Inference for Linear Conditional Moment Inequalities*

Isaiah Andrews Jonathan Roth Ariel Pakes

October 31, 2022

Abstract

We show that moment inequalities in a wide variety of economic applications have a particular linear conditional structure. We use this structure to construct uniformly valid confidence sets that remain computationally tractable even in settings with nuisance parameters. We first introduce least favorable critical values which deliver non-conservative tests if all moments are binding. Next, we introduce a novel conditional inference approach which ensures a strong form of insensitivity to slack moments. Our recommended approach is a hybrid technique which combines desirable aspects of the least favorable and conditional methods. The hybrid approach performs well in simulations calibrated to Wollmann (2018), with favorable power and computational time comparisons relative to existing alternatives.

Keywords: Moment Inequalities, Subvector Inference, Uniform Inference

JEL Codes: C12

1 Introduction

Moment inequalities are a useful tool in a wide range of fields in empirical economics. As described in recent reviews by Ho & Rosen (2017) and Molinari (2020), moment inequalities

*We thank Tim Armstrong, Gary Chamberlain, Ivan Canay, Jiafeng Chen, Kirill Evdokimov, Jerry Hausman, Bulat Gafarov, Hiroaki Kaido, Adam McCloskey, Francesca Molinari, Whitney Newey, Ashesh Rambachan, Bas Sanders, Jesse Shapiro, Brit Sharoni, Xiaoxia Shi, Joerg Stoye, Chris Walker, and participants at several seminars for helpful comments, and thank Thomas Wollmann for helpful discussion of his application. We are grateful to Xiaoxia Shi and Matt Thirkettle for sharing code and providing advice on its implementation. Andrews gratefully acknowledges financial support from the NSF under Grant 1654234. Roth gratefully acknowledges financial support from an NSF Graduate Research Fellowship under Grant DGE1144152.
Andrews: iandrews@fas.harvard.edu. Roth: jonathan_roth@brown.edu. Pakes: apakes@fas.harvard.edu

can be used to exploit the most direct implications of utility or profit maximization for inference in both single-agent settings and games. They can also be used to weaken parametric, behavioral, measurement, and selection assumptions in a range of problems. Inference using moment inequalities raises practical challenges, however, particularly when there are nuisance parameters (e.g. coefficients on control variables) that are not of direct interest.

A first challenge is obtaining tests that are computationally tractable. Many available moment inequality methods rely on test inversion over a grid for the full parameter vector (including the nuisance parameters), but the computational costs of such approaches grow exponentially in the dimension of the parameter vector. This has necessitated the development of alternative approaches that either profile out (i.e. optimize over) the nuisance parameters in the computation of the test statistic (e.g., Bugni et al. 2017) or use computational shortcuts to form projection confidence sets without computing the test for all values of the nuisance parameter (e.g., Kaido et al. 2019*a*). Nevertheless, computation can still be challenging when the dimension of the nuisance parameters is moderate or large.

A second challenge is obtaining tests with good power. When there are nuisance parameters, tests for the parameter of interest can be obtained via projection, but this can lead to conservative tests with poor power (see Bugni et al. 2017, Kaido et al. 2019*a*). Moreover, the power of many existing procedures can be negatively affected by the inclusion of non-binding moments, yet it may not be clear *ex ante* which of the moments implied by economic theory will be binding. This has prompted a variety of approaches to eliminate or reduce the sensitivity of moment inequality tests to slack moments including work by D. Andrews & Soares (2010), D. Andrews & Barwick (2012), Romano et al. (2014), Chernozhukov et al. (2015), Bugni et al. (2017), and Belloni et al. (2018), among many others.

In this paper, we show that a variety of applications of moment inequalities have a particular structure that can be exploited to address these challenges. Specifically, we study settings with moment inequalities of the form $E[Y_i(\beta_0) - X_i(\beta_0)\delta|Z_i] \leq 0$, where β_0 is the parameter of interest, δ is a nuisance parameter, and $X_i(\beta_0)$ is a function of Z_i. That is, we study conditional moment inequalities that (a) are linear in the nuisance parameters δ, and (b) have conditional variance (given the instruments Z_i) that does not depend on the nuisance parameters. In Section 2, we highlight several recent applications of moment inequalities that have this structure, including interval-valued regression and revealed preference models in industrial organization.

Under this linear conditional structure, the profiled studentized max statistic can be represented as a linear program, and can thus be computed efficiently even when the dimension of the nuisance parameters is large. Linear conditional structure is also helpful for deriving tractable critical values, since it implies that the asymptotic variance of the moments (conditional on the instruments) does not depend on the value of the nuisance parameters. These features allow us to construct profiling-based confidence sets that rely on test inversion only for the target parameter and not for the nuisance parameters, and thus are computationally tractable even when the dimension of the nuisance parameters is large. We exploit this linear conditional structure to develop two tests that have different desirable properties, as well as a third hybrid approach that combines the two and is our preferred approach.

Our first approach is based on the least-favorable (LF) asymptotic distribution of our test statistic. We show that the distribution of the test statistic is increasing (in the sense of first-order stochastic dominance) in the mean of the moments, and thus the least-favorable distribution under the null corresponds with the case where the mean of all of the moments is zero.[1] It is then straightforward to calculate a critical value under the least-favorable distribution via simulation. The LF test has exact asymptotic size when all of the moments are simultaneously binding in population, and thus avoids conservativeness from projection in this case. A downside of the LF test, however, is that its power can be negatively affected by the inclusion of slack moments.

To address sensitivity to slack moments, we introduce a second test based on a novel conditioning argument. We condition on the Lagrange multipliers in the optimization to compute the test statistic, which intuitively correspond with the set of binding moments in sample after profiling out the nuisance parameters. We show that the set of values of the moments for which a particular Lagrange multiplier is optimal is a polyhedron, and we then derive critical values using results from Lee et al. (2016) on polyhedral conditioning events. We prove that the resulting conditional test is insensitive to slack moments in the strong sense that, as a subset of the moments becomes arbitrarily slack, the conditional test converges to the test that drops these moments ex-ante. A downside of the conditional test, however, is that it may have poor power in settings where multiple moments are approximately equally violated. Finally, given the different relative strengths of the LF and conditional approaches, we introduce a hybrid approach that combines the LF and conditional approaches, while avoiding the conservativeness of Bonferroni approaches.

The critical values for all of our tests are based on a normal approximation to the distribution of the moments conditional on the instruments. If this normal approximation holds

[1]This presumes that the set of data-generating processes considered allows for the possibility that all moments bind simultaneously. If not, then the distribution used for our critical value is an upper bound on the least-favorable distribution under the null.

exactly with known variance, our proposed tests control size in finite samples. In Section 4 we provide regularity conditions under which size control in this finite sample normal model translates to uniform asymptotic size control over a large class of data-generating distributions. A desirable feature of our proposed tests is that they they achieve uniform asymptotic size control without having to specify a sequence of tuning parameters that converges at a certain rate. Nevertheless, our tests do require the researcher to make some choices. To use the hybrid test, the researcher must specify the size of the "first-stage" least favorable test κ, although this choice only affects the power of the test and not its asymptotic validity.[2] Additionally, although conditional moment inequalities can imply an infinite number of unconditional moments, our tests only exploit the implications of k unconditional moments that must be specified by the researcher. We provide heuristic guidance on the choice of the k moments in Section 5.1.

To explore the numerical performance of our methods, we apply our techniques in simulations calibrated to Wollmann (2018)'s study of the US auto bailout. We consider designs with up to ten nuisance parameters, and find that our proposed tests remain computationally tractable and have good size control in all specifications. The power of the hybrid test is similar to or better than that of the LF and conditional tests in all specifications, and we thus recommend the hybrid approach among our proposed procedures. We also find that the hybrid test has power dominating that of the projection-based tests of D. Andrews & Soares (2010) and Kaido et al. (2019*a*) in all specifications for which we are able to compute these tests, and computation time for the hybrid can be over 10 times faster than for either of the projection-based approaches. The hybrid approach is also competitive with the sCC and sRCC tests proposed in concurrent work by Cox & Shi (2022), although neither approach dominates the other across all specifications in terms of power or computational speed.

Related Literature. Cox & Shi (2022) consider the class of linear conditional moment inequalities introduced in this paper and propose tests based on a profiled quasi-likelihood ratio (QLR) statistic, whereas our tests are based on the profiled studentized max statistic. Cox & Shi (2022) and the present paper independently developed conditional testing approaches, but due to the difference in test statistics, the conditioning events and resulting tests are different. As discussed in Section 6, we find in our Monte Carlo simulations that our preferred test (the hybrid) has non-nested power with those proposed by Cox & Shi (2022), which accords with the intuition that tests based on the max and QLR statistics

[2] We recommend using $\kappa = \alpha/10$, and implement this choice in our simulations, following the recommendation for the two-step procedure in Romano et al. (2014).

direct power towards different parts of the parameter space.

Subvector inference for moment inequalities with linear parameters is also considered in Cho & Russell (2021), Gafarov (2019) and Flynn (2019). The setting in these papers differs from ours in that they consider unconditional moment inequalities, whereas we consider conditional moments; our paper also differs in that we allow the target parameters to potentially enter the moments non-linearly. One advantage of our approach relative to these previous papers is that we do not require a linear independence constraint qualification (LICQ) assumption, which restricts what moments can bind in population; see Section 4 for further discussion.[3] Another related paper is Kaido & Santos (2014), who consider efficient estimation and inference for the support function in settings with convex moment inequalities, which nests the problem of subvector estimation/inference in moment inequality models where all parameters enter linearly. Their approach, however, relies on a Slater constraint qualification that, for example, rules out moment equalities cast as inequalities. Our approach is thus complementary, since we do not require such a constraint qualification but also do not provide any formal efficiency results.

Our approach uses a profiled maximum statistic, and thus is also related to other profiling-based methods for moment inequalities. The profiling-based approach in Bugni et al. (2017) differs from ours in that it accommodates unconditional moment inequalities and does not require that the parameters enter the moments linearly. However, the linear structure that we consider enables highly-tractable computation since the profiled test statistic is computed with a linear program, and also enables us to develop tests that are uniformly asymptotically valid without relying on drifting sequences of tuning parameters. Belloni et al. (2018) build on the approach of Bugni et al. (2017) to develop methods for subvector inference with high-dimensional unconditional moments. Fang et al. (2021) propose a test based on the solution to a linear program that is applicable for a large class of problems that nests a high-dimensional version of the conditional linear inequalities considered in this paper, although at the cost of either introducing a sample-size dependent tuning parameter or obtaining a conservative test. Alternative approaches to subvector inference in moment inequality models include projection-based methods (e.g., Kaido et al. 2017); sub-sampling approaches (e.g., Romano & Shaikh 2008); and quasi-posterior Monte Carlo methods (Chen et al. 2018).[4] We emphasize that the aforementioned methods do not impose the specific

[3] Cho & Russell (2021) show that LICQ can be guaranteed to hold by adding a stochastic perturbation to the moments, at the expense of obtaining inference on an outer set of the sharp identified set.

[4] The approach of Chen et al. (2018) delivers inference on the identified set, rather than on points within the identified set.

linear conditional structure considered in this paper, and thus are applicable in a much wider class of problems. We provide comparisons to the profiling-based approach of Cox & Shi (2022) as well as two projection-based methods in our Monte Carlo simulations.

One important limitation of our approach is that — while we assume that conditional moment inequalities are satisfied — we consider tests that exploit only a fixed number (k) of the implied unconditional inequalities. This contrasts with papers that consider asymptotics in which the number of moments grows with the sample size, such as D. Andrews & Shi (2013) for full-vector inference, and Chernozhukov et al. (2015) and Belloni et al. (2018) for subvector inference.[5] An interesting open question is whether the tests proposed in this paper can be extended to the setting with a diverging number of moments. See Section 2 below for additional discussion.

2 Linear Conditional Moment Inequalities

We assume that we observe independent and identically distributed data D_i, $i=1,...,n$ drawn from an unknown distribution $P\in\mathcal{P}$, for a class $\mathcal{P}$ of distributions. The true values of the parameters (β,δ) are assumed to satisfy the conditional moment inequalities

$$E_{P_{D|Z}}[Y_i(\beta)-X_i(\beta)\delta|Z_i]\leq 0 \text{ almost surely,} \tag{1}$$

where Z_i is a subvector of D_i, $Y_i(\beta)=y(D_i,\beta)\in\mathbb{R}^k$ and $X_i(\beta)=x(Z_i,\beta)\in\mathbb{R}^{k\times p}$ for known functions $y(\cdot,\cdot)$ and $x(\cdot,\cdot)$, and $P_{D|Z}$ denotes the conditional distribution of D_i given Z_i. We are interested in β, while $\delta\in\mathbb{R}^p$ is a nuisance parameter. Specifically, we want to test that a given value β_0 belongs to the identified set for β, $\widetilde{H}_0:\beta_0\in B_I(P)$, where

$$B_I(P)=\left\{\beta\text{: there exists } \delta \text{ such that } E_{P_{D|Z}}[Y_i(\beta)-X_i(\beta)\delta|Z_i]\leq 0 \text{ almost surely}\right\} \tag{2}$$

is the set of values β such that there exists δ which makes (1) hold. For the remainder of the paper we omit the phrase "almost surely" for brevity. We call restrictions of the form (1) *linear conditional moment inequalities*. They have two key properties: first, the nuisance parameter δ enters linearly and, second, the Jacobian of the moments with respect to δ, $-X_i(\beta)$, is non-random conditional on Z_i. This structure implies that the variance of the moments conditional on Z_i does not depend on δ.

[5] Flynn (2019) considers a continuum of unconditional moment inequalities.

It is helpful to compare (1) to the linear regression model

$$Y_i^* = X_i^{*\prime}\delta + \varepsilon_i \text{ where } E_{P_{D|X^*}}[\varepsilon_i|X_i^*] = 0 \tag{3}$$

for $Y_i^* \in \mathbb{R}$ and $X_i^* \in \mathbb{R}^p$. Specifically, (1) implies

$$Y_i(\beta) = X_i(\beta)\delta + \varepsilon_i(\beta) \text{ where } E_{P_{D|Z}}[\varepsilon_i(\beta)|Z_i] \leq 0, \tag{4}$$

where $Y_i(\beta) \in \mathbb{R}^k$ and $X_i(\beta) \in \mathbb{R}^{k\times p}$. Linear conditional moment inequalities thus generalize the traditional regression model to (a) relax the conditional moment restriction on the errors ε_i to an inequality, (b) allow the possibility that there are instruments Z_i beyond the regressors X_i, (c) allow a vector-valued outcome, and (d) allow β to enter the moments non-linearly.

2.1 Examples of Linear Conditional Moment Inequalities

Linear conditional moment inequalities appear in a variety of economic applications.

Example 1 Linear conditional moment inequalities arise naturally from the linear regression model (3), and its instrumental variables generalization, when we observe only bounds on the outcome Y_i^*. Consider the model

$$Y_i^* = W_i\beta + V_i'\delta + \varepsilon_i,\ E_{P_{D|Z}}[\varepsilon_i|Z_i] = 0 \tag{5}$$

where V_i is a function of Z_i while W_i may be endogenous. For instance, β may be a causal effect of interest whereas V_i represents a set of control variables. This is a linear instrumental variables model where the error is mean-independent of the instrument.

As in e.g. Manski & Tamer (2002), suppose that rather than observing Y_i^* we instead observe bounds Y_i^L and Y_i^U where $Y_i^L \leq Y_i^* \leq Y_i^U$ with probability one. The model (5) implies that $E_{P_{D|Z}}[Y_i^L - W_i\beta - V_i'\delta|Z_i] \leq 0$ and $E_{P_{D|Z}}[W_i\beta + V_i'\delta - Y_i^U|Z_i] \leq 0$, so we obtain conditional moment inequalities. To cast these inequalities into our framework, suppose we are interested in inference on β, and for any vector of non-negative functions of the instruments $f(Z_i)$ let $Y_i(\beta) = (Y_i^L - W_i\beta, W_i\beta - Y_i^U)' \otimes f(Z_i)$, and $X_i = (V_i', -V_i')' \otimes f(Z_i)$, for "$\otimes$" the Kroneker product. This yields the moments $E_{P_{D|Z}}[Y_i(\beta) - X_i\delta|Z_i] \leq 0$, as desired.[6] $\triangle$

[6] Our approach to this application relies on the conditional moment restriction $E_{P_{D|Z}}[\varepsilon_i|Z_i] = 0$. As discussed by Ponomareva & Tamer (2011), this means that the identified set may be empty if the linear model is incorrect. For $Z_i = (W_i, V_i')'$, Beresteanu & Molinari (2008) assume only that $E_P[\varepsilon_i Z_i] = 0$ and conduct inference on the (necessarily nonempty) set of best linear predictors. Bontemps et al. (2012) study identification and inference, including specification tests, for a class of linear models with unconditional moment restrictions.

Example 2 Katz (2007) studies the impact of travel time on supermarket choice. Katz assumes that utility is additively separable in the basket of goods bought (B_i), the travel time to the supermarket chosen ($T_{i,s}$), and the cost of the basket ($\pi(B_i,s)$). Normalizing coefficient on cost to one, agent i's realized utility is

$$U_i(B_i,s)=U_i(B_i)+C_s'\delta-(\beta+\nu_i)T_{i,s}-\pi(B_i,s),$$

where C_s are observed characteristics of the supermarket, $T_{i,s}$ is the travel time for i going to s, and $\beta+\nu_i$ is its impact on utility, where ν_i has mean zero given supermarket characteristics and travel times.

Katz assumes travel times and store characteristics are known to the shopper. For $\tilde{s}$ a supermarket with $T_{i,\tilde{s}}>T_{i,s}$ that also marketed B_i, he divides the difference $U_i(B_i,s)-U_i(B_i,\tilde{s})$ by $T_{i,s}-T_{i,\tilde{s}}$ and notes that a combination of expected utility maximization and revealed preference implies that $E_{P_{D|Z}}[Y_i(\beta)-X_i\delta|Z_i]\leq 0$, for

$$Y_i(\beta)\equiv-\beta-\frac{[\pi(B_i,s)-\pi(B_i,\tilde{s})]}{T_{i,s}-T_{i,\tilde{s}}},X_i\equiv-\frac{C_s'-C_{\tilde{s}}'}{T_{i,s}-T_{i,\tilde{s}}},\text{ and } Z_i\equiv(T_{i,s},T_{i,\tilde{s}},C_s',C_{\tilde{s}}')'.$$

Together with an analogous inequality which uses a store closer to the agent, Katz obtains both upper and lower bounds for β. $\triangle$

Example 3 Wollmann (2018) considers the bailout of GM and Chrysler's commercial truck divisions during the 2008 financial crisis and asks what would have happened had they instead been allowed to either fail or merge with another firm. This example is the basis for our simulations below.

Merger analysis focuses on price differences pre- and post-merger. Wollmann notes that some commercial truck production is modular (it is possible to connect different cab types to different trailers), so some products would likely have been repositioned after the change in the environment. To analyze product repositioning he requires estimates for the fixed costs of marketing a product. His estimated demand and cost systems enable him to estimate counterfactual profits from adding or deleting products. Assuming firms maximize expected profits, differences in expected profits from adding or subtracting products imply bounds on fixed costs.

To illustrate, let $J_{f,t}$ be the set of models that firm f marketed in year t and let $J_{f,t}\setminus j$ be that set excluding product j, while $\Delta\pi(J_{f,t},J_{f,t}\setminus j)$ is the difference in expected profits between marketing $J_{f,t}$ and $J_{f,t}\setminus j$. The fixed cost to firm f of marketing product j at time t is given by $(\delta_{c,f}+\delta_g g_j)$ if the product was not marketed previously $(j \notin J_{f,t-1})$, and $\beta(\delta_{c,f}+\delta_g g_j)$ if it was previously marketed. Here $\delta_{c,f}$ is a firm-specific intercept, g_j is the weight of product j, δ_g is the cost of adding additional weight (assumed common across firms), and β captures the cost savings of marketing a pre-existing product. We can write the fixed cost as $X^*_{j,f,t}(\beta)\delta$, where $X^*_{j,f,t}(\beta)$ contains a firm indicator and the product's weight, possibly multiplied by β depending on whether $j \in J_{j,f,t-1}$. For $Z_{f,t}$ a set of variables known to the firm when marketing decisions were made, including the variables used to form $X^*_{j,f,t}(\beta)$,

$$E_{P_{D|Z}}[Y_{j,f,t}-X_{j,f,t}(\beta)\delta|Z_{f,t}] \geq 0 \text{ for all } j, \tag{6}$$

by the firm's equilibrium conditions, where

$$Y_{j,f,t} \equiv \Delta\pi(J_{f,t},J_{f,t}\backslash j)\cdot 1\{j \in J_{f,t}, j \in J_{f,t-1}\}, \quad X_{j,f,t}(\beta) \equiv X^*_{f,j,t}(\beta)\cdot 1\{j \in J_{f,t}, j \in J_{f,t-1}\}$$

and $1\{A\}$ is an indicator for the event A. Additional inequalities can be added for marketing a product that was not marketed in the prior period, for withdrawing products, and for combining the withdrawal of one product with adding another. △

Cox & Shi (2022) note that moment inequalities in Eizenberg (2014) and Gandhi et al. (2019) also have linear conditional structure. Further recent examples appear in Ho & Pakes (2014), Morales et al. (2019), Rambachan & Roth (2022), and Rambachan (2021).

2.2 Simplifications from Linear Conditional Structure

In addition to arising frequently in applications, the structure of linear conditional moment inequalities can be exploited to develop simple and computationally tractable tests of (1). We begin by describing an asymptotic framework frequently used to test moment inequalities, and some challenges it generates. We then describe how linear conditional structure can be used to circumvent some of these issues. We focus on the intuition here, deferring formal results to the following sections.

Unconditional asymptotics Conditional moment inequalities are often tested indirectly. In particular, (1) implies that $E_P[Y_i(\beta)-X_i(\beta)\delta] \leq 0$. To test $\widetilde{H}_0: \beta_0 \in B_I(P)$, we may therefore test that there exists a value of δ such that $E_P[Y_i(\beta_0)-X_i(\beta_0)\delta] \leq 0$. Letting $Y_{n,0} = \frac{1}{\sqrt{n}}\sum_i Y_i(\beta_0)$ and $X_{n,0} = \frac{1}{\sqrt{n}}\sum_i X_i(\beta_0)$, the central limit theorem implies that for each δ, $Y_{n,0}-X_{n,0}\delta-\mu_{U,n,0}(\delta) \to_d N(0,\Sigma_{U,0}(\delta))$, for $\mu_{U,n,0}(\delta) = \sqrt{n}E_P[Y_i(\beta_0)-X_i(\beta_0)\delta]$ and

$\Sigma_{U,0}(\delta) = Var_P(Y_i(\beta_0) - X_i(\beta_0)\delta)$. This suggests the approximation

$$Y_{n,0} - X_{n,0}\delta \approx^d N(\mu_{U,n,0}(\delta), \Sigma_{U,0}(\delta)), \tag{7}$$

where $\approx^d$ denotes approximate equality in distribution. The normal approximation (7) may be used to test $H_0^{joint}(\delta_0) : \mu_{U,n,0}(\delta_0) \leq 0$, which jointly restricts (β, δ). This allows a projection test of $\widetilde{H}_0 : \beta_0 \in B_I(P)$, which rejects if and only if we reject $H_0^{joint}(\delta_0)$ for all δ_0. Simple projection tests can be quite conservative, however, which has motivated approaches based on the joint limiting distribution across different values of δ (e.g. Kaido et al. 2019*b*).

Even if we are happy to use the projection method, projection tests based on (7) are complicated by the dependence of the variance matrix $\Sigma_{U,0}(\delta_0)$ on the value of δ_0, since critical values for tests of $H_0^{joint}(\delta_0)$ will typically depend on δ_0 as well. When the nuisance parameter δ has even moderate dimension, calculating the critical value for many values of δ_0 can become computationally burdensome, necessitating careful attention to algorithms to mitigate the computational cost (e.g., Kaido et al. 2019*b*).

Conditional asymptotics Linear conditional structure allows an alternative asymptotic approximation, which avoids complications discussed above by conditioning on the sequence of realized instrument values $\{Z_i\} = \{Z_i\}_{i=1}^{\infty}$. For $\mu_i(\beta, P_{D|Z}) = E_{P_{D|Z}}[Y_i(\beta)|Z_i]$ and $\mu_{n,0} = \frac{1}{\sqrt{n}}\sum_i \mu_i(\beta_0, P_{D|Z})$, the Lindeberg-Feller central limit theorem implies that under mild conditions $Y_{n,0} - \mu_{n,0}|\{Z_i\} \to_d N(0, \Sigma_0)$, where $\Sigma_0 = E_P[Var_{P_{D|Z}}(Y_i(\beta_0)|Z_i)]$. Since $X_{n,0}$ is non-random conditional on $\{Z_i\}$, this suggests the approximation

$$Y_{n,0} - X_{n,0}\delta | \{Z_i\} \approx^d N(\mu_{n,0} - X_{n,0}\delta, \Sigma_0). \tag{8}$$

Importantly, and in contrast to (7), the variance Σ_0 in (8) does not depend on the value of δ. This substantially simplifies the problem of constructing tests. Further, since $X_{n,0}$ is non-stochastic conditional on $\{Z_i\}$, (8) holds jointly across values of δ.

To construct tests based on this conditional approximation, observe that if $\widetilde{H}_0 : \beta_0 \in B_I(P)$ holds, then there exists (almost surely) a value of δ such that $\mu_{n,0} - X_{n,0}\delta \leq 0$. The null $\widetilde{H}_0 : \beta_0 \in B_I(P)$ thus implies the null $H_0 : \mu_{n,0} \in \mathcal{M}_{n,0}$, where

$$\mathcal{M}_{n,0} = \{\mu \in \mathbb{R}^k : \text{ there exists } \delta \text{ such that } \mu - X_{n,0}\delta \leq 0\}$$

is non-stochastic conditional on $\{Z_i\}$.[7] Equation (8) with $\delta = 0$ further implies that

[7] In fact, $\widetilde{H}_0$ implies that $\mu_{n,0} \in \mathcal{M}_{n,0} \cap \mathcal{M}_{n,0,\mathcal{P}_{D|Z}}$, where $\mathcal{P}_{D|Z}$ is the family of conditional distributions

$Y_{n,0}|\{Z_i\} \approx^d N(\mu_{n,0}, \Sigma_0)$, so testing H_0 reduces, asymptotically, to testing a restriction on the mean of a multivariate normal vector.

Indirect Tests While indirect tests of $\widetilde{H}_0 : \beta_0 \in B_I(P)$ are natural, they can entail a loss of consistency. The original null hypothesis $\widetilde{H}_0 : \beta_0 \in B_I(P)$ implies that there exists a δ such that $\frac{1}{\sqrt{n}}\sum_i (E_{P_{D|Z}}[Y_i(\beta_0)|Z_i] - X_i(\beta_0)\delta) \otimes f(Z_i) \leq 0$ for all non-negative functions $f(Z_i)$, whereas $H_0 : \mu_{n,0} \in \mathcal{M}_{n,0}$ only tests that this is satisfied for $f(Z_i) = 1$.[8] Indeed, conditional moment inequalities based on continuously distributed instruments Z_i generate an infinite number of unconditional inequalities, as discussed in e.g. D. Andrews & Shi (2013), Armstrong (2014), Chernozhukov et al. (2015), and Chetverikov (2018). As a result, the tests we develop do not in general yield consistent tests when the instruments are continuously distributed. This contrasts with the aforementioned papers, which develop consistent tests by checking an (asymptotically) infinite number of moment restrictions.

Inference based on a finite, researcher-selected set of inequalities nonetheless appears widespread in applications, and is the approach adopted in all the empirical applications discussed above save Gandhi et al. (2019). This raises the question of how to select the finite set of moments (i.e, which restrictions to include in Y_i), which we discuss informally in Section 5.1 below. Whether one can go further, either characterizing an optimal selection of moments or combining our results with those in the previous literature on conditional moment inequalities to ensure consistent inference in settings with continuously distributed Z_i, is an interesting question for future work.

3 Inference Procedures in the Normal Model

We now introduce our tests. Motivated by the asymptotic approximation (8), we begin with tests of $H_0 : \mu_{n,0} \in \mathcal{M}_{n,0}$ in the exact normal model

$$Y_{n,0} \sim N(\mu_{n,0}, \Sigma_0) \text{ for known } \Sigma_0. \tag{9}$$

The next section presents sufficient conditions for feasible versions of our tests, based on non-normal data and estimates of Σ_0, to uniformly control asymptotic size.

implied by $\mathcal{P}$, while $\mathcal{M}_{n,0,\mathcal{P}_{D|Z}} = \left\{\frac{1}{\sqrt{n}}\sum_i E_{P_{D|Z}}[Y_i(\beta_0)|Z_i] | P_{D|Z} \in \mathcal{P}_{D|Z}\right\}$. For tractability, we focus on the implied null that $\mu_{n,0} \in \mathcal{M}_{n,0}$ rather than $\mu_{n,0} \in \mathcal{M}_{n,0} \cap \mathcal{M}_{n,0,\mathcal{P}_{D|Z}}$. This yields valid but potentially conservative tests if $X_{n,0}\delta \notin \mathcal{M}_{n,0,\mathcal{P}_{D|Z}}$ for all δ, i.e. if $\mathcal{P}_{D|Z}$ does not allow all moments to simultaneously bind; see Section 3.2 for additional discussion.

[8] Note that if one starts with (Y_i, X_i) satisfying (1), then $E_{P_{D|Z}}[\tilde{Y}_i - \tilde{X}_i\delta|Z_i] \leq 0$ for $(\tilde{Y}_i, \tilde{X}_i) = (Y_i, X_i) \otimes f(Z_i)$ and any non-negative finite instrument function $f(Z_i)$. Thus, a key restriction imposed in our framework is that the researcher chooses a finite set of instruments with which to interact the initial moments.

3.1 Test Statistic

Given $Y_{n,0} \sim N(\mu_{n,0}, \Sigma_0)$ for known Σ_0, we construct tests for the hypothesis $H_0 : \mu_{n,0} \in \mathcal{M}_{n,0}$, that is, that there exists some δ such that $\mu_{n,0} - X_{n,0}\delta \leq 0$. We eliminate the nuisance parameter δ by using the profiled max statistic,

$$\hat{\eta}_{n,0} = \min_{\delta} \max_{j} \left\{ e_j'(Y_{n,0} - X_{n,0}\delta)/\sigma_{0,j} \right\}$$

for e_j the jth standard basis vector and $\sigma_{0,j} = \sqrt{e_j' \Sigma_0 e_j}$.[9] Our test statistic thus profiles the maximum-criterion statistic (S_3 in the notation of D. Andrews & Soares (2010)). By a profiled test statistic, we mean one that optimizes over the nuisance parameter δ to find the value that makes the test statistic as small as possible. Specifically, note that $\max_j \left\{ e_j'(Y_{n,0} - X_{n,0}\delta)/\sigma_{0,j} \right\}$ calculates the maximum studentized violation of the sample moments at a given δ, so $\hat{\eta}_{n,0}$ corresponds to the maximum violation at the value of δ that makes this violation the smallest. One could profile test statistics other than the max statistic — e.g. Cox & Shi (2022) study profiled QLR statistics and Bugni et al. (2017) study profiled modified method of moments (MMM) statistics (among others) — but it will be helpful for our analysis that the profiled max statistic admits an equivalent representation as the solution to the linear program,

$$\hat{\eta}_{n,0} = \min_{\eta, \delta} \eta \text{ subject to } Y_{n,0} - X_{n,0}\delta \leq \eta \cdot \sigma_0, \tag{10}$$

for $\sigma_0 = (\sigma_{0,1}, \ldots, \sigma_{0,k})'$. This allows for tractable computation of $\hat{\eta}_{n,0}$ even when the dimension of δ is large, and the linear structure plays a key role in the construction of our tests.

3.1.1 Dual representation of the test statistic

To derive critical values, we will make use of the dual representation of the linear program (10). Standard results in linear programming (e.g., Chapter 7.4 of Schrijver (1986)) imply that when $\hat{\eta}_{n,0} > -\infty$ it is the solution of the dual linear program,[10]

$$\hat{\eta}_{n,0} = \max_{\gamma} \gamma' Y_{n,0} \text{ s.t. } \gamma \geq 0, \gamma' X_{n,0} = 0, \gamma' \sigma_0 = 1. \tag{11}$$

[9] We define $\frac{c}{0} = \infty$ for all $c > 0$.

[10] Observe that $\hat{\eta}_{n,0}$ is equal to $-\infty$ if and only if $\min_\delta \max_j e_j' X_{n,0}\delta = -\infty$, in which case H_0 is satisfied regardless of the value of $\mu_{n,0}$, so the testing problem is trivial. Finiteness of $\hat{\eta}_{n,0}$ implies that $X_{n,0}$ does not have full row rank, for instance because $k > p$.

Moreover, the maximum is obtained at one of the finite set of vertices of the feasible set. Intuitively, the set of feasible values $F(X_{n,0},\sigma_0)=\{\gamma\geq 0|\gamma' X_{n,0}=0,\gamma'\sigma_0=1\}$ is a polyhedron, i.e. a convex set with flat sides, and a vertex corresponds with a "corner" of this set. More formally, as described in e.g. Schrijver (1986, Section 8.5), $\gamma\in F(X_{n,0},\sigma_0)$ is a vertex if it can be realized as a unique solution to (11) for some value of $Y_{n,0}$:

Definition 1 *The set of vertices $V(X_{n,0},\sigma_0)$ of $F(X_{n,0},\sigma_0)$ is*

$$V(X_{n,0},\sigma_0)=\left\{\gamma\in F(X_{n,0},\sigma_0):\exists y\in\mathbb{R}^k \text{ such that } \gamma' y>\tilde{\gamma}' y \text{ for all } \tilde{\gamma}\in F(X_{n,0},\sigma_0)\setminus\{\gamma\}\right\}.$$

As a simple example, if $\Sigma=I$ and $X_{n,0}=0$, then $V(X_{n,0},\sigma_0)$ is the set of standard basis vectors in $\mathbb{R}^k$. In Lemma A.1 in the appendix, we give an alternative characterization of the set of vertices, which shows that $\gamma\in F(X_{n,0},\sigma_0)$ is a vertex if and only if γ is the solution to the system of equations defined by a full-rank subset of the constraints in (11). Since there are a finite number of constraints in (11), this immediately implies that $V(X_{n,0},\sigma_0)$ is finite. It is neither necessary nor recommended to enumerate all of the elements of $V(X_{n,0},\sigma_0)$ to compute our test statistic and critical values (see Section 5 for details on computation), but this representation will be useful for explaining our approach.

The dual representation for $\hat{\eta}_{n,0}$ implies that in the finite sample normal model the test statistic $\hat{\eta}_{n,0}$ is the maximum of a multivariate normal vector, $\hat{\eta}_{n,0}=\max_{\gamma\in V(X_{n,0},\sigma_0)}\gamma' Y_{n,0}=\max\{\gamma_{(1)}' Y_{n,0},\ldots,\gamma_{(J)}' Y_{n,0}\}$, for $\gamma_{(1)},\ldots,\gamma_{(J)}$ the elements of $V(X_{n,0},\sigma_0)$. Our critical values will then be based on properties of the maximum of a correlated Gaussian vector.

3.2 Least Favorable Tests

Our first test is based on the "least-favorable" value of $\mu_{n,0}$ under the null hypothesis H_0. Recall that $\hat{\eta}_{n,0}=\max_{\gamma\in V(X_{n,0},\sigma_0)}\gamma' Y_{n,0}$. Hence

$$\hat{\eta}_{n,0}=\max_{\gamma\in V(X_{n,0},\sigma_0)}\{\gamma'\mu_{n,0}+\gamma'(Y_{n,0}-\mu_{n,0})\}\leq\max_{\gamma\in V(X_{n,0},\sigma_0)}\gamma'\mu_{n,0}+\max_{\gamma\in V(X_{n,0},\sigma_0)}\gamma'(Y_{n,0}-\mu_{n,0}).$$

Under H_0, however, there exists δ such that $\mu_{n,0}-X_{n,0}\delta\leq 0$. Since every $\gamma\in V(X_{n,0},\sigma_0)$ is feasible in (11) by construction, we also have that $\gamma' X_{n,0}=0$ and $\gamma\geq 0$ for all $\gamma\in V(X_{n,0},\sigma_0)$. It follows that under the null, $\gamma'\mu_{n,0}=\gamma'(\mu_{n,0}-X_{n,0}\delta)\leq 0$ for all $\gamma\in V(X_{n,0},\sigma_0)$. Combined with the previous display, this implies that under H_0,

$$\hat{\eta}_{n,0}\leq\max_{\gamma\in V(X_{n,0},\sigma_0)}\gamma'(Y_{n,0}-\mu_{n,0}). \tag{12}$$

Since $Y_{n,0} - \mu_{n,0} \sim N(0, \Sigma_0)$, we define the least-favorable critical value $c_{\alpha,LF} = c_{\alpha,LF}(X_{n,0}, \sigma_0)$ as the $1-\alpha$ quantile of $\max_{\gamma \in V(X_{n,0},\sigma_0)} \gamma'\xi$ for $\xi \sim N(0, \Sigma_0)$ and consider the test that rejects when $\hat{\eta}_{n,0}$ exceeds this critical value, $\phi_{LF} = 1\{\hat{\eta}_{n,0} > c_{\alpha,LF}\}$. It follows immediately from the inequality (12) that under the finite sample normal model $E[\phi_{LF}] \leq \alpha$ whenever $H_0: \mu_{n,0} \in \mathcal{M}_{n,0}$ holds. Moreover, the inequality (12) reduces to an equality if $\gamma'\mu_{n,0} = 0$ for all $\gamma \in V(X_{n,0}, \sigma_0)$, as for example occurs if $\mu_{n,0} = 0$ or more generally if $\mu_{n,0} = X_{n,0}\delta$ for some δ, in which case $E[\phi_{LF}] = \alpha$. Thus, the LF test has exact size in the finite sample normal model if it is possible for all moments to bind simultaneously. We note, however, that this may not be possible for some data-generating processes (e.g., if certain pairs of moments correspond to upper and lower bounds that cannot simultaneously bind), in which case the least favorable test may have size strictly less than α.[11]

Sensitivity to slack moments An undesirable feature of the LF test is that it may be sensitive to the inclusion of slack moments. That is, the power of the test may be negatively affected if one includes in $Y_{n,0}$ moments that are very far from binding (i.e. elements j with $\mu_{n,0,j} \ll 0$). The reason is that the critical value $c_{\alpha,LF}$ is based on the distribution of the test statistic when $\mu_{n,0} = 0$, and thus generally increases when adding additional moments, even though the test statistic $\hat{\eta}_{n,0}$ will generally not be affected by the inclusion of very slack moments. Motivated by this fact, D. Andrews & Soares (2010), D. Andrews & Barwick (2012), Romano et al. (2014), and related papers propose techniques that use information from the data to either select moments or shift the mean of the distribution from which the critical values are calculated. This yields tests with higher power in cases where many of the moments are slack. Unfortunately, applying these existing methods in our setting breaks the linear structure, and hence the computational advantages from using linear programming, which motivates us to introduce an alternative approach.

3.3 Conditional Test

We next introduce a test that is less sensitive to the inclusion of slack moments than the LF test while also exploiting the linear conditional structure in our context. This test is based on the distribution of $\hat{\eta}_{n,0}$ conditional on the identity of the optimal vertex in the dual problem, $\hat{\gamma} = \operatorname{argmax}_{\gamma \in V(X_{n,0},\sigma_0)} \gamma' Y_{n,0}$.[12] For simplicity of exposition, we begin by assuming

[11] In such cases, where $0 \notin \mathcal{M}_{n,0} \cap \mathcal{M}_{n,0,\mathcal{P}_{D|Z}}$ for $\mathcal{M}_{n,0,\mathcal{P}_{D|Z}}$ as defined in footnote 7, tests based on the critical value $c_{\alpha,LF} + \psi$ for $\psi = \max_{\mu_{n,0} \in \mathcal{M}_{n,0} \cap \mathcal{M}_{n,0,\mathcal{P}_{D|Z}}} \max_{\gamma \in V(X_{n,0},\sigma_0)} \gamma'\mu_{n,0}$ will also control size. These tests have (weakly) improved power since $\psi \leq 0$ by definition. The adjustment factor ψ depends on the class of conditional data generating processes $\mathcal{P}_{D|Z}$ considered, however, so we focus on results using $c_{\alpha,LF}$ for simplicity.

[12] $\hat{\gamma}$ depends on n and β_0, but we leave this dependence implicit for simplicity of notation.

that $\hat{\gamma}$ is unique, in the sense that $\text{argmax}_{\gamma\in V(X_{n,0},\sigma_0)}\gamma' Y_{n,0}$ is a singleton; we will discuss the case of a non-unique dual below.[13] If $\hat{\gamma}'\Sigma_0\hat{\gamma}=0$ then we define the conditional test to reject if and only if $\hat{\eta}_{n,0}>0$. For the remainder of this section, we thus assume that $\hat{\gamma}'\Sigma_0\hat{\gamma}>0$. For any $\gamma\in V(X_{n,0},\sigma_0)$, note that $\hat{\gamma}=\gamma$ only if $\gamma' Y_{n,0}\geq\tilde{\gamma}' Y_{n,0}$ for all $\tilde{\gamma}\in V(X_{n,0},\sigma_0)$. Hence, $\hat{\gamma}=\gamma$ is optimal only if $Y_{n,0}$ lies in the polyhedron $\{y|(\gamma-\tilde{\gamma})' y\geq 0,\forall\tilde{\gamma}\in V(X_{n,0},\sigma_0)\}$. This representation allows us to characterize the distribution of $\hat{\eta}_{n,0}$ conditional on $\hat{\gamma}=\gamma$ using Lemma 5.1 in Lee et al. (2016), which characterizes the behavior of Gaussian random variables conditional on polyhedral events.

Lemma 1 *Let $S_{n,0,\gamma}=\left(I-\frac{\Sigma_0\gamma\gamma'}{\gamma'\Sigma_0\gamma}\right)Y_{n,0}$. Then under (9),*

$$\hat{\eta}_{n,0}|\{\hat{\gamma}=\gamma,S_{n,0,\gamma}=s\}\sim TN(\gamma'\mu_{n,0},\gamma'\Sigma_0\gamma,[\mathcal{V}_{n,0}^{lo},\mathcal{V}_{n,0}^{up}]), \tag{13}$$

where $TN(\mu,\sigma^2,[a,b])$ denotes the $N(\mu,\sigma^2)$ distribution truncated to $[a,b]$,

$$\mathcal{V}_{n,0}^{lo}=\max_{\substack{\tilde{\gamma}\in V(X_{n,0},\sigma_0):\\ \gamma'\Sigma_0\gamma>\gamma'\Sigma_0\tilde{\gamma}}}\frac{\gamma'\Sigma_0\gamma\cdot\tilde{\gamma}' s}{\gamma'\Sigma_0\gamma-\gamma'\Sigma_0\tilde{\gamma}},\ \mathcal{V}_{n,0}^{up}=\min_{\substack{\tilde{\gamma}\in V(X_{n,0},\sigma_0):\\ \gamma'\Sigma_0\gamma<\gamma'\Sigma_0\tilde{\gamma}}}\frac{\gamma'\Sigma_0\gamma\cdot\tilde{\gamma}' s}{\gamma'\Sigma_0\gamma-\gamma'\Sigma_0\tilde{\gamma}}, \tag{14}$$

and we define $\mathcal{V}_{n,0}^{lo}=-\infty$ and $\mathcal{V}_{n,0}^{up}=\infty$, respectively, when we optimize over the empty set.

Recall that under H_0, $\gamma'\mu_{n,0}\leq 0$ for all $\gamma\in V(X_{n,0},\sigma_0)$. Additionally, Lemma A.1 in Lee et al. (2016) shows that the $TN(\mu,\sigma^2;[a,b])$ distribution is increasing in μ in the sense of first order stochastic dominance. It follows that the distribution on the right-hand side of (13) is weakly dominated by the $TN(0,\hat{\gamma}'\Sigma_0\hat{\gamma},[\mathcal{V}_{n,0}^{lo},\mathcal{V}_{n,0}^{up}])$ distribution under the null. We therefore base our test on this distribution. Letting $\bar{c}_{\alpha,C}$ be the $1-\alpha$ quantile of the $TN(0,\hat{\gamma}'\Sigma_0\hat{\gamma},[\mathcal{V}_{n,0}^{lo},\mathcal{V}_{n,0}^{up}])$ distribution, we define the conditional critical value as $c_{\alpha,C}=c_{\alpha,C}(Y_{n,0},X_{n,0},\Sigma_0)=\max\{\bar{c}_{\alpha,C},0\}$ and reject if $\hat{\eta}_{n,0}$ exceeds it, $\phi_C=1\{\hat{\eta}_{n,0}>c_{\alpha,C}\}$.[14] It follows immediately that ϕ_C controls size conditionally in the finite sample normal model, with $E[\phi_C|\hat{\gamma}=\gamma,S_{n,0,\gamma}]\leq\alpha$ whenever $\mu_{n,0}\in\mathcal{M}_{n,0}$.[15] Unconditional size control follows by the law of iterated expectations.

---

[13]Our asymptotic results in the next section impose a sufficient condition for uniqueness to hold with probability one asymptotically.

[14]The censoring of the critical value at 0 is unnecessary for size control in the finite-sample normal model, but simplifies asymptotic arguments. It is also substantively reasonable as it prevents the test from rejecting when all of the moment inequalities are satisfied in sample ($\hat{\eta}_{n,0}\leq 0$).

[15]As for the least favorable test, if $X_{n,0}\delta\notin\mathcal{M}_{n,0}\cap\mathcal{M}_{n,0,\mathcal{P}_{D|Z}}$ for all δ, we can potentially use smaller critical values, replacing $\bar{c}_{\alpha,C}$ with the $1-\alpha$ quantile of a $TN(\psi_{\hat{\gamma}},\hat{\gamma}'\Sigma_0\hat{\gamma},[\mathcal{V}_{n,0}^{lo},\mathcal{V}_{n,0}^{up}])$ distribution for $\psi_{\hat{\gamma}}=\max_{\mu_{n,0}\in\mathcal{M}_{n,0}\cap\mathcal{M}_{n,0,\mathcal{P}_{D|Z}}}\hat{\gamma}'\mu_{n,0}$. As before, $\psi_{\hat{\gamma}}$ will depend on the specification of $\mathcal{P}_{D|Z}$, and we focus on tests based on $\bar{c}_{\alpha,C}$ for simplicity.

Example (uncorrelated moments) Consider the case where $Y_{n,0} \sim N(\mu_{n,0}, I)$, and $X_{n,0} = 0$, so that there is no nuisance parameter δ. Then $V(X_{n,0}, \sigma_0)$ is simply the set of standard basis vectors, so $\hat{\eta}_{n,0} = \max_j e_j' Y_{n,0}$ is the maximum component of $Y_{n,0}$. In this case $\mathcal{V}_{n,0}^{lo}$ corresponds to the second-largest component of $Y_{n,0}$, i.e. $\max_{j \neq \hat{j}} e_j' Y_{n,0}$, for $\hat{j}$ the location of the maximum, and $\mathcal{V}_{n,0}^{up} = \infty$. The conditional test thus rejects if $\hat{\eta}_{n,0}$ exceeds the $1-\alpha$ quantile of the standard normal distribution truncated to $[\mathcal{V}_{n,0}^{lo}, \infty]$.

Non-unique dual solutions. So far we have assumed the existence of a unique dual solution, $\hat{\gamma} = \gamma$. If Σ_0 is not full-rank, however, then there may be multiple solutions to the dual problem with positive probability.[16] In Appendix B, we consider a version of the conditional test that, when the dual solution is non-unique, calculates $(\mathcal{V}_{n,0}^{lo}, \mathcal{V}_{n,0}^{up})$ via (14) by selecting an element of the dual solution set, $\gamma = h(\hat{\gamma})$. We show that in the finite sample normal model, with probability 1 the critical values do not depend on how the optimal vertex is chosen, so the test obtained does not depend on the choice of $h(\cdot)$. Further, we show in Appendix B that this test controls size in the finite-sample normal model. Our sufficient conditions for uniform asymptotic size control in Section 4 below imply that the dual solution will be unique with probability tending to 1, however, so we focus primarily on the case where the dual solution is unique.

Insensitivity to Slack Moments In contrast with the LF test, the conditional test has the desirable property that it is insensitive to the inclusion of slack moments. Specifically, our next result shows that the conditional test is insensitive to slack moments in the strong sense that as a moment becomes arbitrarily slack the conditional test converges to the conditional test that drops that moment ex-ante. Intuitively, this happens because (under mild conditions) sufficiently slack moments make no contribution to $\hat{\eta}_{n,0}$, $\mathcal{V}_{n,0}^{lo}$, or $\mathcal{V}_{n,0}^{up}$, and so have no impact on the conditional test. To state this result formally, define $Y_{n,0}^{j,d} = Y_{n,0} - e_j \cdot d$ as a version of $Y_{n,0}$ which decreases the jth moment by d. Let $Y_{n,0}^{-j}$ collect the rows of $Y_{n,0}$ other than the jth, and define $X_{n,0}^{-j}$ and Σ_0^{-j} accordingly. Define $\hat{\eta}_{n,0}^{j,d}$ and $\hat{\eta}_{n,0}^{-j}$ as versions of $\hat{\eta}_{n,0}$ based on $\left(Y_{n,0}^{j,d}, X_{n,0}, \Sigma_0\right)$ and $\left(Y_{n,0}^{-j}, X_{n,0}^{-j}, \Sigma_0^{-j}\right)$, respectively, and let $\phi_C^{j,d}$ and ϕ_C^{-j} denote the corresponding tests.

Lemma 2 *For any $Y_{n,0}$ such that $\gamma' Y_{n,0} \neq \tilde{\gamma}' Y_{n,0}$ for all distinct $\gamma, \tilde{\gamma} \in V(X_{n,0}, \sigma_0)$ and $\hat{\eta}_{n,0}^{-j} \neq c_{\alpha,C}\left(Y_{n,0}^{-j}, X_{n,0}^{-j}, \Sigma_0^{-j}\right)$, we have $\lim_{d \to \infty} \phi_C^{j,d} = \phi_C^{-j}$.*

---

[16] Since the dual objective is $\hat{\eta}_{n,0} = \max\{\gamma_{(1)}' Y_{n,0}, \ldots, \gamma_{(J)}' Y_{n,0}\}$ and $\gamma_{(j)} \neq \gamma_{(j')}$ for $j \neq j'$, the dual has a unique solution with probability 1 so long as Σ_0 is full rank.

The conditions of Lemma 2 hold for Lebesgue almost every $Y_{n,0}$, and hold with probability 1 under (9) provided that $\gamma'\Sigma_0\gamma > 0$ and $(\gamma-\tilde{\gamma})'\Sigma_0(\gamma-\tilde{\gamma}) > 0$ for all distinct $\gamma,\tilde{\gamma} \in V(X_{n,0},\sigma_0)$, so that the variables $\gamma' Y_{n,0}$ have positive variance and are not perfectly correlated with one another. The only other tests we are aware of that both control size in the finite-sample normal model and are unaffected by the inclusion of arbitrarily slack moments in the sense of Lemma 2 are those of Cox & Shi (2022).

Power with Multiple Violated Moments. Although the conditional test exhibits a desirable insensitivity to the inclusion of slack moments, it may exhibit poor power in cases where two (or more) moments are approximately equally violated. This is most easily seen in the example of uncorrelated moments from above, where $\mathcal{V}^{lo}_{n,0}$ corresponds with the value of the second-largest sample moment, and the critical value is the $1-\alpha$ quantile of the standard normal distribution truncated to $[\mathcal{V}^{lo}_{n,0},\infty]$. If two moments are approximately equally violated, then the largest and second largest sample moments ($\hat{\eta}_{n,0}$ and $\mathcal{V}^{lo}_{n,0}$, respectively) may be close together, so the conditional test need not reject even if both of these are large. This phenomenon is highlighted in parts of the parameter space in our simulations in Section 6.

3.4 Hybrid Tests

To mitigate the possible power losses of the conditional test when multiple moments are approximately equally violated, we next introduce a hybrid test that combines the least favorable and conditional approaches. For some $0<\kappa<\alpha$, we define the size-α hybrid test to reject whenever the size-κ least favorable test does. If the least favorable test does not reject, we then consider a size-$\frac{\alpha-\kappa}{1-\kappa}$ test that conditions on both $\hat{\gamma}=\gamma$ and the event that the least-favorable test did not reject. Specifically, the same argument used to prove Lemma 1 yields that

$$\hat{\eta}_{n,0}|\{\hat{\gamma}=\gamma,S_{n,0,\gamma}=s,\phi_{LF,\kappa}=0\}\sim TN(\gamma'\mu_{n,0},\gamma'\Sigma_0\gamma,[\mathcal{V}^{lo}_{n,0},\mathcal{V}^{up,H}_{n,0}]),$$

where $\mathcal{V}^{up,H}_{n,0}=\min\{\mathcal{V}^{up}_{n,0},c_{\alpha,LF}\}$. We then construct the second-stage critical value $\bar{c}_{\frac{\alpha-\kappa}{1-\kappa},H}=\bar{c}_{\frac{\alpha-\kappa}{1-\kappa},H}(Y_{n,0},X_{n,0},\Sigma_0)$ analogously to the conditional critical value $c_{\frac{\alpha-\kappa}{1-\kappa},C}$ except using the modified truncation point $\mathcal{V}^{up,H}_{n,0}$. Letting $c_{\frac{\alpha-\kappa}{1-\kappa},H}=\min\{c_{\kappa,LF},\bar{c}_{\frac{\alpha-\kappa}{1-\kappa},H}\}$, the hybrid test is then $\phi_H=1\{\hat{\eta}_{n,0}>c_{\frac{\alpha-\kappa}{1-\kappa},H}\}$. Observe that the critical value for the hybrid test approaches that of the LF test as $\kappa\to\alpha$, while it approaches that of the conditional test as $\kappa\to 0$.

As argued above, the first-stage LF test for the hybrid rejects with probability not exceeding κ under the null in the finite-sample normal model. Likewise, by arguments analogous to those for the conditional test, the second stage test rejects with probability no more than $\frac{\alpha-\kappa}{1-\kappa}$ conditional on the first stage not rejecting. It follows that when $\mu_{n,0} \in \mathcal{M}_{n,0}$, the hybrid test rejects with probability

$$E[\phi_{LF,\kappa}]+(1-E[\phi_{LF,\kappa}])E\left[\hat{\eta}_{n,0}>\bar{c}_{\frac{\alpha-\kappa}{1-\kappa},H}|\phi_{LF,\kappa}=0\right]\leq\kappa+(1-\kappa)\frac{\alpha-\kappa}{1-\kappa}=\alpha,$$

and so controls size in the finite sample normal model.

The hybrid test proposed above always rejects whenever a simple Bonferroni combination of a size-κ LF test and size-$(\alpha-\kappa)$ conditional test would reject, and can reject in cases where the simple Bonferroni does not. The proposed method improves upon the simple Bonferroni approach in two ways, first modifying the second-stage test to condition on the event that the LF test does not reject (which truncates the distribution above and so reduces the critical value), and then using a size $\frac{\alpha-\kappa}{1-\kappa} > \alpha-\kappa$ critical value. This helps to reduce the conservativeness usually associated with Bonferroni approaches.

Sensitivity to Slack Moments The hybrid test will be sensitive to the inclusion of slack moments via its dependence on the LF critical values. However, this sensitivity will be small when κ is close to zero, since in this case the critical values will tend to be close to those of the conditional test, which as shown above do not depend on the inclusion of slack moments. Similar to Romano et al. (2014), we consider $\kappa=\alpha/10$ in our simulations below.

4 Asymptotic Validity

We conduct our analysis conditional on a sequence of values for the instruments, $\{Z_i\}=\{Z_i\}_{i=1}^{\infty}$, where the data are independent but potentially not identically distributed conditional on $\{Z_i\}$, $D_i \perp\!\!\!\perp D_{i'}|\{Z_j\}$ for all $i\neq i'$. Recall that $\mathcal{P}_{D|Z}$ is the class of conditional distributions for D_i given Z_i, and let $B_I(P_{D|Z})$ denote the conditional identified set for β given $\{Z_i\}$,

$$B_I(P_{D|Z})=\left\{\beta\colon \text{ there exists } \delta \text{ } s.t. \text{ } E_{P_{D|Z}}[Y_i(\beta)-X_i(\beta)\delta|Z_i]\leq 0 \text{ for all } i\right\}.$$

Note that for $B_I(P)$ as defined in (2), $B_I(P)\subseteq B_I(P_{D|Z})$ for almost every $\{Z_i\}$. We provide conditions under which our tests uniformly control asymptotic rejection probabilities over $P_{D|Z}\in\mathcal{P}_{D|Z}$ and $\beta_0\in B_I(P_{D|Z})$. For brevity, we will leave the conditioning on $\{Z_i\}$ implicit when this is without loss of clarity.

Our first assumption is that, conditional on Z_i, $Y_i(\beta_0)$ can be written as a known linear transformation of a vector $U_i(\beta_0)$, whose average conditional variance given Z_i converges uniformly to a bounded and full-rank limit.

Assumption 1 *Suppose that we can write* $Y_i(\beta_0) = TU_i(\beta_0) + \zeta_i(\beta_0)$, *where* T *is a known* $k \times l$ *matrix while* $\zeta_i(\beta_0) \in \mathbb{R}^k$ *is known and non-random conditional on* $\{Z_i\}$. *Further suppose that, (i), for some* $\Omega\left(P_{D|Z}, \beta_0\right)$,

$$\lim_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} \left\| \frac{1}{n}\sum_{i=1}^{n} Var_{P_{D|Z}}(U_i(\beta_0)|Z_i) - \Omega\left(P_{D|Z}, \beta_0\right) \right\| \to 0 \tag{15}$$

and, (ii), for $\bar{\lambda} > 0$ *a finite constant,* $\Omega\left(P_{D|Z}, \beta_0\right) \in \mathbf{\Omega}_{\bar{\lambda}}$ *for all* $P_{D|Z} \in \mathcal{P}_{D|Z}$, $\beta_0 \in B_I(P_{D|Z})$, *where*

$$\mathbf{\Omega}_{\bar{\lambda}} = \{\Omega | \bar{\lambda}^{-1} \le \lambda_{\min}(\Omega) \le \lambda_{\max}(\Omega) \le \bar{\lambda}\}$$

is the set of matrices with minimal and maximal eigenvalues bounded by $\bar{\lambda}^{-1}$ *and* $\bar{\lambda}$.

Note that if the variance of $Y_i(\beta_0)$ is full-rank (as in Examples 2 and 3 above) then the moments can trivially be written as $Y_i(\beta_0) = TU_i(\beta_0) + \zeta_i(\beta_0)$ for $T = I$, $U_i(\beta_0) = Y_i(\beta_0)$, and $\zeta_i(\beta_0) = 0$. The structure in Assumption 1 also commonly arises in moment inequality settings where the variance of $Y_i(\beta_0)$ is not full-rank. For example, consider the case of interval-valued regression (Example 1 above) where the upper- and lower-bounds of the interval are perfectly collinear, $Y_i^U = Y_i^L + c$ for fixed constant c. Then $Y_i(\beta_0) = TU_i(\beta_0) + \zeta_i(\beta_0)$ with $T = [I, -I]'$, $U_i(\beta_0) = Y_i^L - W_i\beta_0$, and $\zeta_i(\beta_0) = [0, -c]'$. Settings with moment equalities represented as inequalities can similarly be expressed in this form — if all the moments are of this form, for example, then we can take $T = [I, -I]'$ and $\zeta_i(\beta_0) = 0$.

Assumption 1 implies that the average conditional variance of $Y_i(\beta_0)$ given Z_i converges, $\frac{1}{n}\sum Var_{P_{D|Z}}(Y_i(\beta_0)|Z_i) \to \Sigma(P_{D|Z}, \beta_0) = T\Omega\left(P_{D|Z}, \beta_0\right)T'$. Although $\Omega\left(P_{D|Z}, \beta_0\right)$ has full rank, $\Sigma(P_{D|Z}, \beta_0)$ may have reduced rank since e.g. the dimension of $\Sigma\left(P_{D|Z}, \beta_0\right)$ may exceed that of $\Omega\left(P_{D|Z}, \beta_0\right)$. We next assume that we have a uniformly consistent estimator for $\Omega(P_{D|Z}, \beta_0)$, and thus for $\Sigma(P_{D|Z}, \beta_0)$.

Assumption 2 $\widehat{\Sigma}_{n,0} = T'\widehat{\Omega}_{n,0}T$, *where* $\widehat{\Omega}_{n,0}$ *is uniformly consistent for* $\Omega\left(P_{D|Z}, \beta_0\right)$,

$$\lim_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} Pr_{P_{D|Z}}\left\{ \left\| \widehat{\Omega}_{n,0} - \Omega\left(P_{D|Z}, \beta_0\right) \right\| > \varepsilon \right\} = 0 \text{ for all } \varepsilon > 0.$$

We discuss sufficient conditions for uniform consistency of $\widehat{\Omega}_{n,0}$ in Appendix C. Note that $\widehat{\Omega}_{n,0}$ depends on the null parameter value β_0 considered, where we again suppress this dependence for brevity of notation.

We further assume that the scaled sample average of $U_i(\beta_0)$ is uniformly asymptotically normal once recentered around its mean. To state this assumption we use the fact that uniform convergence in distribution is equivalent to uniform convergence in bounded Lipschitz metric (see e.g. Theorem 1.12.4 of van der Vaart and Wellner, 1996).

Assumption 3 *For BL_1 the class of real-valued functions which are bounded in absolute value by one and have Lipschitz constant bounded by one, $U_{n,0} = \frac{1}{\sqrt{n}}\sum U_i(\beta_0)$, $\pi_i(\beta_0)=E_{P_{D|Z}}[U_i(\beta_0)|Z_i]$, $\pi_{n,0}=\frac{1}{\sqrt{n}}\sum_i \pi_i(\beta_0)$, and $\xi_{P_{D|Z}} \sim N\left(0,\Omega\left(P_{D|Z},\beta_0\right)\right)$,*

$$\lim_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} \sup_{f\in BL_1} \left| E_{P_{D|Z}}[f(U_{n,0}-\pi_{n,0})] - E\left[f\left(\xi_{P_{D|Z}}\right)\right]\right| = 0.$$

Under Assumption 1, the following lower-level condition is sufficient for Assumption 3.

Lemma 3 *Under Assumption 1, if for all $\varepsilon>0$*

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} \frac{1}{n}\sum_{i=1}^{n} E_{P_{D|Z}}\left[\|U_i(\beta_0)-\pi_i(\beta_0)\|^2 1\left\{\|U_i(\beta_0)-\pi_i(\beta_0)\|>\varepsilon\sqrt{n}\right\}|Z_i\right]=0,$$

then Assumption 3 holds.

Our final assumption, which is needed for the conditional and hybrid approaches, restricts T and $X_{n,0}$. Before stating this assumption, we note that the structure imposed by Assumption 1 allows us to consider a subset of the vertices $V(X_{n,0},\sigma_0)$ discussed in the previous section. Intuitively, the optimal vertex $\hat{\gamma}$ corresponds to a vector of Lagrange multipliers for the primal problem (10), and thus $\hat{\gamma}$ must satisfy the complementary slackness conditions. Assumption 1 then implies that certain vertices can never be optimal when the test rejects – for example, if the matrix T encodes moment equalities as inequalities, then the positive and negative copies of a given moment cannot bind simultaneously unless $\hat{\eta}_{n,0}=0$, in which case our tests do not reject. The following lemma shows that we can ignore such "never-optimal" vertices when establishing size control.

Lemma 4 *Suppose Assumption 1 holds, and let $\hat{\sigma}_{n,0}=\sqrt{diag(\widehat{\Sigma}_{n,0})}\in\mathbb{R}^k$. Then:*

1. *$V(X_{n,0},\hat{\sigma}_{n,0}) = \{\lambda_{(1)}(X_{n,0},\hat{\sigma}_{n,0})\gamma_{(1)}(X_{n,0}), ..., \lambda_{(J)}(X_{n,0},\hat{\sigma}_{n,0})\gamma_{(J)}(X_{n,0})\}$, where the $\lambda_{(j)}(\cdot,\cdot)$ are scalar functions of X and σ, while $\gamma_{(1)}(X_{n,0}),...,\gamma_{(J)}(X_{n,0})$ are the elements of $V(X_{n,0},\upsilon)$ for $\upsilon=\sqrt{Diag(TT')}$.*

2. *Let* $\Upsilon_{n,0} = \{Tu + \zeta_{n,0} | u \in \mathbb{R}^l\}$*, where* $\zeta_{n,0} = \frac{1}{\sqrt{n}}\sum_i \zeta_i(\beta_0)$*. Let* $V_\dagger(X_{n,0},\hat{\sigma}_{n,0})$ *be the subset of* $V(X_{n,0},\hat{\sigma}_{n,0})$ *corresponding with the indices* j *such that there exists some* $\sigma > 0$ *and some* $y \in \Upsilon_{n,0}$ *such that* $\lambda_{(j)}(X_{n,0},\sigma)\gamma_{(j)}(X_{n,0}) \in argmax_{\tilde{\gamma}\in V(X_{n,0},\sigma)}\tilde{\gamma}'y$ *and* $\lambda_{(j)}(X_{n,0},\sigma)\gamma_{(j)}(X_{n,0})'y > 0$*. Suppose* $V_\dagger(X_{n,0},\hat{\sigma}_{n,0})$ *is non-empty.*[17] *Then for any* $\alpha < 0.5$*, the LF, Conditional, and Hybrid tests constructed using* $V(X_{n,0},\hat{\sigma}_{n,0})$ *reject only if their analogs constructed using* $V_\dagger(X_{n,0},\hat{\sigma}_{n,0})$ *also reject.*

With the definition of $V_\dagger(X_{n,0},\hat{\sigma}_{n,0})$ in hand, we can now state our final assumption.

Assumption 4 *For* n *sufficiently large and all* β_0*,* $X_{n,0} \in \mathcal{X}^*$ *for* $\mathcal{X}^*$ *a closed set such that*

$$\inf_{\Omega\in\mathbf{\Omega}_{\bar{\lambda}}} \inf_{X\in\mathcal{X}^*} \inf_{\gamma,\tilde{\gamma}\in V_\dagger(X,\sigma(\Omega)),\gamma\neq\tilde{\gamma},c\in\mathbb{R}_{\geq 0}} (\gamma - c\cdot\tilde{\gamma})'T\Omega T'(\gamma - c\cdot\tilde{\gamma}) > 0,$$

where $\sigma(\Omega) = \sqrt{Diag(T\Omega T')}$*.*

Together with the structure for the variance matrix Σ imposed in Assumption 1, Assumption 4 ensures that (i) $\gamma' Y_{n,0}$ has nonvanishing asymptotic variance for all dual vertices $\gamma \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})$, and (ii) for distinct dual vertices γ and $\tilde{\gamma}$ in $V_\dagger(X_{n,0},\hat{\sigma}_{n,0})$, $\gamma' Y_{n,0}$ and $\tilde{\gamma}' Y_{n,0}$ are not perfectly positively correlated asymptotically. The former implies that $\hat{\eta}_{n,0}$ is continuously distributed in large samples, while the latter ensures that the dual problem $\max_{\gamma\in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})}\gamma' Y_{n,0}$ has a unique solution with probability tending to one.

In Appendix D, we provide lower-level sufficient conditions for Assumption 4 in settings where either $\Sigma(P_{D|Z},\beta_0)$ is full-rank or degeneracy in $\Sigma(P_{D|Z},\beta_0)$ arises from matching moments of opposite signs (e.g. moment equalities cast as inequalities). In these settings, we show that Assumption 4 holds automatically when $X_{n,0}$ is constant up to scale (as occurs, e.g., in the difference-in-differences setting of Rambachan & Roth (2022)). When $X_{n,0}$ is non-constant, a sufficient condition is that $X_{n,0}$ lies in a set $\mathcal{X}$ such that the distance between distinct vertices of $V(X,\upsilon)$ is bounded away from zero over $X \in \mathcal{X}$, where again $\upsilon = \sqrt{diag(TT')}$. Intuitively, this assumption requires that distinct vertices in $V(X_{n,0},\upsilon)$ not "converge to each other."

We also note that we do not require any additional assumptions about how $V(X,\sigma)$ depends on σ, since the proof of Lemma 4 shows that σ affects $V(X,\sigma)$ only through a continuous re-scaling of the vertices of $V(X,\upsilon)$. This enables us to establish size control when $\sigma_{n,0}$ is replaced with a consistent estimate $\hat{\sigma}_{n,0}$ without further assumptions.

[17] If not, then $\hat{\eta}_{n,0} \leq 0$ with probability 1, and thus none of our tests ever rejects for $\alpha < 0.5$.

It is worth highlighting that Assumption 4 involves the variance of $Y_{n,0}$ but not its mean $\mu_{n,0}$. This contrasts with linear independence constraint qualification (LICQ) assumptions that have been considered in other work (e.g., Cho & Russell 2021, Gafarov 2019), which restrict the set of moments that can bind in population and thus the value of $\mu_{n,0}$ (see Kaido et al. (2021) for discussion). In the simplest case without nuisance parameters ($X_{n,0}=0$), for example, Assumption 4 holds if all of the elements of $Y_{n,0}$ have positive variance and are not perfectly correlated, whereas a standard LICQ condition would impose that $\mu_{n,0}$ has a unique maximum element.[18] We explore the connections between LICQ and Assumption 4 more formally in Appendix F, where we show that LICQ implies that there is a unique solution to a "population version" of the dual for $\hat{\eta}_{n,0}$, whereas Assumption 4 only implies uniqueness of the sample version of the problem (but not necessarily the population version). The tests proposed in Cox & Shi (2022), as well as our LF test, do not require Assumption 4 for uniform asymptotic validity, and thus may be attractive in settings where the researcher is not comfortable with this assumption.

Under these assumptions, feasible versions of our tests, based on the observed $(Y_{n,0},X_{n,0})$, and the estimated variance $\widehat{\Sigma}_{n,0}$, are uniformly asymptotically valid.

Proposition 1 *Under Assumptions 1, 2, and 3 the least favorable test is uniformly asymptotically valid for $\alpha<0.5$,*

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} Pr_{P_{D|Z}}\left\{\hat{\eta}_{n,0}>c_{\alpha,LF}\left(X_{n,0},\widehat{\Sigma}_{n,0}\right)\right\}\leq\alpha.$$

Proposition 2 *Under Assumptions 1, 2, 3, and 4, the conditional and hybrid tests are uniformly asymptotically valid for $\alpha<0.5$,*

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} Pr_{P_{D|Z}}\left\{\hat{\eta}_{n,0}>c_{\alpha,C}\left(Y_{n,0},X_{n,0},\widehat{\Sigma}_{n,0}\right)\right\}\leq\alpha,$$

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} Pr_{P_{D|Z}}\left\{\hat{\eta}_{n,0}>c_{\frac{\alpha-\kappa}{1-\kappa},H}\left(Y_{n,0},X_{n,0},\widehat{\Sigma}_{n,0}\right)\right\}\leq\alpha.$$

5 Implementation

We next provide practical guidance on implementing the tests described above. We also provide Matlab code to facilitate implementation.[19]

[18] Rambachan & Roth (2022) show that in a special setting where β_0 enters the moments linearly, a population version of LICQ implies that our conditional test has optimal local asymptotic power.

[19] The code is available at https://github.com/jonathandroth/LinearMomentInequalities/.

5.1 Choice of Moments

Researchers can use our methods whenever their model implies conditional moment inequalities of the form (1). As discussed in Section 2.2, if the model (1) holds for a given (Y,X) pair, then it also holds if Y and X are interacted with any non-negative function of the instruments – i.e., if we replace Y and X with $\tilde{Y} = Y \otimes f(Z)$ and $\tilde{X} = X \otimes f(Z)$. An important choice in implementing our methods is thus the choice of the k moments (i.e., the choice of Y). A formal analysis of how to optimally choose the k moments is beyond the scope of this paper, but we offer some heuristic guidance.

Intuitively, including more informative moments can tighten the identified set based on the included moments, but including too many moments relative to the sample size can harm the quality of the normal approximation. Including uninformative moments (that are not infinitely slack) can also reduce the finite-sample power of our tests. The multivariate Berry-Esseen theorem (e.g. Bentkus 2003) suggests that the normal approximation to the distribution of the sample average should perform well when the number of moments included is sufficiently small relative to the sample size.[20] As a heuristic, Cox & Shi (2022) suggest that one should ensure there are at least 15 observations per cell in cases where the instruments $f(Z)$ are binary indicators for whether Z falls in a particular cell. In our Monte Carlo simulations below, where the instrument functions are continuous, we find that our proposed tests have good size control with 500 observations and up to 110 moments, although we caution that the quality of the normal approximation may depend on the specific data-generating process.

Regarding the choice of *which* k moments to use, researchers should include the moments that they think will be most informative about the parameter of interest. Note that interacting an original set of moments with an instrument function $f(Z)$ will only add identifying information to the extent that $f(Z)$ is correlated with Y and X, since if (Y,X) and $f(Z)$ are uncorrelated $E_P[f(Z)(Y - X\delta)] = E_P[f(Z)]E_P[Y - X\delta] \propto E_P[Y - X\delta]$, so adding the interaction does not shrink the set of values where the moment inequalities are satisfied on average. Heuristically, researchers should therefore include instrument functions that are likely to be strongly related to (Y,X).[21] Consistent with this intuition, Ho & Pakes

[20]Specifically, as discussed in Chernozhukov et al. (2017), we need the dimension of the moments (k) to be smaller than $o(n^{\frac{2}{7}})$ for the approximation to hold uniformly over all convex sets. If the moments are of the form $Y = TU$, as in Assumption 1, then the relevant dimension is $dim(U)$ rather than $dim(Y)$.

[21]As noted in Section 2.2 above, our approach does not deliver consistent tests in settings with continuously distributed Z_i. Hence, to derive general optimality results one would have to go beyond our finite-dimensional analysis. Armstrong (2014, 2018) and Chetverikov (2018) establish convergence rates for inference on the full parameter in partially identified settings, including rate-optimality results for procedures using particular kernel-based instruments and bandwidths. Their analysis could provide

(2014) use instrument functions based on the distance of an individual to a hospital, since their Y and X relate to individuals' choices of hospitals, and distance to the hospital is known to be an important determinant of hospital choice; see Section VI.B of Ho & Pakes (2014) for an intuitive discussion of how economic knowledge can inform the choice of moments. We also emphasize that applied researchers frequently conduct inference based on a finite set of unconditional moments implied by conditional moment inequalities, so the use of our methods does not introduce a new choice relative to this common practice in empirical work.

5.2 Forming confidence sets

Researchers often wish to compute confidence sets for the target parameter β. This can be achieved by discretizing the parameter space for β as $\{\beta_{(1)},...,\beta_{(L)}\}$ and testing the null hypothesis $H_0:\beta=\beta_{(l)}$ for each l using the tests described above. A confidence set can then be formed by collecting the grid points for which the test fails to reject. If the researcher is interested in a subvector of β – e.g. the first component of β is of interest, whereas the remaining components are nuisance parameters that enter the moments non-linearly – then the researcher can first form a confidence set for the full parameter vector β, and then obtain a confidence set for the parameter of interest by projection. We emphasize that test inversion is only required for β, and not for the nuisance parameters δ, which can lead to substantial computational simplifications when the dimension of δ is large. For the remainder of the section, we focus on the implementation of our tests for a particular null value β_0.

5.3 Estimating the conditional covariance

Our tests require an estimate of the average conditional variance, $\Omega_0=E_P[Var(U_i(\beta_0)|Z_i)]$. We briefly describe how a matching procedure proposed by Abadie et al. (2014) can be used to estimate Ω_0 when the data are i.i.d. across i; see Chetverikov (2018) and Horowitz & Spokoiny (2001) for alternative estimators. Let $\widehat{\Sigma}_Z$ be the sample variance of Z_i.[22] For each i, find the nearest neighbor using the Mahalanobis distance for Z_i:

$$\ell_Z(i)=\text{argmin}_{j\in\{1,...,n\},j\neq i}(Z_i-Z_j)'\widehat{\Sigma}_Z^{-1}(Z_i-Z_j).$$

a natural starting point for the study of asymptotic optimality in our setting. We thank Tim Armstrong for bringing these connections to our attention.

[22]The matching procedure described below assumes that $\widehat{\Sigma}_Z$ is non-singular. In certain applications, such as in our Monte Carlo, elements of Z_i may be linearly dependent by construction, leading $\widehat{\Sigma}_Z$ to be singular. In this case conditioning on a maximal linearly independent subset of Z_i is equivalent to conditioning on the full vector, so one can drop dependent elements from Z_i until $\widehat{\Sigma}_Z$ is non-singular.

For ease of exposition we assume that Z_i has at least one continuously distributed dimension, so that $\ell_Z(i)$ is unique for all i.[23] The estimate of Ω_0 is then:

$$\widehat{\Omega}_{n,0} = \frac{1}{2n}\sum_{i=1}^{n}\left(U_i(\beta_0) - U_{\ell_{Z(i)}}(\beta_0)\right)\left(U_i(\beta_0) - U_{\ell_{Z(i)}}(\beta_0)\right)'. \quad (16)$$

Appendix C provides regularity conditions under which $\widehat{\Omega}_{n,0}$ is uniformly consistent for Ω_0.

5.4 Computation of test statistic and critical values

To test the null hypothesis for a particular null value β_0, one needs to compute the test statistic $\hat{\eta}_{n,0}$ and the critical value for the relevant test ($c_{\alpha,LF}$,$c_{\alpha,C}$, or $c_{\alpha,H}$). We discuss computation of each component in turn.

5.4.1 Computing $\hat{\eta}_{n,0}$

The test statistic $\hat{\eta}_{n,0}$ can be computed by solving the linear program (10). This can be achieved using standard software, such as Matlab's `linprog` command. We recommend using the dual-simplex method in Matlab, which conveniently returns both the optimal value $\hat{\eta}_{n,0}$ as well as the optimal vector of Lagrange multipliers $\hat{\gamma}$, which is used for computing the conditional and hybrid critical values.

5.4.2 Computing LF critical values

Recall that the LF critical value $c_{\alpha,LF}$ is the $1-\alpha$ quantile of $\max_{\gamma \in V(X_{n,0},\hat{\sigma}_{n,0})} \gamma'\xi$ for $\xi \sim N(0,\widehat{\Sigma}_{n,0})$. By duality results for linear programming, we have that

$$\hat{\eta}(\xi) = \max_{\gamma \in V(X_{n,0},\hat{\sigma}_{n,0})} \gamma'\xi = \left(\min_{\eta,\delta} \eta \text{ subject to } \xi - X_{n,0}\delta \leq \eta \cdot \hat{\sigma}_{n,0}\right),$$

where $\hat{\sigma}_{n,0} = \sqrt{Diag(\widehat{\Sigma}_{n,0})}$. To compute $c_{\alpha,LF}$, one can simulate $\xi_{(1)},...,\xi_{(S)} \sim N(0,\widehat{\Sigma}_{n,0})$, compute $\hat{\eta}(\xi_{(s)})$ using the linear program in the previous display and then take the $1-\alpha$ quantile of $\hat{\eta}(\xi_{(1)}),...,\hat{\eta}(\xi_{(S)})$.[24] We use $S = 1000$ in our simulations below.

5.4.3 Computing conditional and hybrid critical values

To compute the conditional and hybrid critical values, one needs to compute $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$. Equation (14) gives an analytical formula for these quantities that involves a minimum

[23] If instead Z_i is entirely discrete, one can estimate $\widehat{\Omega}_{n,0}$ using the average of the sample conditional variances across Z_i cells.

[24] To increase computational speed and stability across different values of β, one can fix $Z_1,...,Z_S \sim N(0,I)$, and then set $\xi_s = \widehat{\Sigma}^{\frac{1}{2}}_{n,0} Z_s$.

and maximum over the set of dual vertices $V(X_{n,0},\hat{\sigma}_{n,0})$. Enumerating all of the vertices is, however, computationally prohibitive when there are many moments or nuisance parameters. Fortunately, we show in Appendix E that there are two computational shortcuts available that allow for computation of $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$ without vertex enumeration. First, when the problem for $\hat{\eta}_{n,0}$ has a non-degenerate solution, $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$ can each be written as the maximum/minimum of a set of at most k easy-to-compute elements.[25] Second, if the problem for $\hat{\eta}_{n,0}$ is degenerate, $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$ can be solved using a computationally-tractable bisection approach. We thus recommend to first check whether the solution to the primal problem (10) is non-degenerate, and if so, use the formula given in Lemma E.1; if not, then use the bisection approach described in Appendix E. We implement this approach in our publicly-available Matlab code, and find that it yields computationally tractable tests with as many as 110 moments and 11 parameters in our simulations below.

5.4.4 Simplifications when target parameters enter the moments linearly

In some settings, we may have inequalities of the form

$$E_{P_{D|Z}}[Y_i - X_{\beta,i}\beta - X_{\delta,i}\delta | Z_i] \le 0,$$

where β is the parameter of interest, δ is again a nuisance parameter, $X_{\beta,i}$ and $X_{\delta,i}$ are non-random conditional on Z_i, and the value of $(Y_i, X_{\beta,i}, X_{\delta,i})$ does not depend on β or δ. This structure arises, for example, in interval-valued regression if we are interested in the coefficient on an exogenous variable. This structure also arises in Rambachan & Roth (2022), who consider bounds on treatment effects in difference-in-differences settings under linear constraints on the possible violations of parallel trends. Moment inequalities of this sort can be cast into the form (1) by setting $Y_i(\beta) = Y_i - X_{\beta,i}\beta$ and $X_i(\beta) = X_{\delta,i}$. The methods described above can thus be applied directly.

The additional linear structure allows for multiple computational shortcuts, however. First, the conditional covariance matrix $E_P[Var_{P_{D|Z}}(Y_i(\beta)|Z_i)]$ does not depend on β, and thus the estimated variance $\widehat{\Sigma}_n$ need only be calculated once, rather than for every candidate value of β.[26] Second, the LF critical value $c_{\alpha,LF}(X_n, \widehat{\Sigma}_n)$ likewise does not depend on the value of β. As a result, a confidence set for the LF test can be computed by solving a linear

[25] The solution to the primal problem is said to be non-degenerate if $W_{n,0,B}$ is invertible, where $W_{n,0} = (\hat{\sigma}_{n,0},\ X_{n,0})$ and B indexes the set of binding moments in the primal. To use this approach, we also require that $e_1' W_{n,0,B} \ge 0$.

[26] We write $\widehat{\Sigma}_n$ instead of $\widehat{\Sigma}_{n,0}$, since the value does not depend on the null hypothesis. We apply an analogous convention for other variables, e.g. writing X_n instead of $X_{n,0}$ and $\hat{\sigma}_n$ instead of $\hat{\sigma}_{n,0}$.

program for each of the upper and lower bounds, without any test inversion at all. For instance, the lower bound of the confidence set for the LF test can be calculated by solving

$$\min_{\beta,\delta} \beta \text{ subject to } Y_n - X_{n,\beta}\beta - X_{n,\delta}\delta \leq c_{\alpha,LF} \cdot \hat{\sigma}_n,$$

where $Y_n = \frac{1}{\sqrt{n}}\sum_i Y_i$, and $X_{n,\beta}$ and $X_{n,\delta}$ are defined analogously. Computation of confidence sets for the conditional and hybrid tests still requires test inversion over a grid for β, but will be faster because $\widehat{\Sigma}_n$ and the first-stage LF critical value for the hybrid need only be computed once.

6 Simulations

6.1 Simulation Design

Our simulations are calibrated to Wollmann (2018)'s study of the bailouts of GM and Chryslers' truck divisions. As discussed in Example 3 above, Wollmann obtains bounds on the fixed cost of marketing a product using moment inequalities derived from revealed preference arguments. The fixed cost to firm f of marketing product j at time t is $\beta(\delta_{c,f} + \delta_g g_j)$ if the product was marketed at time $t-1$, and $\delta_{c,f} + \delta_g g_j$ otherwise. Consistent with (1), the parameter $\delta = (\delta_g, \{\delta_{c,f}\})$ enters the moments linearly for a fixed value of β.

The moments we consider take the form of the example given in equation (6) for the case where a product was marketed in both periods. To illustrate how performance varies with the number of parameters, we consider specifications where the intercept $\delta_{c,f}$ is constant across firms, specifications where it is allowed to vary across three groups of firms, and specifications where each of the nine firms in the data has its own intercept. In each case, we average the moment inequalities involving $\delta_{c,f}$ across firms assumed to have the same coefficient. We also vary the instruments used. See Appendix G for details on the exact construction of the moments. Overall, the number of moments varies between 6 and 110 across our specifications.

We consider inference on three parameters of interest: the cost of marketing the truck of mean weight when it was not marketed in the prior year;[27] the incremental cost of changing the weight of a product, δ_g; and the non-linear parameter β, where $1-\beta$ represents the proportional cost savings from marketing a product that was previously marketed relative to

[27] When we assume $\delta_{c,f}$ is common across firms this is $\delta_c + \delta_g \mu_g$, where μ_g is the population average weight of trucks. When we allow the estimated δ_c parameters to vary across groups, we estimate $l'\delta$, for $l = (\frac{1}{G}, \ldots, \frac{1}{G}, \mu_g)'$, where G denotes the number of groups and $\delta = (\delta_{c,1}, \ldots, \delta_{c,G}, \delta_g)'$. Note that since the simulation DGP holds the true value of δ_c constant across groups, the true value of the parameter is the same in all specifications.

a new product. For the first two target parameters, which can be written in the form $l'\delta$, we hold β fixed at its true value and treat the component of δ orthogonal to $l'\delta$ as the nuisance parameter. This allows us to examine performance in the linear case discussed in Section 5.4. In Wollman's setting the parameter β might be calibrated based on industry knowledge about the relative cost of marketing a new versus pre-existing product. As discussed in Section 5.2, if we instead treated β as unknown we could form joint confidence sets for β along with the linear combination of interest and obtain confidence sets for the linear parameter alone by projection. For inference on β we treat the entire vector δ as a nuisance parameter. Overall, the number of unknown parameters varies between 2 and 11 across our specifications.

We calibrate the data-generating process in our simulations using moments reported in Wollmann – see Appendix G for details. In each simulation draw, we generate data from a cross-section of 500 independent markets.[28] This is substantially larger than the 27 observations used by Wollmann, but allows us to consider specifications with a widely varying number of moments. All results are based on 500 simulations.

We consider the performance of the LF, Conditional, and Hybrid tests and compare these to several benchmarks. First, we compare to a studentized-max-statistic-based projection test which we label the least favorable projection, or LFP, test. Second, we compute the sCC and sRCC tests proposed in Cox & Shi (2022). The sRCC test, which is a refinement of the sCC test, can be computationally difficult when there are many parameters. For the specifications with 10+ parameters and 100+ moments, we therefore report an upper bound for the power of the sRCC test using the fact that the refinement to the sCC test can only matter when the test statistic falls in a certain range.[29] Third, we compute the projection tests of D. Andrews & Soares (2010), AS and Kaido et al. (2019*b*, KMS) using the EAM algorithm implemented in Matlab by Kaido et al. (2017). The AS and KMS tests can be computationally taxing when there are many parameters, and at present, the Matlab implementation of KMS by Kaido et al. (2017) is only written for settings where the parameters enter in an additively separable way. We therefore compute the AS and KMS tests only for the specifications when the parameters enter linearly and there are fewer than 10 parameters. See Appendix G for additional details on the implementation of these comparisons.

[28] The data in Wollmann (2018) are a time-series but his variance estimates assume no serial correlation, so we adopt a simulation design consistent with this.

[29] Specifically, the sRCC test always rejects when the sCC test does, and can only differ from the sCC test when one moment is active ($k=1$) and the test statistic falls between the $1-\alpha$ and $1-\alpha/2$ quantiles of the chi-squared distribution. When there are 10 or more parameters, we thus report the power of the test that rejects whenever either the sCC test rejects or the refinement could potentially lead the sRCC test to reject.

6.2 Results

Table 1 reports the maximum null rejection probability (size) over a conservative estimate of the identified set. Since we do not have an analytical characterization of the identified set, we approximate it by the set satisfying the sample (unconditional) moment inequalities based on a simulation run with five million observations. To ensure that our estimate of the identified set is conservative, we follow Chernozhukov et al. (2007) and add a correction factor to the moments of $\log(n)/\sqrt{n} \approx .003$. Our estimate of the identified set is thus conservative due to both (a) the Chernozhukov et al. (2007) correction factor and (b) the use of unconditional rather than conditional moment inequalities. All of the procedures nevertheless approximately control size on this set, with rejection probabilities never exceeding 0.08 for any of the procedures.

We next turn to comparisons of power. Figure 1 shows the rejection rates for each of our three main tests in the simulation design where the target parameter is the cost of the mean-weight truck. The vertical dashed lines denote conservative estimates of the bounds of the identified set, and the remaining curves show the probability that each of the tests rejects given a null value of the parameter of interest (holding fixed the DGP). Since the rejection probability is near-zero for all procedures in the interior of the identified set, we omit the portion of the x-axis well inside the identified set bounds so as to focus on the most relevant parts of the parameter space; the omitted part is grayed out in Figure 1 and subsequent figures.

Overall, the figure indicates that the hybrid approach performs best among our three procedures, with rejection probabilities comparable to or above those of the LF and conditional approaches at all points in the parameter space. To understand the superior performance of the hybrid approach, it is worth highlighting that the rejection curves for the LF and conditional approaches cross: in some specifications, the conditional approach has power substantially above that of the LF test at all parameter values (e.g. panel (e) of Figure 1). In other specifications, however, the conditional approach exhibits poor power relative to the LF test in some areas of the parameter space – e.g., in the area above the identified set in panel (d) of Figure 1. We have confirmed that in this simulation design for some parameter values there are two vertices which are optimal with approximately equal probability in this part of the parameter space, which as discussed in Section 3 can lead to poor power for the conditional test. Indeed, this feature can even lead the power curves for the conditional approach to be non-monotonic, since moving farther away from the identified set can push the mean values of a pair of vertices closer together. The hybrid approach has similar power to the conditional approach in most of the parameter space, while mitigating the issues in regions of the parameter space where multiple vertices are close to binding, thus leading to better performance overall. Appendix Figures G.1-G.2 show results when the parameter of interest is δ_g or β: the qualitative patterns are similar, with the hybrid exhibiting power comparable to or above the other two methods throughout the parameter space.

Table 2 provides a comparison of our three procedures relative to the other benchmarks. We report the median excess length for confidence sets formed based on each approach, where excess length is defined as the length of the confidence set minus the length of the identified set. For reference, we also report the length of the identified set. We find that the median excess length of the hybrid confidence set is below that for the AS and KMS sets in all specifications. The median excess length for the hybrid is also better or equal to that for the sCC and sRCC sets in most specifications, although the sRCC set outperforms the hybrid for three of the specifications with target parameter β.[30] The ranking of the hybrid and sRCC approaches in these results differs from that in the simulations in Cox & Shi (2022), who find better performance for sRCC. One potential factor is that the hybrid test is based on the max statistic whereas the sRCC test uses a QLR statistic, so the hybrid may be more powerful in settings where one moment is violated to a large extent, whereas the sRCC test may be more powerful when several moments are locally violated. Finally, it is worth highlighting that all of the procedures considered have better power than the LFP test in nearly all specifications. Appendix Figures G.3-G.7 display comparisons of the full power curves of the hybrid relative to the LFP, sCC, sRCC, AS, and KMS tests.

In our simulations the excess length of KMS intervals sometimes exceeds that of AS intervals. This is potentially surprising, since by construction the KMS test should reject whenever the AS test rejects, and thus should yield confidence intervals with uniformly shorter excess length. In practice, however, the bounds of the projected confidence intervals are approximated using a finite number of objective evaluations of the Evaluation-Approximation-Maximization algorithm studied by KMS, and thus are subject to optimization error. As a consequence of these optimization errors we find the median excess length of AS to be slightly smaller than that of KMS in two of our specifications (although by less than 2%). We have verified in an example where these issues arise that providing the EAM algorithm for AS with the optimal solution for KMS as a starting point leads to an AS interval that is a superset of the KMS interval. For simplicity, however, we report results from applying the EAM algorithm for AS directly.[31]

[30] Appendix Figures G.3-G.5 show a comparison of the power curves of the hybrid and the sCC and sRCC tests. The figures show that for several specifications the rejection curves for the hybrid and sRCC tests cross.

[31] We also found that reducing the objective tolerance to half the default value reduced (but did not fully

Lastly, Table 3 reports runtimes in minutes to calculate confidence sets for each parameter, averaging over 20 runs on a 2022 MacStudio (with M1 Ultra processor, 64GM RAM) without parallelizing the test inversion. Perhaps the most remarkable feature of the table is that our proposed tests are computationally tractable even in settings with as many as 11 parameters and 110 moments. Our preferred test, the hybrid, has runtimes under 5 minutes for all specifications in panels (a) and (b), where all of the parameters enter the moments linearly, and under 2 hours in all specifications in panel (c), where the target parameter enters the moments non-linearly. We emphasize that these runtimes could be further improved by parallelizing the test inversion.

We highlight a few noteworthy comparisons of runtimes across both procedures and specifications. First, the runtime of the hybrid test can be either faster or slower than the runtime of the sCC and sRCC tests proposed by Cox & Shi (2022) depending on the specification.[32] The hybrid test is faster in the majority of simulations where all parameters enter the moments linearly; this is because the LF test used in the first-stage of the hybrid is particularly fast for these specifications, as the LF confidence set can be calculated without any test inversion (see Section 5.4). The Cox & Shi (2022) tests are faster in most of the specifications in panel (c), where the target parameter enters the moments non-linearly and thus the LF critical value must be re-calculated for each candidate value of β, with the exception of the specification with the most moments and parameters in which the hybrid is faster. Second, the runtimes for the hybrid tests are faster than for the AS and KMS projection tests in nearly all specifications, with larger differences in settings with more moments/parameters.[33] In the specification in the fourth row of panel (b), for example, the hybrid test is over 14 times faster than both AS and KMS.[34] It is intuitive that the

eliminate) this issue, but were unable to reduce the tolerance further owing to computational constraints.

[32]The refinement for the sRCC test is needed relatively rarely, and thus the reported runtimes for the sRCC and sCC test are identical to two decimal places.

[33]Runtimes between the hybrid and sCC/sRCC tests are directly comparable, since both tests use test inversion over the same grid. Comparing runtimes between the hybrid and AS/KMS projection confidence sets is somewhat more difficult, since the former depends on the grid resolution while the latter depend on the stopping criteria for the EAM algorithm. Given that the EAM algorithm relies on several stopping criteria (see Kaido et al. 2017, p. 8), it is not entirely obvious how to align these parameters so that the computational accuracy of the tests is comparable. Note, however, that if the lower bound for the AS confidence set computed by the EAM algorithm is larger than that for the KMS confidence set, then the computational error in the former must be at least as large as the difference between the two computed endpoints. In the specification corresponding with the first row in Table 3, this difference is larger than the grid resolution used for the hybrid test in 13 percent of the cases, which provides suggestive evidence that the computational errors of the two approaches are often of a similar order of magnitude.

[34]We ran a single iteration of AS for the specification with 10 parameters and 38 moments, which took 5.5 hours to complete (and the EAM algorithm for the upper bound reached the maximum of 1000

computation time is faster for the hybrid since it exploits the linear conditional structure present in our setting, whereas the EAM algorithm used to calculate the AS/KMS CIs is designed for a larger class of potentially non-linear problems and thus does not make use of this additional structure. Third, both the conditional and hybrid tests are somewhat slower when the target parameter is δ_g (panel b) relative to the cost of the mean-weight truck (panel a). The reason is that the primal solution for $\hat{\eta}_{n,0}$ is often degenerate, and thus we must use the slower bisection method to calculate the $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$, as described in Appendix E.

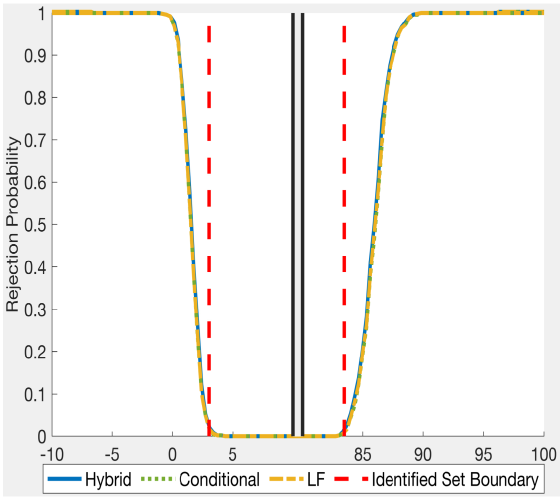
7 Conclusion

This paper considers the problem of inference based on linear conditional moment inequalities, which arise in a wide variety of economic applications. Using linear conditional structure, we develop inference procedures which remain both computationally tractable and powerful in the presence of nuisance parameters. We find good performance for our procedures under a variety of simulation designs based on Wollmann (2018), with especially good performance for our recommended hybrid procedure.

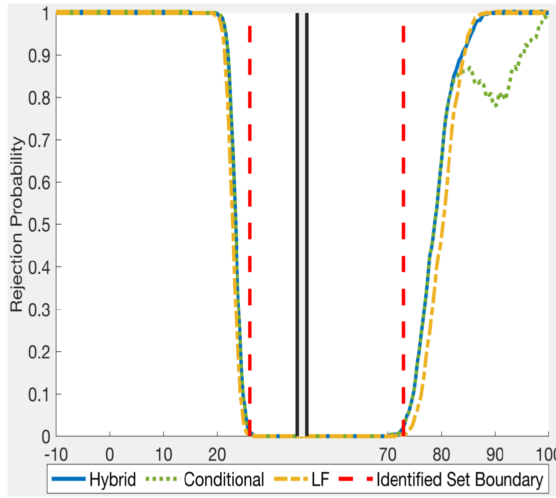
iterations without converging).

Figure 1: Rejection probabilities for 5% tests of fixed cost for truck of mean weight

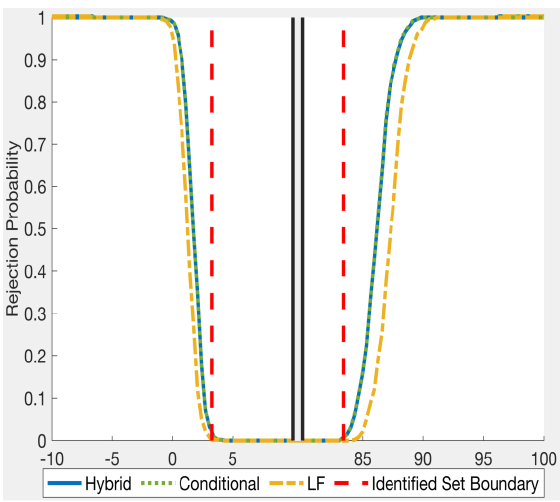
(a) 2 Parameters, 6 Moments



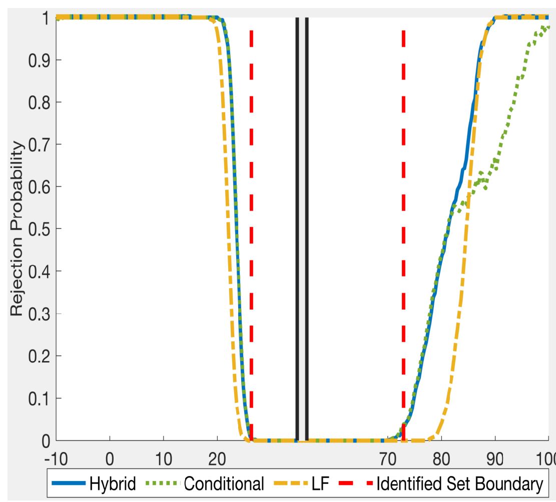
(b) 2 Parameters, 14 Moments



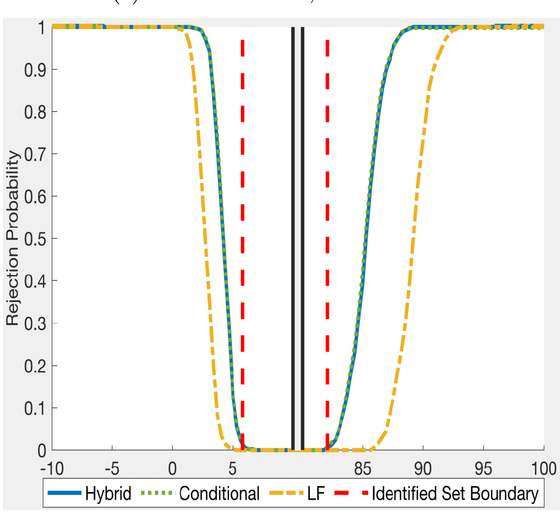
(c) 4 Parameters, 14 Moments



(d) 4 Parameters, 38 Moments



(e) 10 Parameters, 38 Moments



(f) 10 Parameters, 110 Moments

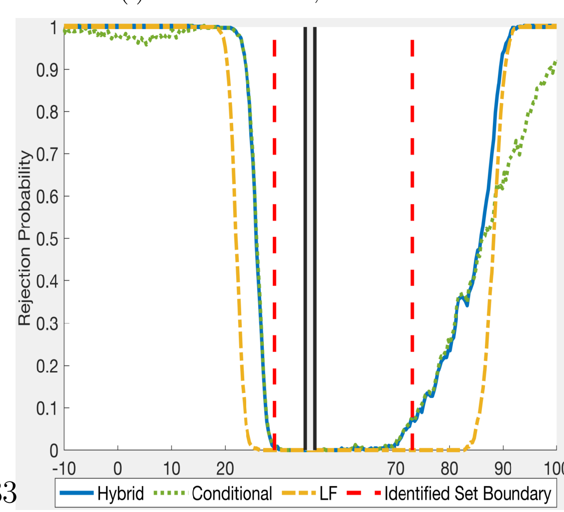


Table 1: Size Comparisons

(a) Parameter: Cost of Mean-Weight Truck

		Max Size							
#Params	#Moments	LF	Cond.	Hybrid	LFP	sCC	sRCC	AS	KMS
2	6	0.02	0.02	0.02	0.00	0.01	0.02	0.02	0.02
2	14	0.00	0.02	0.02	0.00	0.01	0.02	0.02	0.02
4	14	0.00	0.02	0.02	0.00	0.01	0.02	0.03	0.05
4	38	0.00	0.04	0.04	0.00	0.01	0.03	0.00	0.00
10	38	0.00	0.02	0.01	0.00	0.01	0.02		
10	110	0.00	0.07	0.07	0.00	0.00	0.00		

(b) Parameter: δ_g

		Max Size							
#Params	#Moments	LF	Cond.	Hybrid	LFP	sCC	sRCC	AS	KMS
2	6	0.04	0.04	0.06	0.01	0.02	0.04	0.03	0.03
2	14	0.02	0.05	0.05	0.00	0.03	0.05	0.02	0.02
4	14	0.03	0.04	0.05	0.00	0.03	0.04	0.04	0.05
4	38	0.00	0.05	0.05	0.00	0.03	0.05	0.07	0.08
10	38	0.00	0.05	0.05	0.00	0.03	0.05		
10	110	0.00	0.03	0.03	0.00	0.02	0.02		

(c) Parameter: β

		Max Size					
#Params	#Moments	LF	Cond.	Hybrid	LFP	sCC	sRCC
3	6	0.00	0.00	0.00	0.00	0.00	0.00
3	14	0.00	0.01	0.01	0.00	0.00	0.01
5	14	0.00	0.01	0.01	0.00	0.01	0.01
5	38	0.00	0.03	0.02	0.00	0.02	0.02
11	38	0.00	0.01	0.01	0.00	0.00	0.01
11	110	0.00	0.05	0.04	0.00	0.01	0.01

Table 2: Excess Length Comparisons

(a) Parameter: Cost of Mean-Weight Truck

			Median Excess Length							
#Params	#Moments	ID Set	LF	Cond.	Hybrid	LFP	sCC	sRCC	AS	KMS
2	6	80.42	3.99	4.08	3.76	5.33	4.73	4.08	4.12	4.14
2	14	46.89	10.30	10.31	8.36	12.57	9.66	8.36	9.67	9.80
4	14	80.13	5.92	4.37	4.37	7.57	5.02	4.37	5.82	5.38
4	38	46.61	16.14	14.49	11.56	18.88	12.86	12.54	15.90	15.41
10	38	76.21	10.21	4.72	4.72	12.71	5.37	4.72		
10	110	43.95	22.24	17.80	14.25	25.50	18.45	18.45		

(b) Parameter: δ_g

			Median Excess Length							
#Params	#Moments	ID Set	LF	Cond.	Hybrid	LFP	sCC	sRCC	AS	KMS
2	6	120.05	4.29	4.20	3.95	6.04	4.95	4.20	4.94	4.74
2	14	120.05	5.41	4.45	4.20	6.93	5.20	4.45	5.31	5.26
4	14	120.07	5.19	4.43	4.18	6.99	5.18	4.43	5.48	5.13
4	38	120.07	6.68	4.43	4.43	7.97	5.43	4.43	6.23	6.08
10	38	120.07	6.58	4.43	4.43	8.09	5.43	4.43		
10	110	120.07	7.69	5.18	5.18	9.11	7.43	7.18		

(c) Parameter: β

			Median Excess Length					
#Params	#Moments	ID Set	LF	Cond.	Hybrid	LFP	sCC	sRCC
3	6	16.89	61.87	42.93	36.62	118.69	60.61	42.93
3	14	1.41	0.55	0.45	0.35	0.76	0.45	0.35
5	14	8.71	7.78	6.01	5.30	10.25	6.36	5.66
5	38	1.31	0.66	0.96	0.45	0.86	0.40	0.35
11	38	2.99	1.01	1.01	0.81	1.41	0.71	0.71
11	110	1.01	0.66	2.57	0.55	0.86	0.45	0.45

Table 3: Computational Time Comparison

(a) Parameter: Cost of Mean-Weight Truck

		Average Runtime in Minutes							
#Params	#Moments	LF	Cond.	Hybrid	LFP	sCC	sRCC	AS	KMS
2	6	0.12	0.24	0.23	0.03	1.06	1.06	3.23	3.63
2	14	0.05	0.22	0.22	0.00	3.05	3.05	0.52	0.95
4	14	0.12	0.42	0.42	0.03	2.48	2.48	22.46	27.95
4	38	0.06	0.38	0.38	0.00	19.39	19.39	18.59	22.44
10	38	0.05	1.40	1.39	0.00	19.16	19.16		
10	110	0.10	0.75	0.79	0.01	208.65	208.65		

(b) Parameter: δ_g

		Average Runtime in Minutes							
#Params	#Moments	LF	Cond.	Hybrid	LFP	sCC	sRCC	AS	KMS
2	6	0.05	5.11	2.14	0.00	0.74	0.74	0.14	0.36
2	14	0.05	2.58	0.67	0.00	2.49	2.49	3.72	2.78
4	14	0.05	4.23	2.35	0.01	2.22	2.22	11.97	19.38
4	38	0.05	6.04	3.83	0.00	13.77	13.77	59.27	55.17
10	38	0.06	6.03	4.04	0.00	13.10	13.10		
10	110	0.10	5.80	3.93	0.01	127.46	127.46		

(c) Parameter: β

		Average Runtime in Minutes					
#Params	#Moments	LF	Cond.	Hybrid	LFP	sCC	sRCC
3	6	47.66	0.47	47.86	0.24	1.32	1.32
3	14	48.23	0.61	48.39	0.41	3.31	3.31
5	14	47.71	6.45	49.59	0.36	2.84	2.84
5	38	52.23	7.80	53.57	1.24	22.47	22.47
11	38	52.52	7.24	55.01	1.13	18.24	18.24
11	110	98.13	14.69	99.59	7.75	251.21	251.21

References

Abadie, A., Imbens, G. W. & Zheng, F. (2014), 'Inference for misspecified models with fixed regressors', *Journal of the American Statistical Association* **109**(508), 1601–1614.

Andrews, D. W. & Barwick, P. J. (2012), 'Inference for parameters defined by moment inequalities: A recommended moment selection procedure', *Econometrica* **80**(6), 2805–2826.

Andrews, D. W., Guggenberger, P. & Cheng, X. (2020), 'Generic results for establishing the asymptotic size of confidence sets and tests', *Journal of Econometrics* **218**(2), 496–531.

Andrews, D. W. & Shi, X. (2013), 'Inference based on conditional moment inequalities', *Econometrica* **81**(2), 609–666.

Andrews, D. W. & Soares, G. (2010), 'Inference for parameters defined by moment inequalities using generalized moment selection', *Econometrica* **78**(1), 119–159.

Andrews, I., Kitagawa, T. & McCloskey, A. (2021), Inference on winners. Working Paper.

Appa, G. (2002), 'On the uniqueness of solutions to linear programs', *The Journal of the Operational Research Society* **53**(10), 1127–1132.

Armstrong, T. B. (2014), 'Weighted ks statistics for inference on conditional moment inequalities', *Journal of Econometrics* **181**(2), 92–116.

Armstrong, T. B. (2018), 'On the choice of test statistic for conditional moment inequalities', *Journal of Econometrics* **203**(2), 241–255.

Armstrong, T. B. & Kolesar, M. (2018), 'Optimal inference in a class of regression models', *Econometrica* **86**(2), 655–683.

Belloni, A., Bugni, F. & Chernozhukov, V. (2018), Subvector inference in PI models with many moment inequalities. Working Paper.

Bentkus, V. (2003), 'On the dependence of the Berry-Esseen bound on dimension', *Journal of Statistical Planning and Inference* **113**(2), 385–402.

Beresteanu, A. & Molinari, F. (2008), 'Asymptotic properties for a class of partially identified models', *Econometrica* **76**(4), 763–814.

Bontemps, C., Magnac, T. & Maurin, E. (2012), 'Set identified linear models', *Econometrica* **80**(3), 1129–1155.

Bugni, F., Canay, I. & Shi, X. (2017), 'Inference for subvectors and other functions of partially identified parameters in moment inequality models', *Quantitative Economics* **8**(1), 1–38.

Chen, X., Christensen, T. & Tamer, E. (2018), 'Monte carlo confidence sets for identified sets', *Econometrica* **86**(6), 1965–2018.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2017), 'Central limit theorems and boostrap in high dimensions', *The Annals of Probability* **45**(4), 2309–2352.

Chernozhukov, V., Hong, H. & Tamer, E. (2007), 'Estimation and confidence regions for parameter sets in econometric models', *Econometrica* **75**(5), 1243–1284.

Chernozhukov, V., Newey, W. & Santos, A. (2015), Constrained conditional moment restriction models. Working Paper.

Chetverikov, D. (2018), 'Adaptive test of conditional moment inequalities', *Econometric Theory* **34**(1), 186–227.

Cho, J. & Russell, T. M. (2021), Simple inference on functionals of set-identified parameters defined by linear moments. Working paper.

Cox, G. & Shi, X. (2022), 'Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models', *The Review of Economic Studies* p. rdac015. **URL:** *https://doi.org/10.1093/restud/rdac015*

Eizenberg, A. (2014), 'Upstream innovation and product variety in the U.S. home pc market', *Review of Economic Studies* **81**(3), 1003–1045.

Fang, Z., Santos, A., Shaikh, A. & Torgovitsky, A. (2021), Inference for large-scale linear systems with known coefficients. Working paper.

Flynn, Z. (2019), Inference based on continuous linear inequalities via semi-infinite programming. Working Paper.

Gafarov, B. (2019), Inference in high-dimensional set-identified affine models. Working Paper.

Gandhi, A., Lu, Z. & Shi, X. (2019), Estimating demand for differentiated products with zeroes in market share data. Working Paper.

Ho, K. & Pakes, A. (2014), 'Hospital choices, hospital prices and financial incentives to physicians', *American Economic Review* **104**(12), 3841–84.

Ho, K. & Rosen, A. (2017), Partial identification in applied research, *in* B. Honore, A. Pakes, M. Piazessi, & L. Samuelson, eds, 'Advances in Economics and Econometrics', Cambridge University Press.

Horowitz, J. L. & Spokoiny, V. G. (2001), 'An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative', *Econometrica* **69**(3), 599–631.

Kaido, H., Molinari, F. & Stoye, J. (2019*a*), 'Confidence intervals for projections of partially identified parameters', *Econometrica* **87**(4), 1397–1432.

Kaido, H., Molinari, F. & Stoye, J. (2019*b*), 'Confidence Intervals for Projections of Partially Identified Parameters', *Econometrica* **87**(4), 1397–1432. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA14075.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14075*

Kaido, H., Molinari, F. & Stoye, J. (2021), 'Constraint qualifications in partial identification', *Econometric Theory* pp. 1–24.

Kaido, H., Molinari, F., Stoye, J. & Thirkettle, M. (2017), 'Calibrated projection in matlab: Users' manual', *arXiv:1710.09707 [econ, stat]* . arXiv: 1710.09707.

Kaido, H. & Santos, A. (2014), 'Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities', *Econometrica* **82**(1), 387–413.

Katz, M. (2007), Supermarkets and zoning laws. Ph.D. dissertation, Harvard University.

Lee, J. D., Sun, D. L., Sun, Y. & Taylor, J. E. (2016), 'Exact post-selection inference, with application to the lasso', *Annals of Statistics* **44**(3), 907–927.

Manski, C. F. & Tamer, E. (2002), 'Inference on regressions with interval data on a regressor or outcome', *Econometrica* **70**(2), 519–546.

Molinari, F. (2020), Microeconometrics with partial identification, *in* S. N. Durlauf, L. P. Hansen, J. J. Heckman & R. L. Matzkin, eds, 'Handbook of Econometrics', Vol. 7A, Elsevier, chapter 5, pp. 355–486.

Morales, E., Sheu, G. & Zahler, A. (2019), 'Extended gravity', *Review of Economic Studies* **86**(6), 2668–2712.

Ponomareva, M. & Tamer, E. (2011), 'Misspecification in moment inequality models: Back to moment equalities?', *Econometrics Journal* **14**(2), 186–203.

Rambachan, A. (2021), 'Identifying Prediction Mistakes in Observational Data', *Working paper* p. 91.

Rambachan, A. & Roth, J. (2022), 'An More Credible Approach to Parallel Trends', *Working paper* .

Romano, J. P. & Shaikh, A. (2008), 'Inference for identifiable parameters in partially identified econometric models', *Journal of Statistical Planning and Inference* **138**(9), 2786–2807.

Romano, J. P., Shaikh, A. & Wolf, M. (2014), 'A practical two-step method for testing moment inequalities', *Econometrica* **82**(5), 1979–2002.

Schrijver, A. (1986), *Theory of Linear and Integer Programming*, Wiley-Interscience.

Van der Vaart, A. (2000), *Asymptotic Statistics*, Cambridge University Press.

Wachsmuth, G. (2013), 'On LICQ and the uniqueness of Lagrange multipliers', *Operations Research Letters* **41**, 78–80.

Wollmann, T. (2018), 'Trucks without bailouts: Equilibrium product characteristics for commercial vehicles', *American Economic Review* **108**(6), 1364–1406.