Review of Economic Studies (2023) **90**, 201–228 https://doi.org/10.1093/restud/rdac015

Advance access publication 15 March 2022

Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models

GREGORY COX
Department of Economics, National University of Singapore

and

XIAOXIA SHI
Department of Economics, University of Wisconsin-Madison

First version received September 2020*; Editorial decision December* 2021*; Accepted March* 2022 (*Eds.*)

We propose a simple test for moment inequalities that has exact size in normal models with known variance and has uniformly asymptotically exact size under asymptotic normality. The test compares the quasi-likelihood ratio statistic to a chi-squared critical value, where the degree of freedom is the rank of the inequalities that are active in finite samples. The test requires no simulation and thus is computationally fast and especially suitable for constructing confidence sets for parameters by test inversion. It uses no tuning parameter for moment selection and yet still adapts to the slackness of the moment inequalities. Furthermore, we show how the test can be easily adapted to inference on subvectors in the common empirical setting of conditional moment inequalities with nuisance parameters entering linearly. User-friendly Matlab code to implement the test is provided.

Key words: Moment inequalities; Uniform inference; Likelihood ratio; Subvector inference; Convex polyhedron; Linear programming

JEL Codes: C12

1. INTRODUCTION

In the past decade or so, inequality testing has become a mainstream inference method used for models where standard maximum likelihood or method of moments are difficult to use, for reasons including multiple equilibria, incomplete data, or complicated dynamic patterns. In such models, inequalities can often be derived from equilibrium conditions and rational decision-making. Inference can then be conducted by inverting tests for these inequalities at each given

The editor in charge of this paper was Francesca Molinari.

parameter value. That is, by testing the inequalities at each parameter value and collecting the values at which the test does not reject to form a confidence set.[1]

Although conceptually simple, conducting inference via test inversion poses considerable computational challenges to practitioners. This is because, in order to get an accurate calculation of the confidence set, one needs to test the inequalities at a set of parameter values that is dense enough in the parameter space. Depending on the application, the number of values that need to be tested can be astronomical and increases exponentially with the dimension of the parameter space. Moreover, existing tests often require simulated critical values that are non-trivial to compute even for a single value of the parameter, let alone repeated for a large number of parameter values.[2]

Besides computational challenges, most existing methods for moment inequality models involve tuning parameter sequences that are required to diverge at a certain rate as the sample size increases. The threshold in the generalized moment selection procedures (e.g. Rosen, 2008; Andrews and Soares, 2010) and the subsample size in subsampling-based methods (e.g. Chernozhukov *et al.*, 2007; Romano and Shaikh, 2012) are notable examples.[3] Appropriate choices often depend on data in complicated ways, and an inappropriate choice can threaten the validity of the test.

Clearly, there are two ways to ease the computational burden: one is to make the inequality test easier for each parameter value, and the other is to reduce the number of parameter values that need to be tested. We contribute to the literature in both. First, we propose a simple test for general moment inequalities that requires no simulation. It simply uses the (quasi-) likelihood ratio statistic (T_n) and a chi-squared critical value, where the data-dependent degrees of freedom comes as a by-product of computing T_n. We call it a conditional chi-squared (CC) test. By not requiring simulation, the test saves computation time hundreds-fold compared to tests involving simulated critical values, where a statistic needs to be computed for each simulated sample. For example, in the simulation experiment reported in Section 5.1, our test is about 200–400 times faster than the recommended testing procedures in Andrews and Barwick (2012) (AB, hereafter) and Romano *et al.* (2014) (RSW, hereafter).

Second, we then consider a conditional moment inequality model where the parameter vector can be partitioned into two subvectors $(\theta',\delta')'$. The subvector θ is the parameter of interest, while δ is the subvector that the researcher is not interested in, commonly referred to as the nuisance parameter. We specialize to the setting where δ enters the moment inequalities linearly and propose a version of the CC test for θ. The subvector test is based on eliminating the nuisance parameters from a system of inequalities. By eliminating the nuisance parameters, one only needs to consider a grid on the space of θ, which can be much lower dimensional than the space of $(\theta',\delta')'$. Thus, the number of parameter values that need to be tested is drastically reduced. For example, in the simulation experiment reported in Section 5.2 below, our subvector test uses only

1. An incomplete list of applications that use inequalities as estimation restrictions includes Tamer (2003), Uhlig (2005), Bajari, Benkard and Levin (2007), Blundell, Gosling, Ichimura and Meghir (2007), Ciliberto and Tamer (2009), Beresteanu, Molchanov and Molinari (2011), Holmes (2011), Baccara, Imrohoroglu, Wilson and Yariv (2012), Chetty (2012), Nevo and Rosen (2012), Kawai and Watanabe (2013), Eizenberg (2014), Huber and Mellace (2015), Pakes, Porter, Ho and Ishii (2015), Magnolfi and Roncoroni (2016), Sheng (2016), Sullivan (2017), He (2017), Iaryczower, Shi and Shum (2018), Wollman (2018), Fack, Grenet and He (2019), and Morales, Sheu and Zahler (2019). For a recent overview of the literature, see for example, Ho and Rosen (2017), Canay and Shaikh (2017), and Molinari (2020).

2. Existing tests for general moment inequalities with simulated critical values include Chernozhukov, Hong and Tamer (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano and Shaikh (2012), and Romano, Shaikh and Wolf (2014). See Canay and Shaikh (2017) and Molinari (2020) for more references.

3. Arguably, the size of a first-stage confidence set or the number of simulation/bootstrap draws are also tuning parameters commonly used to test moment inequalities.

10 seconds to compute a confidence interval (CI) in a specification with a four-dimensional δ and 32 moment inequalities.

In both contexts, the CC test is simulation and tuning parameter free. Its critical value is simply the chi-squared critical value with degrees of freedom equal to the rank of the active moment inequalities, where we call a moment inequality *active* if it holds with equality at the restricted estimator of the moments.[4] In a normal model with known variance, the test is shown to have exact size in finite sample. That is, its worst-case rejection probability under the null hypothesis is equal to its nominal significance level. In an asymptotically normal model, it is shown to be uniformly asymptotically valid. Moreover, it automatically adapts to the slackness of the moment inequalities despite the absence of a deliberate moment selection step. Combining uniform asymptotic validity and adaptation to slackness is challenging, especially without using a tuning parameter. The CC test is able to achieve the combination thanks to a tractable and adaptive bound on the *conditional* distribution of the test statistic. Specifically, the intuition for the uniform asymptotic validity is that the conditional upper quantiles of the likelihood ratio statistic are asymptotically bounded by the quantiles of the chi-squared distribution with k degrees of freedom, given the event that k inequalities are active. The intuition for the adaptiveness is that very slack inequalities tend not to become active and hence not to count toward the degree of freedom used in the critical value calculation.

The idea of simple chi-squared critical values for testing inequalities appeared as early as in Bartholomew (1961) and Rogers (1986) for testing one-sided alternatives against a *simple* null but was only recently proved to be valid for a composite null in Mohamad, van Zwet, Cator and Goeman (2020) in a normal model. We extend Mohamad *et al.* (2020) in four ways: (1) we allow an intercept in the inequalities defining the null hypothesis and thus generalize the null hypothesis from a cone to a polyhedron. This is important for moment inequality models as, in the limit, the null hypothesis may not be a cone when some inequalities are close to binding; (2) we design a simple but novel refinement to make the test size exact; (3) we prove the test is uniformly asymptotically valid in moment inequality models; and (4) we show how to feasibly extend the test to the subvector inference context in the presence of nuisance parameters that enter the moments linearly. Extensions (1)–(3) rely on technical contributions described in a Supplementary Appendix. We highlight them briefly here, as they may be useful in other contexts. The finite sample validity of the refinement relies on a careful partition of the state space (see Lemmas 3 and 4) combined with an inequality on the tail of the truncated normal distribution (see Lemma 6). The uniform asymptotic validity relies on a lemma guaranteeing convergence of an arbitrary sequence of polyhedra to a limiting polyhedron along a subsequence (see Lemma 9).

The idea of eliminating nuisance parameters from linear moment inequalities is first suggested in Guggenberger, Hahn and Kim (2008), where they introduce Fourier–Motzkin elimination, a classical algorithm for eliminating nuisance parameters from linear inequalities, to the literature and propose a Wald-type test on the resulting inequalities. Yet two main difficulties hinder the application of this idea: (1) numerical calculation of the Fourier–Motzkin elimination in general is an NP-hard computational problem and (2) the estimated coefficients in front of the nuisance parameters enter the resulting inequalities via a non-differentiable function and could undermine the validity of testing procedures applied directly to them. The first difficulty is circumvented because the CC test only relies on the rank of the active inequalities, and results from the convex analysis literature (see Lemmas 1 and 2) allow us to compute the rank of the active inequalities without carrying out Fourier–Motzkin elimination. The second difficulty is circumvented by

4. Active inequalities are the sample counterpart of binding inequalities, which hold with equality at the population expectation of the moments. An inequality that is not active is referred to as inactive. An inequality that is not binding is referred to as slack.

considering models where the moment inequalities hold conditional on a vector of instrumental variables, a class of models first proposed by Andrews, Roth and Pakes (2019).

Andrews *et al.* (2019) (hereafter ARP) have the closest setting with our article. They propose a test based on the largest standardized sample moment. In the most basic version, their test uses a conditional critical value from a truncated normal distribution. This basic version involves no simulation or tuning parameter and as a result is easy to compute. However, the basic version has poor power properties that prompt them to recommend a hybrid test. The hybrid test uses a simulated critical value as well as a tuning parameter that determines the size of a first-stage least favourable test.

There are a few papers in the literature that propose methods to mitigate the computational challenges described above. Kaido, Molinari and Stoye (2019) cast the problem of finding the bounds of the projection CI of each parameter into a non-linear non-convex constrained optimization problem and provide a novel algorithm to solve this optimization problem more efficiently. Our simple inequality test is complementary to Kaido *et al.* (2019)'s algorithm in that we make testing for each value hundreds-fold easier while their algorithm reduces the number of values that need to be tested. Bugni *et al.* (2017) propose a profiling method that simplifies computation in the same way as the subvector confidence set proposed in this article, by reducing the search from the space of the whole parameter vector to that of a low-dimensional subvector. The difference is that our subvector test, by taking advantage of the linearity of the model, is much easier to compute than Bugni *et al.* (2017)'s test, which applies more generally. Chen, Christensen and Tamer (2018) propose a quasi-Bayesian method that can also be applied to subvector inference in moment inequality models, as well as a simple method that applies to scalar parameters of interest. Gafarov (2019) and Cho and Russell (2020) propose CIs for affine functionals of parameters defined by affine moment inequalities, both by approximating the null distributions of regularized support functions of the identified set.

A couple of other papers aim to reduce the sensitivity of testing to tuning parameters. AB refine the procedure of Andrews and Soares (2010) (AS, hereafter) by computing an optimal moment selection threshold that maximizes a weighted average power (WAP) and a size correction. Using the optimal threshold and the size correction provided in that paper, one no longer needs to choose a tuning parameter. Computationally, it is the same as AS if one has 10 or fewer moment inequalities and can use the tables of optimal tuning and size correction values in the article. It is much more computationally demanding otherwise. RSW replace the moment selection step of the previous literature with a confidence set for the slackness parameter and employ a Bonferroni correction to take into account the error rate of this confidence set. There is still a tuning parameter, the confidence level of the first step, but this tuning parameter no longer affects the asymptotic size of the test. Computationally, using the same number of bootstrap draws, it is slightly more costly than AS due to the first-step confidence set construction. The recommended tests in AB and RSW are our points of comparison in the simulation experiments in Section 5.1, where we show that our simple test saves computational cost hundreds-fold, while having competitive size and power.

The remainder of this article proceeds as follows. Section 2 describes our setup and several examples. Section 3 describes how to implement the full-vector and subvector CC tests. Section 4 states the theoretical properties that the tests have. Section 5 reports the simulation results. Section 6 concludes. The Supplementary Appendix contains the proofs and additional results.

2. SETUP AND EXAMPLES

This section describes the setup for full-vector and subvector moment inequality testing, together with several examples.

2.1. *Moment inequality model: full-vector inference*

Consider a d_m-dimensional moment function, $m(W_i,\theta)$, that depends on a vector parameter of interest, θ. Let Θ denote the parameter space for θ, and denote the data by $\{W_i\}_{i=1}^n$ with joint distribution F. We assume the moments satisfy a vector of linear inequalities given by

$$A\mathbb{E}_F\overline{m}_n(\theta)\le b, \tag{1}$$

where A is a $d_A\times d_m$ matrix, b is a $d_A\times 1$ vector, and $\overline{m}_n(\theta)=n^{-1}\sum_{i=1}^n m(W_i,\theta)$. The moment inequalities identify the true parameter value up to the identified set,[5]

$$\Theta_0(F)=\{\theta\in\Theta : A\mathbb{E}_F\overline{m}_n(\theta)\le b\}. \tag{2}$$

The specification of a moment inequality model given by (1) is very general. Other papers in the moment inequality literature, such as AS, specify moment inequalities of the form

$$\mathbb{E}_F m_1(W_i,\theta)\le \mathbf{0} \text{ and } \mathbb{E}_F m_2(W_i,\theta)=\mathbf{0}, \tag{3}$$

where $m_1(W_i,\theta)$ denotes a d_{m_1}-vector of moments that satisfy inequalities and $m_2(W_i,\theta)$ denotes a d_{m_2}-vector of moments that satisfy equalities. By including a coefficient matrix A and an intercept b, (1) covers the specification in (3) with $b=\mathbf{0}$ and

$$A=\begin{pmatrix} I_{d_{m_1}} & \mathbf{0}_{d_{m_1}\times d_{m_2}} \\ \mathbf{0}_{d_{m_2}\times d_{m_1}} & -I_{d_{m_2}} \\ \mathbf{0}_{d_{m_2}\times d_{m_1}} & I_{d_{m_2}} \end{pmatrix}, \tag{4}$$

where $d_A=d_{m_1}+2d_{m_2}$. Introducing A and b is convenient because it allows us to succinctly cover both equalities and inequalities. It also readily accommodates models with upper and lower bounds with a deterministic gap in between.[6] Below, we assume the variance–covariance matrix of the moments is invertible. Introducing A and b is useful because it specifies the inequalities as a linear combination of a "core" set of moments, and only the core set of moments needs to have an invertible variance–covariance matrix.

Moment inequalities have become widely used in practice as the reference list given in the first paragraph of the introduction shows. We mention two recent examples here.

Example 1. *He (2017) uses a moment inequality model to estimate preferences of applicants in a school admission problem under a matching mechanism called the Boston mechanism. To fix ideas, consider a simple case with three schools, a, b, and c. Each applicant i submits a rank-ordered list (r_i^1, r_i^2, r_i^3) to the mechanism. The Boston mechanism first assigns as many applicants as possible to the top-ranked school in their list while respecting the capacity constraints of the schools. The unassigned applicants are considered by their second-ranked school for the remaining school seats, if any. The process continues until all seats are filled or all students assigned. The Boston mechanism is not strategy-proof in that applicants, instead of submitting their true preference ranking, can benefit from an untruthful rank-ordered list.*

He (2017) aims to answer an important policy question: does switching to a strategy-proof mechanism make the less sophisticated applicants better off? The question necessitates He to

5. The quantities A and b may depend on θ and the sample size n, a dependence that we keep implicit for simplicity unless otherwise needed. If the dependence is made explicit, the formula for $\Theta_0(F)$ becomes $\{\theta\in\Theta : A(\theta)\mathbb{E}_F\overline{m}_n(\theta)\le b(\theta)\}$.

6. For example, $\mathbb{E}[\bar{W}_n]-1\le\theta\le\mathbb{E}[\bar{W}_n]$ can be written in our notation with $m(w,\theta)=\theta-w$, $A=\binom{1}{-1}$ and $b=\binom{0}{1}$.

allow in his model less sophisticated applicants who do not form correct beliefs about admission probabilities. He allows them to form individualized beliefs which become incidental parameters that preclude full identification. However, He shows that the model can uniquely predict the probability of some rank-ordered lists and bound that of other rank-ordered lists using functions that do not involve beliefs. For example,

$$\Pr((r_i^1, r_i^2, r_i^3) = (0,0,0)) = \Pr(u_{ia} < u_{i0}, u_{ib} < u_{i0}, u_{ic} < u_{i0}) \text{ or equivalently} \tag{5}$$
$$\mathbb{E}[1\{(r_i^1, r_i^2, r_i^3) = (0,0,0)\}] = g_{000}(\theta) := \int 1\{u_a < u_0, u_b < u_0, u_c < u_0\} dF(u_0, u_a, u_b, u_c|\theta),$$

where u_{is} is the utility of being admitted to school s, u_{i0} is the utility of the outside option, and $F(\cdot|\theta)$ is the joint distribution of $(u_{i0}, u_{ia}, u_{ib}, u_{ic})$ assumed to be known up to the parameter θ. Also,

$$\Pr((r_i^1, r_i^2, r_i^3) = (a,b,c)) \le \Pr(u_{ia} \ge u_{i0}, u_{ia} \ge \min\{u_{ib}, u_{ic}\}), \text{ or equivalently} \tag{6}$$
$$\mathbb{E}[1\{(r_i^1, r_i^2, r_i^3) = (a,b,c)\}] \le g_{abc}(\theta) := \int 1\{u_a > u_0, u_a > \min\{u_b, u_c\}\} dF(u_0, u_a, u_b, u_c|\theta),$$

where the preference restriction on the right-hand side of the first line means that school a cannot be unacceptable (i.e. worse than the outside option) or be the least favourite. While everyone who submits (a,b,c) must have those preferences, not everyone who has such preferences will submit (a,b,c). For example, if an applicant expects the admission probability at school a is too low even when it is top-ranked, she may rank it bottom. Thus, we have an inequality instead of an equality in (6). Observing a data set of $\{r_i^1, r_i^2, r_i^3\}$, one can use moment equalities and inequalities like (5) and (6) to conduct inference on θ. In this example, our formulation (1) using the A matrix and the b intercept is particularly useful because there are equalities and inequalities, the probabilities on the left-hand side sum up to one, and the probabilities may have both upper and lower bounds.

Example 2. *Morales et al. (2019) use moment inequalities to estimate a model of international trade to quantify the importance of extended gravity, which is the dependence of an exporting firm's entry cost to a new market on its previous exporting to similar markets. Due to extended gravity, the set of markets the firm exports to has important dynamic implications for future entry costs. Thus, it becomes necessary to consider a dynamic discrete choice model with the choice set of each firm being the power set of potential markets. The sheer size of this choice set makes it difficult, if not impossible, to estimate the model using traditional maximum likelihood methods.*

Morales et al. (2019) form moment inequalities by comparing the equilibrium profit with the profit after a single-period perturbation from the optimal strategy. More specifically, they obtain moment inequalities of the form

$$\mathbb{E}[(\pi_{ijj't} + \delta\pi_{ijj't+1})\mathcal{I}(Z_{it})1\{d_{ijt}(1-d_{ij't}) = 1\}] \ge \mathbf{0}, \tag{7}$$

where i is the firm index, j and j' are indices for destination markets, t is time, $d_{ijt} = 1$ indicates that firm i exports to market j at time t and $d_{ijt} = 0$ indicates otherwise, $\pi_{ijj't}$ is the loss of static profit at time t due to switching the time t export from country j to j', $\pi_{ijj't+1}$ is the loss of static profit at time $t+1$ due to the same switch, δ is the discount factor, Z_{it} is a vector of instrumental variables, and $\mathcal{I}(Z_{it})$ is a vector of nonnegative functions of Z_{it}. Profit is parameterized to reflect revenue and costs with the extended gravity parameters as part of the cost function. Morales et al. (2019) use the AS test on a grid of parameter values to compute a joint confidence set for a five-dimensional parameter.

2.2. *Conditional moment inequality model: subvector inference*

We next consider a conditional moment inequality model with nuisance parameters entering linearly. The model has three differences from the full-vector setup: (1) there are nuisance parameters, denoted by δ, that enter the moments linearly, (2) the inequalities hold conditionally on exogenous variables, $\{Z_i\}_{i=1}^n$, and (3) the coefficients on δ depend only on the exogenous variables, $\{Z_i\}_{i=1}^n$. In mathematical terms, the model is given by

$$\mathbb{E}_{F_Z}[B_Z\overline{m}_n(\theta)-C_Z\delta|Z]\leq d_Z,\ a.s. \tag{8}$$

where $Z=\{Z_i\}_{i=1}^n$ is a sample of instrumental variables (each Z_i is taken to be a subvector of W_i without loss of generality), B_Z, C_Z, and d_Z are $k\times d_m$, $k\times p$, and $k\times 1$ matrix-valued measurable functions of Z, δ is a vector of unknown nuisance parameters, θ is a vector of unknown parameters of interest, and F_Z denotes the conditional distribution of $\{W_i\}_{i=1}^n$ given $\{Z_i\}_{i=1}^n$. The subscript Z is used to denote dependence on $Z_1,\ldots,Z_n$. The quantities B_Z, C_Z, and d_Z are also allowed to depend on θ and the sample size n (while k and p are fixed), but we keep that implicit for notational simplicity. Similar to the full-vector case, this model can succinctly cover both equalities and inequalities by an appropriate choice of B_Z, C_Z, and d_Z, as well as accommodating upper and lower bounds with a gap between the bounds that depends on $\{Z_i\}_{i=1}^n$.

The model (8) was first spotted as an interesting class of models by ARP. It is a special case of the conditional moment inequality models considered in Andrews and Shi (2013). We note two special features of this setup: (1) the nuisance parameter δ enters linearly and (2) the coefficients on δ depend only on the exogenous variables $\{Z_i\}_{i=1}^n$. ARP recognized that, while these features significantly restrict the generality of the full-vector model, they are common in many empirical models: exogenous covariates are frequently used to incorporate heterogeneity and/or to control for confounders. We use these features to develop a conditional subvector inference procedure, which, like our full-vector test, is tuning parameter and simulation free. Here, we describe three examples of models that fit this setup.

Example 3. *Manski and Tamer (2002) consider an interval regression model:*

$$Y_i^*=X_i'\theta+Z_{ci}'\delta+\varepsilon_i, \tag{9}$$

where Y_i^* is a dependent variable, X_i is a vector of possibly endogenous variables, Z_{ci} is a vector of exogenous covariates including the constant. There is a vector of excluded instrumental variables Z_{ei} that satisfies $\mathbb{E}[\varepsilon_i|Z_i]=0$, where $Z_i=(Z_{ci}',Z_{ei}')'$. The outcome Y_i^* is not observed. Instead, Y_{Li} and Y_{Ui} are observed such that $Y_i^*\in[Y_{Li},Y_{Ui}]$. The imperfect observation of Y_i^* may be caused by missing data or survey design where respondents are given a few brackets to select from instead of asked to give a precise answer.

Let $\mathcal{I}(Z_i)$ be a finite non-negative vector of instrumental functions. Then, we have

$$\mathbb{E}\left[\left(\begin{matrix}(Y_{Li}-X_i'\theta_0)\mathcal{I}(Z_i)\\-(Y_{Ui}-X_i'\theta_0)\mathcal{I}(Z_i)\end{matrix}\right)-\left(\begin{matrix}\mathcal{I}(Z_i)Z_{ci}'\\-\mathcal{I}(Z_i)Z_{ci}'\end{matrix}\right)\delta_0\middle|Z_i\right]\leq\mathbf{0}, \tag{10}$$

which yields a model of the form (8) with $B_Z=I$, $W_i=(Y_{Li},Y_{Ui},X_i',Z_i')'$, $m(W_i,\theta)=\left(\begin{smallmatrix}(Y_{Li}-X_i'\theta)\mathcal{I}(Z_i)\\-(Y_{Ui}-X_i'\theta)\mathcal{I}(Z_i)\end{smallmatrix}\right)$, $C_Z=n^{-1}\sum_{i=1}^n\left(\begin{smallmatrix}\mathcal{I}(Z_i)Z_{ci}'\\-\mathcal{I}(Z_i)Z_{ci}'\end{smallmatrix}\right)$, and $d_Z=\mathbf{0}$.

Example 4. *Gandhi, Lu and Shi (2019) consider a generalized interval regression model to conduct inference for an aggregate demand function when observed market shares of differentiated products have many zero values. Mathematically, the latent inverse demand model is of the form*

$$\psi(Y_i^*, X_i, \theta_0) = Z_{ci}'\delta_0 + \varepsilon_i, \ \ \mathbb{E}[\varepsilon_i | Z_i] = 0, \tag{11}$$

where ψ is a known function, but Y_i^*, the expectation of the market share in market i, is unobserved. Under an assumption on the source of the zeroes, Gandhi et al. (2019) construct bounds, $\psi_i^U(\theta_0)$ and $\psi_i^L(\theta_0)$, such that

$$\mathbb{E}[\psi_i^L(\theta_0)|Z_i] \leq \mathbb{E}[\psi(Y_i^*, X_i, \theta_0)|Z_i] \leq \mathbb{E}[\psi_i^U(\theta_0)|Z_i]. \tag{12}$$

Then, analogously to the previous example, we have

$$\mathbb{E}\left[\left(\begin{array}{c} \psi_i^L(\theta_0)\mathcal{I}(Z_i) \\ -\psi_i^U(\theta_0)\mathcal{I}(Z_i) \end{array} \right) - \left(\begin{array}{c} \mathcal{I}(Z_i)Z_{ci}' \\ -\mathcal{I}(Z_i)Z_{ci}' \end{array} \right) \delta_0 \middle| Z_i \right] \leq \mathbf{0}, \tag{13}$$

where $\mathcal{I}(Z_i)$ is a finite non-negative vector of instrumental functions of $Z_i = (Z_{ci}', Z_{ei}')'$. This yields a model of the form (8), where $B_Z = I$, W_i contains Z_i as well as the variables used to construct ψ_i^L and ψ_i^U, $m(W_i, \theta) = \left(\begin{smallmatrix} \psi_i^L(\theta)\mathcal{I}(Z_i) \\ -\psi_i^U(\theta)\mathcal{I}(Z_i) \end{smallmatrix} \right)$, $C_Z = n^{-1}\sum_{i=1}^n \left(\begin{smallmatrix} \mathcal{I}(Z_i)Z_{ci}' \\ -\mathcal{I}(Z_i)Z_{ci}' \end{smallmatrix} \right)$, and $d_Z = \mathbf{0}$.

In Section 5.2, we consider a Monte Carlo example of a special case of this model where we also provide more details on the bound construction. In the application of Gandhi et al. (2019), control variables (Z_{ci}) are essential for the validity of the instruments.

Example 5. *Eizenberg (2014) studies the portable PC market to quantify the welfare effect of eliminating a product. Central to the question is the fixed cost of providing the product. Eizenberg uses the revealed preference approach to construct bounds, L_i and U_i, for the fixed cost of product i. Let Z_i be a vector of product characteristics (including the constant). One can consider the following conditional moment inequality model:*

$$\begin{aligned} &\mathbb{E}\left[(L_i - P(Z_i)'\gamma_0)\mathcal{I}(Z_i)|Z_i\right] \leq \mathbf{0} \\ &\mathbb{E}\left[(-U_i + P(Z_i)'\gamma_0)\mathcal{I}(Z_i)|Z_i\right] \leq \mathbf{0}, \end{aligned} \tag{14}$$

where $P(Z_i)$ is a vector of known functions of Z_i and $\mathcal{I}(Z_i)$ is a vector of nonnegative instrumental functions. The function $P(Z_i)'\gamma_0$ captures the (observed) heterogeneity of fixed costs across products. Using our method, one can construct CIs for each element of γ_0 and any linear combinations of γ_0 such as the average derivative.

Suppose the parameter of interest is the average derivative with respect to the first element of Z_i: $\theta_0 = \gamma_0'\zeta$, where $\zeta = n^{-1}\sum_{i=1}^n \partial P(Z_i)/\partial z_1$. One can first reparametrize the model so that θ_0 becomes an element of the parameter. To do so, let $\zeta_\perp$ denote a $d_\gamma - 1 \times d_\gamma$ matrix with orthonormal rows that are all orthogonal to ζ, where d_γ is the dimension of ζ_0. Let $\delta_0 = \zeta_\perp\gamma_0$. Then $\gamma_0 = \theta_0\zeta/\|\zeta\| + \zeta_\perp'\delta_0$. Plug it into (14), and we obtain the reparametrized model:

$$\begin{aligned} &\mathbb{E}\left[(L_i - P(Z_i)'\zeta\theta_0/\|\zeta\|)\mathcal{I}(Z_i) - \mathcal{I}(Z_i)P(Z_i)'\zeta_\perp'\delta_0|Z_1\right] \leq \mathbf{0} \\ &\mathbb{E}\left[(-U_i + P(Z_i)'\zeta\theta_0/\|\zeta\|)\mathcal{I}(Z_i) + \mathcal{I}(Z_i)P(Z_i)'\zeta_\perp'\delta_0|Z_i\right] \leq \mathbf{0}, \end{aligned} \tag{15}$$

which falls into the framework of (8) where $B_Z = I$, W_i contains Z_i as well as the variables used to construct L_i and U_i, $m(W_i, \theta) = \left(\begin{smallmatrix} (L_i - P(Z_i)'\zeta\theta/\|\zeta\|\mathcal{I}(Z_i) \\ -(U_i - P(Z_i)'\zeta\theta/\|\zeta\|)\mathcal{I}(Z_i) \end{smallmatrix} \right)$, $C_Z = n^{-1}\sum_{i=1}^n \left(\begin{smallmatrix} \mathcal{I}(Z_i)P(Z_i)'\zeta_\perp' \\ -\mathcal{I}(Z_i)P(Z_i)'\zeta_\perp' \end{smallmatrix} \right)$, and $d_Z = \mathbf{0}$.

Two additional examples that fit into our subvector framework are Katz (2007) and Wollman (2018) as reviewed in ARP.

3. CC TEST: IMPLEMENTATION

In this section, we define a new test, called the CC test, for the inequalities specified in (1) and (8). It is called a CC (conditional chi-squared) test because it uses a critical value that is a quantile of the chi-squared distribution, where the degree of freedom depends on the active inequalities. We give instructions for implementing the test, which shows that it is easy to implement and has low computational cost.

3.1. *Full-vector test*

We use the inequalities specified in (1) to test hypotheses on θ. Like most papers in the literature, including AS, AB, and RSW, we conduct inference for the true parameter θ_0 by test inversion. That is, for a given significance level $\alpha \in (0,1)$, one constructs a test $\phi_n(\theta,\alpha)$ for $H_0 : \theta = \theta_0$, where $\phi_n(\theta,\alpha) = 1$ indicates rejection and $\phi_n(\theta,\alpha) = 0$ indicates a failure to reject. One then obtains the confidence set for θ_0 by calculating

$$CS_n(1-\alpha) = \{\theta \in \Theta : \phi_n(\theta,\alpha) = 0\}. \tag{16}$$

In practice, $CS_n(1-\alpha)$ is calculated by testing $H_0 : \theta = \theta_0$ on a grid of values of $\theta \in \Theta$.

We introduce a new test that is easy to compute, requiring no tuning parameters or simulations. It uses the (quasi-) likelihood ratio statistic,

$$T_n(\theta) = \min_{\mu : A\mu \le b} n(\overline{m}_n(\theta) - \mu)' \widehat{\Sigma}_n(\theta)^{-1} (\overline{m}_n(\theta) - \mu), \tag{17}$$

where $\widehat{\Sigma}_n(\theta)$ denotes an estimator of $\mathrm{Var}_F(\sqrt{n}\overline{m}_n(\theta))$, the variance–covariance matrix of the standardized moments. When $\{W_i\}_{i=1}^n$ is i.i.d., we can take

$$\widehat{\Sigma}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} (m(W_i,\theta) - \overline{m}_n(\theta))(m(W_i,\theta) - \overline{m}_n(\theta))'. \tag{18}$$

When $\{W_i\}_{i=1}^n$ is not i.i.d., we can define $\widehat{\Sigma}_n(\theta)$ to account for the clustering or autocorrelation in $\{W_i\}_{i=1}^n$.

The test uses a data-dependent critical value that is based on the rank of the rows of A corresponding to the inequalities that are active in finite sample. To define it rigorously, let $\hat{\mu}$ be the solution to the minimization problem in (17). This is the restricted estimator for the moments. It can be calculated using a quadratic programming algorithm. Let a_j' denote the jth row of A and let b_j denote the jth element of b for $j = 1, 2, \ldots, d_A$. Let

$$\widehat{J} = \{j \in \{1,2,\ldots,d_A\} : a_j'\hat{\mu} = b_j\}, \tag{19}$$

which is the set of indices for the active inequalities. For a set $J \subseteq \{1,2,\ldots,d_A\}$, let A_J be the submatrix of A formed by the rows of A corresponding to the elements in J. Let $\mathrm{rk}(A_J)$ denote the rank of A_J, and let $\hat{r} = \mathrm{rk}(A_{\widehat{J}})$. Note that for test inversion, $\hat{\mu}$, $\hat{J}$, and $\hat{r}$ need to be recalculated for every value of θ.

The critical value of the CC test is the $100(1-\alpha)\%$ quantile of $\chi^2_{\hat{r}}$, the chi-squared distribution with $\hat{r}$ degrees of freedom, denoted by $\chi^2_{\hat{r},1-\alpha}$. We denote the CC test by

$$\phi_n^{\mathrm{CC}}(\theta,\alpha) = 1\left\{T_n(\theta) > \chi^2_{\hat{r},1-\alpha}\right\}, \tag{20}$$

where CC stands for "conditional chi-squared" indicating that the test uses the chi-squared critical value conditional on the active inequalities.[7] We show the validity of the CC test below. The intuition is that $T_n(\theta)$ (asymptotically) follows the $\chi^2_{\hat{r}}$ distribution conditional on $\hat{r}$ when all inequalities are binding (i.e. $A\mathbb{E}_F\overline{m}_n(\theta)=b$), and is stochastically dominated by the $\chi^2_{\hat{r}}$ distribution when some of the inequalities are slack.

The CC test does not reject when $\hat{r}=0$, and rejects with probability at most α when $\hat{r}>0$. Thus, an upper bound on its (asymptotic) null rejection probability is $(1-\Pr(\hat{r}=0))\alpha$. This shows that the CC test can be somewhat conservative unless the moment inequality model contains an equality.[8]

We also propose a second simple test that eliminates the conservativeness. We call this the refined CC (RCC) test. We define the RCC test by adjusting the critical value when $\hat{r}=1$ from the $100(1-\alpha)\%$ quantile to the $100(1-\hat{\beta})\%$ quantile of the χ^2_1 distribution, where $\hat{\beta}$ varies between α and 2α depending on how far from being active the additional inequalities are. We construct $\hat{\beta}$ geometrically to make the refinement exactly restore the size of the test. To save space, we provide the details for implementing the refinement in Supplementary Appendix A.1.

It helps to illustrate the CC and RCC tests in a simple two-inequality example.

Illustration 1. *Consider an example where $d_m=2$, $A=I$, $b=\mathbf{0}$, and $\Sigma_n(\theta)=I$. We omit θ from the notation for ease of exposition. Thus, we are testing $H_0:\mathbb{E}_F\overline{m}_n\leq\mathbf{0}$ using the statistic $\sqrt{n}\overline{m}_n$, which asymptotically follows a bivariate standard normal distribution.*

On the space of $\sqrt{n}\overline{m}_n$, the rejection region for the CC test is illustrated by the shaded region in Figure 1. In this example, the likelihood ratio statistic is the squared distance between $\sqrt{n}\overline{m}_n$ and the third quadrant of the plane. If $\sqrt{n}\overline{m}_n$ lies in the second or fourth quadrants of the plane, one inequality is active and the χ^2_1 quantile is used. If $\sqrt{n}\overline{m}_n$ lies in the first quadrant of the plane, two inequalities are active and the χ^2_2 quantile is used.[9] The critical values for the RCC test are illustrated using a dashed line where they deviate from the CC test. From the figure, we can see that the RCC test deviates from the CC test only when the number of active inequalities is one (in the second and fourth quadrants of the plane).

To end this subsection, Algorithm 1 presents pseudo-code that can be used to compute the CC test. The pseudo-code is implemented in user-friendly Matlab code provided in the replication files. The implementation requires a tolerance (tol) to account for numerical imprecision in the quadratic programming used to compute $T_n(\theta)$. We use 10^{-8} in the Monte Carlo simulations.

Remark. Algorithm 1 makes clear some of the convenient features of the implementation of the CC test. We list them here for emphasis. (1) The CC test does not require any tuning parameters or simulations to implement. (2) The CC test is simple to code. (3) There is also a third convenient feature of the implementation that is less clear from Algorithm 1, which is that the inequalities do not need to be "reduced" before implementing the test. Often, in practice, a collection of inequalities contains redundant inequalities, or inequalities that are implied by the

7. The conditional aspect of our critical value gives it an apparent resemblance with the critical value of the conditional test in ARP. However, the resemblance is only superficial. Like any conditional test, what is important is the statistic that is conditioned on. That statistic is the set of active inequalities in our case, while it is the second largest standardized sample moment in ARP's case.

8. When the model contains an equality, the two inequalities representing the equality are always active, and thus $\Pr(\hat{r}=0)=0$.

9. The discontinuity in the critical value illustrated in Figure 1 is similar to the discontinuity in the recommended generalized moment selection function (their $\varphi^{(1)}$) in AB that occurs whenever a moment is at the threshold of being selected.

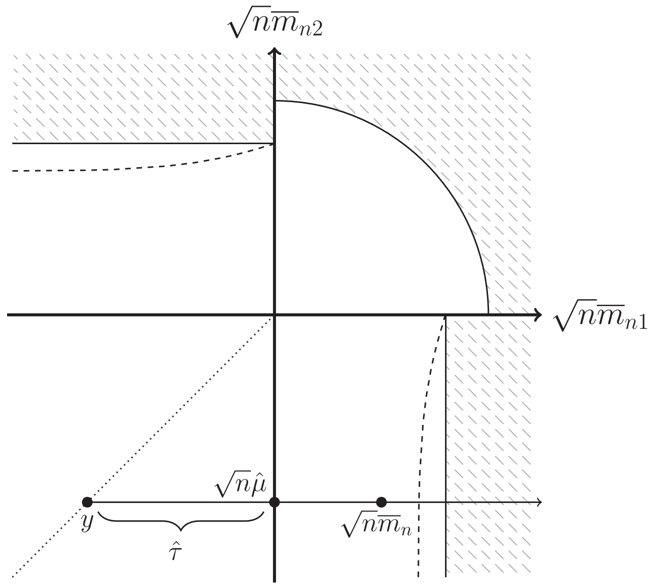

FIGURE 1

Geometric representation of the CC test (shaded) and the RCC test (dashed) in Illustration 1.

Algorithm 1: Pseudo-code for implementing the CC test.		
1:	**%Compute the CC Test**	
2:	$T_n(\theta), \hat{\mu}$	$\leftarrow \min_{\mu: A\mu \leq b} n(\overline{m}_n(\theta)-\mu)'\widehat{\Sigma}_n(\theta)^{-1}(\overline{m}_n(\theta)-\mu)$
3:	$\widehat{J}$	$:= \{j=1,\ldots,d_A : a_j'\hat{\mu}=b_j\}$
4:	$A_{\widehat{J}}$	$\leftarrow \widehat{J}, A$
5:	$\hat{r}$	$:= \mathrm{rk}(A_{\widehat{J}})$
6:	$\phi_n^{\mathrm{CC}}(\theta,\alpha)$	$:= 1\{T_n(\theta) > \max\{\chi^2_{\hat{r},1-\alpha}, \mathrm{tol}\}\}$.

other inequalities. The CC test is invariant to the inclusion of redundant inequalities. In contrast, other tests for moment inequalities, including AS, AB, and RSW, are not invariant, and thus are improved by reducing the collection of inequalities by removing the redundant ones before implementing the tests. □

3.2. *Subvector test*

Next we use the inequalities in (8) to test hypotheses on θ. For a given value, θ_0, testing $H_0: \theta = \theta_0$ amounts to testing the following hypothesis:

$$H_0: \exists\delta \text{ such that } B_Z \mathbb{E}_{F_Z}[\overline{m}_n(\theta_0)|Z] - C_Z\delta \leq d_Z,\ a.s. \tag{21}$$

In this subsection, we define a subvector version of the CC test for (21).

Directly testing (21) is difficult because it requires checking the validity of the inequality for all values of δ. We construct our test using an equivalent form of (21) that eliminates δ:

$$H_0: A_Z \mathbb{E}_{F_Z}[\overline{m}_n(\theta_0)|Z] \leq b_Z, \tag{22}$$

for some matrix A_Z and vector b_Z that are deterministic functions of C_Z, B_Z, and d_Z. The existence of such a transformation is well-known in the theory of linear inequalities, dating back to Fourier (1826). It has been noted in the moment inequality literature by Guggenberger *et al.* (2008) but has not been used in practice to the best of our knowledge. One significant obstacle is that calculating A_Z and b_Z is computationally difficult except in small dimensions. The key innovation in our approach is to conduct the CC test on (22) *without* calculating A_Z and b_Z.

The subvector CC (sCC) test for (21) is the full-vector CC test based on (22). It uses the test statistic

$$T_n(\theta) = \min_{\mu : A_Z \mu \leq b_Z} n(\overline{m}_n(\theta) - \mu)' \widehat{\Sigma}_n(\theta)^{-1} (\overline{m}_n(\theta) - \mu), \tag{23}$$

where $\widehat{\Sigma}_n(\theta)$ is an estimator of the conditional variance: $\Sigma_n(\theta) = \mathrm{Var}(\sqrt{n}\overline{m}_n(\theta)|Z)$, discussed in more detail below. The critical value of the sCC test is $\chi^2_{\hat{r},1-\alpha}$, where $\hat{r}$ is the rank of the active inequalities, defined as in the full-vector CC test applied to the problem in (23).

The key to computing $T_n(\theta)$ without computing A_Z and b_Z is to recognize that

$$T_n(\theta) = \min_{\delta, \mu : B_Z \mu - C_Z \delta \leq d_Z} n(\overline{m}_n(\theta) - \mu)' \widehat{\Sigma}_n(\theta)^{-1} (\overline{m}_n(\theta) - \mu). \tag{24}$$

One can calculate $T_n(\theta)$ by quadratic programming, where $(\delta', \mu')'$ is the decision variable. Let $(\hat{\delta}', \hat{\mu}')'$ be the solution to the minimization problem.

There are multiple ways to define A_Z and b_Z for (22) to be equivalent to (21). The Fourier–Motzkin algorithm noted in Guggenberger *et al.* (2008) is one of them. Another that is particularly convenient for our purpose is to take convex combinations of the inequalities. If we let $h \in \mathbb{R}^k$ denote a vector of non-negative weights that sum to one, then the convex combination of the inequalities in (21) is given by

$$h' B_Z \mathbb{E}_{F_Z}[\overline{m}_n(\theta_0)|Z] - h' C_Z \delta \leq h' d_Z. \tag{25}$$

When $h' C_Z = \mathbf{0}$, the δ parameter is eliminated from the inequalities. It follows from Gale's Theorem[10] that it is sufficient to consider the set of all inequalities (25) indexed by

$$h \in \mathcal{H} := \{h \in \mathbb{R}^k : h \geq \mathbf{0}, C_Z' h = \mathbf{0}, 1' h = 1\}. \tag{26}$$

To connect this result to (22), note that $\mathcal{H}$ defines a convex polyhedron in $\mathbb{R}^k$. Every element of a convex polyhedron is a convex combination of its extreme points, or *vertices*. Thus, it is sufficient to consider the vertices of $\mathcal{H}$. That is, a particular value θ_0 satisfies (21) if and only if θ_0 satisfies (25) for all h that are vertices of $\mathcal{H}$. Equivalently, if we take $H(C_Z)$ to denote a matrix where each row is a vertex of $\mathcal{H}$, then defining

$$A_Z = H(C_Z) B_Z \text{ and } b_Z = H(C_Z) d_Z \tag{27}$$

renders (22) equivalent to (21). This result is formally stated in Lemma 1 in Supplementary Appendix A.2.

10. See Theorem 2.7 in Gale (1960). Gale's Theorem is considered by some authors (e.g. Bachem and Kern, 1992, Theorem 4.1) to be a variant of Farkas' Lemma, a result that may be familiar to readers who have worked on non-negative solutions to linear systems of equations.

Algorithm 2: Pseudo-code for the sCC test using vertex enumeration of $\mathcal{H}$.

1: **%Compute the sCC Test**
2: $T_n(\theta), \hat{\mu} \quad \leftarrow \min_{\delta,\mu: B_Z\mu - C_Z\delta \le d_Z} n(\widehat{m}_n(\theta)-\mu)'\widehat{\Sigma}_n(\theta)^{-1}(\overline{m}_n(\theta)-\mu)$
3: $H(C_Z) \quad \leftarrow \mathcal{H}$ using a vertex enumeration algorithm, e.g. **con2vert.m** in Matlab (ref. Kleder (2020))
4: $A_Z, b_Z \quad \leftarrow H(C_Z)B_Z, H(C_Z)d_Z$
5: $\widehat{J} \quad := \{j=1,\ldots,d_A : e_j'A_Z\hat{\mu} = e_j'b_Z\}$
6: $\hat{r} \quad := \text{rk}(I_{\widehat{J}}A_Z)$
7: $\phi_n^{\text{sCC}}(\theta,\alpha) \quad := 1\{T_n(\theta) > \max\{\chi^2_{\hat{r},1-\alpha}, \text{tol}\}\}$.

Thus, to calculate A_Z and b_Z, we could enumerate the vertices of $\mathcal{H}$. While vertex enumeration seems simple, it can be computationally challenging when k and/or p are large. (Experience suggests even moderate values of k and p can lead to computational challenges.) As noted in various textbooks, including Sierksma and Zwols (2015), there is no polynomial time algorithm for vertex enumeration available in general. However, in our context there is a computationally simpler way to compute $\hat{r}$ without A_Z or b_Z. The key is to note that the active inequalities correspond to a face of $\mathcal{H}$—the one that takes convex combinations of the active inequalities. This can be computed numerically using linear programs using B_Z, C_Z, and d_Z. Then, the rank of the vectors in the face of $\mathcal{H}$ is equal to $\hat{r}$. We provide the details of this calculation in Supplementary Appendix A.3.

Next, we discuss the conditional variance estimator $\widehat{\Sigma}_n(\theta)$. The *conditional* variance is the appropriate variance matrix to be estimated because the inequalities hold conditionally on Z and the theoretical properties of the tests are derived using the conditional distribution of $\overline{m}_n(\theta_0)$ given Z. Here, we describe a conditional variance matrix estimator for discrete Z_i in the context of i.i.d. data. In the Supplementary Appendix, we define another conditional variance matrix estimator for continuous Z_i.

Suppose Z_i takes on a finite number of values in a set, $\mathcal{Z}$. A straightforward estimator of $\text{Var}(\sqrt{n}\overline{m}_n(\theta)|Z)$ is the weighted average of the sample variances of $m(W_i,\theta)$ within each category of Z_i:

$$\widehat{\Sigma}_n(\theta) = \sum_{\ell\in\mathcal{Z}} \frac{n_\ell}{n}\frac{1}{n_\ell - 1}\sum_{i=1}^{n}(m(W_i,\theta)-\overline{m}_n^\ell(\theta))(m(W_i,\theta)-\overline{m}_n^\ell(\theta))'1\{Z_i=\ell\}, \tag{28}$$

where $n_\ell = \sum_{i=1}^n 1\{Z_i=\ell\}$ and $\overline{m}_n^\ell(\theta) = \frac{1}{n_\ell}\sum_{i=1}^n m(W_i,\theta)1\{Z_i=\ell\}$. As we show in Supplementary Appendix B.4, sufficient conditions for the consistency of this estimator involve boundedness of the fourth moment of $m(W_i,\theta)$ and the assumption that every Z_i value occurs twice or more in the sample $\{Z_i\}_{i=1}^n$ eventually. This is the estimator used in our Monte Carlo simulations in Section 5.2.

To end this subsection, Algorithm 2 presents pseudo-code that can be used to compute the sCC test using vertex enumeration of $\mathcal{H}$. An algorithm that avoids vertex enumeration of $\mathcal{H}$ is available in the Supplementary Appendix. In the algorithm, e_j denotes the jth unit basis vector in $\mathbb{R}^{d_A}$, and I_J denotes the submatrix of the $d_A \times d_A$ identity matrix containing the rows with indices in J. The pseudo-code is implemented in user-friendly Matlab code provided in the replication files. The implementation requires a tolerance to account for numerical imprecision in the quadratic programming used to compute $T_n(\theta)$. We use 10^{-8} in the Monte Carlo simulations. Note that the sCC test has the same convenient implementation features listed in the remark on Algorithm 1.

The third feature, that the inequalities do not need to be "reduced" before implementing the tests, is especially convenient for the sCC test because algorithms for the vertex enumeration used to calculate A_Z and b_Z often deliver redundant inequalities, and these do not have to be removed before implementing the sCC test.

4. THEORETICAL PROPERTIES

Next, we consider the theoretical properties of the CC and sCC tests. We show that both the full-vector and subvector CC tests are uniformly asymptotically valid when the moments are asymptotically normal. Moreover, we make precise the adaptiveness of the tests to slackness of the moment inequalities. In Supplementary Appendix B, we state the theoretical properties of the RCC and sRCC tests.

4.1. *Full-vector*

Now, we state the asymptotic properties of the CC test when the moments are asymptotically normal and the variance–covariance matrix is estimated. For expositional purposes, we focus on the independent and identically distributed (i.i.d.) data case here, while results in Supplementary Appendix B cover more general cases. With i.i.d. data, we can estimate the variance matrix with $\widehat{\Sigma}_n(\theta)$ defined in (18). We show that the CC test has correct asymptotic size uniformly over a large class of data generating processes.

The following assumption defines the class of data generating processes allowed. Here, $|\cdot|$ denotes the matrix determinant, and ϵ and M are fixed positive constants that do not depend on F or θ.

Assumption 1. *For all $F\in\mathcal{F}$ and $\theta\in\Theta_0(F)$, the following hold.*

(a) $\{W_i\}_{i=1}^n$ *are i.i.d. under* F.
(b) $\sigma^2_{F,j}(\theta):=\mathrm{Var}_F(m_j(W_i,\theta))>0$ *for* $j=1,\dots,d_m$.
(c) $|\mathrm{Corr}_F(m(W_i,\theta))|>\epsilon$, *where* $\mathrm{Corr}_F(m(W_i,\theta))$ *is the correlation matrix of the random vector* $m(W_i,\theta)$ *under* F.
(d) $\mathbb{E}_F|m_j(W_i,\theta)/\sigma_{F,j}(\theta)|^{2+\epsilon}\le M$ *for* $j=1,\dots,d_m$.

Remarks. (1) This set of assumptions is commonly made in the moment inequality literature (see e.g. Andrews and Guggenberger, 2009, AS, or Kaido *et al.*, 2019). Part (a) assumes i.i.d. for simplicity but is not essential for the results. One can use our method on data with cluster, spatial, or temporal dependence, after changing $\widehat{\Sigma}_n(\theta)$ to a variance estimator that appropriately accommodates the dependence. In that case, the validity of our procedure follows from Theorem 5 in Supplementary Appendix B. Part (b) is innocuous as it simply requires the moment functions be non-constant in W_i. Parts (a), (b), and (d) together imply asymptotic normality of the sample moments via a Lyapunov central limit theorem.

(2) Part (c) requires uniform invertibility of the correlation matrix, which is imposed because we use the inverse of $\widehat{\Sigma}_n(\theta)$ in the test statistic. While this rules out perfectly correlated moments and near-perfectly correlated moments, perfectly correlated moments can be handled in specification (1) by an appropriate choice of A and b provided the perfect correlation is known. For example, in Example 1, suppose one reaches the moment inequalities:

$$\mathbb{E}[1\{(r_i^1,r_i^2,r_i^3)=(0,0,0)\}-g_{000}(\theta)]\le 0$$

$$\mathbb{E}[-1\{(r_i^1,r_i^2,r_i^3)=(0,0,0)\}+g_{000}(\theta)]\le 0$$

$$\mathbb{E}[1\{(r_i^1,r_i^2,r_i^3)=(a,b,c)\}-g_{abc}(\theta)]\leq 0$$
$$\vdots$$
$$\mathbb{E}[1\{(r_i^1,r_i^2,r_i^3)=(b,0,0)\}-g_{b00}(\theta)]\leq 0$$
$$\mathbb{E}[1\{(r_i^1,r_i^2,r_i^3)=(c,0,0)\}-g_{c00}(\theta)]\leq 0. \tag{29}$$

These moment inequalities are collinear both because the first is the negative of the second and because the probabilities of all rank-order lists add up to 1. The invertibility requirement can still be satisfied by defining

$$m(W_i,\theta)=\begin{pmatrix}1\{(r_i^1,r_i^2,r_i^3)=(0,0,0)\}\\ 1\{(r_i^1,r_i^2,r_i^3)=(a,b,c)\}\\ \vdots\\ 1\{(r_i^1,r_i^2,r_i^3)=(b,0,0)\}\end{pmatrix}, A=\begin{pmatrix}1 & 0 & \dots & 0\\ -1 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \dots & 1\\ -1 & -1 & \dots & -1\end{pmatrix}, \text{ and } b=\begin{pmatrix}g_{000}(\theta)\\ -g_{000}(\theta)\\ g_{abc}(\theta)\\ \vdots\\ g_{b00}(\theta)\\ g_{c00}(\theta)-1\end{pmatrix}.$$

Note that $m(W_i,\theta)$ is the core set of moments, one for each possible rank-order list, omitting the last one. This is similar to dealing with perfect multi-collinearity in a linear regression with binary variables. □

Let $D_F(\theta)$ denote the diagonal matrix formed by $\sigma^2_{F,j}(\theta):j=1,\dots,d_m$. For $J\subseteq\{1,\dots,d_A\}$, let I_J denote the rows of the identity matrix corresponding to the indices in J.[11] We define a reduced test that only uses a subset of the inequalities. For $J\subseteq\{1,...,d_A\}$, let $\phi^{\text{CC}}_{n,J}(\theta,\alpha)$ denote the CC test defined with A_J and b_J instead of A and b, where b_J denotes the subvector of b formed by the elements of b corresponding to the indices in J. This test is a useful point of comparison when the inequalities not in J are very slack.

The following theorem states the asymptotic properties of the CC test.

Theorem 1. *Suppose Assumption 1 holds.*

(a) $\limsup_{n\to\infty}\sup_{F\in\mathcal{F}}\sup_{\theta\in\Theta_0(F)}\mathbb{E}_F\phi^{\text{CC}}_n(\theta,\alpha)\leq\alpha$.

(b) *For a sequence $\{(F_n,\theta_n):F_n\in\mathcal{F},\theta_n\in\Theta_0(F_n)\}_{n=1}^\infty$ such that $A(\theta_n)D_{F_n}(\theta_n)\to A_\infty$, for some matrix A_∞ and for all $J\subseteq\{1,\dots,d_A\}$, $\text{rk}(I_JA(\theta_n)D_{F_n}(\theta_n))=\text{rk}(I_JA_\infty)$ for all n, if there is a $J\subseteq\{1,\dots,d_A\}$ such that for all $j\notin J$, $\sqrt{n}(a_j'\mathbb{E}_{F_n}\overline{m}_n(\theta_n)-b_j)\to-\infty$ as $n\to\infty$, then*

$$\lim_{n\to\infty}\Pr_{F_n}\left(\phi^{\text{CC}}_n(\theta_n,\alpha)\neq\phi^{\text{CC}}_{n,J}(\theta_n,\alpha)\right)=0.$$

Proof. The theorem follows from the proofs of Theorems 4 and 5 in Supplementary Appendix B, where the only modification needed is to change $\hat{\beta}$ to α. Theorem 5 states the analogous results for the RCC test, and the proof of Theorem 4 verifies the assumptions of Theorem 5 under Assumption 1. □

11. Note that I_JA is an alternate notation for A_J.

Remarks. (1) Part (a) shows that the CC test is asymptotically uniformly valid. Part (b) shows that when the slack inequalities get very slack, the CC test reduces to a version of it that does not use those inequalities. Put another way, the CC test adapts to the slackness of the inequalities. We call this property "irrelevance of distant inequalities" or IDI. AS introduced a generalized moment selection procedure that achieves the IDI property via a sequence of tuning parameters. Note that if θ and F are fixed and A and b do not depend on n, then the condition in part (b) holds, where J equals the set of all binding inequalities.[12]

(2) Theorem 1 and the other results in this article are stated in terms of hypothesis tests. However, they can be extended to results on the coverage probability of confidence sets defined by test inversion in a standard way. Specifically, under the conditions of Theorem 1(a), we have

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{F}} \inf_{\theta_0\in\Theta_0(F)} \Pr_F\left(\theta_0 \in CS_n^{\mathrm{CC}}(1-\alpha)\right) \geq 1-\alpha, \tag{30}$$

where $CS_n^{\mathrm{CC}}(1-\alpha)=\{\theta\in\Theta:\phi_n^{\mathrm{CC}}(\theta,\alpha)=0\}$ is the confidence set formed by inverting the CC test.

(3) Theorems 3–5 in Supplementary Appendix B state the theoretical properties of the RCC test. Theorem 3 shows that the RCC test has finite sample exact size and possesses a finite sample version of the IDI property when the moments are normally distributed and the variance matrix is known. Theorems 4 and 5 state the asymptotic validity and the IDI property of the RCC test. In addition, these two theorems also show that the RCC test is asymptotically size exact under appropriate conditions, which means that the limit of the worst case rejection probability is equal to α. It follows from this and a relationship between the CC and the RCC test given in equation (40) in Supplementary Appendix A that, under the same conditions, the asymptotic size of the CC test is between $\alpha/2$ and α. $\square$

4.2. *Subvector*

Now, we state the asymptotic properties of the sCC test when the moments are asymptotically normal and the conditional variance–covariance matrix is estimated. For expositional purposes, we focus on the independent and identically distributed (i.i.d.) data case here with discrete Z_i. Results in the Supplementary Appendix cover more general cases. With discrete Z_i, we can estimate the conditional variance–covariance matrix with $\widehat{\Sigma}_n(\theta)$ defined in (28). We show that the sCC test has correct asymptotic size uniformly over a large class of data generating processes.

The following assumption defines the class of data generating processes allowed. Fix a realization of $\{Z_i\}_{i=1}^n$ and denote it by z. Let $\mathcal{F}_z$ be a collection of conditional distributions F_z for $\{W_i\}_{i=1}^n$ given $\{Z_i\}_{i=1}^n=z$. Let $\sigma_{j|z}^2(\theta):=n^{-1}\sum_{i=1}^n \mathrm{Var}_{F_z}(m_j(W_i,\theta)|z)$ and

$$D_{|z}(\theta)=\mathrm{diag}(\sigma_{1|z}^2(\theta),\ldots,\sigma_{d_m|z}^2(\theta)). \tag{31}$$

Here, $\mathrm{eig}_{\min}(V)$ denotes the minimum eigenvalue of a matrix V, and ϵ_0 and M_0 are fixed positive constants that do not depend on F_z or θ.

Assumption 2. *For all $F_z\in\mathcal{F}_z$ and $\theta\in\Theta$, the following hold.*

(a) *Under F_z, $\{W_i\}_{i=1}^n$ are independent across i, and for all $i_1,i_2\in\{1,\ldots,n\}$, the distributions of W_{i_1} and W_{i_2} are identical if $z_{i_1}=z_{i_2}$.*

12. Technically, since F is the joint distribution of $\{W_i\}_{i=1}^n$, we need the marginal distribution of each W_i to be fixed.

(b) $\sigma_{j|z}^2(\theta) > 0$ *for all* $j = 1, \ldots, d_m$ *and for all* n.
(c) $\mathrm{eig}_{\min}(n^{-1}\sum_{i=1}^n [\mathrm{Var}_{F_z}(D_{|z}^{-1/2}(\theta)m(W_i,\theta)|z)]) > \epsilon_0$ *for all* n.
(d) $n^{-1}\sum_{i=1}^n \mathbb{E}_{F_z}((m_j(W_i,\theta)/\sigma_{j|z}(\theta))^4|z) < M_0$ *for all* $j = 1, \ldots, d_m$ *and for all* n.
(e) $\{z_i\}_{i=1}^n$ *contains at least two instances of each value eventually as* $n \to \infty$.

Remark. Part (a) is the i.i.d. assumption for conditional distributions. Part (b) simply requires the moment functions be nonconstant in W_i after conditioning on Z_i. Part (c) requires that the average conditional variance of $m(W_i,\theta)$ be invertible uniformly over θ and $F_z \in \mathcal{F}_z$. This is required since we use the quasi-likelihood ratio statistic which involves inverting $\widehat{\Sigma}_n(\theta)$. Part (d) requires $m(W_i,\theta)$ to have bounded fourth moment conditional on Z_i. This is used both to derive the asymptotic normality of $\overline{m}_n(\theta)$ using the Lindeberg–Feller central limit theorem under a sequence of values of F_z, and to show the consistency of $\widehat{\Sigma}_n(\theta)$. Part (e) can be established for a probability-one set of z values by the strong law of large numbers. □

The following theorem states the asymptotic properties of the sCC test. In the theorem, A_z and b_z are from the nuisance-parameter-eliminated inequalities, defined in (27). As before, let $\phi_{n,J}^{\mathrm{sCC}}(\theta,\alpha)$ denote the CC test defined with $I_J A_z$ and $I_J b_z$ instead of A_z and b_z, for any $J \subseteq \{1,\ldots,d_A\}$. Also let $\Theta_0(F_z) := \{\theta \in \Theta : A_z(\theta)\mathbb{E}_{F_z}\overline{m}_n(\theta) \leq b_z(\theta)\}$ denote the identified set for the given z values.

Theorem 2. *Suppose Assumption 2 holds.*

(a) $\limsup_{n\to\infty} \sup_{F_z \in \mathcal{F}_z} \sup_{\theta \in \Theta_0(F_z)} \mathbb{E}_{F_z}(\phi_n^{\mathrm{sCC}}(\theta,\alpha)|z) \leq \alpha$.
(b) *Consider a sequence* $\{(F_{z,n},\theta_n) : F_{z,n} \in \mathcal{F}_z, \theta_n \in \Theta_0(F_{z,n})\}_{n=1}^\infty$. *Suppose there exists a sequence of positive definite* $d_m \times d_m$ *matrices,* $\{D_n\}$, *such that* $\Lambda_n A_z(\theta_n) D_n \to \bar{A}_0$ *for some* $d_A \times d_m$ *matrix* $\bar{A}_0$, *where* Λ_n *is the diagonal* $d_A \times d_A$ *matrix whose jth diagonal entry is one if* $e_j' A_z(\theta_n) = 0$ *and* $\|e_j' A_z(\theta_n) D_n\|^{-1}$ *otherwise. Also suppose for every* $J \subseteq \{1,\ldots,d_A\}$, $\mathrm{rk}(I_J A_z(\theta_n) D_n) = \mathrm{rk}(I_J \bar{A}_0)$. *If, for some* $J \subseteq \{1,\ldots,d_A\}$, $\sqrt{n} e_j' \Lambda_n (A_z(\theta_n)\mathbb{E}_{F_{z,n}}\overline{m}_n(\theta_n) - b_z(\theta_n)) \to -\infty$ *as* $n \to \infty$ *for all* $j \notin J$, *then*

$$\lim_{n\to\infty} \Pr_{F_{z,n}}(\phi_n^{\mathrm{sCC}}(\theta_n,\alpha) \neq \phi_{n,J}^{\mathrm{sCC}}(\theta_n,\alpha)) = 0.$$

Proof. In Supplementary Appendix B, Theorem 6(a,b) verifies the assumptions of Corollary 2 under Assumption 2. Given the assumptions of Corollary 2, the conclusions of Theorem 2 follow from the proof of Theorem 5(a,c) with $\hat{\beta}$ replaced by α in the same way that Corollary 2 follows from Theorem 5. □

Remarks. (1) Part (a) shows that the sCC test is asymptotically uniformly valid. Part (b) shows an IDI property of the sCC test: when the slack inequalities get very slack, the sCC test reduces to a version of it that does not use those inequalities. In the special case that A_z and b_z do not depend on n, and if F_z and θ are fixed, then the condition in part (b) is satisfied with J equal to the set of all binding inequalities.[13]

(2) The condition in part (b) depends on $A_z = H(C_z)B_z$ and $b_z = H(C_z)d_z$, which are the inequalities after eliminating the nuisance parameters. It is unclear whether an alternative sufficient condition can be formulated that depends only on the original inequalities. In a model with a scalar parameter of interest, Rambachan and Roth (2020) use a linear

13. Technically, since F_z is the joint conditional distribution of $\{W_i\}_{i=1}^n$, we need the marginal conditional distributions of W_i given $Z_i = z_i$ to be fixed for all possible values of z_i.

independence constraint qualification to show that the test in ARP reduces to the one-sided t-test at an endpoint of the identified set. A similar constraint qualification may be helpful in formulating a sufficient condition for part (b) that depends only on the original inequalities.

(3) The invertibility requirement on $\Sigma_n(\theta)$ guides the choice of instrumental functions, $\mathcal{I}(Z_i)$, in Examples 3–5. In those examples, the instrumental functions are used to increase the number of moment inequalities in order to sharpen identification. The instrumental functions in Andrews and Shi (2013) serve the same purpose. Like in Andrews and Shi (2013), appropriate functions are indicators of cells defined by Z_i. However, unlike Andrews and Shi (2013), we do not recommend using cells of multiple levels of fineness. For example, when $Z_i \in \{0,1\}^2$, we do not recommend using both $1\{Z_i=(0,1)'\}, 1\{Z_i=(0,0)'\}, 1\{Z_i=(1,1)'\}, 1\{Z_i=(1,0)'\}$ and $1\{Z_i \in \{(0,1),(0,0)\}\}, 1\{Z_i \in \{(1,0),(1,1)\}\}$. This is because the subsequent moments are linearly dependent, causing $\Sigma_n(\theta)$ to be singular. We thus recommend choosing a partition of the space of Z_i and using the indicator of all cells in that partition. For example, when $Z_i \in \{0,1\}^2$, use $\mathcal{I}(Z_i)=(1\{Z_i=(0,1)'\}, 1\{Z_i=(0,0)'\}, 1\{Z_i=(1,1)'\}, 1\{Z_i=(1,0)'\})'$. Our simulations suggest this choice of the instruments leads to no loss in terms of the power of the test or the size of the confidence set.

The need to choose instrumental functions is common in conditional moment inequality models. A complete cost and benefit analysis is beyond the scope of this article, but we can make some general observations. (a) A finer partition yields sharper identification, meaning that $\Theta_0(F_z)$ is smaller. (b) A finer partition also means fewer observations per cell, potentially implying a worse normal approximation. A crude rule of thumb is to ensure that the smallest cell in the partition contains 15 or more observations.

(4) Corollaries 1 and 2 in Supplementary Appendix B state the theoretical properties of the sRCC test. In addition to being asymptotically valid and possessing the IDI property, the sRCC test is size exact under appropriate conditions, which means that the limit of the rejection probability is equal to α. Also, Corollary 1 states finite sample properties of the sRCC test when the moments are normally distributed. It states that the sRCC test has finite sample exact size and a finite sample version of the IDI property. □

5. MONTE CARLO SIMULATIONS

We have shown that the CC and RCC tests have a variety of desirable properties, including convenient implementation and theoretical properties such as validity and adaptativeness to slackness. In this section, we use Monte Carlo simulations to compare the CC and RCC tests to alternative moment inequality tests in terms of size, power, and computational cost. We consider two sets of Monte Carlo simulations, one to evaluate the performance of the CC and RCC tests in a general moment inequality model without nuisance parameters, and the second to evaluate the performance of the sCC and sRCC tests in Example 4. In these simulations, no test should be expected to dominate any other in terms of power. Still, we find that the CC and RCC tests are at least competitive in terms of size and power and dominate in terms of computational cost.

5.1. *Full-vector simulations*

Our first set of simulations takes the generic moment inequality design from AB. This design allows a variety of correlation structures across moments and thus can approximate a wide range of applications.

We briefly describe the Monte Carlo design here and refer readers to Section 6 of AB for further details. Consider the moment inequality model

$$E[\theta - W_i] \leq \mathbf{0}, \tag{32}$$

and the null hypothesis $H_0 : \theta = \mathbf{0}$, where W_i is a k-dimensional random vector. Let the data $\{W_i\}_{i=1}^n$ be i.i.d. with sample size n. Let W_i be distributed with mean μ and variance–covariance matrix Ω. Three choices of Ω are considered: Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}. For Ω_{Zero}, the moments are uncorrelated. For Ω_{Pos}, the moments are positively correlated. For Ω_{Neg}, some pairs of moments are strongly negatively correlated while other pairs of moments are positively correlated. The exact numerical specifications of these matrices for different k's are in Section 4 of AB and Section S7.1 of the supplementary material of AB.

We consider separately cases with $k \leq 10$ and cases with $k \geq 10$. With $k \leq 10$, we compare the CC and the RCC tests to the recommended tests in AB and RSW. More specifically, we compare to the bootstrap-based adjusted quasi-likelihood ratio (AQLR) test in AB and two two-step procedures in RSW, one using their T_n^{qlr} statistic and the other using their $T_n^{\max}$ statistic.[14] With $k \geq 10$, we only compare the CC and RCC tests to the RSW tests as the AB test is no longer computationally feasible. The RSW tests are implemented using 499 bootstrap draws and with a first-step significance level of 0.005. The AB test is implemented using 1000 bootstrap draws. These are the recommended values in RSW and AB, respectively.

5.1.1. $k \leq 10$. We approximate the size of the tests using the maximum null rejection probability (MNRP) over a set of μ values that satisfies $\mu \geq \mathbf{0}$ for each combination of Ω and k. These μ values are taken from AB, whose calculations suggest that these points are capable of approximating the size of the tests. We also compute a WAP for easy comparison. The WAP is the simple average of a set of carefully chosen points in the alternative space. We take these points also from AB, who design them to reflect cases with various degrees of violation or slackness for each of the inequalities. These μ values are given in Section 4 of AB and Section S7.1 of the supplementary material of AB. Besides WAP, we also report size-corrected WAP (ScWAP), which is obtained by adding a (positive or negative) number to the critical value where the number is set to make the size-corrected MNRP equal to the nominal level.

Table 1 shows the MNRP and WAP results when W_i is normally distributed with known Ω. In this case, only the RCC test should have exact size. The CC test should be somewhat under-sized especially with small k. The results are consistent with these theoretical predictions. The MNRP of the RCC test is within simulation error of 5%, and the MNRP of the CC test is somewhat below the MNRP of the RCC test. The MNRP's of the AB and the RSW1 tests appear to be more different from 5% than the RCC test, while the MNRP of the RSW2 test is close to 5%.

Table 2 shows the results when W_i is normally distributed with estimated Ω. In this case, none of the tests have exact size. The RCC test still has very good MNRP at $k = 2$ but has noticeably larger MNRP (up to 7.4% from 5%) when $k = 10$ with $\Omega = \Omega_{\text{Neg}}$. This may reflect the difficulty in estimating Ω with a small sample size ($n = 100$). The AB test and the RSW1 test continue to have good size, while the RSW2 test now exhibits some over-rejection when $k = 10$ and $\Omega = \Omega_{\text{zero}}$.

In terms of WAP, the RCC test has weakly higher ScWAP than both RSW tests in all but one case in Table 2 (estimated Ω), and in all but two cases in Table 1 (known Ω). The RCC test has

14. We use the AB test for comparison because it is tuning parameter free (in the sense that AB propose and use an optimal choice of the AS tuning parameter), and we use RSW's two-step tests for comparison because they should be insensitive to reasonable choices of their tuning parameters.

TABLE 1

Finite sample maximum null rejection probabilities and size-corrected average power of nominal 5% tests (normal distribution, known Ω, $n=100$)

Test	$k=10$ MNRP	WAP	ScWAP	Time	$k=4$ MNRP	WAP	ScWAP	Time	$k=2$ MNRP	WAP	ScWAP	Time
					$\Omega=\Omega_{\text{Neg}}$							
RCC	0.051	0.61	0.61	0.003	0.052	0.62	0.62	0.003	0.051	0.62	0.62	0.003
CC	0.051	0.61	0.60	0.003	0.049	0.60	0.61	0.003	0.046	0.58	0.60	0.003
AB	0.046	0.53	0.55	1.11	0.051	0.59	0.59	1.07	0.059	0.65	0.64	1.40
RSW1	0.054	0.58	0.56	0.551	0.056	0.60	0.59	0.538	0.052	0.64	0.63	0.701
RSW2	0.050	0.23	0.23	0.014	0.052	0.34	0.34	0.013	0.052	0.50	0.49	0.014
					$\Omega=\Omega_{\text{Zero}}$							
RCC	0.052	0.63	0.63	0.003	0.052	0.65	0.65	0.003	0.051	0.68	0.68	0.003
CC	0.050	0.62	0.62	0.003	0.045	0.62	0.64	0.003	0.038	0.61	0.66	0.004
AB	0.043	0.65	0.66	1.08	0.050	0.67	0.67	1.07	0.056	0.69	0.67	1.39
RSW1	0.053	0.61	0.60	0.545	0.056	0.63	0.62	0.539	0.052	0.65	0.65	0.699
RSW2	0.053	0.54	0.52	0.014	0.052	0.62	0.62	0.014	0.049	0.66	0.66	0.014
					$\Omega=\Omega_{\text{Pos}}$							
RCC	0.051	0.76	0.75	0.003	0.053	0.75	0.74	0.003	0.051	0.72	0.71	0.003
CC	0.038	0.72	0.76	0.003	0.033	0.68	0.74	0.003	0.032	0.62	0.69	0.003
AB	0.042	0.78	0.80	1.05	0.051	0.75	0.75	1.03	0.059	0.72	0.70	1.34
RSW1	0.053	0.77	0.77	0.547	0.056	0.73	0.71	0.534	0.052	0.67	0.66	0.700
RSW2	0.052	0.77	0.77	0.014	0.052	0.74	0.74	0.013	0.049	0.68	0.69	0.014

Notes: CC, RCC, AB, RSW1, and RSW2 denote the CC test, the refined CC test, the AQLR test with bootstrap critical value in AB, the two-step test in RSW based on the QLR statistic, and that based on the Max statistic, respectively. MNRP, WAP, ScWAP, and Time denote maximum null rejection probability, weighted average power, size-corrected WAP, and average computation time used in seconds in each Monte Carlo simulation. Cases with different k and knowledge status of Ω may have been assigned to different machines and their computation times are not comparable. But times across tests are comparable. The AB test and the RSW tests use 1000 and 499 bootstrap draws, respectively. The results for the CC, RCC, and RSW2 tests are based on 5000 simulations, while those for the AB and RSW1 tests are based on 2000 simulations for feasibility.

higher ScWAP than the AB test in four out of nine cases in both Tables 1 and 2. The ScWAP of all the tests, except RSW2, are quite close to each other, with differences between them no greater than 6 percentage points. The ScWAP of RSW2 is close to the other tests in all cases except when the moments have negative correlations ($\Omega=\Omega_{\text{Neg}}$), when they are much lower, especially for $k=10$.

On the computational side, the AB test and the RSW1 test are 200–400 times as costly as the RCC test, and the RSW2 test is 4–9 times as costly as the RCC test, as shown in the Time columns in the tables. Also note that the AB test is computed using 1000 bootstrap draws while the RSW tests using 499 bootstrap draws. Increasing the number of bootstrap draws would increase their computational costs proportionally.

The theoretical properties of the CC tests rely in an essential way on the normality or asymptotic normality of the moments. In Supplementary Appendix F, we report results when W_i is not normal to investigate the sensitivity of these simulations to the data distribution. Supplementary Appendix Tables 5 and 7 report results when W_i has a student t distribution with 3 degrees of freedom, denoted by $t(3)$. This distribution is chosen to investigate the consequences of thick tails on the tests. Supplementary Appendix Tables 6 and 8 report results when W_i has a mixed normal distribution, taken to be the equal probability mix of $N(-2,1)$ and $N(2,4)$ scaled to have unit variance. This distribution is chosen to investigate the consequences of a skewed and bimodal distribution on the tests. Neither non-normal distribution leads to significant changes in the qualitative findings of Tables 1 and 2.

TABLE 2
Finite sample maximum null rejection probabilities and size-corrected average power of nominal 5% tests (normal distribution, estimated Ω, $n = 100$)

	$k=10$				$k=4$				$k=2$			
Test	MNRP	WAP	ScWAP	Time	MNRP	WAP	ScWAP	Time	MNRP	WAP	ScWAP	Time
					$\Omega = \Omega_{\text{Neg}}$							
RCC	0.074	0.63	0.54	0.003	0.058	0.63	0.61	0.004	0.053	0.62	0.61	0.003
CC	0.074	0.63	0.54	0.003	0.058	0.61	0.59	0.004	0.048	0.59	0.60	0.003
AB	0.046	0.51	0.53	1.23	0.049	0.58	0.58	1.56	0.056	0.64	0.63	1.13
RSW1	0.054	0.55	0.53	0.569	0.053	0.58	0.58	0.736	0.050	0.62	0.62	0.533
RSW2	0.053	0.24	0.23	0.026	0.051	0.34	0.33	0.026	0.051	0.49	0.48	0.023
					$\Omega = \Omega_{\text{Zero}}$							
RCC	0.069	0.65	0.59	0.003	0.053	0.66	0.65	0.004	0.051	0.68	0.68	0.003
CC	0.069	0.64	0.57	0.003	0.049	0.63	0.63	0.004	0.039	0.61	0.66	0.003
AB	0.043	0.62	0.64	1.23	0.048	0.66	0.67	1.55	0.053	0.68	0.67	1.14
RSW1	0.052	0.58	0.58	0.570	0.053	0.62	0.61	0.732	0.050	0.64	0.64	0.541
RSW2	0.062	0.54	0.50	0.026	0.053	0.61	0.60	0.026	0.050	0.64	0.64	0.024
					$\Omega = \Omega_{\text{Pos}}$							
RCC	0.056	0.77	0.75	0.003	0.054	0.75	0.74	0.004	0.051	0.71	0.71	0.003
CC	0.043	0.73	0.74	0.003	0.034	0.68	0.73	0.004	0.035	0.63	0.69	0.003
AB	0.044	0.78	0.79	1.19	0.049	0.74	0.75	1.50	0.055	0.71	0.70	1.07
RSW1	0.053	0.76	0.75	0.566	0.053	0.71	0.71	0.731	0.052	0.66	0.66	0.528
RSW2	0.056	0.76	0.74	0.026	0.052	0.73	0.72	0.027	0.050	0.67	0.67	0.023

Notes: CC, RCC, AB, RSW1, and RSW2 denote the CC test, the refined CC test, the AQLR test with bootstrap critical value in AB, the two-step test in RSW based on the QLR statistic, and that based on the Max statistic, respectively. MNRP, WAP, ScWAP, and Time denote maximum null rejection probability, weighted average power, size-corrected WAP, and average computation time used in seconds in each Monte Carlo simulation. Cases with different k and knowledge status of Ω may have been assigned to different machines and their computation times are not comparable. But times across tests are comparable. The AB test and the RSW tests use 1000 and 499 bootstrap draws, respectively. The results for the CC, RCC, and RSW2 tests are based on 5000 simulations, while those for the AB and RSW1 tests are based on 2000 simulations for feasibility.

It is worth noting that, in this context, different tests direct power to different alternatives, and there is not a test that is uniformly most powerful. The ScWAP comparison hides important power variations across different directions. To investigate these power variations, Figure 2 reports simulated power curves for the tests in the $k = 2$ case with normally distributed moments, estimated Ω, and $\Omega = \Omega_{\text{Zero}}$. The power curves are functions of the true mean vector, $\mu = (\mu_1, \mu_2)$. When either μ_1 or μ_2 is negative, the corresponding inequality is violated, and we expect higher power. As the figure shows, the RCC test has better power when only one inequality is violated, while the AB and the RSW1 tests have better power when both inequalities are violated. We expect this pattern extends to cases with more inequalities and more general variance–covariance matrices: the AB or RSW1 tests have better power when all or most of the inequalities are violated, while the RCC test has better power when few inequalities are violated. If a researcher has a preference for tests that direct power in a particular direction, they can choose a test based on this pattern. Otherwise, when such a preference is not present, the RCC test is at least competitive with the other tests in terms of size and power.

5.1.2. $k \geq 10$. One advantage of the CC and RCC tests is that they remain feasible when the number of inequalities (k) and the sample size (n) are both large. In this subsection, we investigate the size and computational time of the RCC, CC, RSW1, and RSW2 tests when both k and n are large. Table 3 reports the results for four pairs of (k, n): $(10, 100)$, $(50, 700)$, $(100, 1600)$,

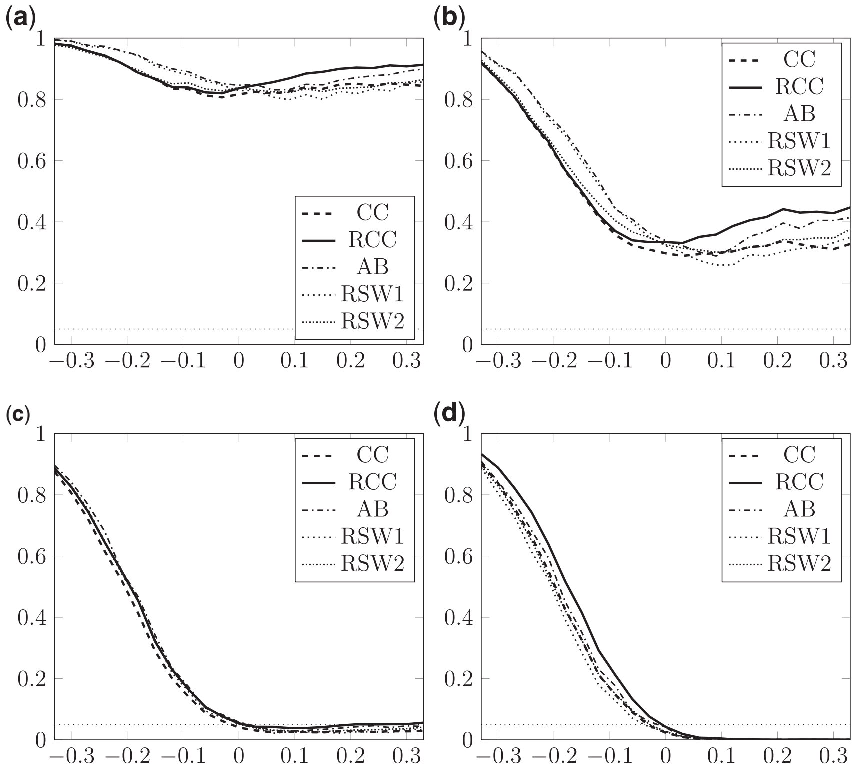

FIGURE 2

Power curves for five tests ($k=2$, normal distribution, $\Omega=\Omega_{\text{Zero}}$, estimated Ω, $n=100$, $\alpha=5\%$)

Notes: CC, conditional chi-squared test; RCC, refined CC test; AB, AQLR test with bootstrap critical value in AB; RSW1 and RSW2, two-step test in RSW based on the QLR statistic and the Max statistic, respectively. The AB test uses 1000 bootstrap draws and the RSW tests uses 499 bootstrap draws. The results for the CC, RCC, and RSW2 tests are based on 5000 simulations, while the results for the AB and RSW1 tests are based on 2000 simulations for computational reasons.

and $(150,2550)$, where the pairs are chosen so that k is approximately proportional to $n/\log(n)$. The message from Table 3 is quite encouraging. The MNRP's of all the tests appear to be stable as we move across columns. The computational time of the RCC and CC tests increases the slowest with k, while that for RSW2 increases the fastest.

The CC and RCC tests have not been proven to control size asymptotically when k grows with n at this rate, but these simulations suggest such a result could be formulated, under the correct assumptions.[15] Intuitively, if the moments are approximately normal, in some sense, then one can appeal to Theorem 3 as a good approximation. We do not pursue this type of result here, but note two challenges to keep in mind. On the theoretical side, a theory of Gaussian approximations for quadratic forms, such as the likelihood ratio statistic, that covers this high-dimensional case is an open question, to the best of our knowledge. On the practical side, a consistent covariance matrix estimator can be difficult to find. A potential way to improve covariance matrix estimation is to assume sparsity or use shrinkage as in Ledoit and Wolf (2012). It would be interesting to study the theoretical properties of the CC and RCC tests in settings with many inequalities, but we leave that to future research.

15. Bai, Santos and Shaikh (2021) have shown that RSW2 controls size asymptotically when k grows with n at this rate.

TABLE 3

Finite sample maximum null rejection probabilities of nominal 5% tests, the large k and n cases (estimated Ω, $\Omega=\Omega_{zero}$)

	$k=10, n=100$		$k=50, n=700$		$k=100, n=1600$		$k=150, n=2550$	
Test	MNRP	Time	MNRP	Time	MNRP	Time	MNRP	Time
				Normal				
RCC	0.069	0.003	0.074	0.004	0.076	0.011	0.081	0.024
CC	0.069	0.003	0.074	0.005	0.076	0.011	0.081	0.024
RSW1	0.052	0.582	0.062	1.07	0.047	3.25	0.051	7.96
RSW2	0.062	0.027	0.056	0.211	0.048	1.15	0.045	4.04
				$t(3)$				
RCC	0.063	0.003	0.069	0.004	0.071	0.010	0.079	0.024
CC	0.063	0.003	0.069	0.005	0.071	0.011	0.079	0.024
RSW1	0.057	0.568	0.054	1.07	0.050	3.23	0.054	7.97
RSW2	0.093	0.026	0.069	0.210	0.061	1.24	0.063	4.05
				Mixed normal				
RCC	0.096	0.003	0.090	0.004	0.089	0.010	0.089	0.024
CC	0.096	0.003	0.090	0.005	0.089	0.010	0.089	0.024
RSW1	0.045	0.567	0.051	1.06	0.059	3.20	0.054	7.92
RSW2	0.086	0.026	0.069	0.210	0.065	1.24	0.058	4.04

Notes: CC, conditional chi-squared test; RCC, refined CC test; RSW1 and RSW2, two-step test in RSW based on the QLR statistic and the Max statistic; MNRP, maximum null rejection probability; Time, average computation time in seconds for the test in each Monte Carlo repetition. The RSW tests use 499 critical value simulations. The results for the CC, RCC and RSW2 tests are based on 5000 simulations, while the results for the RSW1 tests are based on 2000 simulations for computational reasons.

5.2. *Subvector inference in interval regression*

To investigate the finite sample performance of the subvector CC and RCC tests, we consider a special case of Example 4, where $Y_i^*=s_i^*$ is the probability of an event of interest. For example, the event can be death by homicide for a random person in county i, or a product being purchased by a random consumer in market i. For simplicity, a simple logit model is assumed for the probability: $s_i^*=\frac{\exp(X_i'\theta_0+Z_{ci}'\delta_0+\varepsilon_i)}{1+\exp(X_i'\theta_0+Z_{ci}'\delta_0+\varepsilon_i)}$, where ε_i is the country or market level unobservable that satisfies $\mathbb{E}[\varepsilon_i|Z_i]=0$. Then, (11) holds with

$$\psi(Y_i^*,X_i,\theta)=\log(s_i^*/(1-s_i^*))-X_i'\theta. \tag{33}$$

The variable s_i^* is unobserved, but we observe $s_{N,i}$, an empirical estimate of s_i^* based on N independent chances for the event of interest to happen. $Ns_{N,i}$ follows a binomial distribution with parameters (N,s_i^*). For example, N could be the population of a county while $s_{N,i}$ is the homicide rate of the county. We use the method introduced in Gandhi *et al.* (2019) to construct $\psi_i^L(\theta)$ and $\psi_i^U(\theta)$ based on $s_{N,i}$. By Gandhi *et al.* (2019), for $N\geq 100$, the following construction satisfies (12):

$$\begin{aligned}\psi_i^U(\theta)&=\log(s_{N,i}+2/N)-\log(1-s_{N,i}+\underline{s})-X_i'\theta \\ \psi_i^L(\theta)&=\log(s_{N,i}+\underline{s})-\log(1-s_{N,i}+2/N)-X_i'\theta,\end{aligned} \tag{34}$$

where $\underline{s}$ is the smaller of 0.05 and half of the minimum possible value of $\min(s_i^*,1-s_i^*)$.[16] We assume that $\underline{s}$ is known and refer the reader to Gandhi *et al.* (2019) for practical recommendations regarding $\underline{s}$.

16. These bounds are not necessarily sharp, but that is not important for our purpose, which is to investigate the statistical performance of the sCC and sRCC tests.

Let the endogenous variable X_i be a scalar, and let Z_{ei} be a scalar excluded instrument. Let there be a d_c-dimensional exogenous covariate $Z_{ci}=(1,Z_{c,2,i},\ldots,Z_{c,d_c,i})'$. To generate the data, we let $N=100$, $\varepsilon_i \sim \min\{\max\{-4,N(0,1)\},4\}$, and let the non-constant elements of Z_i be mutually independent Bernoulli variables with success probability 0.5. Let $X_i=1\{Z_{ei}+\varepsilon_i/2>0\}$. Also let $\theta_0=-1$, $\delta_0=(0,-1,\mathbf{0}'_{d_c-2})'$, parameters chosen so that the identified set for θ_0 does not change when d_c is varied from 2 to 4. The value -1 is chosen to match the typical sign of a price coefficient and to normalize the scale for presentation purposes. Here, since δ_0 is the nuisance parameter, d_c is the number of nuisance parameters.

Given this data generating process, the lowest and the highest possible values for s_i^* are respectively

$$\frac{\exp(-6)}{1+\exp(-6)}=0.0025 \text{ and } \frac{\exp(4)}{1+\exp(4)}=0.982.$$

Thus, $\underline{s}=0.00125$. Given $\underline{s}$ and N, we calculate numerically that the identified set of θ_0 is approximately $[-1.203,-0.757]$. Details of the calculation are given in Supplementary Appendix G.1.

For instrumental functions, we use

$$\mathcal{I}(Z_i)=(1\{(Z_{ei},Z_{c,2,i},\ldots,Z_{c,d_c,i})=z\})_{z\in\{0,1\}^{d_c}}. \tag{35}$$

Thus, when $d_c=2$ (or 3, 4), there are 4 (or 8, 16) instrumental functions, which give us 8 (or 16, 32) moment inequalities.

We consider 5000 Monte Carlo repetitions. In each repetition, we generate an i.i.d. data set, $\{s_{N,i},X_i,Z_i\}_{i=1}^n$, for two sample sizes, $n=500$ and $n=1000$. For each repetition, we compute an implied CI for the sRCC test and the sCC test. We also include the hybrid test of ARP (with their recommended tuning parameter and number of simulation draws) for comparison. For all tests, the CI endpoints are computed with an accuracy to the third digit using an algorithm that always yields an interval. The details for computing the CIs are given in Supplementary Appendix G.2.

Table 4 reports the average CI, average excess length (= length of CI − length of identified set), as well as average computation time for the CI. As the table shows, the average computation time of the sCC and the sRCC tests are identical to each other up to 1/10 of a second in all cases. This is mainly because the vertex enumeration required to compute the refinement is easy to do for $d_c=2$, and the refinement is rarely needed for $d_c=3$ and $d_c=4$. The sRCC and sCC tests are faster than the ARP hybrid test in all cases. The relative computational cost of the ARP hybrid test seems to improve as the model gets bigger, but it remains more than 14 times as costly as the subvector CC and RCC tests when $d_c=4$.

In terms of length, the sRCC and sCC CIs are similar to each other, and are shorter on average than the ARP hybrid test for all cases. As we move from $d_c=2$ to $d_c=4$, the model contains more and more non-informative moment inequalities since the added control variables do not contribute in the data generating process. All tests are negatively affected by these non-informative inequalities to various degrees.

Figure 3 reports the rejection rates of the tests for $H_0:\theta_0=\theta$ plotted against θ values in $[-2.5,0.5]$.[17] The shaded area indicates the identified set for θ_0. As we can see, the rejection rates for the points in the identified set are less than or equal to 5% in all cases. For $d_c=3$ and $d_c=4$, all tests show some under-rejection at the boundary of the identified set, while the under-rejection of the ARP hybrid test appears to be less. It is encouraging to see that the under-rejection

17. For all three tests, a point is considered rejected if it is outside the CI. We found in our simulations that these rejection rates are slightly lower than those obtained directly point by point.

TABLE 4
Average value, length, and computation time (in seconds) of confidence intervals

	$n=500$			$n=1000$		
	CI	Excess length	Time	CI	Excess length	Time
			$d_c=2$, 8 moment inequalities			
sRCC	[−1.774, −0.339]	0.989	2.5	[−1.609 ,−0.440]	0.723	2.4
sCC	[−1.780, −0.332]	1.00	2.5	[−1.615, −0.433]	0.736	2.4
ARP hybrid	[−1.998, −0.264]	1.29	111	[−1.736, −0.395]	0.895	109
			$d_c=3$, 16 moment inequalities			
sRCC	[−1.852, −0.293]	1.11	7.1	[−1.659, −0.404]	0.809	4.7
sCC	[−1.852, −0.293]	1.11	7.1	[−1.659, −0.404]	0.809	4.7
ARP hybrid	[−2.219, −0.123]	1.65	199	[−1.883, −0.287]	1.15	120
			$d_c=4$, 32 moment inequalities			
sRCC	[−1.921, −0.254]	1.22	6.5	[−1.718, −0.366]	0.906	10
sCC	[−1.921, −0.254]	1.22	6.5	[−1.718, −0.366]	0.906	10
ARP hybrid	[−2.596, −0.011]	2.14	97	[−2.104, −0.180]	1.48	145

Notes: The identified set for θ_0 is $[-1.203, -0.757]$. The computation times across different (n, d_c) cases are not comparable because they may have been performed by different computers on the computer cluster. The computation of different tests within each (n, d_c) case is always completed on the same computer. Thus the computation times across tests are comparable.

does not translate to poor power. The power of the sCC and sRCC tests are nearly identical to each other and are higher than the power of the ARP hybrid test except in the area of θ immediately next to the identified set, consistent with the excess length results in Table 4. However, we note that this comparison is specific to this example, and the power comparison may change with other examples or data generating processes.

6. CONCLUSION

This article proposes the CC test for moment inequality models. This test compares a quasi-likelihood ratio statistic to a chi-squared critical value, where the number of degrees of freedom is the rank of the active inequalities. This test has many desirable properties, including being tuning parameter and simulation free, adaptive to slackness, easy to code, and invariant to redundant inequalities. We show that, with an easy refinement, it has exact size in normal models and has uniformly asymptotically exact size in asymptotically normal models. We also propose a version of the test for subvector inference with conditional moment inequalities and when the nuisance parameters enter linearly. Simulations show the CC and subvector CC tests have a computational advantage over alternatives while being competitive in terms of size and power.

Acknowledgments. We acknowledge helpful feedback from Donald Andrews, Isaiah Andrews, Xiaohong Chen, Whitney Newey, Adam Rosen, Jonathan Roth, Matthew Shum, Jörg Stoye, the participants of the 2nd Econometrics Jamboree at the UC Berkeley, the 2nd CEMMAP UCL/Vanderbilt Joint Conference, the 2019 China Meeting of the Econometric Society, the 2019 Asia meeting of the Econometric Society, the 2020 World Congress of the Econometric Society, the 2021 NAWM of the Econometric Society, the 2021 NASM of the Econometric Society, and econometrics seminars at Columbia University, the National University of Singapore, the UCSD, the UCLA, the University of Wisconsin at Madison, Harvard/MIT, NSD of Beijing University, the University of Alabama, Simon Fraser University, and Syracuse University. We also thank the editor Francesca Molinari and three anonymous referees at the *Review of Economic Studies* for many helpful comments. Shi dedicates this article to the loving memory of her late mother Guizhi Ma who passed away unexpectedly in December 2019 at the age of 66.

Data Availability Statement

No new data were generated or analysed in support of this research. Replication files for the simulations in this article are available in a Zenodo repository. The permanent link for the repository is https://doi.org/10.5281/zenodo.5989149.

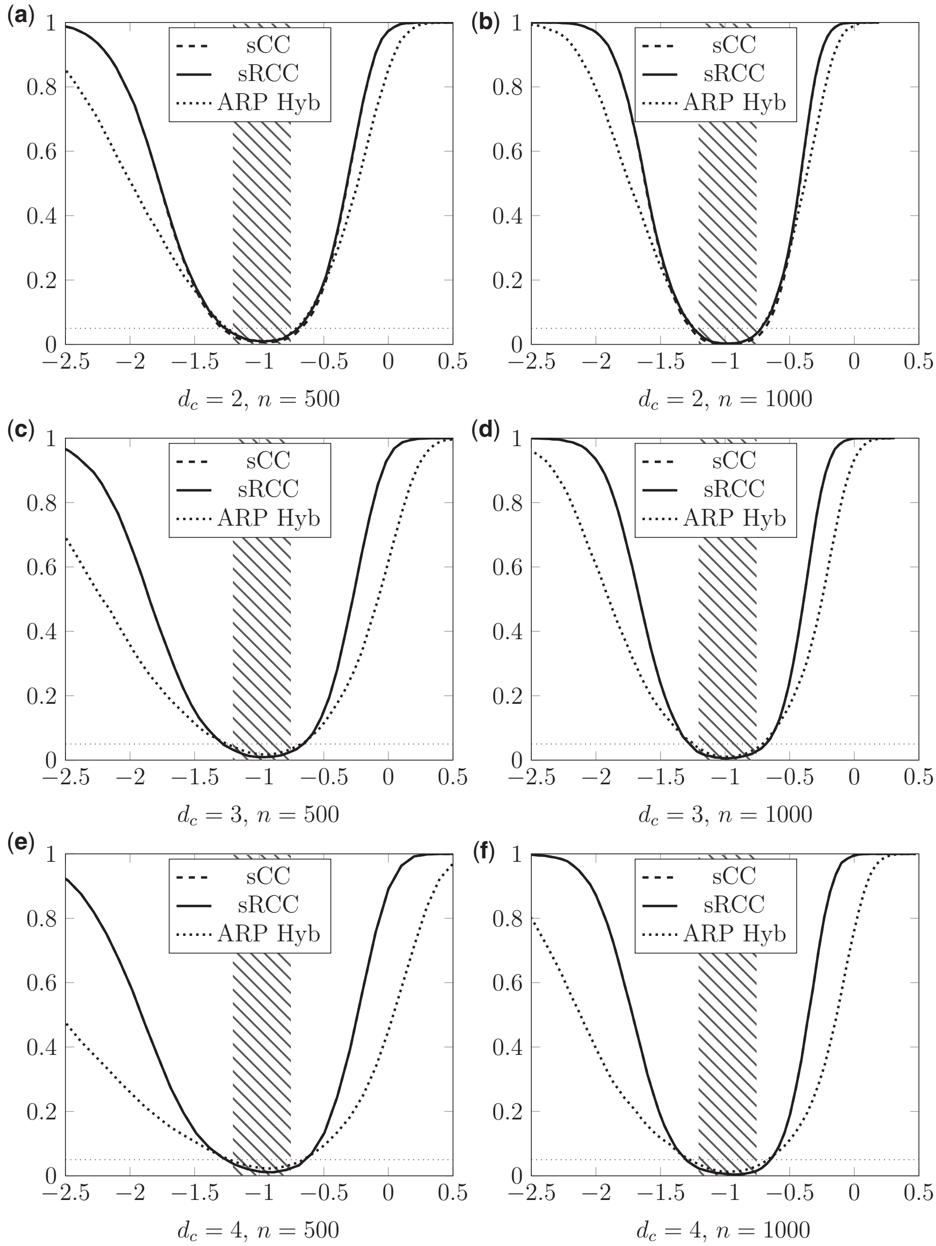

FIGURE 3

Rejection rates of the sCC, sRCC, and ARP hybrid tests for $d_c \in \{2,3,4\}$ and for sample size $n \in \{500, 1000\}$ with nominal size 5%.

REFERENCES

ANDREWS, D. AND BARWICK, P. (2012), "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure", *Econometrica*, **80**, 2805–2862.

ANDREWS, D. AND GUGGENBERGER, P. (2009), "Validity of Subsampling and "Plug-In Asymptotic" Inference for Parameters Defined by Moment Inequalities", *Econometric Theory*, **25**, 669–709.

ANDREWS, D. AND SOARES, G. (2010), "Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection", *Econometrica*, **78**, 119–157.

ANDREWS, D. W. AND SHI, X. (2013), "Inference Based on Conditional Moment Inequalities", *Econometrica*, **81**, 609–666.

ANDREWS, I., ROTH, J. AND PAKES, A. (2019), "Inference for Linear Conditional Moment Inequalities" (Unpublished Manuscript, Department of Economics, Harvard University).

BACCARA, M., IMROHOROGLU, A., WILSON, A. J. AND YARIV, L. (2012), "A Field Study on Matching with Network Externalities", *American Economic Review*, **102**, 1773–1804.

BACHEM, A. AND KERN, W. (1992), *Linear Programming Duality: An Introduction to Oriented Matroids* (Berlin, Heidelberg: Springer).

BAI, Y., SANTOS, A. AND SHAIKH, A. (2021), "A Two-Step Method for Testing Many Moment Inequalities", *Journal of Business & Economic Statistics*, 1–11.

BAJARI, P., BENKARD, C. L. AND LEVIN, J. (2007), "Estimating Dynamic Models of Imperfect Competition", *Econometrica*, **75**, 1331–1370.

BARTHOLOMEW, D. J. (1961), "A Test of Homogeneity of Means under Restricted Alternatives", *Journal of the Royal Statistical Society. Series B (Methodological)*, **23**, 239–281.

BERESTEANU, A., MOLCHANOV, I. AND MOLINARI, F. (2011), "Sharp Identification Regions in Models with Convex Moment Predictions", *Econometrica*, **79**, 1785–1821.

BLUNDELL, R., GOSLING, A., ICHIMURA, H. AND MEGHIR, C. (2007), "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition using Bounds", *Econometrica*, **75**, 323–363.

BUGNI, F., CANAY, I. AND SHI, X. (2017), "Inference for Functions of Partially Identified Parameters in Moment Inequality Models", *Quantitative Economics*, **8**, 1–38.

BUGNI, F. A. (2010), "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set", *Econometrica*, **78**, 735–753.

CANAY, I. A. (2010), "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity", *Journal of Econometrics*, **156**, 408–425.

CANAY, I. A. AND SHAIKH, A. (2017), "Practical and Theoretical Advances for Inference in Partially Identified Models", in Honoré, B., Pakes, A., Piazzesi, M. and Samuelson, L. (eds) *Advances in Economics and Econometrics*: Vol. 2: Eleventh World Congress (Cambridge, UK: Econometric Society Monographs, Cambridge University Press) 271–306.

CHEN, X., CHRISTENSEN, T. M. AND TAMER, E. (2018), "Monte Carlo Confidence Sets for Identified Sets", *Econometrica*, **86**, 1965–2018.

CHERNOZHUKOV, V., HONG, H. AND TAMER, E. (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, **75**, 1243–1284.

CHETTY, R. (2012), "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply", *Econometrica*, **80**, 969–1018.

CHO, J. AND RUSSELL, T. M. (2020), "Simple Inference on Functionals of Set-Identified Parameters Defined by Linear Moments" (Unpublished Manuscript, Department of Economics, University of Toronto).

CILIBERTO, F. AND TAMER, E. (2009), "Market Structure and Multiple Equilibria in Airline Markets", *Econometrica*, **77**, 1791–1828.

EIZENBERG, A. (2014), "Upstream Innovation and Product Variety in the U.S. Home PC Market", *The Review of Economic Studies*, **81**, 1003–1045.

FACK, G., GRENET, J. AND HE, Y. (2019), "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions", *American Economic Review*, **109**, 1486–1529.

FOURIER, J. B. J. (1826), "Solution d'une question particulière du calcul des inégalités", *Nouveau Bulletin des sciences par la Société philomathique de Paris*, p. 99.

GAFAROV, B. (2019), "Inference in High-Dimensional Set-Identified Affine Models" (Unpublished Manuscript, Department of Agricultural and Resource Economics, University of California, Davis).

GALE, D. (1960), *The Theory of Linear Economic Models*, 1st edn (New York: McGraw-Hill Book Company).

GANDHI, A. K., LU, Z. AND SHI, X. (2019), "Estimating Demand for Differentiated Products with Zeroes in Market Share Data" (Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison).

GUGGENBERGER, P., HAHN, J. AND KIM, K. (2008), "Specification Testing under Moment Inequalities", *Economics Letters*, **99**, 375–378.

HE, Y. (2017), "Gaming the Boston School Choice Mechanism in Beijing" (Unpublished Manuscript, Department of Economics, Rice University).

HO, K. AND ROSEN, A. (2017), "Partial Identification in Applied Research: Benefits and Challenges", in Honoré B., Pakes, A., Piazzesi, M. and Samuelson, L. (eds) *Advances in Economics and Econometrics*, Vol. 2: Eleventh World Congress (Cambridge, UK: Econometric Society Monographs, Cambridge University Press) 307-359.

HOLMES, T. J. (2011), "The Diffusion of Wal-Mart and Economies of Density", *Econometrica*, **79**, 253–302.

HUBER, M. AND MELLACE, G. (2015), "Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints", *The Review of Economics and Statistics*, **97**, 398–411.

IARYCZOWER, M., SHI, X. AND SHUM, M. (2018), "Can Words Get in the Way? The Effect of Deliberation in Collective Decision-Making", *Journal of Political Economy*, **126**, 688–734.

KAIDO, H., MOLINARI, F. AND STOYE, J. (2019), "Confidence Intervals for Projections of Partially Identified Parameters", *Econometrica*, **87**, 1397–1432.

KATZ, M. (2007), "Supermarkets and Zoning Laws" (Ph.D. Dissertation, Harvard University).
KAWAI, K. AND WATANABE, Y. (2013), "Inferring Strategic Voting", *American Economic Review*, **103**, 624–662.
KLEDER, M. (2020), "Con2vert - Constraints to Vertices" (https://www.mathworks.com/matlabcentral/fileexchange/7894-con2vert-constraints-to-vertices). MATLAB Central File Exchange. Retrieved 9 June 2020.
LEDOIT, O. AND WOLF, M. (2012), "Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices", *The Annals of Statistics*, **40**, 1024–1060.
MAGNOLFI, L. AND RONCORONI, C. (2016), "Estimation of Discrete Games with Weak Assumptions on Information" (Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison).
MANSKI, C. F. AND TAMER, E. (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome", *Econometrica*, **70**, 519–546.
MOHAMAD, D. A., VAN ZWET, E. W., CATOR, E. A. AND GOEMAN, J. J. (2020), "Adaptive Critical Value for Constrained Likelihood Ratio Testing", *Biometrica*, **107**, 677–688.
MOLINARI, F. (2020), "Microeconometrics with Partial Identification Econometrics with Partial Identification", in Durlauf, S., Hansen, L., Heckman, J. and Matzkin, R. (eds) *Handbook of Econometrics,* Vol. 7A, 1st edn (Amsterdam: North Holland).
MORALES, E., SHEU, G. AND ZAHLER, A. (2019), "Extended Gravity", *The Review of Economic Studies*, **86**, 2668–2712.
NEVO, A. AND ROSEN, A. (2012), "Identification with Imperfect Instruments", *The Review of Economics and Statistics*, **94**, 659–671.
PAKES, A., PORTER, J., HO, K. AND ISHII, J. (2015), "Moment Inequalities and their Applications", *Econometrica*, **83**, 315–334.
RAMBACHAN, A. AND ROTH, J. (2020), "An Honest Approach to Parallel Trends" (Unpublished manuscript, Department of Economics, Harvard University).
ROGERS, A. J. (1986), "Modified Lagrange Multiplier Tests for Problems with One-Sided Alternatives", *Journal of Econometrics*, **31**, 341–361.
ROMANO, J., SHAIKH, A. AND WOLF, M. (2014), "A Practical Two-Step Method for Testing Moment Inequalities", *Econometrica*, **82**, 1979–2002.
ROMANO, J. P. AND SHAIKH, A. M. (2008), "Inference for Identifiable Parameters in Partially Identified Econometric Models", *Journal of Statistical Planning and Inference*, **138**, 2786–2807. Special Issue in Honor of Theodore Wilbur Anderson, Jr. on the Occasion of his 90th Birthday.
ROMANO, J. P. AND SHAIKH, A. M. (2012), "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap", *The Annals of Statistics*, **40**, 2798–2822.
ROSEN, A. (2008), "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities", *Journal of Econometrics*, **146**, 107–117.
SHENG, S. (2016), "A Structural Econometric Analysis of Network Formation Games" (Unpublished Manuscript, Department of Economics, University of California Los Angeles).
SIERKSMA, G. AND ZWOLS, Y. (2015), *Linear and Integer Optimization: Theory and Practice*, 3rd edn (Boca Raton: CRC Press, Taylor & Francis Group).
SULLIVAN, C. J. (2017), "The Ice Cream Split: Empirically Distinguishing Price and Product Space Collusion" (Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison).
TAMER, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria", *The Review of Economic Studies*, **70**, 147–165.
UHLIG, H. (2005), "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure", *Journal of Monetary Economics*, **52**, 381–419.
WOLLMAN, T. G. (2018), "Trucks Without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles", *American Economic Review*, **108**, 1364–1406.