,*

$$\Phi_{\Omega,\Lambda}^* := \underset{\varphi \in \Phi}{\mathrm{argmax}} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu - \int \left(\int \varphi f_\theta d\nu - \alpha \right) d\Lambda(\theta) \neq \emptyset$$

   and any $\varphi \in \Phi_{\Omega,\Lambda}^*$ can be written as

$$\varphi = \begin{cases} 1 & , \text{ if } \int f_\theta d\Omega(\theta) > \int f_\theta d\Lambda(\theta) \\ \varkappa & , \text{ if } \int f_\theta d\Omega(\theta) = \int f_\theta d\Lambda(\theta) \\ 0 & , \text{ if } \int f_\theta d\Omega(\theta) < \int f_\theta d\Lambda(\theta) \end{cases}$$

   for some $\varkappa \in \Phi$.

3. *It holds that (16) is equal to*

$$\inf_{\Lambda \in \mathcal{M}_0} \sup_{\varphi \in \Phi} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu - \int \left(\int \varphi f_\theta d\nu - \alpha \right) d\Lambda(\theta).$$

---

[21]See Theorem 2 in Fan (1953).

4. *Let*

$$\mathcal{M}_0^* = \operatorname*{argmin}_{\Lambda \in \mathcal{M}_0} \sup_{\varphi \in \Phi} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu - \int \left(\int \varphi f_\theta d\nu - \alpha \right) d\Lambda(\theta).$$

Assume that $\mathcal{M}_0^* \neq \emptyset$ and that the event

$$\left\{ y \in \mathcal{Y} : \int f_\theta d\Omega(\theta) = \int f_\theta d\Lambda^*(\theta) \right\}$$

has ν-measure zero for all $\Lambda^* \in \mathcal{M}_0^*$, such that for any $\Lambda^* \in \mathcal{M}_0^*$ there is a ν-a.e. unique element in $\Phi_{\Omega,\Lambda^*}^*$ given by

$$\varphi_{\Omega,\Lambda^*}^* = \mathbb{1}\left\{ \int f_\theta d\Omega(\theta) > \int f_\theta d\Lambda^*(\theta) \right\}.$$

Then, for any $\Lambda_1^*, \Lambda_2^* \in \mathcal{M}_0^*$ and any $\tilde{\varphi}_1, \tilde{\varphi}_2 \in \Phi_\Omega^*$

$$\varphi_{\Omega,\Lambda_1^*}^* = \varphi_{\Omega,\Lambda_2^*}^* = \tilde{\varphi}_1 = \tilde{\varphi}_2 \ \nu\text{-a.e.}$$

5. *If $\Lambda^* \in \mathcal{M}_0^*$ with $\Lambda^*(\Theta_0) > 0$, then $\tilde{\Lambda}^* = \Lambda^*/\Lambda^*(\Theta_0)$ is a least favorable distribution.*

6. *If there exists a least favorable distribution Λ_Ω^*, then $\operatorname{cv}_\Omega \Lambda_\Omega^* \in \mathcal{M}_0^*$, where cv_Ω denotes the critical value of the Neyman-Pearson test of $\int f_\theta d\Lambda_\Omega^*$ against $\int f_\theta d\Omega$.*

Proof: We prove each part of the lemma as follows:

1. This follows by similar arguments to those in the proof of Lemma 1.

2. Rewrite the criterion function as follows

$$\int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu - \int \left(\int \varphi f_\theta d\nu - \alpha \right) d\Lambda(\theta)$$

$$= \int \varphi \left(\int f_\theta d\Omega(\theta) - \int f_\theta d\Lambda(\theta) \right) d\nu + \alpha \Lambda(\Theta_0) - \int \varphi_{ah} \int f_\theta d\Omega(\theta) d\nu.$$

This function is maximized by any $\varphi \in \Phi$ of the claimed form.

3. By 2. and its proof,

$$\inf_{\Lambda \in \mathcal{M}_0} \sup_{\varphi \in \Phi} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu - \int \left(\int \varphi f_\theta d\nu - \alpha \right) d\Lambda(\theta)$$

$$= \inf_{\Lambda \in \mathcal{M}_0} \int \left(\int f_\theta d\Omega(\theta) - \int f_\theta d\Lambda(\theta) \right)_+ d\nu + \alpha \Lambda(\Theta_0) - \int \varphi_{ah} \int f_\theta d\Omega(\theta) d\nu,$$

where $(a)_+ = \max\{0,a\}$ for any $a \in \mathbb{R}$. The claim now follows along the same lines as the proof of Theorem 4 in Krafft and Witting (1967).

4. For any $\Lambda^* \in \mathcal{M}_0^*$ and any $\tilde{\varphi} \in \Phi_\Omega^*$, we have

$$\int (\tilde{\varphi} - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu$$

$$= \sup_{\varphi \in \Phi_\alpha} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu$$

$$= \inf_{\Lambda \in \mathcal{M}_0} \sup_{\varphi \in \Phi} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu - \int \left(\int \varphi f_\theta d\nu - \alpha \right) d\Lambda(\theta)$$

$$= \sup_{\varphi \in \Phi} \int (\varphi - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu - \int \left(\int \varphi f_\theta d\nu - \alpha \right) d\Lambda^*(\theta)$$

$$\geq \int (\tilde{\varphi} - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu - \int \left(\int \varphi^* f_\theta d\nu - \alpha \right) d\Lambda^*(\theta)$$

$$\geq \int (\tilde{\varphi} - \varphi_{ah}) \int f_\theta d\Omega(\theta) d\nu,$$

where we have used $\tilde{\varphi} \in \Phi_\Omega^*$ in the first equality, part 3. in the second equality, $\Lambda^* \in \mathcal{M}_0^* \neq \emptyset$ in the third equality, the definition of the supremum together with $\tilde{\varphi} \in \Phi_\alpha \subset \Phi$ in the first inequality and that $\tilde{\varphi} \in \Phi_\alpha$ in the last inequality. This implies that $\tilde{\varphi} \in \Phi_{\Omega,\Lambda^*}^*$ and thus $\tilde{\varphi} = \varphi_{\Omega,\Lambda^*}^* \nu$-a.e. Since $\tilde{\varphi}$ and Λ^* are arbitrary, the desired result follows.

5./6. The problem in part 3. leads to the same set of solutions as the problem given in equations (22) and (23) in Krafft and Witting (1967). The proof therefore follows along the same lines as the proof of Theorem 12 in Krafft and Witting (1967). $\square$

Lemma 4

The following statements hold:

1. *$\tilde{\phi}$ is Lipschitz continuous such that*

$$|\tilde{\phi}(\widetilde{\Lambda}) - \tilde{\phi}(\widetilde{\Lambda}')| \leq \sqrt{M_0 \max\{1-\alpha,\alpha\}} \|\widetilde{\Lambda} - \widetilde{\Lambda}'\|_2$$

for any $\widetilde{\Lambda}, \widetilde{\Lambda}' \geq 0$.

2. For any $\widetilde{\Lambda} \geq 0$,

$$\left(\alpha - \int \varphi_{\widetilde{\Lambda}} f_{\tilde{\theta}_i} d\nu \right)_{i=1}^{M_0} \in \partial \tilde{\phi}(\widetilde{\Lambda}).$$

Proof: We prove each part of the lemma as follows:

1. Take any $\widetilde{\Lambda}, \widetilde{\Lambda}' \geq 0$. Then, by standard bounds for the supremum and the Cauchy-Schwarz inequality,

$$|\tilde{\phi}(\widetilde{\Lambda}) - \tilde{\phi}(\widetilde{\Lambda}')| \leq \sup_{\varphi \in \Phi} \left| \sum_{i=1}^{M_0} \{\tilde{\lambda}_i - \tilde{\lambda}_i'\} \left(\int \varphi f_{\tilde{\theta}_i} d\nu - \alpha \right) \right|$$

$$\leq \|\widetilde{\Lambda} - \widetilde{\Lambda}'\|_2 \sup_{\varphi \in \Phi} \sqrt{\sum_{i=1}^{M_0} \left(\int \varphi f_{\tilde{\theta}_i} d\nu - \alpha \right)^2}.$$

The supremum on the right hand side is achieved by either $\varphi = 1$ or $\varphi = 0$ and thus

$$\sup_{\varphi \in \Phi} \sqrt{\sum_{i=1}^{M_0} \left(\int \varphi f_{\tilde{\theta}_i} d\nu - \alpha \right)^2} = \sqrt{M_0 \max\{1 - \alpha, \alpha\}}.$$

2. By definition of $\tilde{\phi}$ as a maximum and Proposition 1, for any $\widetilde{\Lambda}, \widetilde{\Lambda}' \geq 0$,

$$\tilde{\phi}(\widetilde{\Lambda}') \geq \int \varphi_{\widetilde{\Lambda}} d\nu - \sum_{i=1}^{M_0} \tilde{\lambda}_i' \left(\int \varphi_{\widetilde{\Lambda}} f_{\tilde{\theta}_i} d\nu - \alpha \right)$$

$$= \int \varphi_{\widetilde{\Lambda}} d\nu - \sum_{i=1}^{M_0} \tilde{\lambda}_i \left(\int \varphi_{\widetilde{\Lambda}} f_{\tilde{\theta}_i} d\nu - \alpha \right) - \sum_{i=1}^{M_0} \{\tilde{\lambda}_i' - \tilde{\lambda}_i\} \left(\int \varphi_{\widetilde{\Lambda}} f_{\tilde{\theta}_i} d\nu - \alpha \right)$$

$$= \tilde{\phi}(\widetilde{\Lambda}) + \sum_{i=1}^{M_0} \{\tilde{\lambda}_i' - \tilde{\lambda}_i\} \underbrace{\left(\alpha - \int \varphi_{\widetilde{\Lambda}} f_{\tilde{\theta}_i} d\nu \right)}_{=:k_i}$$

and thus $\tilde{\phi}(\widetilde{\Lambda}') \geq \tilde{\phi}(\widetilde{\Lambda}) + k^{\intercal}(\widetilde{\Lambda}' - \widetilde{\Lambda})$, for all $\widetilde{\Lambda}' \geq 0$, i.e., $k \in \partial \tilde{\phi}(\widetilde{\Lambda})$. $\square$

Lemma 5
The following statements hold:

1. ϕ is Lipschitz continuous such that

$$|\phi(\bar\Omega)-\phi(\bar\Omega')|\le\sqrt{M_1}\|\bar\Omega-\bar\Omega'\|_2$$

for any $\bar\Omega,\bar\Omega'\in\Delta_{M_1}$.

2. For any $\bar\Omega\in\Delta_{M_1}$,

$$\left(\int(\varphi_{\bar\Omega}^*-\varphi_{ah})f_{\theta_i}d\nu\right)_{i=1}^{M_1}\in\partial\phi(\bar\Omega).$$

Proof: We prove each part of the lemma as follows:

1. Take any $\bar\Omega,\bar\Omega'\in\Delta_{M_1}$. Then, by standard bounds for the supremum and the Cauchy-Schwarz inequality,

$$|\phi(\bar\Omega)-\phi(\bar\Omega')|\le\sup_{\varphi\in\Phi_\alpha}\left|\sum_{i=1}^{M_1}\{\omega_i-\omega_i'\}\int(\varphi-\varphi_{ah})f_{\theta_i}d\nu\right|$$

$$\le\|\bar\Omega-\bar\Omega'\|_2\sup_{\varphi\in\Phi_\alpha}\sqrt{\sum_{i=1}^{M_1}\left(\int(\varphi-\varphi_{ah})f_{\theta_i}d\nu\right)^2}.$$

The supremum on the right hand side can be bounded above using

$$\sup_{\varphi\in\Phi}\sqrt{\sum_{i=1}^{M_1}\left(\int(\varphi-\varphi_{ah})f_{\theta_i}d\nu\right)^2}\le\sqrt{M_1}.$$

2. By definition of ϕ as a maximum, for any $\bar\Omega,\bar\Omega'\in\Delta_{M_1}$,

$$\phi(\bar\Omega')\ge\sum_{i=1}^{M_1}\omega_i'\int(\varphi_{\bar\Omega}^*-\varphi_{ah})f_{\theta_i}d\nu$$

$$=\sum_{i=1}^{M_1}\omega_i\int(\varphi_{\bar\Omega}^*-\varphi_{ah})f_{\theta_i}d\nu+\underbrace{\sum_{i=1}^{M_1}(\omega_i'-\omega_i)\int(\varphi_{\bar\Omega}^*-\varphi_{ah})f_{\theta_i}d\nu}_{=:\gamma_i}$$

$$=\phi(\bar\Omega)+\gamma^\intercal(\bar\Omega'-\bar\Omega)$$

and thus $\gamma\in\partial\phi(\bar\Omega)$. $\square$

43

A.3 Details for the Applications Considered in Section 6

For all applications we use 300,000 simulation draws to evaluate rejection probabilities. The choice of $\{h_k\}$ is common across all applications. In particular,

$$
h_k = \begin{cases}
0.01 & \text{if } \min_{i=1,\dots,M_1}\{\gamma_i\} < -0.02 \\
0.001 & \text{if } -0.02 \leq \min_{i=1,\dots,M_1}\{\gamma_i\} < -0.002 \\
0.0001 & \text{otherwise,}
\end{cases}
$$

where $\gamma = (\gamma_1,\dots,\gamma_{M_1})$ is defined in step 2.(a) of the algorithm in Section 4.2. Furthermore, the number of iterations for the outer loop is set equal to 1,000.

A.3.1 Details for the Homoskedastic Linear IV Model

The draws of Q are obtained through baseline draws of S and T, given by two independent sets of draws from $\mathcal{N}(0, I_k)$, which are standardized across all simulation draws. The conditional critical values for the CLR test are obtained using 1,000,000 simulation draws. The choice of $\{\tilde{h}_k\}$ is common for the two designs (fixed-Ω and fixed-Σ). In particular,

$$
\tilde{h}_k = \begin{cases}
0.01 & \text{if } \max_{i=1,\dots,M_0}\{\tau_i\} > 0.02 \\
0.001 & \text{if } 0.02 \geq \max_{i=1,\dots,M_0}\{\tau_i\} > 0.002 \\
0.0001 & \text{otherwise,}
\end{cases}
$$

where $\tau = (\tau_1, \dots, \tau_{M_0})$ is defined in step 2.(a) of the algorithm in Section 4.1. Furthermore, the number of iterations for the inner loop is set equal to 1,000. In both designs, we take $\Theta_0^f = \{(\beta, \lambda) : \beta = 0, \lambda \in \{0, 2, \dots, 150\}\}$ and $\Theta_1^f = \{(b/\sqrt{\lambda}, \lambda) : b \in \{-3.5, -3, \dots, -0.5, 0.5, 1, \dots, 3.5\}, \lambda \in \{0.1, 10, 20, \dots, 170\}\}$. And in the fixed-$\Sigma$ design, we take $\bar{\Theta}_0 = \{(\beta, \lambda) : \beta = 0, \lambda \in \{1, 5, 10, 15, 20, 30, 40, 50, 70, \dots, 150, 175, \dots, 300\}\}$ and

$$
\begin{aligned}
\bar{\Theta}_1 = &\{(\beta, \lambda) : \beta \in \{-40, -30, -20, -10, -2.5, -1, 1, 6, 20, 30\}, \lambda = 1\} \\
&\cup \{(b/\sqrt{\lambda}, \lambda) : b \in \{-40, -30, -20, -10, -5, -1, 1, 5, 10, 20, 30\}, \lambda = 5\} \\
&\cup \{(b/\sqrt{\lambda}, \lambda) : b \in \{-40, -30, -20, -10, -6, -1, 1, 5, 10, 20, 30\}, \lambda = 10\} \\
&\cup \{(b/\sqrt{\lambda}, \lambda) : b \in \{-40, -30, -20, -10, -7.5, -2, 2, 10, 20, 30\}, \lambda = 15\} \\
&\cup \{(b/\sqrt{\lambda}, \lambda) : b \in \{-30, -10, -5, -3, 3, 7, 10, 20, 40\}, \lambda = 20\} \\
&\cup \{(b/\sqrt{\lambda}, \lambda) : b \in \{-3, -1, 2, 4, 6, 8\}, \lambda = 30\}
\end{aligned}
$$

$$\cup \{(b/\sqrt{\lambda},\lambda): b \in \{-3,2,4,6,8\}, \ \lambda = 40\}$$
$$\cup \{(b/\sqrt{\lambda},\lambda): b \in \{-3,2,4\}, \ \lambda \in \{50,70,...,150,175,...,300\}\}.$$

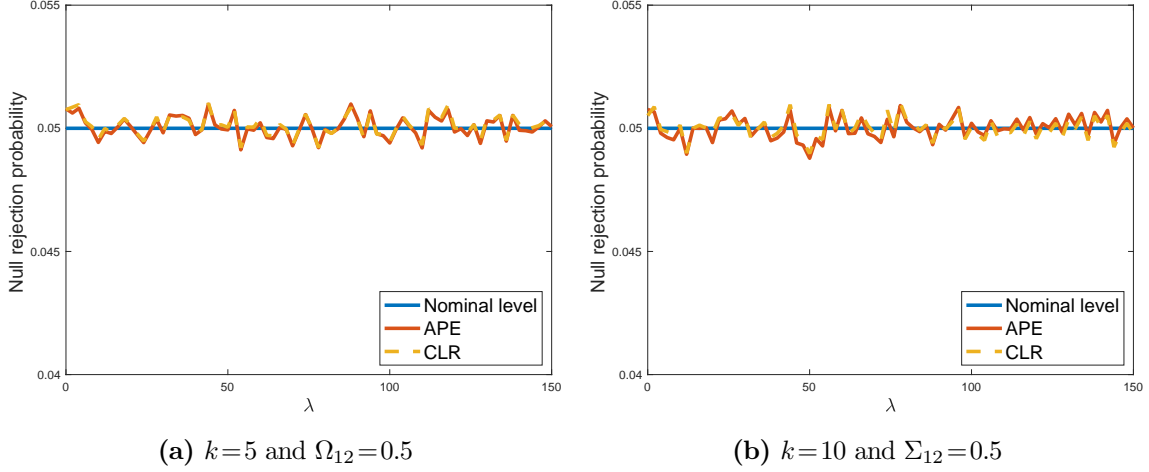**(a)** $k=5$ and $\Omega_{12}=0.5$ **(b)** $k=10$ and $\Sigma_{12}=0.5$

Figure 9: Null rejection probabilities of tests underlying APEs and CLR test.

Figure 9 shows the null rejection probabilities of the tests underlying the APEs and of the CLR test on Θ_0^f. Panel (a) reports the results for $k=5$ and $\Omega_{12}=0.5$ and panel (b) for $k=10$ and $\Sigma_{12}=0.5$. All probabilities are computed using 300,000 simulation draws, which are independent across the values in Θ_0^f. In each panel, the test underlying the corresponding APE and the CLR test yield virtually identical null rejection probabilities. Although the rejection probabilities of both tests can exceed the nominal level, we know that for the CLR test any such excess is solely due to simulation error since the CLR test controls size by construction. Because the two tests coincide so closely in both designs, we conclude that the tests underlying the APEs effectively control size.

A.3.2   Details for the Boundary Problem

The baseline draws used to obtain the draws of Y are standardized across all simulations and also symmetrized. The parameter $\tilde{h}_k$ is set equal to 0.01 for all k and the number of iterations for the inner loop is set equal to 1,000. As mentioned in the main text, we follow Elliott et al. (2015) in discretizing Θ_0 in terms of "base" distributions. Our "base" distributions are uniform distributions for δ on the following intervals: $[0,0.00001]$, $[0,0.04]$, $[1.99,2.01]$, $[0,0.5]$, $[0.5,1]$, ..., $[12,12.5]$. We take $\Theta_0^f = \{(\beta,\delta): \beta=0, \delta \in \{0,0.1,...,7\}\}$ and $\Theta_1^f = \{(\beta,\delta): \beta \in \{-3.5,-3,...,-0.5,0.5,1,...,3.5\}, \ \delta \in \{0,0.5,...,8\}\}$.

Figure 10 shows the null rejection probabilities of the test underlying the APE and the
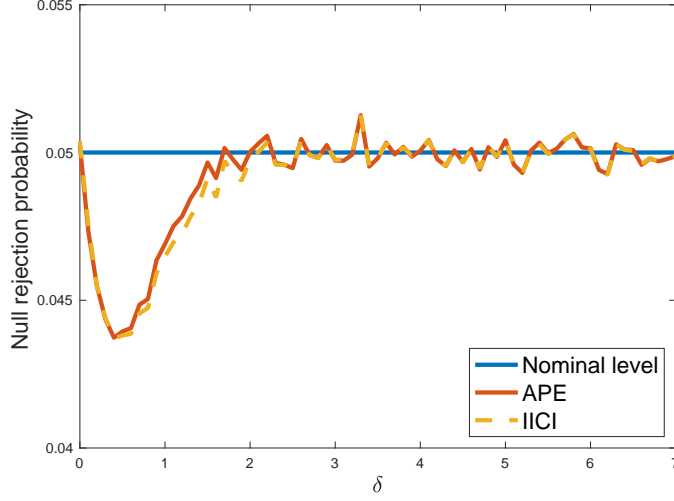
Figure 10: Null rejection probabilities of test underlying APE and test implied by IICI.

test implied by the IICI on Θ_0^f. All probabilities are computed using 300,000 simulation draws, which are independent across the values in Θ_0^f. Although the null rejection probabilities for both tests can exceed the nominal level, we know that for the test implied by the IICI this is only due to simulation error given that Cox (2024) proves that the IICI has uniformly correct coverage. Since the maximal null rejection probability of the test underlying the APE is very close to α and the null rejection probability of the test implied by the IICI over the entire grid, we conclude that the test underlying the APE effectively controls size.
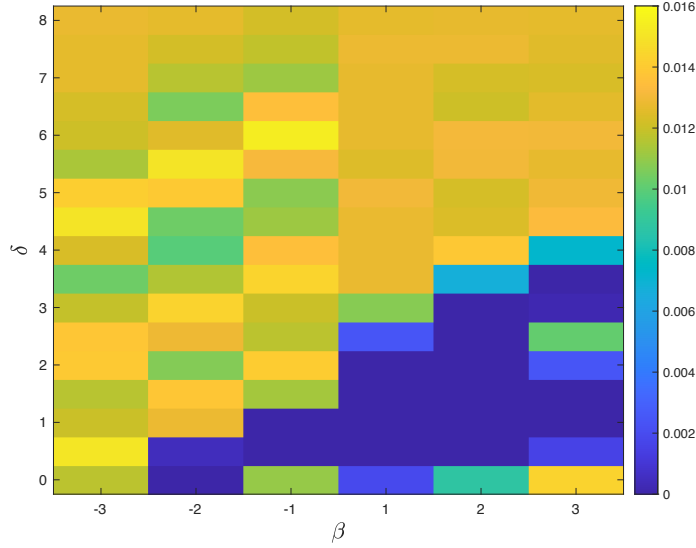

Figure 11: Weights underlying the APE for the test implied by the IICI

46

Figure 11 shows the weights underlying the APE corresponding to the test implied by the IICI. Apparently the test implied by the IICI does not prioritize power at positive alternatives $\beta > 0$ when the value of the nuisance parameter δ is small and $\rho = 0.7$. Interestingly, the region of the alternative parameter space receiving little weight by the APE also roughly corresponds to where we see the largest power differences between the APE and the power of the test implied by the IICI in Figure 8.

References

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20:46–63.

Andrews, D. W. K. and Guggenberger, P. (2010). Applications of subsampling, hybrid and size-correction methods. *Journal of Econometrics*, 158:285–305.

Andrews, D. W. K. and Guggenberger, P. (2019). Identification- and singularity-robust inference for moment condition models. *Quantitative Economics*, 10:1703–1746.

Andrews, D. W. K., Marmer, V., and Yu, Z. (2019). On optimal inference in the linear IV model. *Quantitative Economics*, 10(2):457–485.

Andrews, D. W. K., Moreira, M. J., and Stock, J. H. (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica*, 74:715–752.

Andrews, D. W. K., Moreira, M. J., and Stock, J. H. (2008). Efficient two-sided nonsimilar invariant tests in IV regression with weak instruments. *Journal of Econometrics*, 146(2):241–254.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62:1383–1414.

Andrews, I., Kitagawa, T., and McCloskey, A. (2024). Inference on winners. *Quarterly Journal of Economics*, 139:305–358.

Andrews, I. and Mikusheva, A. (2016). Conditional inference with a functional nuisance parameter. *Econometrica*, 84:1571–1612.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81:27–60.

Cavaliere, G., McCloskey, A., Pedersen, R. S., and Rahbek, A. (2025). Uniform Critical Values for Likelihood Ratio Tests in Boundary Problems. Working paper.

Cox, G. F. (2024). A simple and adaptive confidence interval when nuisance parameters satisfy an inequality. arXiv:2409.09962v1.

Elliott, G. and Müller, U. K. (2014). Pre and post break parameter inference. *Journal of Econometrics*, 180:141–157.

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

Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences*, 39(1):42–47.

Fernández, A. A., Blanchet, J., Olea, J. L. M., Qiu, C., Stoye, J., and Tan, L. (2025). Approximate least-favorable distributions and nearly optimal tests via stochastic mirror descent. Working Paper.

Guggenberger, P. and Huang, J. (2024). On the numerical approximation of minimax regret rules via fictitious play. Working Paper.

Guggenberger, P., Kleibergen, F., and Mavroeidis, S. (2019). A more powerful subvector Anderson Rubin test in linear instrumental variables regression. *Quantitative Economics*, 10(2):487–526.

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

Ketz, P. and McCloskey, A. (2025). Short and simple confidence intervals when the directions of some effects are known. *Review of Economics and Statistics*, 107:820–834.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70:1781–1803.

Krafft, O. and Witting, H. (1967). Optimale tests und ungünstigste verteilungen. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 7(4):289–302.

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

Luenberger, D. G. (1997). *Optimization by vector space methods*. John Wiley & Sons.

Montiel Olea, J. L. (2020). Admissible, similar tests: A characterization. *Econometric Theory*, 36(2):347–366.

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

Moreira, H. and Moreira, M. J. (2019). Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors. *Journal of Econometrics*, 213:398–433.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71(1027-1048).

Moreira, M. J., Sharifvaghefi, M., and Ridder, G. (2021). Optimal invariant tests in an instrumental variables regression with heteroskedastic and autocorrelated errors. arXiv:1705.00231v2.

Müller, U. K. (2011). Efficient Tests under a Weak Convergence Assumption. *Econometrica*, 79(2):395–435.

Nesterov, Y. et al. (2018). *Lectures on convex optimization*, volume 137. Springer.

Rüschendorf, L. (2014). *Mathematische statistik*, volume 62. Springer.

Van de Sijpe, N. and Windmeijer, F. (2023). On the power of the conditional likelihood ratio and related tests for weak-instrument robust inference. *Journal of Econometrics*, 235:82–104.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54:426–482.

Wang, W. and Carreira-Perpinan, M. A. (2013). Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. arXiv:1309.1541.