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

Gregory Cox[†] Xiaoxia Shi[‡]

June 10, 2021

Abstract

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.

Keywords: Moment Inequalities, Uniform Inference, Likelihood Ratio, Subvector Inference, Convex Polyhedron, Linear Programming

*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 2020 World Congress of the Econometric Society, the 2021 Winter Meeting of the Econometric Society, and econometrics seminars at Columbia University, the National University of Singapore, the UCSD, the UCLA, and the University of Wisconsin at Madison.

[†]Department of Economics, National University of Singapore (ecsgfc@nus.edu.sg)

[‡]Department of Economics, University of Wisconsin-Madison (xshi@ssc.wisc.edu)

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 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 nontrivial 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) and Andrews and Soares (2010)) and the subsample size in subsampling-based methods (e.g. Chernozhukov et al. (2007) and 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

[1] An incomplete list of applications that use inequalities as estimation restrictions includes Tamer (2003), Uhlig (2005), Bajari et al. (2007), Blundell et al. (2007), Ciliberto and Tamer (2009), Beresteanu et al. (2011), Holmes (2011), Baccara et al. (2012), Chetty (2012), Nevo and Rosen (2012), Kawai and Watanabe (2013), Eizenberg (2014), Huber and Mellace (2015), Pakes et al. (2015), Magnolfi and Roncoroni (2016), Sheng (2016), Sullivan (2017), He (2017), Iaryczower et al. (2018), Wollman (2018), Fack et al. (2019), and Morales et al. (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 et al. (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano and Shaikh (2012), and Romano et al. (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.

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 come as a by-product of computing T_n. We call it a conditional chi-squared 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 conditional chi-squared 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 10 seconds to compute a confidence interval in a specification with a 4-dimensional δ and 32 moment inequalities.

In both contexts, the conditional chi-squared 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. In particular, when all but one inequality get increasingly slack, the test asymptotes to one that ignores all the slack inequalities, which coincides with the uniformly most powerful test for the limiting model.

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

[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.

simple null, but was only recently proved to be valid for a composite null in Mohamad et al. (2020) in a normal model. We extend Mohamad et al. (2020) in four ways: (a) 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; (b) we design a simple but novel refinement to make the test size-exact; (c) we prove the test is uniformly asymptotically valid in moment inequality models; and (d) 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 (a)-(c) rely on technical contributions described in the 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 1 and 2) combined with an inequality on the tail of the truncated normal distribution (see Lemma 4). 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 7).

The idea of eliminating nuisance parameters from linear moment inequalities is first suggested in Guggenberger et al. (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: (a) numerical calculation of the Fourier-Motzkin elimination in general is an NP-hard computational problem, and (b) 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 conditional chi-squared test only relies on the rank of the active inequalities, and results from the convex analysis literature (see Lemmas 12 and 13) allow us to compute the rank of the active inequalities without carrying out Fourier-Motzkin elimination. The second difficulty is circumvented by considering models where the moment inequalities hold conditional on a vector of instrumental variables, a class of models first proposed by Andrews et al. (2019).

Andrews et al. (2019) (hereafter ARP) has the closest setting with our paper. 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 favorable test.

There are a few papers in the literature that propose methods to mitigate the computational challenges described above. Kaido et al. (2019) cast the problem of finding the bounds of the projection confidence interval of each parameter into a nonlinear nonconvex 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 paper, 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 et al. (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.

A couple of other papers aim to reduce the sensitivity of testing to tuning parameters. AB refines the procedure of Andrews and Soares (2010) (AS, hereafter) by computing an optimal moment selection threshold that maximizes a weighted average power 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 paper. 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 paper proceeds as follows. Section 2 describes our setup and several examples. Section 3 describes how to implement the full-vector and subvector conditional chi-squared tests. Section 4 states theoretical properties that the tests have. Section 5 reports the simulation results. Section 6 concludes. An 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

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

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]

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

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

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

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

[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 $\{\theta \in \Theta : A(\theta)\mathbb{E}_F\overline{m}_n(\theta) \leq b(\theta)\}$.

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

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 3 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 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,

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

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,

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

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 favorite. 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

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

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 5-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: (a) there are nuisance parameters, denoted by δ, that enter the moments linearly, (b) the inequalities hold conditionally on exogenous variables, $\{Z_i\}_{i=1}^n$, and (c) 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 $k \times 1$ matrices, 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 $Z_1, ..., Z_n$. 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: (i) the nuisance parameter δ enters linearly, and (ii) 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

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

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

Example 4. *Gandhi et al. (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*

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

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

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

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, $m(W_i, \theta) = \begin{pmatrix} \psi_i^L(\theta)\mathcal{I}(Z_i) \\ -\psi_i^U(\theta)\mathcal{I}(Z_i) \end{pmatrix}$, $C_Z = n^{-1} \sum_{i=1}^n \begin{pmatrix} \mathcal{I}(Z_i) Z_{ci}' \\ -\mathcal{I}(Z_i) Z_{ci}' \end{pmatrix}$, *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 confidence intervals 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: $\theta_0 = \gamma_0' \bar{P}_{1,n}$, where $\bar{P}_{1,n} = n^{-1} \sum_{i=1}^n \partial P(Z_i)/\partial z_1$. One can rewrite (14) as

$$\begin{aligned} \mathbb{E}\left[(L_i - \theta_0)\mathcal{I}(Z_i) - \mathcal{I}(Z_i)(P(Z_i) - \bar{P}_{1,n})'\gamma_0 | Z_i\right] &\leq \mathbf{0} \\ \mathbb{E}\left[(-U_i + \theta_0)\mathcal{I}(Z_i) + \mathcal{I}(Z_i)(P(Z_i) - \bar{P}_{1,n})'\gamma_0 | Z_i\right] &\leq \mathbf{0}, \end{aligned} \tag{15}$$

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, $m(W_i, \theta) = \begin{pmatrix} (L_i - \theta)\mathcal{I}(Z_i) \\ -(U_i - \theta)\mathcal{I}(Z_i) \end{pmatrix}$, $C_Z = n^{-1} \sum_{i=1}^n \begin{pmatrix} \mathcal{I}(Z_i)(P(Z_i) - \bar{P}_{1,n}) \\ -\mathcal{I}(Z_i)(P(Z_i) - \bar{P}_{1,n}) \end{pmatrix}$, 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 Conditional Chi-Squared Tests: Implementation

In this section we define a new family of tests, called conditional chi-squared tests, for the inequalities specified in (1) and (8). They are called conditional chi-squared tests because they use 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 tests, which shows that they are easy to implement and have low computational cost.

3.1 Full-Vector Tests

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 two new tests, one being a refinement of the other. Both are easy to compute, requiring no tuning parameters or simulations. Both use the (quasi-) likelihood ratio statistic,

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

where $\widehat{\Sigma}_n(\theta)$ denotes an estimator of $\text{Var}_F(\sqrt{n}\overline{m}_n(\theta))$, 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 $\{W_i\}_{i=1}^n$.

Both tests use data-dependent critical values that are based on the rank of the rows of A corresponding to the inequalities that are active in finite sample. To define them 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 $\text{rk}(A_J)$ denote the rank of A_J, and let $\hat{r} = \text{rk}(A_{\widehat{J}})$. 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 first simple 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 first simple test by

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

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 (that is, $A\mathbb{E}_F\overline{m}_n(\theta) = b$), 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.

[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.

We propose a second simple test that eliminates the conservativeness. We call this the RCC (refined CC) test. We define the RCC test by adjusting the quantile of the χ^2_1 distribution when $\hat{r} = 1$. Instead of the $100(1-\alpha)\%$ quantile, the RCC test uses a $100(1-\hat{\beta})\%$ quantile, where $\hat{\beta}$ varies between α and 2α depending on how far from active the additional (inactive) inequalities are. We now construct $\hat{\beta}$ carefully so that the refinement exactly restores the size of the test.

When $\hat{r} = 1$, suppose without loss of generality that the first inequality is active and satisfies $a_1 \neq 0$.[8] Next, for each $j = 2, ..., d_A$, let

$$\hat{\tau}_j = \begin{cases} \frac{\sqrt{n}\|a_1\|_{\widehat{\Sigma}_n(\theta)}\left(b_j - a_j'\hat{\mu}\right)}{\|a_1\|_{\widehat{\Sigma}_n(\theta)}\|a_j\|_{\widehat{\Sigma}_n(\theta)} - a_1'\widehat{\Sigma}_n(\theta)a_j} & \text{if } \|a_1\|_{\widehat{\Sigma}_n(\theta)}\|a_j\|_{\widehat{\Sigma}_n(\theta)} \neq a_1'\widehat{\Sigma}_n(\theta)a_j \\ \infty & \text{otherwise} \end{cases}, \tag{21}$$

where $\|a\|_\Sigma = (a'\Sigma a)^{1/2}$. This $\hat{\tau}_j$ is a normalized measure of the inactivity of the jth inequality. It is essentially $b_j - a_j'\hat{\mu}$ normalized using the ratio of the Euclidean norms of $\hat{\Sigma}_n^{1/2}(\theta)a_1$ and $\hat{\Sigma}_n^{1/2}(\theta)a_j$ and the angle between the two.[9] Then let

$$\hat{\tau} = \inf_{j \in \{2,...,d_A\}} \hat{\tau}_j. \tag{22}$$

This is a measure of the minimum inactivity of the inactive inequalities. This quantity is easy to compute and has a nice geometric interpretation that is illustrated in Illustration 1 below.

Now we can define

$$\hat{\beta} = \begin{cases} 2\alpha\Phi(\hat{\tau}) & \text{if } \hat{r} = 1 \\ \alpha & \text{otherwise,} \end{cases} \tag{23}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (cdf). When a second inequality is close to being active, $\hat{\tau}$ is close to 0 and then $\hat{\beta}$ is close to α. When all the other inequalities are far from active, then $\hat{\tau}$ is very large and $\hat{\beta}$ is close to 2α. We define the RCC test for $H_0 : \theta = \theta_0$ to be

$$\phi_n^{\mathrm{RCC}}(\theta, \alpha) = 1\{T_n(\theta) > \chi^2_{\hat{r},1-\hat{\beta}}\}. \tag{24}$$

Note that for test inversion, both $\hat{r}$ and $\hat{\beta}$ need to be recalculated for every value of θ since they may depend on θ via A, b, and $\widehat{\Sigma}_n(\theta)$.

[8] In this case, other inequalities may be active too since we do not rule out the possibility that A contains redundant or zero rows. But this is possible only if the other active inequalities are collinear with a_1.

[9] Note that $a_1'\widehat{\Sigma}_n(\theta)a_j = \|a_1\|_{\widehat{\Sigma}_n(\theta)}\|a_j\|_{\widehat{\Sigma}_n(\theta)}\cos\gamma$, where γ stands for the angle.

Since $\hat{\tau} \in [0, \infty]$, $\hat{\beta} \in [\alpha, 2\alpha]$. Thus we have the following comparison of the CC and the RCC tests:

$$\phi_n^{\text{RCC}}(\theta, \alpha/2) \leq \phi_n^{\text{CC}}(\theta, \alpha) \leq \phi_n^{\text{RCC}}(\theta, \alpha). \tag{25}$$

Moreover, when an equality is being tested, at least two inequalities are always active, in which case we have $\hat{\beta} = \alpha$, and the RCC test reduces to the CC test.

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. The critical values for the RCC test are illustrated using a dashed line where they deviate from the CC test.[10]

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). In that case, a smaller critical value is used that depends on how far from active the other inequality is, measured using $\hat{\tau}$. The quantity $\hat{\tau}$ has the following geometric interpretation: the point $\sqrt{n}\hat{\mu}$ is the projection of $\sqrt{n}\overline{m}_n$ onto a face of the polyhedron defined by the inequalities. Continue that line into the interior of the polyhedron until you reach a point, y, that is equidistant between two inequalities. In the figure, the set of points that are equidistant between two inequalities is represented by the dotted line, which is the 45-degree line. Then $\hat{\tau}$ is the distance between $\sqrt{n}\hat{\mu}$ and y. This geometric interpretation extends to more complicated examples with more inequalities or non-orthogonal inequalities.

The reason the refinement still controls size is that we condition on the event that $\sqrt{n}\overline{m}_n$ belongs to the ray that starts at y and emanates through $\sqrt{n}\hat{\mu}$ and $\sqrt{n}\overline{m}_n$. It is sufficient to control the conditional rejection probability for every such ray. By conditioning on the ray, the denominator of the conditional rejection probability is $\Phi(\hat{\tau})$, which allows us to adjust α up to $\hat{\beta}$.

[10] 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.

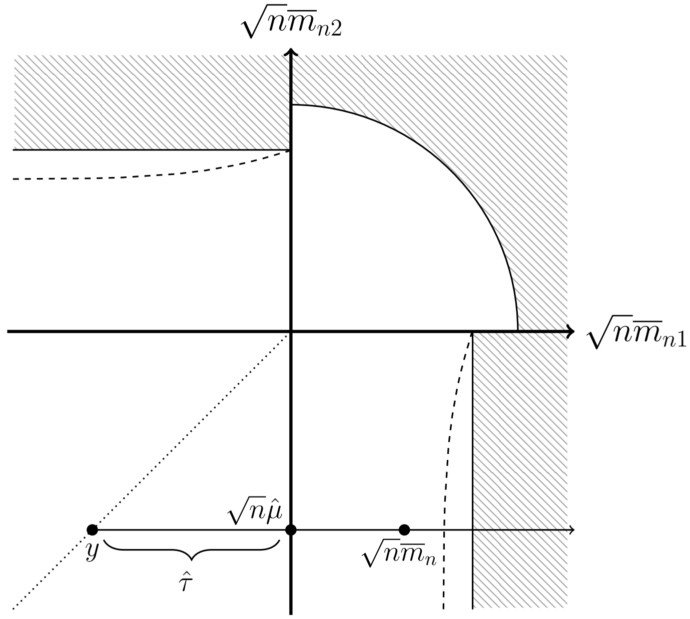


Figure 1: Geometric representation of the CC test (shaded) and the RCC test (dashed) in Illustration 1.

It is also helpful to see the CC tests in a simple model with a scalar parameter of interest, one upper bound, and one lower bound: $E[Y^L] \leq \theta \leq E[Y^U]$. This setup has been considered, for example, in Stoye (2009). For simplicity, suppose Y^L and Y^U are independent and have unit variance. Let $\bar{Y}_n^L$ and $\bar{Y}_n^U$ be the sample average of Y^L and Y^U, respectively. Then it is not difficult to find that (when $\hat{\Delta}_n := \sqrt{n}(\bar{Y}_n^U - \bar{Y}_n^L) > -z_{1-\alpha/2}$) the $100(1-\alpha)\%$ CC confidence interval is $[\bar{Y}_n^L - z_{1-\alpha/2}/\sqrt{n}, \bar{Y}_n^U + z_{1-\alpha/2}/\sqrt{n}]$, and also that the RCC confidence interval is the set of θ values that satisfy $\bar{Y}_n^L - z_{1-\alpha\Phi(\sqrt{n}(\bar{Y}_n^U-\theta)\vee 0)}/\sqrt{n} \leq \theta \leq \bar{Y}_n^U + z_{1-\alpha\Phi(\sqrt{n}(\theta-\bar{Y}_n^L)\vee 0)}/\sqrt{n}$, where $\vee$ is the maximum operator. Solving numerically, we find that the RCC confidence interval is $[\bar{Y}_n^L - c_\alpha/\sqrt{n}, \bar{Y}_n^U + c_\alpha/\sqrt{n}]$ where c_α depends on $\hat{\Delta}_n$. For example, when $\alpha = 0.05$, c_α declines smoothly from 1.96 to 1.67 and then to 1.65 as $\hat{\Delta}_n$ varies from -1.96 to 0 and then to 1. Thus, the refinement brings about a big improvement in this simple setup.

To end this subsection, Algorithm 1 presents pseudo-code that can be used to compute the CC and RCC tests. 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 tests. We list them here for emphasis. (a) The CC tests do not require any tuning parameters or simulations to implement. (b) The CC tests are simple to code. (c) There

Algorithm 1: Pseudo-code for implementing the CC and RCC tests.

1: **%Compute the CC Test**
2: $T_n(\theta), \hat{\mu} \quad \leftarrow \min_{\mu : A\mu \le b} n(\overline{m}_n(\theta) - \mu)' \widehat{\Sigma}_n(\theta)^{-1} (\overline{m}_n(\theta) - \mu)$
3: $\widehat{J} \quad := \{j = 1, \ldots, d_A : a_j' \hat{\mu} = b_j\}$
4: $A_{\widehat{J}} \quad \leftarrow \widehat{J}, A$
5: $\hat{r} \quad := \mathrm{rk}(A_{\widehat{J}})$
6: $\phi_n^{\mathrm{CC}}(\theta, \alpha) := 1\{T_n(\theta) > \max\{\chi^2_{\hat{r}, 1-\alpha}, \mathrm{tol}\}\}.$
7:
8: **%Compute the RCC Test**
9: Implement lines 2-5, and then
10: **if** $\hat{r} = 1$ and $\chi^2_{1,1-2\alpha} \le T_n(\theta) \le \chi^2_{1,1-\alpha}$ **then**
11: (suppose $a_1' \hat{\mu} = b_1$ and $\|a_1\| \neq 0$)
12: **for** $j = 2, \ldots, d_A$ **do**
13: $\hat{\tau}_j := \begin{cases} \frac{\sqrt{n} \|a_1\|_{\widehat{\Sigma}_n(\theta)} \left(b_j - a_j' \hat{\mu}\right)}{\|a_1\|_{\widehat{\Sigma}_n(\theta)} \|a_j\|_{\widehat{\Sigma}_n(\theta)} - a_1' \widehat{\Sigma}_n(\theta) a_j} & \text{if } \|a_1\|_{\widehat{\Sigma}_n(\theta)} \|a_j\|_{\widehat{\Sigma}_n(\theta)} \neq a_1' \widehat{\Sigma}_n(\theta) a_j \\ \infty & \text{otherwise} \end{cases}$
14: **end for**
15: $\hat{\tau} \quad := \inf_{2,\ldots,d_A} \hat{\tau}_j$
16: $\hat{\beta} \quad := 2\alpha \Phi(\hat{\tau})$
17: $\phi_n^{\mathrm{RCC}}(\theta, \alpha) := 1\{T_n(\theta) > \max\{\chi^2_{1,1-\hat{\beta}}, \mathrm{tol}\}\}$
18: **else**
19: $\phi_n^{\mathrm{RCC}}(\theta, \alpha) := 1\{T_n(\theta) > \max\{\chi^2_{\hat{r},1-\alpha}, \mathrm{tol}\}\}.$
20: **end if**

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 other inequalities. The CC tests are 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 Tests

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:

$$H_0 : \exists \delta \text{ such that } \; B_Z \mathbb{E}_{F_Z}[\overline{m}_n(\theta_0) | Z] - C_Z \delta \le d_Z, \; a.s. \tag{26}$$

In this subsection, we define subvector versions of the conditional chi-squared tests for (26).

Directly testing (26) 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 (26) that eliminates δ:

$$H_0 : A_Z \mathbb{E}_{F_Z}[\overline{m}_n(\theta_0)|Z] \leq b_Z, \tag{27}$$

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 conditional chi-squared test on (27) *without* calculating A_Z and b_Z, as we describe next.

The subvector CC (sCC) test for (26) is the full-vector CC test based on (27). It uses the test statistic

$$T_n(\theta) = \min_{\mu : A_Z \mu \leq b_Z} n(\overline{m}_n(\theta) - \mu)' \widehat{\Sigma}_n(\theta)^{-1} (\overline{m}_n(\theta) - \mu), \tag{28}$$

where $\widehat{\Sigma}_n(\theta)$ is an estimator of the conditional variance: $\Sigma_n(\theta) = \text{Var}(\sqrt{n}\overline{m}_n(\theta)|Z)$, 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 (28).

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

$$T_n(\theta) = \min_{\delta, \mu : B_Z \mu - C_Z \delta \leq d_Z} n(\overline{m}_n(\theta) - \mu)' \widehat{\Sigma}_n(\theta)^{-1} (\overline{m}_n(\theta) - \mu). \tag{29}$$

One can calculate $T_n(\theta)$ without knowing A_Z or b_Z by quadratic programming, where $(\delta', \mu')'$ is the decision variable. Let $(\hat{\delta}', \hat{\mu}')'$ be the solution to the minimization problem.

Before we describe how to compute $\hat{r}$ without A_Z or b_Z, we briefly describe what A_Z and b_Z are. There are multiple ways to define A_Z and b_Z for (27) to be equivalent to (26). 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 nonnegative weights that sum to one, then the convex combination of the inequalities in (26) is given by

$$h' B_Z \mathbb{E}_{F_Z}[\overline{m}_n(\theta_0)|Z] - h' C_Z \delta \leq h' d_Z. \tag{30}$$

When $h' C_Z = \mathbf{0}$, the δ parameter is eliminated from the inequalities. It follows from Gale's

Theorem[11] that it is sufficient to consider the set of all inequalities (30) indexed by

$$h \in \mathcal{H} := \{h \in \mathbb{R}^k : h \geq \mathbf{0}, C_Z' h = \mathbf{0}, 1'h = 1\}. \tag{31}$$

To connect this result to (27), 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 (26) if and only if θ_0 satisfies (30) 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

$$A_Z = H(C_Z)B_Z \text{ and } b_Z = H(C_Z)d_Z \tag{32}$$

renders (27) equivalent to (26). This result is formally stated in Lemma 12 in Appendix C.1.

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. We proceed to describe how to compute $\hat{r}$ without A_Z or b_Z.

To compute $\hat{r}$, we define the active inequalities. For any $h \in \mathcal{H}$, we say that the inequality in (30) is active if $h'B_Z\hat{\mu} = h'd_Z$, where $\hat{\mu}$ is calculated from (29). Accordingly, let

$$\mathcal{H}_0 = \{h \in \mathcal{H} : (B_Z\hat{\mu} - d_Z)'h = \mathbf{0}\} \tag{33}$$

denote the subset of $\mathcal{H}$ that characterizes the active inequalities. In fact, $\mathcal{H}_0$ is always a face of $\mathcal{H}$ due to the definition of $\hat{\mu}$. By the definition of $\hat{r}$ and A_Z, $\hat{r}$ is the maximum number of linearly independent vectors of the form $B_Z'h$, where h is a vertex of $\mathcal{H}_0$. The key is to recognize that we do not need to enumerate the vertices of $\mathcal{H}_0$ to calculate $\hat{r}$. Instead, we only have to calculate the maximum number of linearly independent vectors in $B_Z'\mathcal{H}_0 = \{B_Z'h : h \in \mathcal{H}_0\}$. Notationally, we call the maximum number of linearly independent vectors in $B_Z'\mathcal{H}_0$ the "rank of $B_Z'\mathcal{H}_0$" and denote it by $\text{rk}(B_Z'\mathcal{H}_0)$.[12] The fact that $\hat{r} = \text{rk}(B_Z'\mathcal{H}_0)$ is stated formally in Lemma 13 in Appendix C.1.

Therefore, to compute $\hat{r}$ one only needs to find $\text{rk}(B_Z'\mathcal{H}_0)$. It turns out that calculating the rank of a polyhedron is much faster computationally than enumerating the vertices.

[11]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 nonnegative solutions to linear systems of equations.

[12]Usually, $\text{rk}(\cdot)$ is defined for matrices. Here we extend the definition to arbitrary sets of vectors.

Here, we present an algorithm based on solving $k+1$ linear programming (LP) problems. For exposition, we assume $\text{rk}(B_Z)=k$, which is true in Examples 3-5, so that the rank of $B_Z\mathcal{H}_0$ is equal to the rank of $\mathcal{H}_0$.[13] Calculating the rank of $\mathcal{H}_0$ is equivalent to finding the dimension of the smallest linear subspace containing $\mathcal{H}_0$, denoted by $\text{span}(\mathcal{H}_0)$. Note that $\mathcal{H}_0$ is defined by linear equalities and inequalities, where the inequalities are given by $h \geq \mathbf{0}$. Some of these inequalities may have to hold with equality due to the other equations in the definition of $\mathcal{H}_0$. That is, for some $j=1,...,k$, $h \in \mathcal{H}_0$ may imply that $h_j = 0$. If we can figure out which of the inequalities have to hold with equality, we can find a system of equations that defines $\text{span}(\mathcal{H}_0)$, and from there figure out the dimension of $\text{span}(\mathcal{H}_0)$.

Thus, the imminent question becomes: for which j does $h \in \mathcal{H}_0$ imply $h_j = 0$? For each $j = 1,...,k$, we can answer this question with a LP problem.[14] For each $j=1,...,k$, calculate

$$\zeta_j = \min_h -h_j \text{ s.t. } h \geq \mathbf{0}, C_Z'h = \mathbf{0}, (B_Z\hat{\mu} - d_Z)'h = \mathbf{0}, 1'h = 1. \tag{34}$$

If $\zeta_j = 0$, then there does not exist an $h \in \mathcal{H}_0$ with $h_j > 0$, which means that the jth inequality has to hold with equality. Let J_0 be the collection of all j's such that $\zeta_j = 0$. Also let I_{J_0} denote the rows of the k-dimensional identity matrix corresponding to indices in J_0. It follows that

$$\text{span}(\mathcal{H}_0) = \{h \in \mathbb{R}^k : I_{J_0}h = \mathbf{0}, C_Z'h = \mathbf{0}, (B_Z\hat{\mu} - d_Z)'h = \mathbf{0}\}. \tag{35}$$

Correspondingly, the rank of $\mathcal{H}_0$ is k minus the rank of the coefficients on the linear equations defining $\text{span}(\mathcal{H}_0)$:

$$\text{rk}(\mathcal{H}_0) = k - \text{rk}\begin{pmatrix} I_{J_0} \\ C_Z' \\ (B_Z\hat{\mu} - d_Z)' \end{pmatrix}. \tag{36}$$

This is how we compute $\hat{r}$, and hence the sCC test, without computing A_Z or b_Z.

While implementing the sCC test does not require computing A_Z or b_Z, this is not the case for the subvector RCC (sRCC) test. The refinement requires knowing A_Z and b_Z. However, note that the refinement makes a difference only when $\hat{r} = 1$ and $T_n(\theta) \in [\chi^2_{1,1-2\alpha}, \chi^2_{1,1-\alpha}]$ (because $\hat{\beta} \in [\alpha, 2\alpha]$). Thus, to implement the sRCC test, we recommend computing $\hat{r}$ and $T_n(\theta)$ first using the method outlined above, and only computing A_Z, b_Z, and the refinement when $\hat{r} = 1$ and $T_n(\theta) \in [\chi^2_{1,1-2\alpha}, \chi^2_{1,1-\alpha}]$. Our experience is that this event is rare when

[13] Appendix C.2 presents an algorithm for the case that $\text{rk}(B_Z) < k$.

[14] Before implementing these LP problems, one should first determine if $\mathcal{H}_0$ is empty. This can be done by solving the LP problem: $f := \min_h -(B_Z\hat{\mu} - d_Z)'h$ *s.t.* $h \geq \mathbf{0}, C_Z'h = \mathbf{0}, 1'h = 1$. If $f > 0$, that indicates that all elements of $A_Z\hat{\mu} - b_Z$ are negative and there is no active inequality. In this case, set $\mathcal{H}_0 = \emptyset$ and $\hat{r} = 0$.

k and p are large enough to make computing A_Z and b_Z challenging.[15] When k and p are small, computing A_Z and b_Z via vertex enumeration is feasible.

Next we give two examples of 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. We describe two conditional variance matrix estimators, one for discrete Z_i and the other for continuous Z_i, both in the context of i.i.d. data.

In the first case, 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:

$$\widehat{\Sigma}_n(\theta) = \sum_{\ell \in \mathcal{Z}} \frac{n_\ell}{n} \frac{1}{n_\ell - 1} \sum_{i=1}^{n} (m(W_i, \theta) - \overline{m}_n^\ell(\theta))(m(W_i, \theta) - \overline{m}_n^\ell(\theta))' 1\{Z_i = \ell\}, \tag{37}$$

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 Appendix D.2, 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.

In the second case, Z_i contains continuous random variables. One can use a nearest neighbor matching estimator similar to that used for the standard error of a regression discontinuity estimator in Abadie, Imbens, and Zheng (2014).[16] Let $\Sigma_{Z,n} = n^{-1} \sum_{i=1}^n (Z_i - \overline{Z}_n)(Z_i - \overline{Z}_n)'$ where $\overline{Z}_n = n^{-1} \sum_{i=1}^n Z_i$. For each i, define the nearest neighbor to be

$$\ell_Z(i) = \text{argmin}_{j \in \{1, \ldots, n\}, j \neq i} (Z_i - Z_j)' \Sigma_{Z,n}^{-1} (Z_i - Z_j). \tag{38}$$

When the argmin is not unique, picking one randomly does not affect the consistency of the resulting estimator. The estimator of $\Sigma_n(\theta)$ is then given by

$$\widehat{\Sigma}_n(\theta) = \frac{1}{2n} \sum_{i=1}^{n} (m(W_i, \theta) - m(W_{\ell_Z(i)}, \theta))(m(W_i, \theta) - m(W_{\ell_Z(i)}, \theta))'. \tag{39}$$

As we show in Appendix D.2, sufficient conditions for the consistency of this matching estimator involves the boundedness of $\{Z_i\}_{i=1}^\infty$ and the Lipschitz continuity of $\text{Var}(m(W_i, \theta)|Z_i = z_i)$

[15] Theoretically, if the moment inequalities are uncorrelated and k of them are binding, then the probability $\hat{r} = 1$ is asymptotically $k2^{-k}$. This is an upper bound for $\Pr(\hat{r} = 1, T_n(\theta) \in [\chi^2_{1,1-2\alpha}, \chi^2_{1,1-\alpha}])$. In finite sample, the number of near-binding moment inequalities also reduces this probability.

[16] This is also the estimator used in ARP.

Algorithm 2: Pseudo-code for the sCC and sRCC tests when $\text{rk}(B_Z) = k$.

1: **%Compute the sCC Test**
2: $T_n(\theta), \hat{\mu} \leftarrow \min_{\delta,\mu: B_Z\mu - C_Z\delta \leq d_Z} n(\widehat{m}_n(\theta) - \mu)'\widehat{\Sigma}_n(\theta)^{-1}(\overline{m}_n(\theta) - \mu)$
3: $f \quad := \min_{h\in\mathcal{H}} -(B_Z\hat{\mu} - d_Z)'h$
4: **if** $f >$ tol **then**
5: $\quad \hat{r} := 0$
6: **else**
7: $\quad$ **for** $j = 1, \ldots, k$ **do**
8: $\quad\quad h_j^m \leftarrow C_Z, B_Z, d_Z, \hat{\mu}$ by (34).
9: $\quad$ **end for**
10: $\quad J_0 := \{j = 1, \ldots, k : h_j^m = 0\}$
11: $\quad I_{J_0} \leftarrow J_0$
12: $\quad \hat{r} \quad := k - \text{rk}\begin{pmatrix} I_{J_0} \\ C_Z' \\ (B_Z\hat{\mu} - d_Z)' \end{pmatrix}$
13: **end if**
14: $\phi_n^{\text{sCC}}(\theta, \alpha) := 1\{T_n(\theta) > \max\{\chi^2_{\hat{r},1-\alpha}, \text{tol}\}\}$.
15:
16: **%Compute the sRCC Test**
17: Implement lines 2-13, and then
18: **if** $\hat{r} = 1$ **and** $T_n(\theta) \in [\chi^2_{1,1-2\alpha}, \chi^2_{1,1-\alpha}]$ **then**
19: $\quad H(C_Z) \leftarrow \mathcal{H}$ using a vertex enumeration algorithm, e.g. **con2vert.m** in Matlab (ref. Kleder (2020))
20: $\quad A_Z, b_Z \leftarrow H(C_Z)B_Z, H(C_Z)d_Z$
21: $\quad$ Suppose $a_1'\hat{\mu} = b_1$ and $\|a_1\| \neq 0$. %Ignore the subscript Z for notational ease.
22: $\quad$ **for** $j = 2, \ldots, d_A$ **do**
23: $\quad\quad \hat{\tau}_j := \begin{cases} \frac{\sqrt{n}\|a_1\|_{\widehat{\Sigma}_n(\theta)}(b_j - a_j'\hat{\mu})}{\|a_1\|_{\widehat{\Sigma}_n(\theta)}\|a_j\|_{\widehat{\Sigma}_n(\theta)} - a_1'\widehat{\Sigma}_n(\theta)a_j} & \text{if } \|a_1\|_{\widehat{\Sigma}_n(\theta)}\|a_j\|_{\widehat{\Sigma}_n(\theta)} \neq a_1'\widehat{\Sigma}_n(\theta)a_j \\ \infty & \text{otherwise} \end{cases}$
24: $\quad$ **end for**
25: $\quad \hat{\tau} \quad := \inf_{2,\ldots,d_A} \hat{\tau}_j$
26: $\quad \hat{\beta} \quad := 2\alpha\Phi(\hat{\tau})$
27: $\quad \phi_n^{\text{sRCC}}(\theta, \alpha) := 1\{T_n(\theta) > \max\{\chi^2_{1,1-\hat{\beta}}, \text{tol}\}\}$
28: **else**
29: $\quad \phi_n^{\text{sRCC}}(\theta, \alpha) := 1\{T_n(\theta) > \max\{\chi^2_{\hat{r},1-\alpha}, \text{tol}\}\}$.
30: **end if**

in z_i.

To end this subsection, Algorithm 2 presents pseudo-code that can be used to compute the sCC and sRCC tests in the case where B_Z has rank k. 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 tests have 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 sRCC test because the vertex enumeration used to calculate A_Z and b_Z often delivers redundant inequalities, and these do not have to be removed before implementing the sRCC test.

4 Theoretical Properties

Next we consider the theoretical properties of the CC tests. We show that both the full-vector and subvector RCC tests are size-exact in finite samples under normality and known variance-covariance matrix. We also show that they are uniformly asymptotically size-exact when the moments are asymptotically normal. Moreover, we make precise the adaptiveness of the tests to slackness of the moment inequalities.

4.1 Finite Sample Properties

When the moments are normally distributed with known variance-covariance matrix, the following theorem states the finite sample properties of the RCC test. We define a reduced test that only uses a subset of the inequalities. For $J \subseteq \{1, ..., d_A\}$, let $\phi_{n,J}^{\text{RCC}}(\theta, \alpha)$ denote the RCC 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.

Theorem 1. *Suppose $\Sigma_n(\theta)$ is an invertible matrix such that $\sqrt{n}(\overline{m}_n(\theta) - \mathbb{E}_F\overline{m}_n(\theta)) \sim N(0, \Sigma_n(\theta))$ and $\widehat{\Sigma}_n(\theta) = \Sigma_n(\theta)$ a.s. for all $\theta \in \Theta$. Then the following hold.*

(a) *For any $\theta \in \Theta_0(F)$, $\mathbb{E}_F\phi_n^{\text{RCC}}(\theta, \alpha) \leq \alpha$.*

(b) *If $A\mathbb{E}_F\overline{m}_n(\theta) = b$ and $A \neq \mathbf{0}$, then $\mathbb{E}_F\phi_n^{\text{RCC}}(\theta, \alpha) = \alpha$.*

(c) *If $J \subseteq \{1, ..., d_A\}$ and $\{\theta_s\}_{s=1}^{\infty} \subseteq \Theta$ is a sequence such that for all $j \notin J$, $a_j \neq 0$ and $(a_j'\mathbb{E}_F\overline{m}_n(\theta_s) - b_j)/\|a_j\| \to -\infty$ as $s \to \infty$, where the dependence of a_j and b_j on s via θ_s is implicit, then*

$$\lim_{s\to\infty} \Pr_F\left(\phi_n^{\text{RCC}}(\theta_s, \alpha) \neq \phi_{n,J}^{\text{RCC}}(\theta_s, \alpha)\right) = 0.$$

Remarks. (1) Part (a) shows the finite sample validity of the RCC test when the moments are normally distributed with known variance. Part (b) shows that the RCC test is size-exact when there is an F and $\theta \in \Theta_0(F)$ under which all the inequalities bind.[17] By size-exact, we mean that the worst case null rejection probability is equal to α. This is compatible with having other F and $\theta \in \Theta_0(F)$ under which the rejection probability is less than α (i.e. under-rejection). Indeed, the RCC test under-rejects when some inequalities do not bind, which is a common feature of all moment inequality tests. The under-rejection does not increase monotonically with the slackness of the slack inequalities, however. Part (c) shows that when the slack inequalities get very slack, the RCC test reduces to a version of it that does not use those inequalities. In particular, if some inequalities are binding while others get very slack, the rejection rate of the RCC test gets close to α. Put another way, the RCC test adapts to the slackness of the inequalities. We call this property "irrelevance of distant inequalities" or IDI. This is especially useful if all but one inequality is very slack, because the reduced test is the one-sided t-test for the sole binding inequality, which is uniformly most powerful.

(2) Several papers, including Kudo (1963) and Wolak (1987), propose a classical test for inequalities that can be applied here. The classical test is based on the $1-\alpha$ quantile of the least favorable distribution of $T_n(\theta)$, which is a mixture of $\chi^2_0, \chi^2_1, \ldots, \chi^2_{d_A}$ distributions. This test also has exact size, but lacks the IDI property. When many inequalities tested are slack, the power of this test can be very low. Besides, the critical value typically requires simulation, which makes it computationally less attractive than the RCC test.

(3) AS introduced a generalized moment selection procedure that achieves an asymptotic version of the IDI property via a sequence of tuning parameters. AB consider a test that simulates a critical value from the asymptotic normal distribution. That test has finite sample exact size (due to size correction), but it does not have the IDI property. The size correction the test uses causes it to respond to very slack inequalities, albeit to a lesser extent than the classical test.

(4) The only other moment inequality test with exact size and the IDI property is the non-hybrid test in ARP. This test is based on the conditional distribution of the maximum standardized element of $\sqrt{n}(A\overline{m}_n(\theta) - b)$ given the second-largest maximum. The test also asymptotes to the one-sided t-test when all but one inequality get increasingly slack. However, the test has undesirable power when multiple inequalities are not well separated, which prompts them to recommend a hybrid test instead.

(5) Theorem 1 and the other results in this paper 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

[17] Using (25), part (b) also implies that the size of the CC test is between $\alpha/2$ and α in this case.

1(a), we have for all $\theta_0 \in \Theta_0(F)$,

$$\Pr_F\left(\theta_0 \in CS_n^{\mathrm{RCC}}(1-\alpha)\right) \geq 1-\alpha, \tag{40}$$

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

(6) The proof of Theorem 1 is challenging. It relies on a careful partition of the space of realizations of the moments according to which inequalities are active (see Lemmas 1 and 2). It then uses a bound on probabilities of translations of sets to bound the rejection probability conditional on each set in the partition (see Lemmas 3 and 4). Mohamad et al. (2020) prove a special case of part (a) for the CC test when the inequalities define a cone. We extend the result to the RCC test and allow the inequalities to define an arbitrary polyhedron, an important extension for moment inequality models. In particular, the design of the refinement is not obvious from Mohamad et. al (2020) and requires a careful study of the geometric properties of the test statistic. □

4.2 Asymptotics

Now we turn to the asymptotic properties of the CC tests when the moments are only 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 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 RCC test has correct asymptotic size uniformly over a large class of data generating processes.

The following assumption defines the set of data generating processes allowed. Here $|\cdot|$ denotes the matrix determinant, and ϵ and M are fixed positive constants that do not dependent 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) $\sigma_{F,j}^2(\theta) := \mathrm{Var}_F(m_j(W_i, \theta)) > 0$ *for $j = 1, \ldots, d_m$.*

(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) $\mathbb{E}_F|m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+\epsilon} \leq M$ *for $j = 1, \ldots, d_m$.*

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 3 in Appendix B. Part (b) is innocuous as it simply requires the moment functions be nonconstant 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:

$$\begin{aligned}
\mathbb{E}[1\{(r_i^1, r_i^2, r_i^3) = (0,0,0)\} - g_{000}(\theta)] &\leq 0 \\
\mathbb{E}[-1\{(r_i^1, r_i^2, r_i^3) = (0,0,0)\} + g_{000}(\theta)] &\leq 0 \\
\mathbb{E}[1\{(r_i^1, r_i^2, r_i^3) = (a,b,c)\} - g_{abc}(\theta)] &\leq 0 \\
&\vdots \\
\mathbb{E}[1\{(r_i^1, r_i^2, r_i^3) = (b,0,0)\} - g_{b00}(\theta)] &\leq 0 \\
\mathbb{E}[1\{(r_i^1, r_i^2, r_i^3) = (c,0,0)\} - g_{c00}(\theta)] &\leq 0. \qquad (41)
\end{aligned}$$

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 multicollinearity 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 $J \subseteq$

$\{1,\ldots,d_A\}$, let I_J denote the rows of the identity matrix corresponding to the indices in J.[18] The following theorem states the asymptotic properties of the RCC test.

Theorem 2. *Suppose Assumption* 1 *holds.*

(a) $\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{\theta\in\Theta_0(F)} \mathbb{E}_F \phi_n^{\mathrm{RCC}}(\theta,\alpha) \le \alpha$.

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* $J \subseteq \{1,\ldots,d_A\}, \mathrm{rk}(I_J A(\theta_n)D_{F_n}(\theta_n)) = \mathrm{rk}(I_J A_\infty)$ *for all* n,

(b) *if* $A_\infty \ne \mathbf{0}$ *and for all* $j \in \{1,...,d_A\}$, $\sqrt{n}(a_j'\mathbb{E}_{F_n}\overline{m}_n(\theta_n) - b_j) \to 0$, *then*

$$\lim_{n\to\infty} \mathbb{E}_{F_n}\phi_n^{\mathrm{RCC}}(\theta_n,\alpha) = \alpha, \text{ and}$$

(c) *if instead there is a* $J \subseteq \{1,\ldots,d_A\}$ *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*

$$\lim_{n\to\infty} \Pr_{F_n}\left(\phi_n^{\mathrm{RCC}}(\theta_n,\alpha) \ne \phi_{n,J}^{\mathrm{RCC}}(\theta_n,\alpha)\right) = 0.$$

Remarks. (1) Part (a) shows that the RCC test is asymptotically uniformly valid. Part (b) shows that when all the inequalities bind or are sufficiently close to binding, the RCC test does not under-reject asymptotically, assuming the rank of combinations of rows of A does not change in the limit. Part (c) shows an asymptotic IDI property of the RCC test: if some inequalities are very slack, the test reduces to the one based only on the not-very-slack inequalities.

(2) If θ and F are fixed and A and b do not depend on n, then the condition in part (c) is satisfied with J equal to the set of all binding inequalities.[19] If, in addition, $A_J \ne \mathbf{0}$, parts (b) and (c) can be combined to show that the RCC test has exact asymptotic size (and hence is asymptotically non-conservative). By exact asymptotic size, we mean that there exists a sequence of $\theta_n \in \Theta_0(F_n)$ such that the limiting rejection probability is equal to α. (In this case, the sequence is just the fixed sequence with $(\theta_n, F_n) = (\theta, F)$ for all n.) This is compatible with the possibility that other sequences, $\theta_n \in \Theta_0(F_n)$, have limiting rejection probability strictly less than α. Indeed, whenever some moment inequalities are local to binding such that their slackness neither converges to zero nor diverges to infinity, the limiting rejection probability will be less than α.

[18] Note that $I_J A$ is an alternate notation for A_J.

[19] 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.

(3) Theorem 2 combines with (25) to imply that the CC test is asymptotically uniformly valid, and, when the RCC test is asymptotically non-conservative, it can only be conservative to a limited extent:

$$\alpha/2 \leq \limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{\theta\in\Theta_0(F)} \mathbb{E}_F \phi_n^{\mathrm{CC}}(\theta,\alpha) \leq \alpha. \tag{42}$$

(4) The outline of the proof of Theorem 2(a) is conceptually simple. The almost sure representation theorem is invoked on the convergence of the moments, and then Theorem 1 is invoked on the limiting experiment. However, the details are quite complicated. A technical complication that arises is that the rank of the inequalities can be lower in the limit than in the finite sample. This is handled by adding additional inequalities so the sequence of polyhedra defined by the inequalities converges to a limiting polyhedron along a subsequence (see Lemma 7 in the appendix). □

4.3 Finite Sample Validity of the sCC and sRCC Tests

The following result states the finite sample properties of the sRCC test assuming normally distributed moments and a known conditional variance matrix. The result is a corollary of Theorem 1. Let z be a realization of Z, and let $\Theta_0(F_z) = \{\theta \in \Theta : \exists\delta \text{ s.t. } B_z\mathbb{E}_{F_z}[\overline{m}_n(\theta)|z] - C_z\delta \leq d_z\}$.[20] Let e_j denote the $\mathbb{R}^{d_A}$-vector with jth element one and all other elements zero. For any $J \subseteq \{1,\ldots,d_A\}$, let $\phi_{n,J}^{\mathrm{sRCC}}(\theta,\alpha)$ denote the sRCC test defined using $I_J A_z$ and $I_J b_z$ in place of A_z and b_z.

Corollary 1. *Suppose $\Sigma_n(\theta)$ is an invertible matrix such that the conditional distribution of $\sqrt{n}(\overline{m}_n(\theta) - \mathbb{E}_{F_Z}\overline{m}_n(\theta))$ given $Z = z$ is distributed $N(\mathbf{0}, \Sigma_n(\theta))$ and $\widehat{\Sigma}_n(\theta) = \Sigma_n(\theta)$ a.s. for all $\theta \in \Theta$. Then the following hold.*

(a) *For any $\theta \in \Theta_0(F_z)$, $\mathbb{E}_{F_z}[\phi_n^{\mathrm{sRCC}}(\theta,\alpha)|z] \leq \alpha$.*

(b) *If $A_z\mathbb{E}_{F_z}[\overline{m}_n(\theta)|z] = b_z$ and $A_z \neq \mathbf{0}$, then $\mathbb{E}_{F_z}[\phi_n^{\mathrm{sRCC}}(\theta,\alpha)|z] = \alpha$.*

(c) *If $J \subseteq \{1,...,d_A\}$ and $\{\theta_s\}_{s=1}^{\infty} \subseteq \Theta$ is a sequence such that for all $j \notin J$, $e_j'A_z \neq \mathbf{0}$ and $e_j'(A_z\mathbb{E}_{F_z}\overline{m}_n(\theta_s) - b_z)/\|e_j'A_z\| \to -\infty$ as $s \to \infty$, where the dependence of A_z and b_z on s via θ_s is implicit, then*

$$\lim_{s\to\infty} \Pr_{F_z}\left(\phi_n^{\mathrm{sRCC}}(\theta_s,\alpha) \neq \phi_{n,J}^{\mathrm{sRCC}}(\theta_s,\alpha)|z\right) = 0.$$

Remarks. (1) Part (a) shows the finite sample validity of the sRCC test under normality. Part (b) shows that the sRCC test is size-exact when there is an F_z under which all the

[20] Technically, z and the objects that are defined given z, including $\Theta_0(F_z)$, depend on n as well. We keep this dependence implicit for simplicity.

inequalities bind. Part (c) states the IDI property of the sRCC test. Since the sRCC test rejects whenever the sCC test does, the corollary implies the validity of the sCC test: $\mathbb{E}_{F_z}[\phi_n^{\text{sCC}}(\theta,\alpha)|z] \leq \alpha$.

(2) A result on the asymptotic properties of the sRCC test is available in Appendix D. It relies on the asymptotic normality of the moments conditional on $Z_1, ..., Z_n$ and a consistent estimator for $\Sigma_n(\theta)$.

(3) The condition in part (c) 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 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 (c) that depends only on the original inequalities.

(4) 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 $\mathcal{I}(Z_i) = (1\{Z_i = (0,1)'\}, 1\{Z_i = (0,0)'\}, 1\{Z_i = (1,1)'\}, 1\{Z_i = (1,0)'\})'$.

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 paper, 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. □

5 Monte Carlo Simulations

In the previous two sections, we have shown that the CC and RCC tests have a variety of desirable properties, including convenient implementation features and theoretical results on size and adaptation 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

$$E[\theta - W_i] \leq \mathbf{0}, \tag{43}$$

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 $W_i \sim (\mu, \Omega)$, where Ω is a correlation matrix and μ is a mean-vector. 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 Supplemental 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 AQLR (adjusted quasi-likelihood ratio) 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.[21] 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

[21] 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.

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 weighted average power (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 Supplemental Material of AB. Besides WAP, we also report size-corrected WAP, 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 weighted average power, 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 higher ScWAP than the AB test in 4 out of 9 cases in both Table 1 and Table 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

Table 1: Finite Sample Maximum Null Rejection Probabilities and Size-Corrected Average Power of Nominal 5% Tests (normal distribution, known 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	.051	.61	.61	.003	.052	.62	.62	.003	.051	.62	.62	.003
CC	.051	.61	.60	.003	.049	.60	.61	.003	.046	.58	.60	.003
AB	.046	.53	.55	1.11	.051	.59	.59	1.07	.059	.65	.64	1.40
RSW1	.054	.58	.56	.551	.056	.60	.59	.538	.052	.64	.63	.701
RSW2	.050	.23	.23	.014	.052	.34	.34	.013	.052	.50	.49	.014
					$\Omega = \Omega_{\text{Zero}}$							
RCC	.052	.63	.63	.003	.052	.65	.65	.003	.051	.68	.68	.003
CC	.050	.62	.62	.003	.045	.62	.64	.003	.038	.61	.66	.004
AB	.043	.65	.66	1.08	.050	.67	.67	1.07	.056	.69	.67	1.39
RSW1	.053	.61	.60	.545	.056	.63	.62	.539	.052	.65	.65	.699
RSW2	.053	.54	.52	.014	.052	.62	.62	.014	.049	.66	.66	.014
					$\Omega = \Omega_{\text{Pos}}$							
RCC	.051	.76	.75	.003	.053	.75	.74	.003	.051	.72	.71	.003
CC	.038	.72	.76	.003	.033	.68	.74	.003	.032	.62	.69	.003
AB	.042	.78	.80	1.05	.051	.75	.75	1.03	.059	.72	.70	1.34
RSW1	.053	.77	.77	.547	.056	.73	.71	.534	.052	.67	.66	.700
RSW2	.052	.77	.77	.014	.052	.74	.74	.013	.049	.68	.69	.014

Note: CC, RCC, AB, RSW1 and RSW denote the conditional chi-squared test, the refined CC test, the adjusted quasi-likelihood ratio 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.

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	.074	.63	.54	.003	.058	.63	.61	.004	.053	.62	.61	.003
CC	.074	.63	.54	.003	.058	.61	.59	.004	.048	.59	.60	.003
AB	.046	.51	.53	1.23	.049	.58	.58	1.56	.056	.64	.63	1.13
RSW1	.054	.55	.53	.569	.053	.58	.58	.736	.050	.62	.62	.533
RSW2	.053	.24	.23	.026	.051	.34	.33	.026	.051	.49	.48	.023
					$\Omega = \Omega_{\text{Zero}}$							
RCC	.069	.65	.59	.003	.053	.66	.65	.004	.051	.68	.68	.003
CC	.069	.64	.57	.003	.049	.63	.63	.004	.039	.61	.66	.003
AB	.043	.62	.64	1.23	.048	.66	.67	1.55	.053	.68	.67	1.14
RSW1	.052	.58	.58	.570	.053	.62	.61	.732	.050	.64	.64	.541
RSW2	.062	.54	.50	.026	.053	.61	.60	.026	.050	.64	.64	.024
					$\Omega = \Omega_{\text{Pos}}$							
RCC	.056	.77	.75	.003	.054	.75	.74	.004	.051	.71	.71	.003
CC	.043	.73	.74	.003	.034	.68	.73	.004	.035	.63	.69	.003
AB	.044	.78	.79	1.19	.049	.74	.75	1.50	.055	.71	.70	1.07
RSW1	.053	.76	.75	.566	.053	.71	.71	.731	.052	.66	.66	.528
RSW2	.056	.76	.74	.026	.052	.73	.72	.027	.050	.67	.67	.023

Note: Same as Table 1.

Table 3: Finite Sample Maximum Null Rejection Probabilities and Size-Corrected Average Power of Nominal 5% Tests (t(3) distribution, known 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	.057	.61	.58	.003	.054	.62	.60	.003	.046	.62	.64	.003
CC	.057	.60	.57	.003	.053	.60	.58	.003	.043	.59	.62	.003
AB	.036	.54	.58	1.08	.048	.64	.64	1.07	.047	.71	.71	1.40
RSW1	.046	.58	.60	.551	.047	.65	.66	.536	.044	.70	.71	.702
RSW2	.048	.24	.25	.014	.048	.36	.37	.013	.047	.54	.55	.014
					$\Omega = \Omega_{\text{Zero}}$							
RCC	.059	.63	.59	.003	.050	.66	.66	.003	.047	.69	.70	.003
CC	.059	.61	.57	.003	.047	.63	.64	.003	.036	.62	.68	.003
AB	.037	.65	.70	1.07	.046	.72	.73	1.05	.052	.75	.74	1.39
RSW1	.045	.59	.60	.541	.047	.67	.68	.529	.046	.71	.72	.700
RSW2	.046	.49	.52	.014	.043	.63	.65	.013	.046	.69	.71	.014
					$\Omega = \Omega_{\text{Pos}}$							
RCC	.054	.76	.75	.003	.051	.75	.75	.003	.047	.72	.74	.003
CC	.045	.72	.73	.003	.033	.68	.74	.003	.032	.63	.71	.003
AB	.040	.82	.85	1.05	.049	.81	.82	1.04	.052	.78	.78	1.34
RSW1	.050	.80	.80	.547	.047	.78	.79	.537	.047	.73	.74	.700
RSW2	.046	.79	.80	.014	.045	.78	.79	.013	.046	.72	.73	.014

Note: Same as Table 1

Table 4: Finite Sample Maximum Null Rejection Probabilities and Size-Corrected Average Power of Nominal 5% Tests (mixed normal distribution, known 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	.049	.61	.62	.003	.051	.63	.62	.003	.055	.63	.61	.003
CC	.049	.60	.61	.003	.049	.61	.61	.003	.051	.59	.59	.003
AB	.049	.53	.55	1.08	.062	.61	.57	1.07	.069	.65	.61	1.40
RSW1	.055	.58	.56	.550	.063	.62	.59	.539	.063	.64	.60	.701
RSW2	.058	.24	.22	.014	.057	.34	.32	.013	.062	.51	.47	.014
					$\Omega = \Omega_{\text{Zero}}$							
RCC	.049	.64	.64	.003	.047	.66	.67	.003	.054	.70	.68	.003
CC	.046	.62	.64	.003	.043	.63	.65	.003	.041	.62	.66	.003
AB	.052	.66	.65	1.07	.064	.69	.65	1.05	.068	.70	.65	1.39
RSW1	.053	.61	.60	.540	.064	.65	.61	.529	.063	.67	.62	.699
RSW2	.058	.57	.53	.014	.057	.64	.62	.013	.064	.67	.64	.014
					$\Omega = \Omega_{\text{Pos}}$							
RCC	.053	.77	.76	.003	.048	.75	.75	.003	.056	.73	.72	.003
CC	.039	.73	.76	.003	.033	.68	.74	.003	.037	.64	.69	.003
AB	.045	.78	.80	1.04	.060	.76	.74	1.03	.069	.73	.68	1.34
RSW1	.051	.77	.77	.547	.058	.74	.72	.533	.063	.69	.64	.700
RSW2	.058	.78	.75	.014	.057	.74	.73	.013	.062	.70	.66	.014

Note: Same as Table 1.

Table 5: Finite Sample Maximum Null Rejection Probabilities and Size-Corrected Average Power of Nominal 5% Tests (t(3) 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	.067	.70	.64	.003	.053	.69	.69	.004	.053	.69	.68	.003
CC	.067	.69	.64	.003	.053	.68	.67	.004	.046	.66	.67	.003
AB	.042	.56	.57	1.23	.058	.64	.63	1.56	.060	.70	.68	1.12
RSW1	.057	.60	.59	.569	.059	.66	.64	.732	.057	.69	.67	.529
RSW2	.079	.29	.24	.026	.078	.40	.35	.026	.078	.56	.50	.023
					$\Omega = \Omega_{\text{Zero}}$							
RCC	.063	.72	.67	.003	.051	.73	.72	.004	.056	.74	.73	.002
CC	.063	.70	.66	.003	.048	.70	.70	.004	.041	.68	.71	.003
AB	.043	.67	.69	1.22	.056	.72	.71	1.55	.066	.74	.69	1.11
RSW1	.057	.64	.63	.566	.059	.69	.67	.730	.057	.71	.68	.528
RSW2	.093	.67	.57	.026	.085	.71	.63	.026	.085	.72	.65	.023
					$\Omega = \Omega_{\text{Pos}}$							
RCC	.053	.80	.79	.003	.052	.79	.79	.004	.053	.77	.75	.002
CC	.043	.77	.78	.003	.037	.73	.78	.004	.032	.69	.75	.003
AB	.050	.80	.80	1.19	.058	.79	.77	1.50	.064	.77	.72	1.07
RSW1	.059	.79	.77	.565	.059	.77	.75	.728	.057	.73	.71	.528
RSW2	.082	.79	.73	.026	.078	.78	.71	.026	.082	.74	.67	.023

Note: Same as Table 1.

Table 6: Finite Sample Maximum Null Rejection Probabilities and Size-Corrected Average Power of Nominal 5% Tests (mixed 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	.096	.64	.47	.003	.067	.63	.57	.004	.066	.63	.58	.003
CC	.096	.63	.46	.003	.066	.61	.56	.004	.063	.60	.56	.003
AB	.047	.49	.50	1.23	.053	.56	.53	1.56	.055	.62	.61	1.12
RSW1	.046	.52	.53	.568	.055	.56	.55	.732	.053	.60	.60	.529
RSW2	.060	.24	.22	.026	.055	.33	.31	.026	.060	.48	.46	.023
					$\Omega = \Omega_{\text{Zero}}$							
RCC	.096	.67	.51	.003	.074	.67	.60	.004	.066	.69	.64	.003
CC	.096	.65	.49	.003	.070	.64	.57	.004	.055	.62	.61	.003
AB	.046	.58	.59	1.22	.056	.64	.61	1.54	.053	.66	.65	1.11
RSW1	.045	.54	.57	.565	.055	.60	.58	.728	.053	.62	.61	.528
RSW2	.086	.52	.41	.026	.072	.58	.51	.026	.067	.61	.56	.023
					$\Omega = \Omega_{\text{Pos}}$							
RCC	.065	.77	.72	.003	.061	.75	.71	.004	.068	.72	.66	.002
CC	.053	.73	.72	.003	.044	.68	.71	.004	.050	.63	.64	.003
AB	.039	.75	.80	1.19	.046	.73	.75	1.50	.055	.70	.68	1.07
RSW1	.045	.73	.76	.565	.046	.70	.72	.728	.053	.65	.64	.527
RSW2	.066	.74	.68	.026	.059	.70	.67	.026	.063	.65	.60	.023

Note: Same as Table 1.

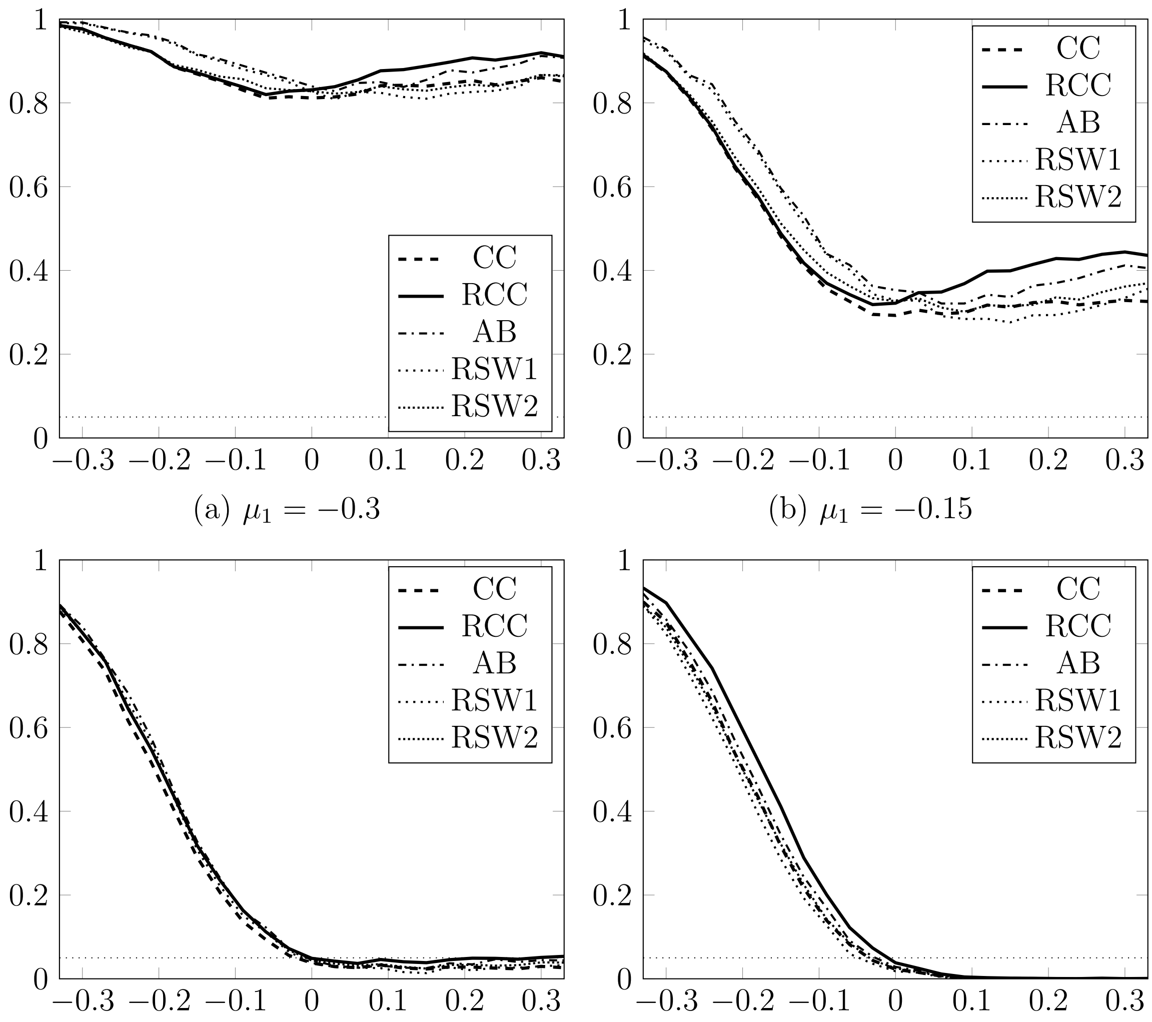
draws while the RSW tests using 499 bootstrap draws. Increasing the number of bootstrap draws would increase their computational costs proportionally.

The theoretical properties of the CC tests rely in an essential way on the normality or asymptotic normality of the moments. Thus, we report results when W_i is not normal to investigate the sensitivity of these simulations to the data distribution. Tables 3 and 5 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. Tables 4 and 6 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.

Tables 5 and 6 show the results for the tests when Ω is estimated. The size performance of the RCC test varies somewhat with the data distribution. It has worse MNRPs under the skewed mixed normal distribution (up to 9.6% for $k = 10$) than under normality (Table 6), while it has better MNRPs under the $t(3)$ distribution than under normality (Table 5). It is noteworthy that the bootstrap-based AB, RSW1, and RSW2 tests have lower worst case MNRP's across Tables 3-6 (6.9%, 6.4%, 9.3%, respectively, compared to 9.6% for the RCC test). This may reflect a type of refinement property for the bootstrap, but more investigation is needed. It also is interesting to note that the over-rejection either disappears or is greatly reduced when the true Ω is used, as shown in Tables 3 and 4. This seems to suggest that the non-normality of the sample moments is less of an issue than estimating the variance matrix in a relatively small sample.

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.

Figure 2: Power Curves for 5 Tests ($k = 2$, normal distribution, $\Omega = \Omega_{\text{Zero}}$, Estimated Ω, $n = 100$, $\alpha = 5\%$)



(a) $\mu_1 = -0.3$

(b) $\mu_1 = -0.15$

(c) $\mu_1 = 0$

(d) $\mu_1 = 0.15$

Note: CC denotes the conditional chi-squared test, RCC denotes the refined CC test, AB denotes the adjusted quasi-likelihood ratio (AQLR) test with bootstrap critical value in AB, and RSW1 and RSW2 denote the 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.

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 7 reports the results for four pairs of (k, n): $(10, 100)$, $(50, 700)$, $(100, 1600)$, and $(150, 2550)$, where the pairs are chosen so that k is approximately proportional to $n/\log(n)$. The message from Table 7 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.

None of these tests have 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. Intuitively, if the moments are approximately normal, in some sense, then one can appeal to Theorem 1 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.

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

$$\psi(Y_i^*, X_i, \theta) = \log(s_i^*/(1 - s_i^*)) - X_i'\theta. \tag{44}$$

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.

Table 7: Finite Sample Maximum Null Rejection Probabilities of Nominal 5% Tests, the Large k and n Cases (Estimated Ω, $\Omega = \Omega_{\text{zero}}$)

	$k=10, n=100$		$k=50, n=700$		$k=100, n=1600$		$k=150, n=2550$	
Test	MNRP	Time	MNRP	Time	MNRP	Time	MNRP	Time
				normal				
RCC	.069	.003	.074	.004	.076	.011	.081	.024
CC	.069	.003	.074	.005	.076	.011	.081	.024
RSW1	.052	.582	.062	1.07	.047	3.25	.051	7.96
RSW2	.062	.027	.056	.211	.048	1.15	.045	4.04
				$t(3)$				
RCC	.063	.003	.069	.004	.071	.010	.079	.024
CC	.063	.003	.069	.005	.071	.011	.079	.024
RSW1	.057	.568	.054	1.07	.050	3.23	.054	7.97
RSW2	.093	.026	.069	.210	.061	1.24	.063	4.05
				mixed normal				
RCC	.096	.003	.090	.004	.089	.010	.089	.024
CC	.096	.003	.090	.005	.089	.010	.089	.024
RSW1	.045	.567	.051	1.06	.059	3.20	.054	7.92
RSW2	.086	.026	.069	.210	.065	1.24	.058	4.04

Note: CC denotes the conditional chi-squared test, RCC denotes the refined CC test, and RSW1 and RSW2 denote the two-step test in RSW based on the QLR statistic and the Max statistic, respectively. MNRP denotes maximum null rejection probability, and Time denotes 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.

Table 8: 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, -.339]	.989	2.5	[-1.609 ,-.440]	.723	2.4
sCC	[-1.780, -.332]	1.00	2.5	[-1.615, -.433]	.736	2.4
ARP Hybrid	[-1.998, -.264]	1.29	111	[-1.736, -.395]	.895	109
			$d_c=3$, 16 moment inequalities			
sRCC	[-1.852, -.293]	1.11	7.1	[-1.659, -.404]	.809	4.7
sCC	[-1.852, -.293]	1.11	7.1	[-1.659, -.404]	.809	4.7
ARP Hybrid	[-2.219, -.123]	1.65	199	[-1.883, -.287]	1.15	120
			$d_c=4$, 32 moment inequalities			
sRCC	[-1.921, -.254]	1.22	6.5	[-1.718, -.366]	.906	10
sCC	[-1.921, -.254]	1.22	6.5	[-1.718, -.366]	.906	10
ARP Hybrid	[-2.596, -.011]	2.14	97	[-2.104, -.180]	1.48	145

Note: The identified set for θ_0 is $[-1.203, -.757]$. 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.

(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):

$$\begin{aligned} \psi_i^U(\theta) &= \log(s_{N,i} + 2/N) - \log(1 - s_{N,i} + \underline{s}) - X_i'\theta \\ \psi_i^L(\theta) &= \log(s_{N,i} + \underline{s}) - \log(1 - s_{N,i} + 2/N) - X_i'\theta, \end{aligned} \tag{45}$$

where $\underline{s}$ is the smaller of 0.05 and half of the minimum possible value of $\min(s_i^*, 1 - s_i^*)$.[22] We assume that $\underline{s}$ is known and refer the reader to Gandhi et al. (2019) for practical recommendations regarding $\underline{s}$.

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 Appendix E.1.

For instrumental functions, we use

$$\mathcal{I}(Z_i) = (1\{(Z_{ei}, Z_{c,2,i}, \ldots, Z_{c,d_c,i}) = z\})_{z \in \{0,1\}^{d_c}}. \tag{46}$$

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 confidence interval 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. The details for computing the confidence intervals are given in Supplemental

[22] 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.

Appendix E.2.

Table 8 reports the average confidence interval (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 compute 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 confidence intervals 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.

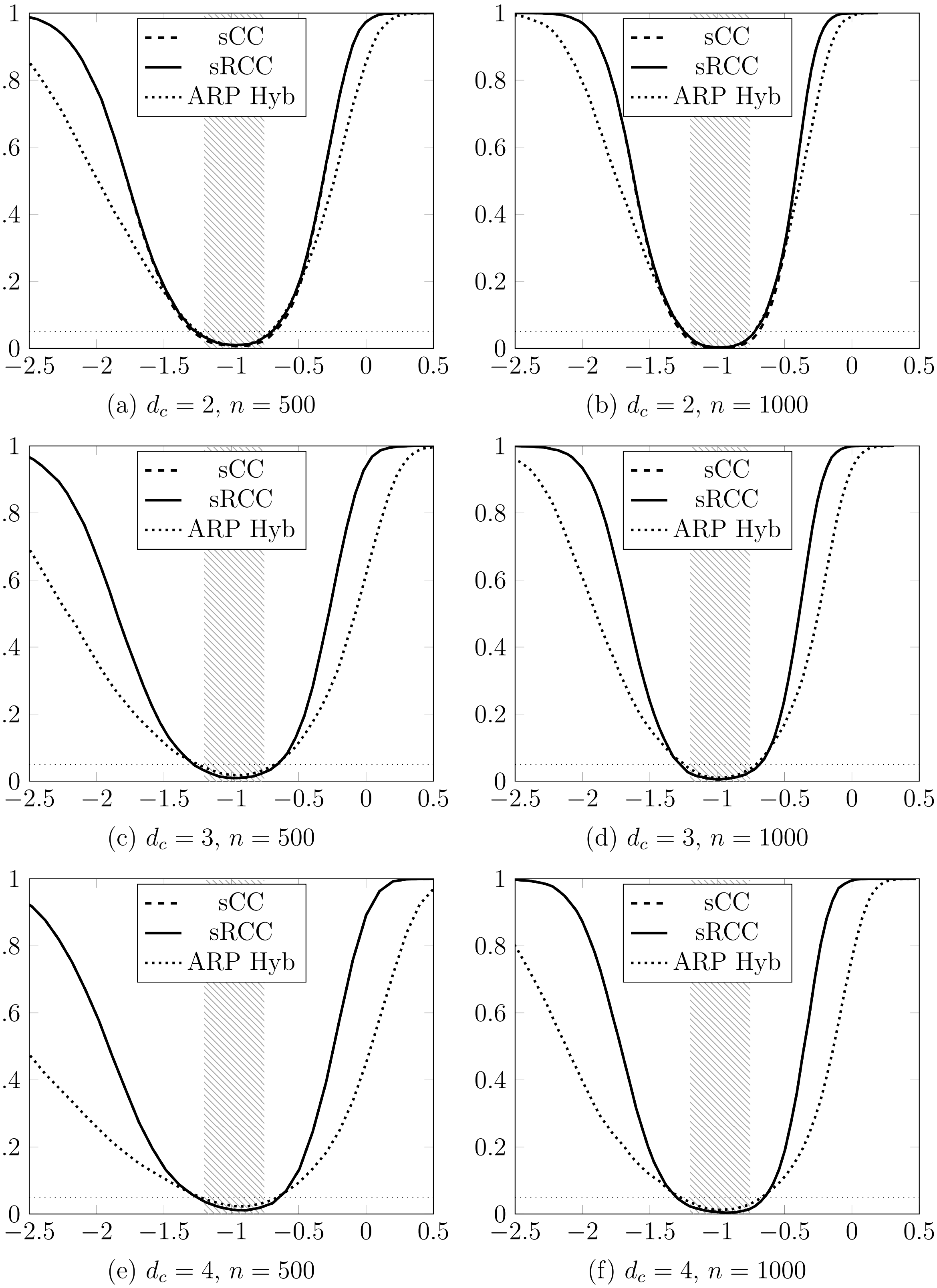
Figure 3 reports the rejection rates of the tests for $H_0 : \theta_0 = \theta$ plotted against θ values in $[-2.5, 0.5]$.[23] The shaded area indicates the identified set for θ_0. As we can see, the rejection rates for the points in the identified set are less than or equal to 5% in all cases. For $d_c = 3$ and $d_c = 4$, all tests shows 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 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 8. 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 paper proposes the refined conditional chi-squared (RCC) 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

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

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%.



(a) $d_c = 2$, $n = 500$

(b) $d_c = 2$, $n = 1000$

(c) $d_c = 3$, $n = 500$

(d) $d_c = 3$, $n = 1000$

(e) $d_c = 4$, $n = 500$

(f) $d_c = 4$, $n = 1000$

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 RCC and subvector RCC tests have a computational advantage over alternatives while being competitive in terms of size and power.

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.

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*. Springer-Verlag Berlin, Heidelberg.

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 B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson (Eds) *Advances in Economics and Econometrics*: Volume 2: Eleventh World Congress, (Econometric Society Monographs, pp. 271-306). Cambridge University Press.

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.

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*, pages 317–319.

Gale, D. (1960). *The Theory of Linear Economic Models*. McGraw-Hill Book Company, 1 edition.

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

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 B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson (Eds) *Advances in Economics and Econometrics*: Volume 2: Eleventh World Congress, (Econometric Society Monographs, pp. 307-359). Cambridge University Press.

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 June 9, 2020.

Kudo, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, 50:403–418.

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). Econometrics with partial identification. In S. Durlauf, L. Hansen, J. Heckman, and R. Matzkin (Eds) *Handbook of Econometrics*: Volume 7A, 1st Edition. 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*. CRC Press, Taylor & Francis Group, 3 edition.

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *econometrica*, 77:1299–1315.

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.

Wolak, F. (1987). An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association*, 82:782–793.

Wollman, T. G. (2018). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, 108:1364–1406.

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

Gregory Cox and Xiaoxia Shi

This supplemental appendix contains proofs and other supporting materials for "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models" (henceforth referred to as the main text) by Gregory Cox and Xiaoxia Shi. The following sections are included:

- Section A contains the proof of Theorem 1 in the main text.
- Section B contains the proof of Theorem 2 in the main text. This section also includes Theorem 3, a general theorem for uniform asymptotic properties of the CC and RCC tests, as well as the proof of Theorem 3. Theorem 3 is used to prove Theorem 2.
- Section C contains supporting materials for Section 3.2 in the main text. This section includes lemmas that reduce the calculation of the sCC test to a rank calculation problem. It also includes an algorithm to carry out the rank calculation in the case not covered in the main text.
- Section D proves the asymptotic validity of the Subvector Tests by verifying the conditional asymptotic normality of the sample moments and the consistency of the two conditional variance matrix estimators proposed in Section 3.2 in the main text.
- Section E provides details for the identified set calculation and the confidence interval calculation in Section 5.2 in the main text.

A Proof of Theorem 1

For this proof, we assume $\Sigma_n(\theta) = nI_{d_m}$. If this is not the case, then the following proof can be applied after premultiplying $\overline{m}_n(\theta)$ by $n^{1/2}\Sigma_n(\theta)^{-1/2}$ and postmultiplying A by $n^{-1/2}\Sigma_n(\theta)^{1/2}$.

Fix θ and let $X = \overline{m}_n(\theta) \sim N(\mu, I_{d_m})$, where $\mu = \mathbb{E}_F\overline{m}_n(\theta)$. Let $C = \{\mu \in \mathbb{R}^{d_m} | A\mu \leq b\}$. The fact that $\theta \in \Theta_0(F)$ implies $\mu \in C$. These simplifications imply that $T_n(\theta) = \|X - \hat{\mu}\|^2$ and

$$\hat{\tau}_j = \begin{cases} \frac{\|a_1\|(b_j - a_j'\hat{\mu})}{\|a_1\|\|a_j\| - a_1'a_j} & \text{if } \|a_1\|\|a_j\| \neq a_1'a_j \\ \infty & \text{otherwise} \end{cases}. \tag{47}$$

The definitions of $\hat{\mu}$, $\widehat{J}$, $\hat{r}$, $\hat{\tau}$, and $\hat{\beta}$ are unchanged. $\hat{\mu}$ is the projection of X onto C. We also denote it by $P_C X$. We also denote $\widehat{J}$ by $J(X)$, $\hat{r}$ by $r(X)$, $\hat{\tau}$ by $\tau(X)$, and $\hat{\beta}$ by $\beta(X)$.

A.1 Auxiliary Lemmas

The proof of Theorem 1 relies on four lemmas.

The first lemma partitions $\mathbb{R}^{d_m}$ according to which inequalities are active. We define some notation for the partition. For any $J \subseteq \{1, ..., d_A\}$, let $J^c = \{1, ..., d_A\}/J$, and let $C_J = \{x \in C : \forall j \in J, a_j' x = b_j, \text{ and } \forall j \in J^c, a_j' x < b_j\}$. Then C_J forms a partition of C. Also let $V_J = \{\sum_{j \in J} v_j a_j : v_j \in \mathbb{R}, v_j \geq 0\}$, and let $K_J = C_J + V_J$.[24] The following lemma shows that K_J forms a partition that characterizes which inequalities are active.

Lemma 1. (a) *If $X \in K_J$, then $X - P_C X \in V_J$ and $P_C X \in C_J$.*

(b) *The set of all K_J for $J \subseteq \{1, ..., d_A\}$ is a partition of $\mathbb{R}^{d_m}$.*

(c) *For every $J \subseteq \{1, ..., d_A\}$, $X \in K_J$ iff $J = J(X)$.*

The next lemma considers the event $\hat{r} = 0$ and partitions that event according to which face of C is closest to the realization of X. Let $J_0 = \{j \in \{1, ..., d_A\} | a_j = 0\}$ and let $J_{00} = \{j \in \{1, ..., d_A\} | a_j = \mathbf{0} \text{ and } b_j = 0\}$. Also let

$$\begin{aligned} \mathcal{J}_1 = & \{J \subseteq \{1, ..., d_A\} | \text{rk}(A_J) = 1, J \cap J_0 = J_{00}, \\ & \text{and if } j \in J, \ell \in J^c, \text{ s.t. } \|a_j\| > 0, \|a_\ell\| > 0, \\ & \text{then } \frac{a_j}{\|a_j\|} \neq \frac{a_\ell}{\|a_\ell\|} \text{ or } \frac{b_j}{\|a_j\|} \neq \frac{b_\ell}{\|a_\ell\|}\}. \end{aligned} \tag{48}$$

Further subdivide

$$\begin{aligned} \mathcal{J}_1^{os} = & \{J \in \mathcal{J}_1 | \text{ if } j, \ell \in J \text{ s.t. } \|a_j\| > 0, \|a_\ell\| > 0, \text{ then } \frac{a_j}{\|a_j\|} = \frac{a_\ell}{\|a_\ell\|}\} \\ \mathcal{J}_1^{ts} = & \{J \in \mathcal{J}_1 | \exists j, \ell \in J \text{ s.t. } \|a_j\| > 0, \|a_\ell\| > 0, \frac{a_j}{\|a_j\|} = -\frac{a_\ell}{\|a_\ell\|} \text{ and } \frac{b_j}{\|a_j\|} = -\frac{b_\ell}{\|a_\ell\|}\}. \end{aligned} \tag{49}$$

The next lemma provides a partition of $C_0 := \cup_{J \subseteq \{1,...,d_A\}: \text{rk}(A_J) = 0} C_J$. (Note that for these sets, $C_J = K_J$.) Let $J_{\neq 0} = \{j = 1, \ldots, d_A : \|a_j\| \neq 0\}$. For each $J \in \mathcal{J}_1^{os}$, let

$$C_J^{\Delta} = \{x \in C_0 | \text{argmin}_{j \in J_{\neq 0}} \|a_j\|^{-1}(b_j - a_j' x) = J \cap J_{\neq 0}\}. \tag{50}$$

[24] When $J = \emptyset$, then $V_J = \{\mathbf{0}_{d_m}\}$.

The set C_J^{Δ} is the set of points in C that are closer to C_J than to any other $C_{\tilde{J}}$ for $\tilde{J} \in \mathcal{J}_1$. It is helpful to picture C_J for $J \in \mathcal{J}_1$ as the faces of a polyhedron, C, and C_J^{Δ} as a partition of C into triangularly shaped sets. Also let

$$C^{|} = C_0 / \left(\cup_{J \in \mathcal{J}_1^{os}} C_J^{\Delta} \right). \tag{51}$$

Lemma 2. (a) $C_0 = C_{J_{00}}$.

(b) *The sets $C^{|}$ and C_J^{Δ} for $J \in \mathcal{J}_1^{os}$ form a partition of C_0.*

(c) *If $A \neq \mathbf{0}_{d_A \times d_m}$, then $C^{|}$ has Lebesgue measure zero.*

(d) $\cup_{J \subseteq \{1,\ldots,d_A\} | \mathrm{rk}(A_J)=1} K_J = \cup_{J \in \mathcal{J}_1^{os} \cup \mathcal{J}_1^{ts}} K_J$.

The next lemma bounds the probabilities of translations of sets in the multivariate normal distribution. Let V denote an arbitrary cone in $\mathbb{R}^r$ for a positive integer r.[25] Let V^* denote the polar cone. That is, $V^* = \{\gamma \in \mathbb{R}^r | \langle y, \gamma \rangle \leq 0 \text{ for all } y \in V\}$. For any $\gamma \in V^*$, let $Y \sim N(\gamma, I_r)$. The following lemma provides a property of probabilities of cones under a translation.

Lemma 3. *For every $\gamma \in V^*$, $\Pr_\gamma(\|Y\|^2 > \chi^2_{r,1-\alpha} | Y \in V) \leq \alpha$, with equality if $\gamma = \mathbf{0}$.*

Lemma 3 states that the probability that a random vector, Y, belongs to the tail of its distribution, conditional on belonging to the cone, V, is less than or equal to α, where the tail is any point outside a sphere of radius $\sqrt{\chi^2_{r,1-\alpha}}$. The key assumption is that the mean of Y must belong to the polar cone, V^*, which translates the distribution away from the cone, V. When $\gamma = 0$, this lemma holds with equality because unconditionally $\|Y\|^2 \sim \chi^2_r$, the tail of which has mass exactly α, and because $\|Y\|^2$ has exactly the same distribution whether or not we condition on $Y \in V$. Lemma 3 follows from Lemma 1 in Mohamad et al. (2020), and thus the proof is omitted.

The following lemma is the key to validity of the refinement to the CC test. It is a bound on translations of sets in the univariate normal distribution.

Lemma 4. *For every $\mu \leq 0$, for every $\tau \geq 0$, and for every $\alpha \in [0, 1]$,*

$$Pr_\mu \left(Z > z_{1-\beta/2} | Z > -\tau \right) \leq \alpha,$$

where $Z \sim N(\mu, 1)$ and $\beta = 2\alpha\Phi(\tau)$, with equality if $\mu = 0$.

[25] A cone is a set, V, such that for all $v \in V$ and for all $\lambda \geq 0$, $\lambda v \in V$.

A.2 Proof of Theorem 1

First, we show part (a). Notice that

$$
\begin{aligned}
&\Pr(\|X - P_C X\|^2 > \chi^2_{\mathrm{rk}(A_{J(X)}),1-\beta(X)}) \\
&= \sum_{J\subseteq\{1,\ldots,d_A\}} \Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{\mathrm{rk}(A_J),1-\beta(X)}) \\
&= \sum_{J\subseteq\{1,\ldots,d_A\}|\mathrm{rk}(A_J)\geq 2} \Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{\mathrm{rk}(A_J),1-\alpha}) && (52)\\
&\quad + \sum_{J\in\mathcal{J}_1^{ts}} \Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{1,1-\alpha}) && (53)\\
&\quad + \sum_{J\in\mathcal{J}_1^{os}} \Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{1,1-\beta(X)}) && (54)\\
&\quad + \sum_{J\subseteq\{1,\ldots,d_A\}|\mathrm{rk}(A_J)=0} \Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{0,1-\alpha}), && (55)
\end{aligned}
$$

where the first equality follows from Lemma 1(b,c), and the second equality uses Lemma 2(d) and the fact that $\beta(X) = \alpha$ whenever $\mathrm{rk}(A_{J(X)}) \neq 1$ or $J \in \mathcal{J}_1^{ts}$. That latter fact follows because for $J \in \mathcal{J}_1^{ts}$ with $X \in K_J$, there exists $j, \ell \in J$ such that $\|a_\ell\|^{-1} a_\ell = -\|a_j\|^{-1} a_j$ and $\|a_\ell\|^{-1} b_\ell = -\|a_j\|^{-1} b_j$, which implies that $b_\ell - a_\ell' P_C X = b_j - a_j' P_C X = 0$ (and therefore $\tau(X) = 0$).

For each J, we consider the span of V_J as a subspace of $\mathbb{R}^{d_m}$. Let P_J denote the projection onto $\mathrm{span}(V_J)$, and M_J denote the projection onto its orthogonal complement. We note that, given J, there exists a $\kappa_J \in \mathrm{span}(V_J)$ such that for every $z \in C_J$, $P_J z = \kappa_J$. This follows because for two $z_1, z_2 \in C_J$, and for any $v \in \mathrm{span}(V_J)$, $\langle z_1 - z_2, v\rangle = 0$, which implies $z_1 - z_2 \perp \mathrm{span}(V_J)$, so that $P_J(z_1 - z_2) = \mathbf{0}_{d_m}$. Thus, for any $X \in K_J$, we can write $P_J X = P_J(X - P_C X) + P_J P_C X = X - P_C X + \kappa_J$, where the second equality follows by Lemma 1(a) and the above discussion. We also write $M_J X = X - P_J X = P_C X - \kappa_J$.

First, let's consider the terms in (55). For J such that $\mathrm{rk}(A_J) = 0$, we have $\mathrm{span}(V_J) = \{\mathbf{0}_{d_m}\}$. Thus, $P_J X = \kappa_J = \mathbf{0}_{d_m}$. This implies that $\|X - P_C X\| = 0$. Therefore,

$$\Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{0,1-\alpha}) = 0. \tag{56}$$

For J such that $\mathrm{rk}(A_J) > 0$, we define a linear isometry from $\mathrm{span}(V_J)$ to $\mathbb{R}^{\mathrm{rk}(A_J)}$. Let B_J be a $d_m \times \mathrm{rk}(A_J)$ matrix whose columns form a basis for $\mathrm{span}(V_J)$. Then $P_J X = B_J(B_J' B_J)^{-1} B_J' X$. The projection matrix $B_J(B_J' B_J)^{-1} B_J'$ is idempotent with rank $\mathrm{rk}(A_J)$, and thus there exists a $d_m \times \mathrm{rk}(A_J)$ matrix with orthonormal columns, Q_J, such that $Q_J Q_J' = B_J(B_J' B_J)^{-1} B_J'$. The linear isometry from $\mathrm{span}(V_J)$ to $\mathbb{R}^{\mathrm{rk}(A_J)}$ is $Q_J(X) = Q_J' X$. This is an isometry because for any $v_1, v_2 \in \text{span}(V_J)$,

$$\begin{aligned}\|v_1 - v_2\|^2 &= (v_1 - v_2)'(v_1 - v_2) \\ &= (v_1 - v_2)'(P_J(v_1 - v_2)) \\ &= (v_1 - v_2)'Q_J Q_J'(v_1 - v_2) \\ &= \|Q_J(v_1) - Q_J(v_2)\|^2, \end{aligned} \tag{57}$$

where the second equality holds because $v_1, v_2 \in \text{span}(V_J)$. Now let $Q_J'V_J = \{Q_J'v : v \in V_J\}$. Then $P_J X - \kappa_J \in V_J$ if and only if $Q_J'(P_J X - \kappa_J) \in Q_J'V_J$ because this isometry is bijective.

Next, we consider the terms in (52) and (53). Notice that

$$\begin{aligned}&\Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{\text{rk}(A_J),1-\alpha}) \\ =&\Pr(M_J X + \kappa_J \in C_J, P_J X - \kappa_J \in V_J, \text{ and } \|P_J X - \kappa_J\|^2 > \chi^2_{\text{rk}(A_J),1-\alpha}) \\ =&\Pr(M_J X + \kappa_J \in C_J) \times \Pr(P_J X - \kappa_J \in V_J \text{ and } \|P_J X - \kappa_J\|^2 > \chi^2_{\text{rk}(A_J),1-\alpha}), \end{aligned} \tag{58}$$

where the first equality uses Lemma 1(a) and the facts that $M_J X + \kappa_J = P_C X$ and $X = P_J X + M_J X$, and the second equality follows from the fact that $P_J X$ is independent of $M_J X$. Applying the isometry, we have

$$\begin{aligned}&\Pr(P_J X - \kappa_J \in V_J \text{ and } \|P_J X - \kappa_J\|^2 > \chi^2_{\text{rk}(A_J),1-\alpha}) \\ =& \Pr(Q_J'(P_J X - \kappa_J) \in Q_J'V_J \text{ and } \|Q_J'(P_J X - \kappa_J)\|^2 > \chi^2_{\text{rk}(A_J),1-\alpha}). \end{aligned} \tag{59}$$

We would like to apply Lemma 3 to this probability. Since $X \sim N(\mu, I)$, we have

$$Q_J'(P_J X - \kappa_J) \sim N(Q_J'(P_J\mu - \kappa_J), Q_J' I Q_J) = N(Q_J'(P_J\mu - \kappa_J), I). \tag{60}$$

Also note that $Q_J'V_J$ is a cone in $\mathbb{R}^{\text{rk}(A_J)}$. The random vector $Q_J'(P_J X - \kappa_J) \sim N(\gamma, I)$ where $\gamma = Q_J'(P_J\mu - \kappa_J)$. The vector γ is in the polar cone because, for all $\tilde{y} \in Q_J'V_J$, there exists a $y = \sum_{j\in J} v_j a_j \in V_J$ such that $\tilde{y} = Q_J'y$, and thus

$$\begin{aligned}\langle \gamma, \tilde{y}\rangle &= \langle Q_J'(P_J\mu - \kappa_J), Q_J'y\rangle \\ &= \langle P_J\mu - \kappa_J, y\rangle \\ &= \langle (\mu - M_J\mu - P_J z), y\rangle \\ &= \langle (\mu - M_J\mu - z + M_J z), y\rangle \\ &= \langle (\mu - z), y\rangle \end{aligned}$$

$$= \sum_{j \in J} v_j(\langle \mu, a_j \rangle - \langle z, a_j \rangle) \leq 0, \tag{61}$$

where z is any element[26] of C_J so that $\kappa_J = P_J z$, the second equality holds because $\langle Q'_J(P_J\mu - \kappa_J), Q'_J y\rangle = y'Q_J Q'_J(P_J\mu - \kappa_J) = y'P_J(P_J\mu - \kappa_J) = y'(P_J\mu - \kappa_J)$, the fifth equality holds because $M_J y = 0$, and the inequality follows because $\langle z, a_j\rangle = b_j \geq \langle \mu, a_j\rangle$, using the facts that $z \in C_J$ and $\mu \in C$.

Therefore, we can apply Lemma 3 to get that, for every $J \subseteq \{1, ..., d_A\}$ such that $\text{rk}(A_J) \geq 1$, we have

$$\begin{aligned}
&\Pr(P_J X - \kappa_J \in V_J \text{ and } \|P_J X - \kappa_J\|^2 > \chi^2_{\text{rk}(A_J),1-\alpha}) \\
&= Pr\left(Q'_J(P_J X - \kappa_J) \in Q'_J V_J \text{ and } \|Q'_J(P_J X - \kappa_J)\|^2 > \chi^2_{\text{rk}(A_J),1-\alpha}\right) \\
&\leq \alpha \Pr(Q'_J(P_J X - \kappa_J) \in Q'_J V_J) \\
&= \alpha \Pr(P_J X - \kappa_J \in V_J),
\end{aligned} \tag{62}$$

where the inequality holds as equality if $\gamma = Q'_J(P_J\mu - \kappa_J) = 0$.

Next, consider the terms in (54). For each $J \in \mathcal{J}_1^{os}$, let $\bar{j} \in J$ such that $\|a_{\bar{j}}\| \neq 0$. Notice that we can take $B_J = a_{\bar{j}}$, so that $P_J = \|a_{\bar{j}}\|^{-2} a_{\bar{j}} a'_{\bar{j}}$, $Q_J = \|a_{\bar{j}}\|^{-1} a_{\bar{j}}$, and $\kappa_J = \|a_{\bar{j}}\|^{-2} a_{\bar{j}} b_{\bar{j}}$. Notice that

$$\begin{aligned}
&\Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{1,1-\beta(X)}) \\
&= \Pr(M_J X + \kappa_J \in C_J, P_J X - \kappa_J \in V_J, \text{ and } \|P_J X - \kappa_J\|^2 > \chi^2_{1,1-\beta(X)}) \qquad (63) \\
&= \Pr(M_J X + \kappa_J \in C_J, Q'_J(P_J X - \kappa_J) \in Q'_J V_J, \text{ and } \|Q'_J(P_J X - \kappa_J)\|^2 > \chi^2_{1,1-\beta(M_J X + \kappa_J)}),
\end{aligned}$$

where the first equality uses the definitions of M_J, P_J, and κ_J, and the second equality uses the isometry Q_J, together with the fact that $\beta(X)$ depends on X only through $M_J X + \kappa_J$ (because the formula for $\tau(X)$ only depends on $P_C X = M_J X + \kappa_J$).

We next note that $Q'_J V_J = [0, \infty)$. This follows because for any $c \geq 0$, $c = Q'_J c a_{\bar{j}} \|a_{\bar{j}}\|$, where $c a_{\bar{j}} \|a_{\bar{j}}\| \in V_J$. Conversely, for any $v = \sum_{\ell \in J} c_\ell a_\ell \in V_J$ for some constants $c_\ell \geq 0$, we have $Q'_J v = \sum_{\ell \in J} c_\ell \|a_{\bar{j}}\|^{-1} a'_{\bar{j}} a_\ell$, where $a'_{\bar{j}} a_\ell \geq 0$ because a_ℓ is either zero or a positive scalar multiple of $a_{\bar{j}}$ by the definition of $\mathcal{J}_1^{os}$.

Also, the fact that $X \sim N(\mu, I)$ implies that $Z := Q'_J(P_J X - \kappa_J) \sim N(\gamma, 1)$, where $\gamma = Q'_J(P_J\mu - \kappa_J) = \|a_{\bar{j}}\|^{-1}(a'_{\bar{j}}\mu - b_{\bar{j}}) \leq 0$. Note that Z is independent of $M_J X$.

[26] If C_J is empty, so that no such z exists, then (58) is zero, and so (62) below is not needed.

Let $z_{1-\alpha}$ denote the $1-\alpha$ quantile of the standard normal distribution. We have

$$\begin{aligned}
&\Pr(M_JX+\kappa_J \in C_J, Q_J'(P_JX-\kappa_J)\in Q_J'V_J, \text{ and } \|Q_J'(P_JX-\kappa_J)\|^2 > \chi^2_{1,1-\beta(M_JX+\kappa_J)}) \\
=&\Pr(M_JX+\kappa_J \in C_J, Z > z_{1-\beta(M_JX+\kappa_J)/2}) \\
=&\mathbb{E}\mathbb{1}(M_JX+\kappa_J\in C_J)\Pr(Z > z_{1-\beta(M_JX+\kappa_J)/2}|M_JX+\kappa_J) \\
\leq&\alpha\mathbb{E}\mathbb{1}(M_JX+\kappa_J\in C_J)\Pr(Z > -\tau(M_JX+\kappa_J)|M_JX+\kappa_J) \qquad (64)\\
=&\alpha\Pr(M_JX+\kappa_J\in C_J, Z>-\tau(M_JX+\kappa_J)) \\
=&\alpha\Pr(M_JX+\kappa_J\in C_J, Q_J'(P_JX-\kappa_J)\in Q_J'V_J) \\
&+\alpha\Pr(M_JX+\kappa_J\in C_J, Q_J'(P_JX-\kappa_J)\in(-\tau(M_JX+\kappa_J),0)) \\
=&\alpha(\Pr(X\in K_J)+\Pr(X\in C_J^\Delta)), \qquad (65)
\end{aligned}$$

where the first equality follows from the events $Z \geq 0$ and $Z^2 > \chi^2_{1,1-\beta(M_JX+\kappa_J)}$ being equivalent to the event $Z > z_{1-\beta(M_JX+\kappa_J)/2}$, the second and third equalities uses the conditional distribution of Z given $M_JX+\kappa_J$, the inequality follows by Lemma 4, the fourth equality follows from splitting the event $Z > -\tau$ into $Z \geq 0$ (equivalent to $Z \in Q_J'V_J$) and $Z \in (-\tau, 0)$, and the final equality follows from the fact that $Q_J'(P_JX-\kappa_J) \in Q_J'V_J$ if and only if $P_JX - \kappa_J \in V_J$, the characterization of K_J using Lemma 1(a), together with the argument that follows.

To show (65), we show that for all $J \in \mathcal{J}_1^{os}$,

$$C_J^\Delta = \{x \in \mathbb{R}^{d_m} | M_Jx+\kappa_J \in C_J \text{ and } Q_J'(P_Jx-\kappa_J)\in(-\tau(M_Jx+\kappa_J),0)\}. \qquad (66)$$

Denote the set on the right hand side of (66) by Υ. We show (1) $x \in C_J^\Delta$ implies $x \in \Upsilon$ and (2) $x \in \Upsilon$ implies $x \in C_J^\Delta$. It is useful to point out that for any x, we can write $Q_J'(P_Jx-\kappa_J) = \|a_{\bar{j}}\|^{-1}(a_{\bar{j}}'x - b_{\bar{j}})$ and $M_Jx+\kappa_J = x - \|a_{\bar{j}}\|^{-2}a_{\bar{j}}(a_{\bar{j}}'x-b_{\bar{j}})$ using the formulas for P_J, Q_J, and κ_J.

(1) Let $x \in C_J^\Delta$. We calculate that $M_Jx+\kappa_J \in C_J$ by showing that equality holds for every $\ell \in J$ and strict inequality holds for every $\ell \notin J$. For any $\ell \in J$ either $\ell \in J_{00}$ or $\ell \in J \cap J_{\neq 0}$. If $\ell \in J_{00}$, $a_\ell'(M_Jx+\kappa_J) = 0 = b_\ell$, so equality holds. If $\ell \in J\cap J_{\neq 0}$,

$$a_\ell'(x - \|a_{\bar{j}}\|^{-2}a_{\bar{j}}(a_{\bar{j}}'x-b_{\bar{j}})) = a_\ell'(x-\|a_{\bar{j}}\|^{-1}\|a_\ell\|^{-1}a_{\bar{j}}(a_\ell'x-b_\ell)) = b_\ell, \qquad (67)$$

where the first equality uses the fact that $\|a_\ell\|^{-1}(b_\ell - a_\ell'x) = \|a_{\bar{j}}\|^{-1}(b_{\bar{j}}-a_{\bar{j}}'x)$ by the definition of C_J^Δ, and the second equality uses the fact that $a_\ell'a_{\bar{j}} = \|a_{\bar{j}}\|\|a_\ell\|$ by the definition of $\mathcal{J}_1^{os}$. Therefore, equality holds for every $\ell \in J$. For any $\ell \in J^c$, we show that strict inequality holds. Either $\ell \in J_0/J_{00}$ or $\ell \in J_{\neq 0}/J$. If $\ell \in J_0/J_{00}$, $a_\ell'(M_J x + \kappa_J) = 0 < b_\ell$.[27] If $\ell \in J_{\neq 0}$,

$$\begin{aligned} a_\ell'(x - \|a_{\bar{j}}\|^{-2} a_{\bar{j}}(a_{\bar{j}}'x - b_{\bar{j}})) =& a_\ell' x - \|a_{\bar{j}}\|^{-2} a_\ell' a_{\bar{j}}(a_{\bar{j}}'x - b_{\bar{j}}) \\ =& a_\ell' x - b_\ell - \|a_{\bar{j}}\|^{-2} a_\ell' a_{\bar{j}}(a_{\bar{j}}'x - b_{\bar{j}}) + b_\ell \\ <& \frac{\|a_\ell\|}{\|a_{\bar{j}}\|}(a_{\bar{j}}'x - b_{\bar{j}}) - \|a_{\bar{j}}\|^{-2} a_\ell' a_{\bar{j}}(a_{\bar{j}}'x - b_{\bar{j}}) + b_\ell \\ =& \frac{(\|a_\ell\|\|a_{\bar{j}}\| - a_\ell' a_{\bar{j}})(a_{\bar{j}}'x - b_{\bar{j}})}{\|a_{\bar{j}}\|^2} + b_\ell \leq b_\ell, \end{aligned} \tag{68}$$

where the first inequality uses the fact that $\|a_{\bar{j}}\|^{-1}(b_{\bar{j}} - a_{\bar{j}}'x) < \|a_\ell\|^{-1}(b_\ell - a_\ell' x)$ by the definition of $x \in C_J^\Delta$ (because ℓ is not in the argmin), and the second inequality uses the fact that $a_{\bar{j}}'x < b_{\bar{j}}$ and $\|a_{\bar{j}}\|\|a_\ell\| \geq a_\ell' a_{\bar{j}}'$. This shows that for every $\ell \in J^c$ the inequality is strict. Therefore, $M_J x + \kappa_J \in C_J$.

We also calculate that $Q_J'(P_J x - \kappa_J) \in (-\tau(M_J x + \kappa_J), 0)$. The fact that $x \in C_0$ implies that $a_{\bar{j}}'x - b_{\bar{j}} < 0$, and so $Q_J'(P_J x - \kappa_J) = \|a_{\bar{j}}\|^{-1}(a_{\bar{j}}'x - b_{\bar{j}}) < 0$. Let $\ell \in \{1, ..., d_A\}/\{\bar{j}\}$.[28] We show that

$$\|a_{\bar{j}}\|^{-1}(a_{\bar{j}}'x - b_{\bar{j}}) > -\tau_j(M_J x + \kappa_J). \tag{69}$$

If $\|a_\ell\|\|a_{\bar{j}}\| - a_\ell' a_{\bar{j}} = 0$, then by definition the right hand side of (69) is $-\infty$. Otherwise, we can plug in $M_J x + \kappa_J = x - \|a_{\bar{j}}\|^{-2} a_{\bar{j}} a_{\bar{j}}'x + \|a_{\bar{j}}\|^{-2} a_{\bar{j}} b_{\bar{j}}$ and rewrite (69) as

$$(a_{\bar{j}}'x - b_{\bar{j}})(\|a_\ell\|\|a_{\bar{j}}\| - a_\ell' a_{\bar{j}}) > -\|a_{\bar{j}}\|^2 (b_\ell - a_\ell'(x - \|a_{\bar{j}}\|^{-2} a_{\bar{j}} a_{\bar{j}}'x + \|a_{\bar{j}}\|^{-2} a_{\bar{j}} b_{\bar{j}})). \tag{70}$$

We can simplify this to show that it holds if and only if

$$\|a_{\bar{j}}\|^{-1}(b_{\bar{j}} - a_{\bar{j}}'x) < \|a_\ell\|^{-1}(b_\ell - a_\ell' x). \tag{71}$$

The fact that $\|a_\ell\|\|a_{\bar{j}}\| \neq a_\ell' a_{\bar{j}}$ implies that $\ell \notin J$ (by the definition of $\mathcal{J}_1^{os}$) and therefore, by the definition of C_J^Δ, (71) holds (because ℓ is not in the argmin). Therefore, (69) holds for every $\ell \in \{1, ..., d_A\}/\{\bar{j}\}$, which implies that

$$Q_J'(P_J x - \kappa_J) = \|a_{\bar{j}}\|^{-1}(a_{\bar{j}}'x - b_{\bar{j}}) > -\tau(M_J x + \kappa_J). \tag{72}$$

This shows that $x \in \Upsilon$.

(2) Let $x \in \Upsilon$. Consider the set $\text{argmin}_{j \in J_{\neq 0}} \|a_j\|^{-1}(b_j - a_j' x)$. We first show that the argmin is equal to $J \cap J_{\neq 0}$. If $\ell \in J_{\neq 0}/J$, an algebraic manipulation similar to above shows

[27] b_ℓ cannot be negative because, by assumption, $\theta \in \Theta_0(F)$, so $\mu \in C$, and therefore C is non-empty.

[28] We note here that $\tau(x)$ is defined for an arbitrary active inequality $\bar{j} \in J \cap J_{\neq 0}$. One can verify that the definition of $\tau(x)$ does not depend on which $\bar{j} \in J \cap J_{\neq 0}$ is selected.

that

$$
\begin{aligned}
&Q_J'(P_J x - \kappa_J) > -\tau(M_J x + \kappa_J) \\
\Rightarrow\ & \|a_{\bar{j}}\|^{-1}(a_{\bar{j}}' x - b_{\bar{j}}) > -\frac{\|a_{\bar{j}}\|(b_\ell - a_\ell'(x - \|a_{\bar{j}}\|^{-2} a_{\bar{j}}(a_{\bar{j}}' x - b_{\bar{j}})))}{\|a_{\bar{j}}\|\|a_\ell\| - a_\ell' a_{\bar{j}}} \\
\iff\ & \|a_\ell\|^{-1}(b_\ell - a_\ell' x) > \|a_{\bar{j}}\|^{-1}(b_{\bar{j}} - a_{\bar{j}}' x), \qquad (73)
\end{aligned}
$$

where the first implication uses the definition of $\tau(x)$ and the "iff" follows from multiplying by $\|a_{\bar{j}}\|\|a_\ell\| - a_\ell' a_{\bar{j}}$ and cancelling the $a_\ell' a_{\bar{j}}$ term. This shows that $\ell \in J_{\neq 0}/J$ cannot be in the argmin. Also consider $\ell \in J \cap J_{\neq 0}$. The definition of $\mathcal{J}_1^{os}$ implies that $\|a_\ell\|^{-1} a_\ell = \|a_{\bar{j}}\|^{-1} a_{\bar{j}}$. Notice that

$$
\begin{aligned}
& 0 = b_{\bar{j}} - a_{\bar{j}}'(M_J x + \kappa_J) = b_\ell - a_\ell'(M_J x + \kappa_J) \\
\iff & 0 = b_{\bar{j}} - a_{\bar{j}}'(x - \|a_{\bar{j}}\|^{-2} a_{\bar{j}}(a_{\bar{j}}' x - b_{\bar{j}})) = b_\ell - a_\ell'(x - \|a_{\bar{j}}\|^{-2} a_{\bar{j}}(a_{\bar{j}}' x - b_{\bar{j}})) \\
\Rightarrow\ & b_\ell = a_\ell'(x - \|a_{\bar{j}}\|^{-2} a_{\bar{j}}(a_{\bar{j}}' x - b_{\bar{j}})) \\
\iff & \|a_\ell\|^{-1}(b_\ell - a_\ell' x) = \|a_\ell\|^{-1} a_\ell' \|a_{\bar{j}}\|^{-2} a_{\bar{j}}(b_{\bar{j}} - a_{\bar{j}}' x) = \|a_{\bar{j}}\|^{-1}(b_{\bar{j}} - a_{\bar{j}}' x), \qquad (74)
\end{aligned}
$$

where the first line holds because $M_J x + \kappa_J \in C_J$, the first "iff" holds by plugging in the formula for $M_J x + \kappa_J$, the implication holds by solving for b_ℓ and cancelling $b_{\bar{j}}$, the second "iff" holds by rearranging and the fact that $a_\ell' a_{\bar{j}} = \|a_\ell\|\|a_{\bar{j}}\|$. We have shown that ℓ should be in the argmin. Therefore, the argmin is equal to $J \cap J_{\neq 0}$.

We also show that $x \in C_0$. Note that $a_{\bar{j}}' x < b_{\bar{j}}$ because $Q_J'(P_J x - \kappa_J) < 0$ and plugging in the formulas for Q_J, P_J, and κ_J. For any other $\ell \in J_{\neq 0}$, we have

$$
\|a_\ell\|^{-1}(b_\ell - a_\ell' x) \geq \|a_{\bar{j}}\|^{-1}(b_{\bar{j}} - a_{\bar{j}}' x) > 0, \qquad (75)
$$

because $\bar{j}$ belongs to the argmin. Thus, $x \in C_0$ because all the inequalities for $\ell \in J_{\neq 0}$ are inactive. Therefore, $x \in C_J^{\Delta}$.

Therefore, we have shown (66), which implies (65).

To finish the proof of part (b), we plug in (56), (58), (59), (62), (63), and (65) into (52), (53), (54), and (55) to get that

$$
\begin{aligned}
&\sum_{J \subseteq \{1,\ldots,d_A\}} \Pr(X \in K_J \text{ and } \|X - P_C X\|^2 > \chi^2_{\mathrm{rk}(A_J),1-\alpha}) \\
\leq & \sum_{J \subseteq \{1,\ldots,d_A\} : \mathrm{rk}(A_J) \geq 2} \alpha \Pr(M_J X + \kappa_J \in C_J) \times \Pr(P_J X - \kappa_J \in V_J)
\end{aligned}
$$

$$+\sum_{J\in\mathcal{J}_1^{ts}}\alpha\Pr(M_JX+\kappa_J\in C_J)\times\Pr(P_JX-\kappa_J\in V_J)$$
$$+\sum_{J\in\mathcal{J}_1^{os}}\alpha(\Pr(X\in K_J)+\Pr(X\in C_J^{\Delta}))$$
$$=\alpha\times\left(\sum_{J\subseteq\{1,\ldots,d_A\}:\text{rk}(A_{J(X)})>0}\Pr(X\in K_J)+\sum_{J\in\mathcal{J}_1^{os}}\Pr(X\in C_J^{\Delta})\right)$$
$$=\alpha(1-\Pr(X\in C^{|}))\leq\alpha, \tag{76}$$

where the first equality uses Lemma 1(a) and the fact that P_JX is independent of M_JX, together with Lemma 2(d), and the second equality uses Lemma 1(b) and Lemma 2(b).

We next prove part (b). Fix $J\subseteq\{1,\ldots,d_A\}$. We first note that when C_J is empty, the inequality in (62) holds with equality because both sides are zero. When $C_J\neq\emptyset$, we show that $\kappa_J=P_J\mu$. Let $z\in C_J$ and for every $\lambda\in[0,1]$ let $\mu_\lambda=\lambda z+(1-\lambda)\mu$. Recall that $A\mu=b$. Thus, for each $\lambda\in(0,1]$

$$a_j'\mu_\lambda=\lambda a_j'z+(1-\lambda)a_j'\mu=\lambda b_j+(1-\lambda)b_j=b_j\text{ for }j\in J,\text{ and} \tag{77}$$
$$a_j'\mu_\lambda=\lambda a_j'z+(1-\lambda)a_j'\mu<\lambda b_j+(1-\lambda)b_j=b_j\text{ for }j\in J^c.$$

This implies that $\mu_\lambda\in C_J$, and hence, for every $\lambda\in(0,1]$, $P_J\mu_\lambda=\kappa_J$. Take $\lambda\to0$ and by the continuity of the projection, $\kappa_J=P_J\mu$. Thus, $\gamma=Q_J'(P_J\mu-\kappa_J)=0$, implying that the inequality in (62) holds with equality. The fact that $\gamma=0$ also implies that the inequality in (64) holds with equality by Lemma 4. The inequality in the last line of (76) holds with equality by Lemma 2(c). This proves part (b).

Part (c). Let $\tilde{C}=\{\mu\in\mathbb{R}^{d_m}|A_J\mu\leq b_J\}$. Let $\tilde{T}_n(\theta)$, $P_{\tilde{C}}X$, $\tilde{J}(X)$, $\tilde{r}(X)$, $\tilde{\tau}(X)$, and $\tilde{\beta}(X)$ be defined with A_J and b_J in place of A and b. For each $L\subseteq J$, also define $\tilde{C}_L$ and $\tilde{K}_L$ similarly. Note that all these objects also depend on s because A and b may depend on θ.

Notice that

$$\Pr\left(\phi_n^{\text{RCC}}(\theta_s,\alpha)\neq\phi_{n,J}^{\text{RCC}}(\theta_s,\alpha)\right)$$
$$=\sum_{L\subseteq\{1,\ldots,d_A\}}\Pr(X\in K_L\text{ and }\phi_n^{\text{RCC}}(\theta_s,\alpha)\neq\phi_{n,J}^{\text{RCC}}(\theta_s,\alpha))$$
$$=\sum_{L\subseteq J}\Pr(X\in K_L\text{ and }\phi_n^{\text{RCC}}(\theta_s,\alpha)\neq\phi_{n,J}^{\text{RCC}}(\theta_s,\alpha))+o(1) \tag{78}$$
$$=\sum_{L\subseteq J}\Pr\left(X\in K_L\text{ and }\chi^2_{\text{rk}(A_L),1-\beta(X)}\geq\|X-P_CX\|^2>\chi^2_{\text{rk}(A_L),1-\tilde{\beta}(X)}\right)+o(1) \tag{79}$$

$$= \sum_{L \subseteq J : \text{rk}(A_L)=1} \Pr\Big(X \in K_L \text{ and } \chi^2_{1,1-\beta(X)} \geq \|X - P_C X\|^2 > \chi^2_{1,1-\tilde{\beta}(X)}\Big) + o(1) \tag{80}$$

$$\to 0, \tag{81}$$

where the first equality follows from Lemma 1(b) and the subsequent equalities and convergence are justified below.

For (78), let $\tilde{X} = X - \mu \sim N(\mathbf{0}, I)$. Fix the value of $\tilde{X}$. We show that for any $L \not\subseteq J$, either (i) $\tilde{X} + \mu \notin K_L$ or (ii) $\mathbb{1}\{\|\tilde{X} + \mu - P_C(\tilde{X} + \mu)\|^2 > \chi^2_{r(\tilde{X}+\mu),1-\beta(\tilde{X}+\mu)}\} = \mathbb{1}\{\|\tilde{X} + \mu - P_{\tilde{C}}(\tilde{X} + \mu)\|^2 > \chi^2_{\tilde{r}(\tilde{X}+\mu),1-\tilde{\beta}(\tilde{X}+\mu)}\}$ eventually as $s \to \infty$. Note that the expression in (ii) is equivalent to $\phi_n^{\text{RCC}}(\theta_s, \alpha) = \phi_{n,J}^{\text{RCC}}(\theta_s, \alpha)$ when evaluated at $X = \tilde{X} + \mu$. Suppose, to reach a contradiction, there exists a subsequence in s such that (i) and (ii) are both false for all s.[29] Let $\ell \in L/J$. Then by Lemma 1(a), we have $a_\ell' P_C(\tilde{X} + \mu) = b_\ell$. It follows that

$$\|\tilde{X} + \mu - P_C(\tilde{X} + \mu)\| \geq \frac{-a_\ell'\left(\tilde{X} + \mu - P_C(\tilde{X} + \mu)\right)}{\|a_\ell\|} \tag{82}$$

$$= \frac{-a_\ell'\tilde{X}}{\|a_\ell\|} - \frac{a_\ell'\mu - b_\ell}{\|a_\ell\|} \to +\infty, \tag{83}$$

where the inequality follows from Cauchy-Schwarz, and the convergence follows by assumption for $\ell \notin J$, using the fact that $\tilde{X}$ is fixed so $\|a_\ell\|^{-1} a_\ell' \tilde{X}$ is bounded. This implies that $\|\tilde{X} + \mu - P_C(\tilde{X} + \mu)\|^2 > \chi^2_{d_A,1-\alpha} \geq \chi^2_{r(\tilde{X}+\mu),1-\beta(\tilde{X}+\mu)}$ eventually as $s \to \infty$.

We next claim that there exists a further subsequence in s and an $\tilde{\ell} \notin J$ such that $a_{\tilde{\ell}}' P_{\tilde{C}}(\tilde{X} + \mu) \geq b_{\tilde{\ell}}$ along the further subsequence. Such an $\tilde{\ell}$ and subsequence must exist because otherwise, $P_{\tilde{C}}(\tilde{X} + \mu) \in C$, which implies that $P_{\tilde{C}}(\tilde{X} + \mu) = P_C(\tilde{X} + \mu)$ (because $C \subseteq \tilde{C}$), and which further implies by Lemma 1(a) that $a_{\tilde{\ell}}' P_{\tilde{C}}(\tilde{X} + \mu) = a_{\tilde{\ell}}' P_C(\tilde{X} + \mu) = b_{\tilde{\ell}}$ for every $\tilde{\ell} \in L$. In that case, we can take $\tilde{\ell} \in L/J$. It follows that

$$\|\tilde{X} + \mu - P_{\tilde{C}}(\tilde{X} + \mu)\| \geq \frac{-a_{\tilde{\ell}}'\left(\tilde{X} + \mu - P_{\tilde{C}}(\tilde{X} + \mu)\right)}{\|a_{\tilde{\ell}}\|} \tag{84}$$

$$\geq \frac{-a_{\tilde{\ell}}'\tilde{X}}{\|a_{\tilde{\ell}}\|} - \frac{a_{\tilde{\ell}}'\mu - b_{\tilde{\ell}}}{\|a_{\tilde{\ell}}\|} \to +\infty, \tag{85}$$

where the inequality follows from Cauchy-Schwarz, and the convergence follows by assumption for $\tilde{\ell} \notin J$, using the fact that $\tilde{X}$ is fixed so $\|a_{\tilde{\ell}}\|^{-1} a_{\tilde{\ell}}' \tilde{X}$ is bounded. This implies that $\|\tilde{X} + \mu - P_{\tilde{C}}(\tilde{X} + \mu)\|^2 > \chi^2_{d_A,1-\alpha} \geq \chi^2_{\tilde{r}(\tilde{X}+\mu),1-\tilde{\beta}(\tilde{X}+\mu)}$ eventually as $s \to \infty$. Therefore, along this subsequence, condition (ii) holds eventually. This contradiction implies that for every

[29] For notational simplicity, we do not introduce notation for this subsequence or any further subsequence.

$L \not\subseteq J$ and for every fixed $\tilde{X}$ either (i) or (ii) holds eventually. Therefore,

$$\begin{aligned}
&\Pr(X \in K_L \text{ and } \phi_n^{\mathrm{RCC}}(\theta_s,\alpha) \neq \phi_{n,J}^{\mathrm{RCC}}(\theta_s,\alpha)\\
=&\Pr(\tilde{X}+\mu \in K_L \text{ and}\\
&\mathbb{1}\{\|\tilde{X}+\mu-P_C(\tilde{X}+\mu)\|^2>\chi^2_{r(\tilde{X}+\mu),1-\beta(\tilde{X}+\mu)}\} \neq \mathbb{1}\{\|\tilde{X}+\mu-P_{\tilde{C}}(\tilde{X}+\mu)\|^2>\chi^2_{\tilde{r}(\tilde{X}+\mu),1-\tilde{\beta}(\tilde{X}+\mu)}\})\\
\to& 0,
\end{aligned}$$

where the equality follows from the fact that X has the same distribution as $\tilde{X}+\mu$, and the convergence follows from the bounded convergence theorem.

For (79), note that for any $L \subseteq J$, if $X \in K_L$, then $P_C X \in C_L$ by Lemma 1(a). We argue that $P_C X = P_{\tilde{C}} X$. By a property of projection onto convex sets, it is sufficient to show that for all $y \in \tilde{C}$, we have $\langle X - P_C X, y - P_C X\rangle \leq 0$.[30] This follows because by the definition of K_L, we can write $X - P_C X = \sum_{j\in L} v_j a_j$ with $v_j \geq 0$, so

$$\langle X - P_C X, y - P_C X\rangle = \sum_{j\in L} v_j(\langle a_j, y\rangle - \langle a_j, P_C X) \leq 0,$$

where the inequality uses the fact that $y \in \tilde{C}$, so $a_j' y \leq b_j$ and $P_C X \in C_L$, so $a_j' P_C X = b_j$. Therefore, $P_C X = P_{\tilde{C}} X$. The fact that $P_{\tilde{C}} X \in C_L$ implies that $P_{\tilde{C}} X \in \tilde{C}_L$ because $C_L \subseteq \tilde{C}_L$. By Lemma 1(b), this implies that $X \in \tilde{K}_L$. Thus, $r(X) = \tilde{r}(X) = \mathrm{rk}(A_L)$, and $T_n(\theta) = \|X - P_C X\|^2 = \|X - P_{\tilde{C}} X\|^2 = \tilde{T}_n(\theta)$. Also, the fact that $\tilde{\beta}(X) \geq \beta(X)$ implies that the only way $\phi_n^{\mathrm{RCC}}(\theta_s,\alpha) \neq \phi_{n,J}^{\mathrm{RCC}}(\theta_s,\alpha)$ is if $\chi^2_{\mathrm{rk}(A_L),1-\beta(X)} \geq \|X - P_C X\|^2 > \chi^2_{\mathrm{rk}(A_L),1-\tilde{\beta}(X)}$.

For (80), we use the fact that $\beta(X) = \tilde{\beta}(X) = \alpha$ whenever $\mathrm{rk}(A_L) \neq 1$.

Finally we show (81). For each $L \subseteq J$, note that $\mathrm{rk}(A_L)$ may depend on s. Fix any subsequence in s such that $\mathrm{rk}(A_L) = 1$ along the subsequence.[31] Write $P_L = Q_L^P Q_L^{P\prime}$ and $M_L = Q_L^M Q_L^{M\prime}$, where Q_L^P is a $d_m \times 1$ vector with unit length, Q_L^M is a $d_m \times (d_m - 1)$ matrix with orthonormal columns, and $Q_L^{P\prime} Q_L^M = \mathbf{0}$. Let $\tilde{X} = (\tilde{x}_1, \tilde{x}_2) \sim N(\mathbf{0}, I)$, where $\tilde{x}_1 \in \mathbb{R}^{d_m-1}$ and $\tilde{x}_2 \in \mathbb{R}$. We can then write

$$X = Q_L^M \tilde{x}_1 + Q_L^P \tilde{x}_2 + \mu. \tag{86}$$

Note that $\tilde{X}$ does not depend on s, while μ, Q_L^M, and Q_L^P may.

For each L, we can rewrite the term in (80) as

$$\int_{\tilde{x}_1}\int_{\tilde{x}_2} 1\{X \in K_L\} 1\{\chi^2_{1,1-\beta(X)} \geq \|X - P_C X\|^2 > \chi^2_{1,1-\tilde{\beta}(X)}\} \phi(\tilde{x}_1)\phi(\tilde{x}_2) d\tilde{x}_2 d\tilde{x}_1$$

[30] See section 3.12 in Luenberger (1969).

[31] For notational simplicity, we do not introduce notation for this subsequence or any further subsequence.

$$= \int_{\tilde{x}_1} 1\{M_L X + \kappa_L \in C_L\} \int_{\tilde{x}_2} g_s(\tilde{x}_1, \tilde{x}_2)\phi(\tilde{x}_2)d\tilde{x}_2\phi(\tilde{x}_1)d\tilde{x}_1, \tag{87}$$

where X is viewed as a function of $(\tilde{x}_1, \tilde{x}_2)$ using (86), $\phi(\cdot)$ is the probability density function of the standard normal distribution of the dimension determined by the dimension of its argument, and

$$g_s(\tilde{x}_1, \tilde{x}_2) = 1\{P_L X - \kappa_L \in V_L\}1\{\chi^2_{1,1-\beta(M_L X+\kappa_L)} \geq \|P_L X - \kappa_L\|^2 > \chi^2_{1,1-\tilde{\beta}(M_L X+\kappa_L)}\}, \tag{88}$$

which uses the same decomposition of $X \in K_L$ as in (58), and the fact that $M_L X$ only depends on $\tilde{x}_1$ (and that $\beta(X)$ and $\tilde{\beta}(X)$ only depend on X through $P_C X = M_L X + \kappa_L$). Fix $\tilde{x}_1$. We show that the inner integral goes to zero as $s \to \infty$.

Fix an arbitrary subsequence in s. We show that there exists a further subsequence such that the inner integral goes to zero. Since $\beta(M_L X + \kappa_L)$ and $\tilde{\beta}(M_L X + \kappa_L)$ do not depend on $\tilde{x}_2$ and both lie in $[\alpha, 2\alpha]$ for all s, there exists a further subsequence along which both converge. Denote the limits by β_∞ and $\tilde{\beta}_\infty$. Also note that $P_L X - \kappa_L = Q^P_L \tilde{x}_2 + P_L\mu - \kappa_L$. Take a further subsequence such that $P_L\mu - \kappa_L$ diverges or converges and such that Q^P_L converges to $Q^P_{L,\infty}$ (since Q^P_L has unit length, it must converge along a subsequence). We consider two cases.

(i) If $P_L\mu - \kappa_L$ diverges, then for every $\tilde{x}_2$, $\|Q^P_L \tilde{x}_2 + P_L\mu - \kappa_L\|^2 \geq (\|P_L\mu - \kappa_L\| - \|Q^P_L \tilde{x}_2\|)^2 \to \infty$, so $g_s(\tilde{x}_1, \tilde{x}_2) = 0$ eventually as $s \to \infty$ along this subsequence. Therefore by the dominated convergence theorem, the inner integral in (87) goes to zero.

(ii) If $P_L\mu - \kappa_L$ converges to some κ_∞, then fix $\tilde{x}_2$ such that $\|Q^P_{L,\infty}\tilde{x}_2 + \kappa_\infty\|^2 \neq \chi^2_{1,1-\beta_\infty}$ and $\|Q^P_{L,\infty}\tilde{x}_2 + \kappa_\infty\|^2 \neq \chi^2_{1,1-\tilde{\beta}_\infty}$. Note that the set of such $\tilde{x}_2$ is a set of probability one with respect to $\tilde{x}_2 \sim N(0,1)$. We show that $g_s(\tilde{x}_1, \tilde{x}_2) = 0$ eventually. Consider $\|a_\ell\|^{-1}a'_\ell(\mu - P_C X)$ for $\ell \notin J$. If there exists an $\ell \notin J$ and a subsequence of s such that $\|a_\ell\|^{-1}a'_\ell(\mu - P_C X) \to \pm\infty$, then

$$\|X - P_C X\| = \|[Q^P_L, Q^M_L]'\tilde{X} + \mu - P_C X\| \tag{89}$$

$$\geq \frac{\pm a'_\ell \left([Q^P_L, Q^M_L]'\tilde{X} + \mu - P_C X\right)}{\|a_\ell\|} \tag{90}$$

$$= \frac{\pm a'_\ell [Q^P_L, Q^M_L]'\tilde{X}}{\|a_\ell\|} \pm \frac{a'_\ell (\mu - P_C X)}{\|a_\ell\|} \to +\infty, \tag{91}$$

where the inequality follows by Cauchy-Schwarz and the convergence follows from the fact that $\|a_\ell\|^{-1}a'_\ell[Q^P_L, Q^M_L]'\tilde{X}$ is bounded. This shows that $\|P_L X - \kappa_L\|^2 = \|X - P_C X\|^2 > \chi^2_{1,1-\alpha} \geq \chi^2_{1,1-\beta(M_L X+\kappa_L)}$ eventually, and therefore $g_s(\tilde{x}_1, \tilde{x}_2) = 0$ eventually along this subsequence. Otherwise, suppose $\|a_\ell\|^{-1}a_\ell'(\mu - P_C X)$ is bounded along a subsequence for all $\ell \notin J$. We show that $\beta_\infty = \tilde{\beta}_\infty$. Let $\bar{j} \in L$ such that $a_{\bar{j}} \neq 0$. Then note that for each $\ell \notin J$,

$$\tau_\ell(X) = \frac{\|a_{\bar{j}}\|(b_\ell - a_\ell' P_C X)}{\|a_{\bar{j}}\|\|a_\ell\| - a_{\bar{j}}' a_\ell} = \frac{(b_\ell - a_\ell' P_C X)/\|a_\ell\|}{1 - a_{\bar{j}}' a_\ell/(\|a_{\bar{j}}\|\|a_\ell\|)} \geq \frac{1}{2}\frac{b_\ell - a_\ell' P_C X}{\|a_\ell\|} \to \infty, \tag{92}$$

where the convergence follows because $\|a_\ell\|^{-1}(b_\ell - a_\ell'\mu) \to \infty$ and $\|a_\ell\|^{-1}(a_\ell'\mu - a_\ell' P_C X)$ is bounded (and if the denominator is zero, then $\tau_\ell(X) = \infty$). Therefore,

$$\tau(X) = \inf_{\ell \neq \bar{j}} \tau_\ell(X) = \min\left(\inf_{\ell \in J; \ell \neq \bar{j}} \hat{\tau}_\ell(X), \inf_{\ell \notin J} \hat{\tau}_\ell(X)\right) = \min\left(\tilde{\tau}(X), \inf_{\ell \notin J} \hat{\tau}_\ell(X)\right). \tag{93}$$

If $\tilde{\tau}(X) \to \infty$, then $\tau(X) \to \infty$ too, and if $\tilde{\tau}(X)$ converges to a finite value, $\tau(X)$ converges to the same value. This shows that $\beta_\infty = \tilde{\beta}_\infty$. Therefore $g_s(\tilde{x}_1, \tilde{x}_2) = 0$ eventually along this subsequence. Since every subsequence has a further subsequence such that $g_s(\tilde{x}_1, \tilde{x}_2) = 0$ eventually, it follows that $g_s(\tilde{x}_1, \tilde{x}_2) = 0$ eventually along the original sequence. Therefore by the dominated convergence theorem, the inner integral in (87) goes to zero.

Since the inner integral in (87) converges to zero in either case (i) or (ii) for every fixed $\tilde{x}_1$, by the dominated convergence theorem, the outer integral converges to zero too along this subsequence. Since every subsequence has a further subsequence such that (87) converges to zero, this shows (81). □

A.3 Proofs of the Auxiliary Lemmas

Proof of Lemma 1. (a) By assumption, $X \in K_J = C_J + V_J$. So, we write $X = X_1 + X_2$, where $X_1 \in C_J$ and $X_2 \in V_J$. Then, $P_C X_1 = X_1$ because $X_1 \in C$ already. We show that $P_C X = X_1$. By a property of projection onto convex sets, it is necessary and sufficient that for all $y \in C$, we have $\langle X - X_1, y - X_1\rangle \leq 0$.[32] This follows because $X_2 = \sum_{j\in J} v_j a_j$ with $v_j \geq 0$, so

$$\langle X_2, y - X_1\rangle = \sum_{j\in J} v_j(\langle a_j, y\rangle - \langle a_j, X_1\rangle) \leq 0, \tag{94}$$

where the inequality uses the fact that $y \in C$, so $a_j' y \leq b_j$ and $X_1 \in C_J$, so $a_j' X_1 = b_j$. Combining these, we get that $P_C X = X_1 \in C_J$ and $X - P_C X = X - X_1 = X_2 \in V_J$.

(b) We first show that every X belongs to some K_J. For every X, $P_C X \in C$, so there exists a J such that $P_C X \in C_J$.

By the inner-product property of projection, we know that for all $y \in C$, $\langle y - P_C X, X -$

[32] See Section 3.12 in Luenberger (1969). Hereafter, call this property of projection onto a convex set the "inner-product property."

$P_C X\rangle \leq 0$. Using this fact, let $z \perp \text{span}(V_J)$. Then, there exists a $\epsilon > 0$ such that $P_C X + \epsilon z$ and $P_C X - \epsilon z$ both belong to C.[33] Then, $\langle \epsilon z, X - P_C X\rangle \leq 0$ and $\langle -\epsilon z, X - P_C X\rangle \leq 0$. These two inequalities imply that $\langle z, X - P_C X\rangle = 0$. Thus, $X - P_C X$ is orthogonal to all vectors, z, which are orthogonal to $\text{span}(V_J)$. This implies that $X - P_C X \in \text{span}(V_J)$.

If $X - P_C X \notin V_J$, then by the separating hyperplane theorem,[34] there exists a direction, $c \in \text{span}(V_J)$ such that $\langle c, X - P_C X\rangle > 0$ and $\langle c, a_j\rangle < 0$ for all $j \in J$. We consider $P_C X + \epsilon c$. We show that for ϵ sufficiently small, (1) $P_C X + \epsilon c \in C$, and (2) $\langle X - P_C X, \epsilon c\rangle > 0$.

(1) For $j \in J$, $\langle P_C X + \epsilon c, a_j\rangle = b_j + \epsilon\langle c, a_j\rangle < b_j$, where the equality follows because $P_C X \in C_J$ and the inequality follows from the definition of c. For $j \in J^c$, $\langle P_C X + \epsilon c, a_j\rangle = \langle P_C X, a_j\rangle + \epsilon\langle c, a_j\rangle$, which is less than b_j for ϵ sufficiently small because $\langle P_C X, a_j\rangle < b_j$.

(2) $\langle X - P_C X, \epsilon c\rangle = \epsilon\langle X - P_C X, c\rangle > 0$ by the definition of c.

This contradicts the inner-product property of projection onto a convex set, and therefore $X - P_C X \in V_J$, and $X \in K_J$.

We next show that no X belongs to two distinct K_J. If $X \in K_J$ and $K_{J'}$, then, by part (a), $P_C X \in C_J$ and $P_C X \in C_{J'}$. But this is a contradiction because the projection onto a convex set is unique, and the C_J form a partition of C.

(c) If $X \in K_J$, then $P_C X \in C_J$, so all the inequalities in J are active. If $X \notin K_J$, then X is in a different $K_{J'}$, for some $J' \neq J$, by part (b). Thus, $J \neq J(X) = J'$. □

Proof of Lemma 2. (a) Note that J_{00} satisfies $\text{rk}(A_{J_{00}}) = 0$. Thus, it is sufficient to show that $C_J = \emptyset$ for all $J \subseteq J_0$ that are not J_{00}. If $J \neq J_{00}$, then either (i) there exists $j \in J_{00}/J$ or (ii) there exists $j \in J/J_{00}$. In the first case, any $x \in C_J$ would have to satisfy $\mathbf{0}'x < 0$, a contradiction. In the second case, any $x \in C_J$ would have to satisfy $\mathbf{0}'x = b_j$, where $b_j \neq 0$, another contradiction.

(b) We first show that the C_J^Δ are disjoint for different $J \in \mathcal{J}_1^{os}$. If $x \in C_{J_1}^\Delta \cap C_{J_2}^\Delta$ for $J_1, J_2 \in \mathcal{J}_1^{os}$, then both

$$J_{\neq 0} \cap J_1 = J_{\neq 0} \cap J_2 = \text{argmin}_{j \in J_{\neq 0}} \|a_j\|^{-1}(b_j - a_j' x) \tag{95}$$

and

$$J_0 \cap J_1 = J_0 \cap J_2 = J_{00}. \tag{96}$$

This implies that $J_1 = J_2$. Part (b) then follows from the definition of $C^{|}$.

[33]This uses the slackness of the inequalities in the definition of C_J.

[34]See Section 11 of Rockafellar (1970) or Section 5.12 in Luenberger (1969).

(c) For any $x \in C^{|}$, let $\tilde{J}_{\neq 0}(x) = \text{argmin}_{j \in J_{\neq 0}} \|a_j\|^{-1}(b_j - a_j' x)$. We show below that

$$\exists j, \ell \in \tilde{J}_{\neq 0}(x) \text{ s.t. } \|a_j\|^{-1} a_j \neq \|a_\ell\|^{-1} a_\ell. \tag{97}$$

That implies that for any $x \in C^{|}$, there exists $j, \ell \in J_{\neq 0}$ such that $\|a_j\|^{-1} a_j \neq \|a_\ell\|^{-1} a_\ell$ and $\|a_j\|^{-1}(b_j - a_j' x) = \|a_\ell\|^{-1}(b_\ell - a_\ell' x)$. Or equivalently,

$$C^{|} \subseteq \cup_{\substack{j, \ell \in J_{\neq 0} \\ \|a_j\| a_\ell \neq \|a_\ell\| a_j}} \{x \in \mathbb{R}^{d_m} | \|a_j\| b_\ell - \|a_\ell\| b_j = (\|a_j\| a_\ell - \|a_\ell\| a_j)' x\}. \tag{98}$$

Since the right hand-side is a finite union of measure-zero subspaces of $\mathbb{R}^{d_m}$, it must be that $C^{|}$ has Lebesgue measure zero, establishing part (c).

Now we show (97). Let $\tilde{J}(x) = J_{00} \cup \tilde{J}_{\neq 0}(x)$. We note that $\tilde{J}_{\neq 0}(x)$ is not empty because $A \neq \mathbf{0}_{d_A \times d_X}$. This implies that $\text{rk}(A_{\tilde{J}(x)}) \geq 1$. Then there are two possibilities: $\text{rk}(A_{\tilde{J}(x)}) \geq 2$ and $\text{rk}(A_{\tilde{J}(x)}) = 1$. In the first case, (97) holds trivially.

In the latter case, we first show that $\tilde{J}(x) \in \mathcal{J}_1$. Suppose there exists $j \in \tilde{J}(x)$ and $\ell \in \{1, ..., d_A\}/\tilde{J}(x)$ such that $\|a_j\| > 0, \|a_\ell\| > 0$, $\frac{a_j}{\|a_j\|} = \frac{a_\ell}{\|a_\ell\|}$, and $\frac{b_j}{\|a_j\|} = \frac{b_\ell}{\|a_\ell\|}$. This implies

$$\|a_\ell\|^{-1}(b_\ell - a_\ell' x) = \|a_j\|^{-1}(b_j - a_j' x), \tag{99}$$

so ℓ should also belong to $\tilde{J}(x)$. Since such a j and ℓ cannot exist, it must be the case that $\tilde{J}(x) \in \mathcal{J}_1$. The fact that $x \in C^{|}$ means that $\tilde{J}(x) \notin \mathcal{J}_1^{os}$. Thus, it must be that $\tilde{J}(x) \in \mathcal{J}_1^{ts}$, which also implies (97). Therefore (97) holds in all cases. This proves part (c).

(d) First note that for every $J \in \mathcal{J}_1^{os} \cup \mathcal{J}_1^{ts}$ we have $\text{rk}(A_J) = 1$. Thus, it is sufficient to show that for every $J \subseteq \{1, ..., d_A\}$ with $\text{rk}(A_J) = 1$ and $J \notin \mathcal{J}_1^{os} \cup \mathcal{J}_1^{ts}$, we have $K_J = \emptyset$.

Note that if $J \cap J_0 \neq J_{00}$, then either (i) there exists $j \in J_{00}/(J_0 \cap J)$ or (ii) there exists $j \in (J \cap J_0)/J_{00}$. In the first case, any $x \in C_J$ would have to satisfy $\mathbf{0}' x < 0$, a contradiction. In the second case, any $x \in C_J$ would have to satisfy $\mathbf{0}' x = b_j$, where $b_j \neq 0$, another contradiction. This implies that C_J, and therefore K_J, is empty.

We next note that if $j \in J$ while $\ell \in \{1, ..., d_A\}/J$ with $\|a_j\| > 0$, $\|a_\ell\| > 0$, $\frac{a_j}{\|a_j\|} = \frac{a_\ell}{\|a_\ell\|}$, and $\frac{b_j}{\|a_j\|} = \frac{b_\ell}{\|a_\ell\|}$, then any $x \in C_J$ should satisfy

$$\|a_\ell\|^{-1}(b_\ell - a_\ell' x) = \|a_j\|^{-1}(b_j - a_j' x) = 0, \tag{100}$$

so ℓ should also belong to J. This contradiction implies that C_J, and therefore K_J, must be empty.

This implies that the only nonempty K_J with $\text{rk}(A_J) = 1$ must belong to $\mathcal{J}_1$. If we suppose that $J \notin \mathcal{J}_1^{os}$, then there must exist $j, \ell \in J$ s.t. $\|a_j\| > 0, \|a_\ell\| > 0$, and $\frac{a_j}{\|a_j\|} \neq \frac{a_\ell}{\|a_\ell\|}$.

However, since $\text{rk}(A_J) = 1$, a_ℓ and a_j must be collinear. This implies that $\frac{a_j}{\|a_j\|} = -\frac{a_\ell}{\|a_\ell\|}$. Then, any $x \in C_J$ must satisfy

$$0 = \|a_\ell\|^{-1}(b_\ell - a_\ell' x) = \|a_j\|^{-1}(b_j - a_j' x). \tag{101}$$

This implies $\|a_\ell\|^{-1} b_\ell = -\|a_j\|^{-1} b_j$, which implies that $J \in \mathcal{J}^{ts}$.

Therefore, the only $J \subseteq \{1, ..., d_A\}$ with $\text{rk}(A_J) = 1$ and $K_J \neq \emptyset$ belong to $\mathcal{J}_1^{os} \cup \mathcal{J}_1^{ts}$. $\square$

Proof of Lemma 4. For every $\lambda \geq 0$, let

$$f(\lambda) = \int_{-\tau}^{\infty} (\alpha - \mathbb{1}\{Z > z_{1-\beta/2}\}) e^{-\frac{1}{2}(Z+\lambda)^2} dZ. \tag{102}$$

Note that $\alpha \leq 1$ implies that $\beta \leq 2\Phi(\tau)$, which in turn implies that $z_{1-\beta/2} \geq -\tau$. We show that $f(\lambda) \geq 0$ for all $\lambda \geq 0$. This is sufficient because

$$\begin{aligned}
&\alpha \Pr{}_\mu(Z \geq -\tau) - Pr_\mu(\{Z \geq z_{1-\beta/2}\}) \\
&= \int_{-\tau}^{\infty} (\alpha - \mathbb{1}\{Z > z_{1-\beta/2}\}) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Z-\mu)^2} dZ \\
&= \frac{f(-\mu)}{\sqrt{2\pi}} \geq 0
\end{aligned} \tag{103}$$

for all $\mu \leq 0$.

Let $f'(\lambda)$ denote the derivative of f. We show that (1) $f(0) \geq 0$ and (2) for all $\lambda \geq 0$, $f'(\lambda) \geq -\left(z_{1-\beta/2} + \lambda\right) f(\lambda)$. Together, these two properties imply that $f(\lambda) \geq 0$ because, if not, then there exists a $\lambda > 0$ such that $f(\lambda) < 0$. Then, by the mean value theorem, there exists a $\tilde{\lambda} \in (0, \lambda)$ such that $f(\tilde{\lambda}) < 0$ and $f'(\tilde{\lambda}) < 0$, which contradicts property (2).

Property (1) holds because

$$\begin{aligned}
\frac{f(0)}{\sqrt{2\pi}} &= \int_{-\tau}^{\infty} (\alpha - \mathbb{1}\{Z > z_{1-\beta/2}\}) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} dZ \\
&= \alpha\Phi(\tau) - (1 - \Phi(z_{1-\beta/2})) = \alpha\Phi(\tau) - \beta/2 = 0.
\end{aligned} \tag{104}$$

This also shows that equality holds when $\mu = 0$.

To show that property (2) holds, we evaluate

$$\begin{aligned}
f'(\lambda) &= \frac{d}{d\lambda} \int_{-\tau}^{\infty} (\alpha - \mathbb{1}\{Z > z_{1-\beta/2}\}) e^{-\frac{1}{2}(Z+\lambda)^2} dZ \\
&= -\int_{-\tau}^{\infty} (Z + \lambda)(\alpha - \mathbb{1}\{Z > z_{1-\beta/2}\}) e^{-\frac{1}{2}(Z+\lambda)^2} dZ
\end{aligned}$$

$$= -\int_{-\tau}^{z_{1-\beta/2}} \alpha(Z+\lambda)e^{-\frac{1}{2}(Z+\lambda)^2}dZ + \int_{z_{1-\beta/2}}^{\infty} (1-\alpha)(Z+\lambda)e^{-\frac{1}{2}(Z+\lambda)^2}dZ$$
$$\geq -\int_{-\tau}^{z_{1-\beta/2}} \alpha(z_{1-\beta/2}+\lambda)e^{-\frac{1}{2}(Z+\lambda)^2}dZ + \int_{z_{1-\beta/2}}^{\infty} (1-\alpha)(z_{1-\beta/2}+\lambda)e^{-\frac{1}{2}(Z+\lambda)^2}dZ$$
$$= -(z_{1-\beta/2}+\lambda)\int_{-\tau}^{\infty} (\alpha - \mathbb{1}\{Z > z_{1-\beta/2}\})e^{-\frac{1}{2}(Z+\lambda)^2}dZ$$
$$= -(z_{1-\beta/2}+\lambda)f(\lambda), \tag{105}$$

where the second equality follows by dominated convergence and the inequality follows from the events $\{Z > z_{1-\beta/2}\}$ and $\{Z \leq z_{1-\beta/2}\}$. □

B Theorem 3 and the Proof of Theorem 2

In this section we prove Theorem 3, a general theorem for uniform asymptotic properties of the CC and RCC tests. Theorem 3 is used to prove Theorem 2.

B.1 Theorem 3: A General Asymptotic Theorem

In this section, we sometimes make explicit the dependence of A and b on θ, denoting them by $A(\theta)$ and $b(\theta)$. The rows of $A(\theta)$ are denoted by $a_j(\theta)$, and submatrices composed of the rows of $A(\theta)$ are denoted by $A_J(\theta)$.

Assumption 2. *The given sequence $\{(F_n, \theta_n) : F_n \in \mathcal{F}, \theta_n \in \Theta_0(F_n)\}_{n=1}^{\infty}$ satisfies, for every subsequence, n_m, there exists a further subsequence, n_q, and there exists a sequence of positive definite $d_m \times d_m$ matrices, $\{D_q\}$ such that:*

(a) *Under the sequence $\{F_{n_q}\}_{q=1}^{\infty}$,*

$$\sqrt{n_q}D_q^{-1/2}(\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_{n_q}}\overline{m}_{n_q}(\theta_{n_q})) \to_d N(\mathbf{0}, \Omega), \tag{106}$$

for some positive definite correlation matrix, Ω, and

$$\|D_q^{-1/2}\widehat{\Sigma}_{n_q}(\theta_{n_q})D_q^{-1/2} - \Omega\| \to_p 0. \tag{107}$$

(b) *$\Lambda_q A(\theta_{n_q})D_q \to \bar{A}_0$ for some $d_A \times d_m$ matrix $\bar{A}_0$, and for every $J \subseteq \{1, ..., d_A\}$, $\mathrm{rk}(I_J A(\theta_{n_q})D_q) = \mathrm{rk}(I_J \bar{A}_0)$, where Λ_q is the diagonal $d_A \times d_A$ matrix whose jth diagonal entry is one if $e_j' A(\theta_{n_q}) = \mathbf{0}$ and $\|e_j' A(\theta_{n_q})D_q\|^{-1}$ otherwise.*

Remark. The matrix D_q typically is the diagonal matrix of variances of the elements of $\sqrt{n_q}\overline{m}_{n_q}(\theta_{n_q})$. In part (a), we allow each diagonal element to go to zero (or infinity) at different rates, to incorporate the cases where different moments are on different scales or where different moments involve time series processes that are integrated at different orders. Andrews and Guggenberger (2009), Andrews and Soares (2010), and Andrews et al. (2020) also use a diagonal normalizing matrix for this purpose. Moreover, the matrix D_q can be non-diagonal, which is useful when the asymptotic variance matrix of $\sqrt{n_q}(\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_n}\overline{m}_{n_q}(\theta_{n_q}))$ is singular but a certain rotation of the vector with proper scaling has a non-singular asymptotic variance matrix.

Part (b) is not required to show the uniform asymptotic validity of the RCC test. It is only used to show asymptotic size-exact and the asymptotic IDI property. The existence of $\bar{A}_0$ follows by the choice of the subsequence, while the rank condition is used to verify Lemma 6, below. □

The following theorem is a general asymptotic theorem used to show the uniform asymptotic properties of the RCC test.

Theorem 3. (a) *Suppose Assumption* 2(a) *holds for all sequences* $\{(F_n, \theta_n) : F_n \in \mathcal{F}, \theta_n \in \Theta_0(F_n)\}_{n=1}^{\infty}$. *Then,*

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{\theta\in\Theta_0(F)} \mathbb{E}_F(\phi_n^{\mathrm{RCC}}(\theta, \alpha)) \leq \alpha.$$

Next consider a sequence $\{(F_n, \theta_n) : F_n \in \mathcal{F}, \theta_n \in \Theta_0(F_n)\}_{n=1}^{\infty}$ *satisfying Assumption* 2(a,b).

(b) *If, along any further subsequence, for all* $j = 1, ..., d_A$, $\sqrt{n_q}e_j'\Lambda_q(A(\theta_{n_q})\mathbb{E}_{F_{n_q}}\overline{m}_{n_q}(\theta_{n_q}) - b(\theta_{n_q})) \to 0$, *and if* $\bar{A}_0 \neq \mathbf{0}_{d_A\times d_m}$, *then,*

$$\lim_{n\to\infty} \mathbb{E}_{F_n}\phi_n^{\mathrm{RCC}}(\theta_n, \alpha) = \alpha.$$

(c) *If, for* $J \subseteq \{1, ..., d_A\}$, *along any further subsequence,* $\sqrt{n_q}e_j'\Lambda_q(A(\theta_{n_q})\mathbb{E}_{F_{n_q}}\overline{m}_{n_q}(\theta_{n_q}) - b(\theta_{n_q})) \to -\infty$ *as* $q \to \infty$, *for all* $j \notin J$, *then*

$$\lim_{n\to\infty} \Pr_{F_n}\left(\phi_n^{\mathrm{RCC}}(\theta_n, \alpha) \neq \phi_{n,J}^{\mathrm{RCC}}(\theta_n, \alpha)\right) = 0.$$

Remarks. (1) Notice that no assumptions are placed on $A(\theta)$ for Theorem 3(a). It can be low-rank or any submatrix of $A(\theta)$ can be local to singular as θ varies. This is achieved by an extra step in the proof that adds inequalities that are redundant in the finite sample but are relevant in the limit (see Lemma 7 below).

(2) If θ_n and F_n are such that $\mathbb{E}_{F_n}\overline{m}_n(\theta_n)$ does not depend on n (for example, if $\{W_i\}_{i=1}^n$ is stationary under F_n with a fixed marginal distribution and $\theta \in \Theta_n(F_n)$ is fixed), and if $A(\theta_n)$ and $b(\theta_n)$ are fixed, then the condition in part (c) is automatically satisfied with J

equal to the set of all binding inequalities. If, in addition, $A_J(\theta_n) \neq \mathbf{0}$, parts (b) and (c) can be combined to show that the RCC test has exact asymptotic size. $\square$

B.2 Auxiliary Lemmas for Theorem 3

The proof of Theorem 3 uses four important lemmas. Lemma 5 establishes a condition under which the projection onto a sequence of polyhedra converges when the coefficient matrix defining the polyhedra converges. The condition is verified in a special context in Lemma 6, which is used to prove part (b) of Theorem 3. The conditions for part (a) are not strong enough for us to apply Lemma 5 because we do not restrict the rank of $A(\theta)$. Nonetheless, Lemma 7 shows that inequalities that are redundant in finite sample but relevant in the limit can be added to guarantee the condition of Lemma 5, and help us to prove part (a) of Theorem 3. Lemma 8 shows that the additional inequalities from Lemma 7 do not change the definition of $\hat{\beta}$.

First we define some notation. For any $d_A \times d_m$ real-valued matrix A and vector $h \in \mathbb{R}^{d_A}_{+,\infty} := [0,\infty]^{d_A}$, let $\text{poly}(A,h) = \{\mu \in \mathbb{R}^{d_m} : A\mu \leq h\}$ denote the polyhedron defined by inequalities with coefficients given by A and constants given by h. Also define

$$\mu^*(x; A, h) = \operatorname*{argmin}_{\mu \in \text{poly}(A,h)} \|x - \mu\|^2. \tag{108}$$

The lemma considers a sequence of $d_A \times d_m$ real-valued matrices $\{A_n\}_{n=1}^{\infty}$ and a sequence of $d_A \times 1$ vectors $h_n \in \mathbb{R}^{d_A}_{+} := [0,\infty)^{d_A}$ such that, as $n \to \infty$, $A_n \to A_0$ and $h_n \to h_0$ for a $d_A \times d_m$ real-valued matrix A_0 and a vector $h_0 \in \mathbb{R}^{d_A}_{+,\infty}$. Also, let $x_n \in \mathbb{R}^{d_m}$ be a sequence of vectors such that $x_n \to x_0 \in \mathbb{R}^{d_m}$ as $n \to \infty$. We say that a sequence of sets, $\text{poly}(A_n, h_n)$, Kuratowski converges to a limit set, $\text{poly}(A_0, h_0)$, denoted by

$$\text{poly}(A_n, h_n) \xrightarrow{K} \text{poly}(A_0, h_0), \tag{109}$$

if (i) for every $x_0 \in \text{poly}(A_0, h_0)$ there exists a sequence $x_n \in \text{poly}(A_n, h_n)$ such that $x_n \to x_0$, and (ii) for every subsequence n_q and for every converging sequence $x_{n_q} \in \text{poly}(A_{n_q}, h_{n_q})$ that converges to a point x_0, we have $x_0 \in \text{poly}(A_0, h_0)$.[35]

Lemma 5. *If* $\text{poly}(A_n, h_n) \xrightarrow{K} \text{poly}(A_0, h_0)$, *then* $\mu^*(x_n; A_n, h_n) \to \mu^*(x_0; A_0, h_0)$.

We denote submatrices of A_n and A_0 formed by the rows with indices in $J \subseteq \{1, ..., d_A\}$ by $A_{J,n}$ and $A_{J,0}$. Important for the following lemma is the fact that every element of h_n is nonnegative for all n.

[35] One can check that this definition of Kuratowski convergence is equivalent to other definitions given in, for example Aubin and Frankowska (1990).

Lemma 6. *If for all $J \subseteq \{1, \ldots, d_A\}$, $\text{rk}(A_{J,n}) = \text{rk}(A_{J,0})$ for all n, then $\text{poly}(A_n, h_n) \xrightarrow{K} \text{poly}(A_0, h_0)$.*

For any $d_A \times d_m$ matrix, A, and for any vector, g, let $J(x; A, g) = \{j \in \{1, \ldots, d_A\} : a_j' \mu^*(x; A, g) = g_j\}$. This generalizes the previous notation for active inequalities to make explicit the dependence on A and g. Also let $[A; B]$ denotes the vertical concatenation of two matrices, A and B.

Lemma 7. *Let A_n be a sequence of $d_A \times d_m$ matrices such that each row is either zero or belongs to the unit circle. Let g_n be a sequence of nonnegative d_A-vectors. Then, there exists a subsequence, n_q, a sequence of $d_B \times d_m$ matrices, B_q, and a sequence of nonnegative d_B-vectors h_q such that the following hold.*

(a) $A_{n_q} \to A_0$, $B_q \to B_0$, $g_{n_q} \to g_0$, *and* $h_q \to h_0$ *(some of the elements of g_0 and h_0 may be $+\infty$, in which case the convergence/divergence occurs elementwise).*

(b) $\text{poly}(A_{n_q}, g_{n_q}) \subseteq \text{poly}(B_q, h_q)$ *for all q.*

(c) *For all q and for all $x \in \text{poly}(A_{n_q}, g_{n_q})$,*

$$\text{rk}(I_{J(x;A_{n_q},g_{n_q})} A_{n_q}) = \text{rk}([I_{J(x;A_{n_q},g_{n_q})} A_{n_q}; I_{J(x;B_q,h_q)} B_q]).$$

(d) $\text{poly}([A_{n_q}; B_q], [g_{n_q}; h_q]) \xrightarrow{K} \text{poly}([A_0; B_0], [g_0; h_0])$ *as $q \to \infty$.*

Suppose that $\bar{j} \in J(x; A, g)$ and $a_{\bar{j}} \neq \mathbf{0}$. If such a $\bar{j}$ does not exist, let $\tau_j(x; A, g) = 0$ for all $j \in \{1, ..., d_A\}$. Otherwise, let

$$\tau_j(x; A, g) = \begin{cases} \frac{\|a_{\bar{j}}\|(g_j - a_j' \mu^*(x; A, g))}{\|a_{\bar{j}}\| \|a_j\| - a_{\bar{j}}' a_j} & \text{if } \|a_{\bar{j}}\| \|a_j\| \neq a_{\bar{j}}' a_j \\ \infty & \text{otherwise.} \end{cases} \tag{110}$$

Let $\tau(x; A, g) = \inf_{j \in \{1, \ldots, d_A\}} \tau_j(x; A, g)$. One can verify that the definition of $\tau(x; A, g)$ does not depend on which $\bar{j} \in J(x; A, g)$ is used to define it, when more than one is available. This definition coincides with the definition of $\hat{\tau}$ or $\tau(X)$, as used in the proof of Theorem 1, making explicit the dependence on A and g.

Lemma 8. *If $\text{poly}(A, g) \subseteq \text{poly}(B, h)$, then $\tau(x; A, g) = \tau(x; [A; B], [g; h])$ for all $x \in \mathbb{R}^{d_m}$.*

B.3 Proof of Theorems 2 and 3

Proof of Theorem 2

We verify the conditions of Theorem 3. We first show that Assumption 1 implies that Assumption 2(a) holds for any sequence $\{(\theta_n, F_n) : F_n \in \mathcal{F}, \theta_n \in \Theta_0(F_n)\}$. Fix an arbitrary sequence $\{(\theta_n, F_n) : F_n \in \mathcal{F}, \theta_n \in \Theta_0(F_n)\}$. Let $\Sigma_n = \mathrm{Var}_{F_n}(m(W_i, \theta_n))$, which does not depend on i due to Assumption 1(a). Let D_n be the diagonal matrix formed by the diagonal elements of Σ_n. By Assumption 1(b), D_n is invertible and thus we can define

$$\Omega_n = D_n^{-1/2} \Sigma_n D_n^{-1/2} = \mathrm{Corr}_{F_n}(m(W_i, \theta_n)). \tag{111}$$

The elements of $\Omega_n \in [-1, 1]$ which is a compact set. Thus, for any subsequence of $\{n\}$, there is a further subsequence $\{n_q\}$ such that

$$\Omega_{n_q} \to \Omega, \tag{112}$$

for some matrix Ω. By Assumption 1(c), Ω is positive definite.

Now consider an arbitrary vector $a \in \mathbb{R}^{d_m}$ such that $a'a = 1$, and consider the sequence of random variables:

$$\begin{aligned}
&n^{1/2} a' D_q^{-1/2} (\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_{n_q}} \overline{m}_{n_q}(\theta_{n_q})) \\
&= n^{-1/2} \sum_{i=1}^{n} a' D_q^{-1/2} (m(W_i, \theta_{n_q}) - \mathbb{E}_{F_{n_q}} m(W_i, \theta_{n_q})) \\
&\to_d N(0, a'\Omega a),
\end{aligned} \tag{113}$$

by the Lindeberg-Feller central limit theorem, where the Lindeberg condition holds because

$$\mathbb{E}_{F_{n_q}} |a' D_q^{-1/2} m(W_i, \theta_{n_q})|^{2+\epsilon} \leq \mathbb{E}_{F_{n_q}} [\sum_{j=1}^{d_m} a_j |m_j(W_i, \theta_{n_q}) / \sigma_{F_{n_q}, j}(\theta_{n_q})|^{2+\epsilon}] \leq M < \infty, \tag{114}$$

where the first inequality holds by the convexity of $g(x) = |x|^{2+\epsilon}$ and the second and the third inequalities hold by Assumption 1(d).

The Cramer-Wold device combined with (113) proves (106) in Assumption 2.

To show (107), consider that

$$D_q^{-1/2} \widehat{\Sigma}_{n_q}(\theta_{n_q}) D_q^{-1/2}$$

$$= n^{-1}\sum_{i=1}^{n} D_q^{-1/2}(m(W_i,\theta_{n_q}) - \mathbb{E}_{F_{n_q}} m(W_i,\theta_{n_q}))(m(W_i,\theta_{n_q}) - \mathbb{E}_{F_{n_q}} m(W_i,\theta_{n_q}))' D_q^{-1/2}$$
$$- D_q^{-1/2}(\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_{n_q}} m(W_i,\theta_{n_q}))(\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_{n_q}} m(W_i,\theta_{n_q}))' D_q^{-1/2}. \quad (115)$$

By Assumptions 1(a) and (d), the law of large numbers for rowwise i.i.d. triangular arrays applies and gives us

$$n^{-1}\sum_{i=1}^{n} D_q^{-1/2}(m(W_i,\theta_{n_q}) - \mathbb{E}_{F_{n_q}} m(W_i,\theta_{n_q}))(m(W_i,\theta_{n_q}) - \mathbb{E}_{F_{n_q}} m(W_i,\theta_{n_q}))' D_q^{-1/2}$$
$$\to_p \Omega, \quad (116)$$

and similarly

$$D_q^{-1/2}(\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_{n_q}} m(W_i,\theta_{n_q})) \to_p \mathbf{0}. \quad (117)$$

Thus, (107) is also verified.

Next, we show that Assumption 1, combined with the additional assumptions in Theorem 2(b), implies Assumption 2(b). First note that each element of Λ_q is either one or $\|e_j' A(\theta_{n_q}) D_q\|^{-1}$. By the common additional condition for Theorem 2(b,c), $\|e_j' A(\theta_{n_q}) D_q\|^{-1} \to \|e_j' A_\infty\|^{-1}$. Note that $e_j' A(\theta_{n_q}) D_q$ cannot go to zero if $e_j' A(\theta_{n_q}) \neq \mathbf{0}$ because that would violate the common additional condition for Theorem 2(b,c) for $J = \{j\}$. Therefore, there exists a further subsequence along which $\Lambda_q \to \Lambda_\infty$ for a positive definite diagonal matrix Λ_∞. Therefore, $\Lambda_q A(\theta_{n_q}) D_q \to \bar{A}_0 = \Lambda_\infty A_\infty$. Also note that for each $J \subseteq \{1, ..., d_A\}$, $\text{rk}(A_J(\theta_{n_q}) D_q) = \text{rk}(I_J A_\infty) = \text{rk}(I_J \bar{A}_0)$, where the first equality follows from the common additional condition for Theorem 2(b,c) and the second equality follows because each row of $\bar{A}_0$ is a positive scalar multiple of the corresponding row of A_∞. This verifies Assumption 2(b).

We also note that along every further subsequence, each diagonal element of Λ_q converges to a positive value. This implies that, for part (b), we have for every $j = 1, ..., d_A$,

$$\sqrt{n_q} e_j' \Lambda_q (A_j(\theta_{n_q}) \mathbb{E}_{F_{n_q}} \overline{m}_{n_q}(\theta_{n_q}) - b(\theta_{n_q})) \to 0. \quad (118)$$

Also, for part (c), we have for every $j \notin J$,

$$\sqrt{n_q} e_j' \Lambda_q (A_j(\theta_{n_q}) \mathbb{E}_{F_{n_q}} \overline{m}_{n_q}(\theta_{n_q}) - b(\theta_{n_q})) \to -\infty. \quad (119)$$

Also, for part (b), $\bar{A}_0 \neq \mathbf{0}$ is implied by $A_\infty \neq \mathbf{0}$ because Λ_∞ is positive definite.

Therefore, Theorem 2 follows from Theorem 3. □

Proof of Theorem 3

We first prove part (a). Let $\{\theta_n, F_n\}_{n=1}^{\infty}$ be an arbitrary sequence satisfying $F_n \in \mathcal{F}$ and $\theta_n \in \Theta_0(F_n)$ for all n. Let $\{n_m\}$ be an arbitrary subsequence of $\{n\}$. It is sufficient to show that there exists a further subsequence, $\{n_q\}$, such that as $q \to \infty$,

$$\liminf_{q\to\infty} \Pr{}_{F_{n_q}} \left(T_{n_q}(\theta_{n_q}) \leq \chi^2_{\hat{r},1-\hat{\beta}} \right) \geq 1-\alpha. \tag{120}$$

Fix an arbitrary subsequence, $\{n_m\}$. By Assumption 2(a), there exists a further subsequence, $\{n_q\}$, a sequence of positive definite matrices, D_q, and a positive definite correlation matrix, Ω_0, such that[36]

$$\sqrt{n_q} D_q^{-1/2}(\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_{n_q}} \overline{m}_{n_q}(\theta_{n_q})) \to_d Y \sim N(\mathbf{0}, \Omega_0), \text{ and} \tag{121}$$

$$D_q^{-1/2} \widehat{\Sigma}_{n_q}(\theta_{n_q}) D_q^{-1/2} \to_p \Omega_0. \tag{122}$$

We introduce some simplified notation. Let $\widehat{\Omega}_q = D_q^{-1/2}\widehat{\Sigma}_{n_q}(\theta_{n_q})D_q^{-1/2}$, $X = \Omega_0^{-1/2} Y \sim N(\mathbf{0}, I)$, $Y_q = \sqrt{n_q} D_q^{-1/2}(\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_{n_q}} \overline{m}_{n_q}(\theta_{n_q}))$, and $X_q = \widehat{\Omega}_q^{-1/2} Y_q$. Equations (121) and (122) imply that

$$X_q \to_d X \sim N(\mathbf{0}, I), \text{ and} \tag{123}$$

$$\widehat{\Omega}_q \to_p \Omega_0. \tag{124}$$

The remainder of the proof proceeds in four steps. (A) In the first step, the problem defined in (17) is transformed to include additional inequalities. (B) In the second step, notation is defined for partitioning $\mathbb{R}^{d_m}$ according to Lemma 1, for both finite q and the limit. (C) In the third step, the almost sure representation theorem is invoked on the convergence in (123) and (124). (D) In the final step, we show that (almost surely) the event $T_{n_q}(\theta_{n_q}) \leq \chi^2_{\hat{r},1-\hat{\beta}}$ eventually implies a limiting event based on X and Ω_0. This limiting event has probability at least $1-\alpha$ from Theorem 1.

(A) Consider the sequence of matrices $A(\theta_{n_q})D_q^{1/2}$. For each q, let Λ_q denote a $d_A \times d_A$ diagonal matrix with positive entries on the diagonal such that each row of $\Lambda_q A(\theta_{n_q})D_q^{1/2}$ is either zero or belongs to the unit circle. Such a Λ_q always exists by taking the diagonal element to be the inverse of the magnitude of the corresponding row of $A(\theta_{n_q})D_q$, if it is nonzero, and one otherwise. Let $g_q = \sqrt{n_q}\Lambda_q(b(\theta_{n_q}) - A(\theta_{n_q})\mathbb{E}_{F_{n_q}}\overline{m}_{n_q}(\theta_{n_q}))$. With this

[36] For notational simplicity, we denote all further subsequences by $\{n_q\}$

notation, we can write

$$T_{n_q}(\theta_{n_q}) = \inf_{y:\Lambda_q A(\theta_{n_q}) D_q^{1/2} y \le g_q} (Y_q - y)'\widehat{\Omega}_q^{-1}(Y_q - y), \tag{125}$$

which adds and subtracts $\mathbb{E}_{F_{n_q}}\overline{m}_{n_q}(\theta_{n_q})$ in the objective and applies the change of variables, $y = \sqrt{n_q} D_q^{-1/2}(\mu - \mathbb{E}_{F_{n_q}}\overline{m}_{n_q}(\theta_{n_q}))$.

We can apply Lemma 7 to $\Lambda_q A(\theta_{n_q}) D_q^{1/2}$ and g_q to get a further subsequence, n_q, a sequence of matrices, B_q, a sequence of vectors, h_q, matrices A_0 and B_0, and vectors g_0 and h_0, satisfying Lemma 7(a-d). Let

$$\bar{A}_q = \begin{bmatrix} \Lambda_q A(\theta_{n_q}) D_q^{1/2} \\ B_q \end{bmatrix} \text{ and } \bar{h}_q = \begin{bmatrix} g_q \\ h_q \end{bmatrix}, \tag{126}$$

and similarly for $\bar{A}_0$ and $\bar{h}_0$. Let $d_{\bar{A}} = d_A + d_B$. We have that

$$T_{n_q}(\theta_{n_q}) = \inf_{y:\bar{A}_q y \le \bar{h}_q} (Y_q - y)'\widehat{\Omega}_q^{-1}(Y_q - y) \tag{127}$$

$$= \inf_{t:\bar{A}_q \widehat{\Omega}_q^{1/2} t \le \bar{h}_q} (X_q - t)'(X_q - t), \tag{128}$$

where the first equation follows from Lemma 7(b) and the second equation follows from the change of variables $t = \widehat{\Omega}_q^{-1/2} y$.

Equation (128) has changed the problem by adding additional inequalities. We verify that the rank of the active inequalities is unchanged. For any positive definite matrix, Ω, let $\bar{J}_q(x, \Omega)$ be the set of indices for the active inequalities in the problem:

$$\inf_{y:\Lambda_q A(\theta_{n_q}) D_q^{1/2} y \le g_q} (x - y)'\Omega^{-1}(x - y). \tag{129}$$

Recall that $\widehat{J}$ is the set of active inequalities for the problem defined in (17), which is equal to $\bar{J}_q(Y_q, \widehat{\Omega}_q)$ by a change of variables. Similarly, let $J_q(x, \Omega)$ be the set of active inequalities in the problem:

$$\inf_{t:\bar{A}_q \Omega^{1/2} t \le \bar{h}_q} (x - t)'(x - t). \tag{130}$$

Also let $t_q^*(x, \Omega)$ denote the unique minimizer. We have that for any $y \in \mathbb{R}^{d_m}$ and for any positive definite Ω,

$$\begin{aligned} \text{rk}(A_{\bar{J}_q(y,\Omega)}(\theta_{n_q})) &= \text{rk}(I_{\bar{J}_q(y,\Omega)} \Lambda_q A(\theta_{n_q}) D_q^{1/2}) \\ &= \text{rk}(I_{J_q(\Omega^{-1/2} y, \Omega)} \bar{A}_q) = \text{rk}(I_{J_q(\Omega^{-1/2} y, \Omega)} \bar{A}_q \Omega^{1/2}), \end{aligned} \tag{131}$$

where the first equality follows because Λ_q is diagonal with positive entries on the diagonal and D_q is positive definite, the second equality follows by Lemma 7(c), and the final equality follows from the fact that Ω is positive definite.

Before proceeding to the next step, we simplify the rank calculation by taking a further subsequence. Notice that for each $J \subseteq \{1, ..., d_{\bar{A}}\}$, $\text{rk}(I_J \bar{A}_q) \in \{1, ..., d_m\}$. We can denote it by r_J^q, and then take a subsequence, n_q, so that for all J, r_J^q does not depend on q. Similarly, we define $r_J^\infty = \text{rk}(I_J \bar{A}_0)$. Note that by the convergence of $\bar{A}_q$ to $\bar{A}_0$, $r_J^q \geq r_J^\infty$ for all J.

(B) For any positive definite $d_m \times d_m$ matrix, Ω, and for every $J \subseteq \{0, 1, ..., d_{\bar{A}}\}$, let

$$
\begin{aligned}
A^q(\Omega) &= \bar{A}_q \Omega^{1/2} \\
a_\ell^{q\prime}(\Omega) &= \ell^{\text{th}} \text{ row of } A^q(\Omega) \\
C^q(\Omega) &= \{x \in \mathbb{R}^{d_m} : a_\ell^{q\prime}(\Omega) x \leq \bar{h}_{\ell,q} \text{ for all } \ell = 1, ..., d_{\bar{A}}\} \\
C_J^q(\Omega) &= \{x \in C^q(\Omega) : a_\ell^{q\prime}(\Omega) x = \bar{h}_{\ell,q} \text{ for all } \ell \in J \text{ and } a_\ell^{q\prime}(\Omega) x < \bar{h}_{\ell,q} \text{ for all } \ell \in J^c\} \\
V_J^q(\Omega) &= \left\{ \sum_{\ell \in J} v_\ell a_\ell^q(\Omega) : v_\ell \in \mathbb{R}, v_\ell \geq 0 \right\}, \text{ and} \\
K_J^q(\Omega) &= C_J^q(\Omega) + V_J^q(\Omega).
\end{aligned} \tag{132}
$$

Furthermore, for every $J \subseteq \{1, ..., d_{\bar{A}}\}$, let $P_J^q(\Omega)$ denote the projection onto $\text{span}(V_J^q(\Omega))$, and let $M_J^q(\Omega)$ denote its orthogonal projection. There exists a $\kappa_J^q(\Omega) \in \text{span}(V_J^q(\Omega))$ such that for every $x \in C_J^q(\Omega)$, $P_J^q(\Omega) x = \kappa_J^q(\Omega)$. This follows because for two $x_1, x_2 \in C_J^q(\Omega)$, and for any $v \in \text{span}(V_J^q(\Omega))$, $v'(x_1 - x_2) = 0$, which implies that $P_J^q(\Omega)(x_1 - x_2) = 0$.

For every given Ω, we can apply Lemma 1 to the objects defined in (132). This implies

(a) if $x \in K_J^q(\Omega)$ then $x - t_q^*(x, \Omega) \in V_J^q(\Omega)$ and $t_q^*(x, \Omega) \in C_J^q(\Omega)$,

(b) the sets $K_J^q(\Omega)$ for all $J \subseteq \{1, \ldots, d_{\bar{A}}\}$ form a partition of $\mathbb{R}^{d_m}$, and

(c) for each $J \subseteq \{1, \ldots, d_{\bar{A}}\}$, we have $x \in K_J^q(\Omega)$ iff $J = J_q(x, \Omega)$.

These properties imply that, for all $x \in K_J^q(\Omega)$, we can write

$$
P_J^q(\Omega) x = P_J^q(\Omega)(x - t_q^*(x, \Omega)) + P_J^q(\Omega) t_q^*(x, \Omega) = x - t_q^*(x, \Omega) + \kappa_J^q(\Omega), \tag{133}
$$

where the second equality follows by (a) and the definition of $\kappa_J^q(\Omega)$. Then, we can also write $M_J^q(\Omega) x = x - P_J^q(\Omega) x = t_q^*(x, \Omega) - \kappa_J^q(\Omega)$.

Let $r^q(x,\Omega) = \text{rk}(\bar{A}_{J_q(x,\Omega),q})$. When $r^q(x,\Omega) = 1$, we can define

$$\tau_j^q(x,\Omega) = \begin{cases} \frac{\|a_{\bar{j}}^q(\Omega)\|(\bar{h}_{j,q} - a_j^q(\Omega)' t_q^*(x,\Omega))}{\|a_j^q(\Omega)\|\|a_{\bar{j}}^q(\Omega)\| - a_j^q(\Omega)' a_{\bar{j}}^q(\Omega)} & \text{if } \|a_j^q(\Omega)\|\|a_{\bar{j}}^q(\Omega)\| \neq a_j^q(\Omega)' a_{\bar{j}}^q(\Omega) \\ \infty & \text{else} \end{cases}, \tag{134}$$

where $\bar{j} \in J_q(x,\Omega)$ such that $a_{\bar{j}}^q(\Omega) \neq \mathbf{0}$. We also let $\tau^q(x,\Omega) = \inf_{j=1,\dots,d_{\bar{A}}} \tau_j^q(x,\Omega)$, and $\beta^q(x,\Omega) = 2\alpha\Phi(\tau^q(x,\Omega))$. When $r^q(x,\Omega) \neq 1$, let $\tau_j^q(x,\Omega) = 0$, so that $\beta^q(x,\Omega) = \alpha$. Note that $\hat{\beta} = \beta^q(X_q, \widehat{\Omega}_q)$, where the addition of extra inequalities via Lemma 7 has no effect on $\hat{\beta}$ or $\hat{\tau}$ because of Lemma 8, where the condition is satisfied by Lemma 7(b).

We define similar notation for the limiting objects. Let $J^\infty = \{\ell \in \{1, ..., d_{\bar{A}}\} : \bar{h}_{\ell,0} < \infty\}$. These are the indices for the inequalities that are "close-to-binding." For any positive definite matrix, Ω, let $A^\infty(\Omega)$ denote the matrix formed by the rows of $\bar{A}_0\Omega^{1/2}$ associated with the indices in J^∞. For notational simplicity, we refer to the rows of $A^\infty(\Omega)$ using $\ell \in J^\infty$ even though the matrix $A^\infty(\Omega)$ has been compressed.

For every $J \subseteq J^\infty$, let

$$\begin{aligned}
&a_\ell^{\infty\prime}(\Omega) = \ell^{\text{th}} \text{ row of } A^\infty(\Omega) \text{ for } \ell \in J^\infty \\
&C^\infty(\Omega) = \{x \in \mathbb{R}^{d_m} : a_\ell^\infty(\Omega)' x \leq \bar{h}_{\ell,0} \text{ for all } \ell \in J^\infty\} \\
&C_J^\infty(\Omega) = \{x \in C^\infty(\Omega) : a_\ell^\infty(\Omega)' x = \bar{h}_{\ell,0}\ \forall \ell \in J \text{ and } a_\ell^\infty(\Omega)' x < \bar{h}_{\ell,0}\ \forall \ell \in J^\infty / J\} \\
&V_J^\infty(\Omega) = \left\{ \sum_{\ell \in J} v_\ell a_\ell^\infty(\Omega) : v_\ell \in \mathbb{R}, v_\ell \geq 0 \right\}, \text{ and} \\
&K_J^\infty(\Omega) = C_J^\infty(\Omega) + V_J^\infty(\Omega).
\end{aligned} \tag{135}$$

Furthermore, for every $J \subseteq J^\infty$, let $P_J^\infty(\Omega)$ denote the projection onto $\text{span}(V_J^\infty(\Omega))$. There exists a $\kappa_J^\infty(\Omega) \in \text{span}(V_J^\infty(\Omega))$ such that for every $x \in C_J^\infty(\Omega)$, $P_J^\infty(\Omega)x = \kappa_J^\infty(\Omega)$. This follows because for two $x_1, x_2 \in C_J^\infty(\Omega)$, and for any $v \in \text{span}(V_J^\infty(\Omega))$, $v'(x_1 - x_2) = 0$, which implies that $P_J^\infty(\Omega)(x_1 - x_2) = \mathbf{0}$.

Let h^∞ denote the vector formed from the elements of $\bar{h}_0$ that are finite. Let $J^\infty(x,\Omega)$ be the indices for the binding inequalities in the problem:

$$\inf_{t : A^\infty(\Omega) t \leq h^\infty} (x - t)'(x - t). \tag{136}$$

Also let $t_\infty^*(x,\Omega)$ denote the unique minimizer. We can apply Lemma 1 to the objects defined in (135). This implies that

(a$^\infty$) if $x \in K_J^\infty(\Omega)$ then $x - t_\infty^*(x,\Omega) \in V_J^\infty(\Omega)$ and $t_\infty^*(x,\Omega) \in C_J^\infty(\Omega)$, (b$^\infty$) the set of all $K_J^\infty(\Omega)$ form a partition of $\mathbb{R}^{d_m}$, and

(c$^\infty$) for each $J \subseteq J^\infty$, we have $x \in K_J^\infty(\Omega)$ iff $J = J^\infty(x,\Omega)$.

Let $r^\infty(x,\Omega) = \text{rk}(A^\infty_{J_\infty(x,\Omega)})$. When $r^\infty(x,\Omega) = 1$, we can define

$$\tau_j^\infty(x,\Omega) = \begin{cases} \frac{\|a_{\bar{j}}^\infty(\Omega)\|(\bar{h}_{j,0} - a_j^\infty(\Omega)' t_\infty^*(x,\Omega))}{\|a_j^\infty(\Omega)\|\|a_{\bar{j}}^\infty(\Omega)\| - a_j^\infty(\Omega)' a_{\bar{j}}^\infty(\Omega)} & \text{if } \|a_j^\infty(\Omega)\|\|a_{\bar{j}}^\infty(\Omega)\| \neq a_j^\infty(\Omega)' a_{\bar{j}}^\infty(\Omega) \\ \infty & \text{else} \end{cases}, \tag{137}$$

where $\bar{j} \in J^\infty(x,\Omega)$ such that $a_{\bar{j}}^\infty(\Omega) \neq 0$. We also let $\tau^\infty(x,\Omega) = \inf_{j=J^\infty} \tau_j^\infty(x,\Omega)$, and $\beta^\infty(x,\Omega) = 2\alpha\Phi(\tau^\infty(x,\Omega))$. When $r^\infty(x,\Omega) \neq 1$, let $\tau_j^\infty(x,\Omega) = 0$, so that $\beta^\infty(x,\Omega) = \alpha$.

Before proceeding to the next step, consider $M_J^q(\Omega_0)$, which is a sequence of projection matrices in $\mathbb{R}^{d_m}$ onto a space of dimension $d_m - r_J^q$. Since the space of such matrices is compact, we can find a subsequence, n_q, such that for all $J \subseteq \{1,...,d_{\bar{A}}\}$, $M_J^q(\Omega_0) \to M_J^N$, where M_J^N is a projection matrix onto a subspace, N_J, of dimension $d_m - r_J^q$.[37] Furthermore, for any sequence of positive definite matrices such that $\Omega_q \to \Omega_0$, we have $M_J^q(\Omega_q) \to M_J^N$. This follows because, if we let E_q denote a $d_m \times r_J^q$ matrix whose columns form an orthonormal basis for $\text{span}(V_J^q(\Omega_0))$ (which is the range of $P_J^q(\Omega_0)$) then for any positive definite matrix, Ω, the columns of $\Omega^{1/2}\Omega_0^{-1/2}E_q$ form a basis for $\text{span}(V_J^q(\Omega))$, which implies that

$$\begin{aligned} M_J^q(\Omega_q) &= I_{d_m} - \Omega_q^{1/2}\Omega_0^{-1/2}E_q(E_q'\Omega_0^{-1/2}\Omega_q\Omega_0^{-1/2}E_q)^{-1}E_q'\Omega_0^{-1/2}\Omega_q^{1/2} \\ &= I_{d_m} - E_q(E_q'E_q)^{-1}E_q' + o(1) = M_J^q(\Omega_0) + o(1). \end{aligned} \tag{138}$$

(C) Next, we invoke the almost sure representation theorem on the convergence in (123) and (124).[38] Then, we can treat the convergence in (123) and (124) as holding almost surely.[39] For the rest of the proof of part (a), consider $A^\infty(\Omega)$, $P_J^\infty(\Omega)$, $\kappa_J^\infty(\Omega)$, and the objects defined in (135) and let the objects without the argument (Ω) denote the objects evaluated at Ω_0. For example, $A^\infty = A^\infty(\Omega_0)$.

We now construct an event, $\Xi \subseteq \mathbb{R}^{d_m}$, such that $\Pr(X \in \Xi) = 1$. For every $L \subseteq J^\infty$, let

$$V_{L+}^\infty = \{x \in V_L^\infty | \forall L' \subseteq L, \text{ if } r_{L'}^q < r_L^\infty \text{ then } M_{L'}^N x \neq 0\}. \tag{139}$$

For each $L \subseteq J^\infty$ such that $r_L^\infty > 0$, let

$$\Xi_L = \{x \in K_L^\infty : P_L^\infty x - \kappa_L^\infty \in V_{L+}^\infty \text{ and } (P_L^\infty x - \kappa_L^\infty)'(P_L^\infty x - \kappa_L^\infty) \neq \chi^2_{r_L^\infty, 1-\beta^\infty(x)}\}. \tag{140}$$

[37]Recall that r_J^q does not depend on q due to the construction of the subsequence $\{n_q\}$.

[38]See van der Vaart and Wellner (1996), Theorem 1.10.3, for the a.s. representation theorem.

[39]This can be formalized by defining random variables, $\ddot{X}_q$, $\ddot{X}$, and $\ddot{\Omega}_q$, satisfying $\ddot{X}_q =_d X_q$, $\ddot{X} =_d X$, $\ddot{\Omega}_q =_d \widehat{\Omega}_q$, $\ddot{X}_q \to_{a.s.} \ddot{X}$, and $\ddot{\Omega}_q \to_{a.s.} \Omega_0$.

Since $r_L^\infty > 0$, $P_L^\infty X \sim N(0, P_L^\infty)$, which is absolutely continuous on span(V_L^∞), and therefore the probability that $P_L^\infty X - \kappa_L^\infty$ lies in any one of the finitely many subspaces, $null(M_{L'}^N) = \{x \in \mathbb{R}^{d_m} : M_{L'}^N x = 0\}$, each with dimension $r_{L'}^q < r_L^\infty$, is zero. Also, $(P_L^\infty X - \kappa_L^\infty)'(P_L^\infty X - \kappa_L^\infty)$ is absolutely continuous because it can be written as the sum of rk(A_L^∞) squared normal random variables. Also, $\chi^2_{r_L^\infty, 1-\beta^\infty(X)}$ depends on X only through $M_L^\infty X$, which is independent of $P_L^\infty X$. Therefore, for each fixed $M_L^\infty X$, the conditional probability that $(P_L^\infty X - \kappa_L^\infty)'(P_L^\infty X - \kappa_L^\infty) = \chi^2_{r_L^\infty, 1-\beta^\infty(M_L^\infty X + \kappa_L^\infty)}$ is zero. This implies that the unconditional probability is also zero. Therefore,

$$\Pr(P_L^\infty X - \kappa_L^\infty \in V_L^\infty / V_{L+}^\infty \text{ or } (P_L^\infty X - \kappa_L^\infty)'(P_L^\infty X - \kappa_L^\infty) = \chi^2_{r_L^\infty, 1-\beta^\infty(X)}) = 0. \tag{141}$$

For $L \subseteq J^\infty$ such that rk$(A_L^\infty) = 0$, let $\Xi_L = K_L^\infty$. Then, let $\Xi = \cup_{L \subseteq J^\infty} \Xi_L$. Therefore, by property (b$^\infty$) and equation (141), $\Pr(X \in \Xi) = 1$.

(D) We consider the set of all sequences such that $x_q \to x_\infty \in \Xi$ and $\Omega_q \to \Omega_0$. By the definition of Ξ and the almost sure convergence of $(X_q, \widehat{\Omega}_q)$, these sequences occur with probability one. Fix such a sequence for the remainder of the proof of part (a). For this step, consider $A^q(\Omega)$, $P_J^q(\Omega)$, and the objects defined in (132), and let the objects without the argument (Ω) denote the objects evaluated at Ω_q.

Below we show that for each sequence,

$$\mathbb{1}\{\|x_q - t_q^*(x_q)\|^2 \leq \chi^2_{r^q(x_q), 1-\beta^q(x_q)}\} \geq \mathbb{1}\{\|x_\infty - t_\infty^*(x_\infty)\|^2 \leq \chi^2_{r^\infty(x_\infty), 1-\beta^\infty(x_\infty)}\} \tag{142}$$

eventually. Notice that by (128) and (130), the left hand side is equal to $\mathbb{1}\{T_{n_q}(\theta_{n_q}) \leq \chi^2_{\hat{r}, 1-\hat{\beta}}\}$ with $(X_q, \widehat{\Omega}_q)$ replaced by (x_q, Ω_q). If (142) holds, then by the bounded convergence theorem,

$$\liminf_{q \to \infty} \Pr{}_{F_{n_q}}(T_{n_q}(\theta_{n_q}) \leq \chi^2_{\hat{r}, 1-\hat{\beta}}) \geq \Pr(\|X - t_\infty^*(X, \Omega_0)\|^2 \leq \chi^2_{r^\infty(X, \Omega_0), 1-\beta^\infty(X, \Omega_0)}). \tag{143}$$

Also,

$$\Pr(\|X - t_\infty^*(X, \Omega_0)\|^2 \leq \chi^2_{r^\infty(X, \Omega_0), 1-\beta^\infty(X, \Omega_0)}) \geq 1 - \alpha \tag{144}$$

by Theorem 1(a), which applies with $n = 1$ and $\overline{m}_n(\theta) = X$ because $t_\infty^*(X, \Omega_0) = P_{C^\infty} X$, where P_{C^∞} is the projection of X onto $C^\infty = \{\mu \in \mathbb{R}^{d_m} : A^\infty(\Omega_0)\mu \leq h^\infty\}$. Together, (143) and (144) imply (120) for the given subsequence, n_q.

To finish the proof of part (a), we prove (142). Let L^∞ be the subset of J^∞ for which $x_\infty \in K_{L^\infty}^\infty$. We show that

$$1 = \sum_{L \subseteq \{1, \ldots, d_{\bar{A}}\}} \mathbb{1}\{x_q \in K_L^q\} = \sum_{L \subseteq L^\infty : r_L^q \geq r_{L^\infty}^\infty} \mathbb{1}\{x_q \in K_L^q\} \tag{145}$$

eventually. Property (b) above implies that the first equality holds at every q. Thus it is sufficient to show the second equality. Note that $t_q^*(x_q) \to t_\infty^*(x_\infty)$ by Lemma 5, using Lemma 7(d) to verify the condition. For the second equality, it is sufficient to show that, for all $L \notin \{L \subseteq L^\infty : r_L^q \geq r_{L^\infty}^\infty\}$, $x_q \notin K_L^q$ eventually. Specifically, we consider three cases: (I) $L \not\subseteq J^\infty$, (II) $L \subseteq J^\infty$ but $L \not\subseteq L^\infty$, (III) $L \subseteq L^\infty$ but $r_L^q < r_{L^\infty}^\infty$.

(I) Let $L \not\subseteq J^\infty$. Then, there exists a $\ell \in L$ such that $\bar{h}_{\ell,q} \to \infty$. Then $a_\ell^{q\prime} t_q^*(x_q) < \bar{h}_{\ell,q}$ eventually because $t_q^*(x_q) \to t_\infty^*(x_\infty)$. This implies that $t_q^*(x_q) \notin C_L^q$, and therefore by (a), $x_q \notin K_L^q$ eventually.

(II) Let $L \subseteq J^\infty$ but $L \not\subseteq L^\infty$. Then, there exists a $\ell \in L$ such that $a_\ell^{\infty\prime} t_\infty^*(x_\infty) < \bar{h}_{\ell,0}$. By the fact that $a_\ell^{q\prime} t_q^*(x_q) \to a_\ell^{\infty\prime} t_\infty^*(x_\infty)$ and $\bar{h}_{\ell,q} \to \bar{h}_{\ell,0}$, we have that $a_\ell^{q\prime} t_q^*(x_q) < \bar{h}_{\ell,q}$ eventually. This implies that $z_q^*(x_q) \notin C_L^q$, and therefore by property (a) above, $x_q \notin K_L^q$ eventually.

(III) Let $L \subseteq L^\infty$ such that $r_L^q < r_{L^\infty}^\infty$. This case is impossible if $r_{L^\infty}^\infty = 0$. Thus we only need to consider $r_{L^\infty}^\infty > 0$. Note that $x_\infty - t_\infty^*(x_\infty) = P_{L^\infty}^\infty x_\infty - \kappa_{L^\infty}^\infty$ by property (a^∞) above. Also, by the definition of Ξ we have $x_\infty \in \Xi_{L^\infty}$, which implies that $x_\infty - t_\infty^*(x_\infty) \in V_{L^\infty +}^\infty$, which in turn means that $M_L^N(x_\infty - t_\infty^*(x_\infty)) \neq \mathbf{0}$. By the convergence, $M_L^q(x_q - t_q^*(x_q)) \to M_L^N(x_\infty - t_\infty^*(x_\infty))$, we have that $M_L^q(x_q - t_q^*(x_q)) \neq \mathbf{0}$ eventually. However, if $x_q \in K_L^q$, then by property (a) above, $x_q - t_q^*(x_q) \in V_L^q$, which implies that $M_L^q(x_q - t_q^*(x_q)) = \mathbf{0}$. This means that $x_q \notin K_L^q$ eventually. Therefore, (145) holds eventually.

We next show that

$$\limsup_{q\to\infty} \beta^q(x_q) \leq \beta^\infty(x_\infty). \tag{146}$$

It is sufficient to show that for every subsequence, there exists a further subsequence such that (146) holds. Thus, it is without loss of generality to suppose that $\beta^q := \beta^q(x_q)$ converges. If $\beta^q \to \alpha$, then (146) holds simply because $\beta^\infty(x_\infty) \geq \alpha$. If $\lim_{q\to\infty} \beta^q > \alpha$, then for every q large enough, there exists a $\bar{j}_q$ such that $a_{\bar{j}_q}^q \neq \mathbf{0}$ and $a_{\bar{j}_q}^q{}' x_q = \bar{h}_{\bar{j}_q,q}$. We can take a further subsequence so that $\bar{j}_q$ does not vary with q (and denote it by $\bar{j}$). Then $\lim_{q\to\infty} a_{\bar{j}}^q = a_{\bar{j}}^\infty \neq \mathbf{0}$ because each row of $\bar{A}_q$ is either zero or belongs to the unit circle. Also, the fact that $\bar{h}_{\bar{j},q} = a_{\bar{j}}^{q\prime} x_q \to a_{\bar{j}}^{\infty\prime} x_\infty = \bar{h}_{\bar{j},0}$ implies that $\bar{j} \in J^\infty$. Thus, $\bar{j}$ can be used to define $\tau_j^\infty(x_\infty)$.

Take $j \in J^\infty$, and consider two cases. (i) For j such that $\|a_{\bar{j}}^\infty\| \|a_j^\infty\| = a_j^{\infty\prime} a_{\bar{j}}^\infty$, we have $\tau_j^\infty(t_\infty^*(x_\infty); \bar{A}_0\Omega^{1/2}, \bar{h}_0) = \infty$. (ii) For j such that $\|a_{\bar{j}}^\infty\| \|a_j^\infty\| \neq a_j^{\infty\prime} a_{\bar{j}}^\infty$, we have

$$\begin{aligned} \tau_j^q(x_q) &= \frac{\|a_{\bar{j}}^q\|(\bar{h}_{j,q} - a_j^{q\prime} t_q^*(x_q))}{\|a_j^q\| \|a_{\bar{j}}^q\| - a_j^{q\prime} a_{\bar{j}}^q} \\ &\to \frac{\|a_{\bar{j}}^\infty\|(\bar{h}_{j,0} - a_j^{\infty\prime} t_\infty^*(x_\infty)}{\|a_j^\infty\| \|a_{\bar{j}}^\infty\| - a_j^{\infty\prime} a_{\bar{j}}^\infty} = \tau_j^\infty(x_\infty), \end{aligned} \tag{147}$$

which uses $a_{\bar{j}}^q \to a_{\bar{j}}^\infty$, $a_j^q \to a_j^\infty$, $\bar{h}_{j,q} \to \bar{h}_{j,0}$, and $t_q^*(x_q) \to t_\infty^*(x_\infty)$ by Lemma 5 and Lemma

7(d). Therefore,

$$\begin{aligned}\lim_{q\to\infty}\tau^q(x_q) &= \lim_{q\to\infty}\inf_{j\in\{1,\dots,d_{\bar{A}}\}}\tau^q_j(x_q)\\ &\le \lim_{q\to\infty}\inf_{\{j\in J^\infty:\|a^\infty_{\bar{j}}\|\|a^\infty_j\|\ne a^{\infty\prime}_j a^\infty_{\bar{j}}\}}\tau^q_j(x_q)\\ &= \inf_{\{j\in J^\infty:\|a^\infty_{\bar{j}}\|\|a^\infty_j\|\ne a^{\infty\prime}_j a^\infty_{\bar{j}}\}}\tau^\infty_j(x_\infty)\\ &= \inf_{j\in J^\infty}\tau^\infty_j(x_\infty)=\tau^\infty(x_\infty).\end{aligned} \tag{148}$$

This shows that (146) holds.

We now verify (142). Notice that

$$\begin{aligned}&\mathbb{1}\{\|x_q-t^*_q(x_q)\|^2\le\chi^2_{r^q(x_q),1-\beta^q(x_q)}\}\\ =&\sum_{J\subseteq\{1,\dots,d_{\bar{A}}\}}\mathbb{1}\{\|x_q-t^*_q(x_q)\|^2\le\chi^2_{r^q(x_q),1-\beta^q(x_q)}\}\mathbb{1}\{x_q\in K^q_J\}\\ =&\sum_{L\subseteq L^\infty:r^q_L\ge r^\infty_{L^\infty}}\mathbb{1}\{\|x_q-t^*_q(x_q)\|^2\le\chi^2_{r^q(x_q),1-\beta^q(x_q)}\}\mathbb{1}\{x_q\in K^q_L\}\\ =&\sum_{L\subseteq L^\infty:r^q_L\ge r^\infty_{L^\infty}}\mathbb{1}\{\|x_q-t^*_q(x_q)\|^2\le\chi^2_{r^q_L,1-\beta^q(x_q)}\}\mathbb{1}\{x_q\in K^q_L\}\\ \ge&\sum_{L\subseteq L^\infty:r^q_L\ge r^\infty_{L^\infty}}\mathbb{1}\{\|x_q-t^*_q(x_q)\|^2\le\chi^2_{r^\infty_{L^\infty},1-\beta^q(x_q)}\}\mathbb{1}\{x_q\in K^q_L\} &(149)\\ \ge&\mathbb{1}\{\|x_\infty-t^*_\infty(x_\infty)\|^2\le\chi^2_{r^\infty_{L^\infty},1-\beta^\infty(x_\infty)}\}\sum_{L\subseteq L^\infty:r^q_L\ge r^\infty_{L^\infty}}\mathbb{1}\{x_q\in K^q_L\} &(150)\\ =&\mathbb{1}\{\|x_\infty-t^*_\infty(x_\infty)\|^2\le\chi^2_{r^\infty_{L^\infty},1-\beta^\infty(x_\infty)}\}\mathbb{1}\{x_\infty\in K^\infty_{L^\infty}\}\\ =&\sum_{J\subseteq J^\infty}\mathbb{1}\{\|x_\infty-t^*_\infty(x_\infty)\|^2\le\chi^2_{r^\infty_J,1-\beta^\infty(x_\infty)}\}\mathbb{1}\{x_\infty\in K^\infty_J\}\\ =&\sum_{J\subseteq J^\infty}\mathbb{1}\{\|x_\infty-t^*_\infty(x_\infty)\|^2\le\chi^2_{r^\infty(x_\infty),1-\beta^\infty(x_\infty)}\}\mathbb{1}\{x_\infty\in K^\infty_J\}\\ =&\mathbb{1}\{\|x_\infty-t^*_\infty(x_\infty)\|^2\le\chi^2_{r^\infty(x_\infty),1-\beta^\infty(x_\infty)}\}, &(151)\end{aligned}$$

where: the first equality follows from property (b); the second equality follows from (145); the third equality follows from property (c); the first inequality follows because $r^q_L\ge r^\infty_{L^\infty}$; the second inequality must hold eventually because, when $r^\infty_{L^\infty}>0$,

$$\|x_q-t^*_q(x_q)\|^2\to\|x_\infty-t^*_\infty(x_\infty)\|^2=\|P^\infty_{L^\infty}x_\infty-\kappa^\infty_{L^\infty}\|^2\ne\chi^2_{r^\infty_{L^\infty},1-\beta^\infty(x_\infty)} \tag{152}$$

and

$$\liminf_{q\to\infty} \chi^2_{r^\infty_{L^\infty},1-\beta^q(x_q)} \geq \chi^2_{r^\infty_{L^\infty},1-\beta^\infty(x_\infty)} \tag{153}$$

(the equality follows from property (a$^\infty$) above because $x_\infty \in K^\infty_{L^\infty}$, the $\neq$ follows because $x \in \Xi_{L^\infty}$, and the $\geq$ follows from (146)), and when $r^\infty_{L^\infty} = 0$, $A^\infty_{L^\infty} = \mathbf{0}$, we have $A^q_{L^\infty} = \mathbf{0}$ eventually (because each row of A^∞ either belongs to the unit circle or is zero), and therefore, $x_q \in K^q_L$ for $L \subseteq K^\infty$ implies $x_q - t^*_q(x_q) = \mathbf{0}$ eventually; the fourth equality follows from (145) and $x_\infty \in K^\infty_{L^\infty}$; the fifth equality follows because all the terms with $J \neq L^\infty$ are zero; the sixth equality follows from (c$^\infty$); and the final equality follows from (a$^\infty$). This verifies (142), proving part (a).

Next, we prove part (b). Consider the sequence $\{F_n, \theta_n\}^\infty_{n=1}$. It is sufficient to show that for every subsequence, n_m, there exists a further subsequence, n_q, such that

$$\lim_{q\to\infty} \Pr{}_{F_{n_q}}(T_{n_q}(\theta_{n_q}) \leq \chi^2_{\hat{r},1-\hat{\beta}}) = 1 - \alpha. \tag{154}$$

The proof follows that of part (a) with the following changes.

(i) The augmentation of the inequalities with additional inequalities defined by B_q and h_q in (126) is no longer needed. We can take $\bar{A}_q = \Lambda_q A(\theta_{n_q}) D_q^{1/2}$ and $\bar{h}_q = g_{n_q}$.

(ii) Note that for each $j \in \{1, ..., d_A\}$, $g_{j,q}$ (which is equal to $\bar{h}_{j,q}$) is either zero or

$$\sqrt{n_q}\frac{b_j(\theta_{n_q}) - a_j(\theta_{n_q})'\mathbb{E}_{F_{n_q}}\overline{m}_n(\theta_{n_q})}{\|a_j(\theta_{n_q})'D_q\|} \to 0, \tag{155}$$

by assumption. Thus, $\bar{h}_0 = \mathbf{0}_{d_A}$.

(iii) Without B_q, (131) still holds without appealing to Lemma 7 by Assumption 2(b).

(iv) By Assumption 2(b), for all $J \subseteq \{1, ..., d_A\}$, $r^q_J = r^\infty_J$ for all q.

(v) The appeal to Lemma 7 below (145) is replaced by Lemma 6 to get $t^*_q(x_q) \to t^*_\infty(x_\infty)$.

(vi) The expression in (146) becomes

$$\lim_{q\to\infty} \beta^q(x_q) = \beta^\infty(x_\infty). \tag{156}$$

To show this, we consider two cases. In the first case, we have $\text{rk}(A^\infty_{L^\infty}) = 0$. Then by (145), $x_q \in K^q_L$ for some L such that $\text{rk}(A^q_L) = 0$ eventually. This implies that $\beta^q(x_q) = \alpha = \beta^\infty(x_\infty)$ eventually. In the second case, we have $\text{rk}(A^\infty_{L^\infty}) \geq 1$. Then by (145), $x_q \in K^q_L$ for some L such that $\text{rk}(A^q_L) \geq 1$ eventually. That is, for large enough q, there exists a $\bar{j}_q$ such that $a^{q}_{\bar{j}_q}{}' t^*_q(x_q) = \bar{h}_{\bar{j}_q,q}$. By a subsequencing argument, we can suppose $\bar{j}_q$ does not depend on q, and denote it by $\bar{j}$. If $\|a^\infty_{\bar{j}}\|\|a^\infty_j\| = a^{\infty}_j{}' a^\infty_{\bar{j}}$, then $\tau^\infty_j(x_\infty) = \infty$ and $\tau^q_j(x_q) = \infty$, using the fact that $\|a_j^q\|\|a_{\bar{j}}^q\| = a_j^{q\prime} a_{\bar{j}}^q$, which follows from the fact that $\text{rk}(A_{\{j,\bar{j}\}}^q) = \text{rk}(A_{\{j,\bar{j}\}}^\infty)$. If $\|a_{\bar{j}}^\infty\|\|a_j^\infty\| \neq a_j^{\infty\prime} a_{\bar{j}}^\infty$, then the convergence in (147) continues to hold by appealing to Lemma 6 instead of Lemma 7(d) to get $t_q^*(x_q) \to t_\infty^*(x_\infty)$. This implies that the inequality in (148) holds with equality because $J^\infty = \{1, ..., d_{\bar{A}}\}$, and for all j such that $\|a_{\bar{j}}^\infty\|\|a_j^\infty\| = a_{\bar{j}}^{\infty\prime} a_j^\infty$, $\tau_j^q(x_q) = \infty$. Therefore, (156) holds.

(vii) Now (142) is satisfied with equality for the following reasons: (1) the inequality in (149) holds with equality because $r_L^q = r_L^\infty$, and thus $r_L^\infty = r_{L^\infty}^\infty$ for all $L \subseteq L^\infty : r_L^q \geq r_{L^\infty}^\infty$, and (2) the inequality in (150) holds with equality eventually using (156).

(viii) An appeal to Theorem 1(b) implies that equality holds in (144). The conditions for Theorem 1(b) are satisfied because $A^\infty \neq \mathbf{0}$ and $h^\infty = \mathbf{0}$.

Combining these changes with the proof of part (a) proves (154), and therefore, part (b).

Finally, we prove part (c). We can apply the proof of part (b) twice. We can define $\tilde{t}_q^*(x)$, $\tilde{r}^q(x)$ and $\tilde{\beta}^q(x)$ in the same way as $t_q^*(x)$, $r^q(x)$ and $\beta^q(x)$ except with A_J and b_J replacing A and b. Consider any sequence $\Omega_q \to \Omega_0$ and $x_q \to x_\infty \in \Xi$. The proof of part (b) applied to the original inequalities (A and b) implies (142) holds with equality eventually. We note that the limiting objects are the same when applied to A_J and b_J because $J^\infty \subseteq J$. This is because the assumption for part (c) implies that the jth element of $g_q = \sqrt{n_q}\Lambda_q(b(\theta_{n_q}) - A(\theta_{n_q})\mathbb{E}_{F_{n_q}}\overline{m}_{n_q}(\theta_{n_q}))$ diverges to $+\infty$ for all $j \notin J$. Therefore, when we apply the proof of part (b) to the reduced inequalities (A_J and b_J), we get that

$$\mathbb{1}\{\|x_q - \tilde{t}_q^*(x_q)\|^2 \leq \chi^2_{\tilde{r}^q(x_q), 1-\tilde{\beta}^q(x_q)}\} = \mathbb{1}\{\|x_\infty - t_\infty^*(x_\infty)\|^2 \leq \chi^2_{r^\infty(x_\infty), 1-\beta^\infty(x_\infty)}\} \tag{157}$$

eventually as $q \to \infty$. Therefore,

$$\mathbb{1}\{\mathbb{1}\{\|x_q - \tilde{t}_q^*(x_q)\|^2 \leq \chi^2_{\tilde{r}^q(x_q), 1-\tilde{\beta}^q(x_q)}\} \neq \mathbb{1}\{\|x_q - t_q^*(x_q)\|^2 \leq \chi^2_{r^q(x_q), 1-\beta^q(x_q)}\}\} = 0 \tag{158}$$

eventually. Therefore, by the bounded convergence theorem,

$$\text{Pr}_{F_{n_q}}\left(\phi_{n_q}^{\text{RCC}}(\theta_{n_q}, \alpha) \neq \phi_{n_q, J}^{\text{RCC}}(\theta_{n_q}, \alpha)\right) \to 0. \tag{159}$$

Since, for every subsequence, n_m, there exists a further subsequence, n_q, such that (159) holds, part (c) of Theorem 3 follows. $\square$

B.4 Proof of Auxiliary Lemmas 5 - 8

Proof of Lemma 5. The assumption implies that there exists a sequence, z_n^*, such that

$$A_n z_n^* \leq h_n \text{ and } z_n^* \to \mu^*(x_0, A_0, h_0) \text{ as } n \to \infty. \tag{160}$$

This implies that

$$\|x_n - \mu^*(x_n, A_n, h_n)\|^2 \leq \|x_n - z_n^*\|^2 \to \|x_0 - \mu^*(x_0, A_0, h_0)\|^2. \tag{161}$$

Taking lim sup on both sides, we get

$$\limsup{}_{n\to\infty}\|x_n - \mu^*(x_n, A_n, h_n)\|^2 \leq \|x_0 - \mu^*(x_0, A_0, h_0)\|^2. \tag{162}$$

Now note that $\mu^*(x_n, A_n, h_n) = \arg\min_{z:A_n z \leq h_n}\|x_n - z\|^2$. This sequence of minimizers is necessarily bounded because otherwise (162) cannot hold. Thus for any subsequence $\{n_m\}$ there is a further subsequence $\{n_q\}$ such that $\mu^*(x_{n_q}, A_{n_q}, h_{n_q}) \to z_\infty$ for some $z_\infty \in \mathbb{R}^{d_m}$. Since $A_{n_q}\mu^*(x_{n_q}, A_{n_q}, h_{n_q}) \leq h_{n_q}$, we have $A_0 z_\infty \leq h_0$. Thus,

$$\lim_{q\to\infty}\|x_{n_q} - \mu^*(x_{n_q}, A_{n_q}, h_{n_q})\|^2 = \|x_0 - z_\infty\|^2 \geq \|x_0 - \mu^*(x_0, A_0, h_0)\|^2. \tag{163}$$

Since the subsequence is arbitrary, this implies that

$$\text{liminf}_{n\to\infty}\|x_n - \mu^*(x_n, A_n, h_n)\|^2 \geq \|x_0 - \mu^*(x_0, A_0, h_0)\|^2. \tag{164}$$

Combining (162) and (164), we have $\lim_{n\to\infty}\|x_n - \mu^*(x_n, A_n, h_n)\|^2 = \|x_0 - \mu^*(x_0, A_0, h_0)\|^2$. This, (163), and the uniqueness of $\arg\min_{z:A_0 z \leq h_0}\|x_0 - z\|^2$ together imply that

$$\mu^*(x_n, A_n, h_n) \to z_\infty = \mu^*(x_0, A_0, h_0) \text{ as } n \to \infty, \tag{165}$$

proving the lemma. □

Proof of Lemma 6. Let $a_{j,0}'$ denote the jth row of A_0 and let $a_{j,n}'$ denote the jth row of A_n. For part (ii) of the definition of convergence, let n_q be a subsequence and z_q be a sequence such that $z_q \in \text{poly}(A_{n_q}, h_{n_q})$ for all q and $z_q \to z_0$ as $q \to \infty$. Then,

$$a_{j,0}' z_0 = \lim_{q\to\infty} a_{j,n_q}' z_q \leq \limsup_{q\to\infty} h_{j,n_q} = h_{j,0}, \tag{166}$$

showing that $z_0 \in \text{poly}(A_0, h_0)$.

For part (i) of the definition of the convergence, let $z_0 \in \text{poly}(A_0, h_0)$, and let $J_0 = \{j = 1, \ldots, d_A : a'_{j,0} z_0 = h_{j,0}\}$. If $J_0 = \emptyset$, then $z_n^* = z_0$ satisfies the requirement by $A_n \to A_0$ and $h_n \to h_0$. If $J_0 \neq \emptyset$ but $\text{rk}(A_{J_0,0}) = 0$, then $a_{j,0} = \mathbf{0}$ for all $j \in J_0$, which implies that $a_{j,n} = \mathbf{0}$ for all $j \in J_0$ by the rank condition stated in the lemma. Then, we can again let $z_n^* = z_0$ and $a'_{j,n} z_n^* = 0 \leq h_{j,n}$ for all $j \in J_0$. Again, $\{z_n^*\}$ satisfies the requirement due to $A_n \to A_0$ and $h_n \to h_0$.

Now suppose that $\text{rk}(A_{J_0,0}) > 0$. The key for the next step is to partition J_0 into two subsets J_0^* and J_0^o. We require the partition to satisfy the following conditions:

(i) J_0^* contains $\text{rk}(A_{J_0,0})$ elements such that $\{a_{j^*,0} : j^* \in J_0^*\}$ has full rank, and for any element in $j^o \in J_0^o$, there exists a unique linear representation $a_{j^o,0} = \sum_{j^* \in J_0^*} w_{j^o,j^*} a_{j^*,0}$, where $w_{j^o,j^*} : j^* \in J_0^*$ are real-valued weights.

(ii) The linear representation satisfies: for any $j^* \in J_0^*$ and $j^o \in J_0^o$ such that $w_{j^o,j^*} \neq 0$, we have $h_{j^*,0} \leq h_{j^o,0}$.

Such a partition always exists. To see why, note that the existence of a partition satisfying (i) is guaranteed by linear algebra. The number of partitions satisfying (i) is finite because J_0 is a finite set. If we choose the partition to be one that minimizes $\sum_{j^* \in J_0^*} h_{j^*,0}$ among those satisfying (i), then the chosen partition also satisfies (ii).

We note that for all n, $\text{rk}(A_{J_0^*,n}) = \text{rk}(A_{J_0})$ implies that for every $j^o \in J_0^o$ and $j^* \in J_0^*$, there exist weights, $w_{j^o,j^*,n}$, such that

$$a_{j^o,n} = \sum_{j^* \in J_0^*} w_{j^o,j^*,n} a_{j^*,n}. \tag{167}$$

Furthermore, we know that if $w_{j^o,j^*,n} \neq 0$, then $w_{j^o,j^*} \neq 0$. This follows because, otherwise, we would have

$$\text{rk}(I_{\{j^o\} \cup (J_0^*/\{j^*\})} A_n) > \text{rk}(I_{(J_0^*/\{j^*\})} A_n) = \text{rk}(I_{(J_0^*/\{j^*\})} A_0) = \text{rk}(I_{\{j^o\} \cup (J_0^*/\{j^*\})} A_0), \tag{168}$$

contradicting the assumed rank condition.

Let $A_{J_0^*,0}$ denote the submatrix of A_0 formed by the rows selected by J_0^*, and let $A_{J_0^*,n}$, $h_{J_0^*,0}$, and $h_{J_0^*,n}$ be defined analogously. Now let D be a $(d_m - |J_0|) \times d_m$ matrix, the rows of which form an orthonormal basis for the orthogonal complement to the space spanned by $\{a_{j,0} : j \in J_0^*\}$. Then the matrix $\begin{pmatrix} A_{J_0^*,0} \\ D \end{pmatrix}$ is invertible, which implies that the matrix $\begin{pmatrix} A_{J_0^*,n} \\ D \end{pmatrix}$ is invertible for large enough n. Let $h^{\wedge}_{J_0^*,n} = \min(h_{J_0^*,n}, h_{J_0^*,0})$, where the minimum is taken element by element. Let

$$z_n^\dagger = \begin{pmatrix} A_{J_0^*,n} \\ D \end{pmatrix}^{-1} \begin{pmatrix} h^\wedge_{J_0^*,n} \\ Dz_0 \end{pmatrix}. \tag{169}$$

It is easy to verify that

$$z_n^\dagger \to \begin{pmatrix} A_{J_0^*,0} \\ D \end{pmatrix}^{-1} \begin{pmatrix} h_{J_0^*,0} \\ Dz_0 \end{pmatrix} = z_0, \text{ and} \tag{170}$$

$$A_{J_0^*,n} z_n^\dagger = h^\wedge_{J_0^*,n} \leq h_{J_0^*,n}. \tag{171}$$

If $a'_{j,n} z_n^\dagger \leq h_{j,n}$ for all $j \in J_0^o$ for large enough n, then (160) holds with $z_n^* = z_n^\dagger$ and we are done. Otherwise, let

$$\lambda_n = \begin{cases} \min\left\{1, \min_{j \in J_0^o : h_{j,0} > 0} \frac{h_{j,n}}{a'_{j,n} z_n^\dagger}\right\} & \text{if } \{j \in J_0^o : h_{j,0} > 0\} \neq \emptyset \\ 1 & \text{otherwise} \end{cases}. \tag{172}$$

This is well-defined for large enough n since $a'_{j,n} z_n^\dagger \to a'_{j,0} z_0 = h_{j,0}$ and thus $a'_{j,n} z_n^\dagger \neq 0$ for large enough n. Also, by definition $\lambda_n \leq 1$, and

$$\lambda_n \to \min_{j \in J_0^o : h_{j,0} > 0} \frac{h_{j,0}}{a'_{j,0} z_0} = 1. \tag{173}$$

Now let

$$z_n^* = \lambda_n z_n^\dagger. \tag{174}$$

Then for any $j \in J_0^o$ such that $h_{j,0} > 0$, we have

$$a'_{j,n} z_n^* \leq h_{j,n} \tag{175}$$

by the definition of λ_n. For any $j \in J_0^o$ such that $h_{j,0} = 0$, we have

$$\begin{aligned} a'_{j,n} z_n^* &= \lambda_n \sum_{j^* \in J_0^*} w_{j,j^*,n} a'_{j^*,n} z_n^\dagger \\ &= \lambda_n \sum_{j^* \in J_0^*} w_{j,j^*,n} \min(h_{j^*,n}, h_{j^*,0}) \\ &= 0 \leq h_{j,n}, \end{aligned} \tag{176}$$

where the first equality follows by the definition of the weights, $w_{j,j^*,n}$, the second equality follows from the definition of $z_n^\dagger$, the third equality follows because, if $w_{j,j^*,n} \neq 0$, then $w_{j,j^*} \neq 0$, and therefore $0 \leq \min(h_{j^*,n}, h_{j^*,0}) \leq h_{j^*,0} \leq h_{j,0} = 0$ by property (ii) of the partition.

Equations (170), (173), and (174) together imply that $z_n^* \to z_0$. This also implies that, for all $j \notin J_0$, $a'_{j,n} z_n^* - h_{j,n} \to a'_{j,0} z_0 - h_{j,0} < 0$ and thus, for large enough n,

$$A_{\{1,\ldots,d_A\}/J_0,n} z_n^* < h_{\{1,\ldots,d_A\}/J_0,n}. \tag{177}$$

This combined with equations (171), $\lambda_n \leq 1$, and (174)-(176) implies that $A_n z_n^* \leq h_n$. Therefore, $\{z_n^*\}$ satisfies the requirement and the lemma is proved. □

Proof of Lemma 7. The proof of Lemma 7 makes use of three additional lemmas which are stated and proved at the end of this subsection. We use $b_{j,q}$ to denote the transpose of the j^{th} row of B_q, and similarly for a_{j,n_q}, $a_{j,0}$, and $b_{j,0}$. An equivalent way to state condition (d) is:

(i) for any further subsequence, n_q, and for every sequence $x_q \in \text{poly}(A_{n_q}, g_{n_q}) \cap \text{poly}(B_q, h_q)$ such that $x_q \to x_0$, $x_0 \in \text{poly}(A_0, g_0) \cap \text{poly}(B_0, h_0)$, and

(ii) for every $x_0 \in \text{poly}(A_0, g_0) \cap \text{poly}(B_0, h_0)$, there exists $x_q \in \text{poly}(A_{n_q}, g_{n_q}) \cap \text{poly}(B_q, h_q)$ such that $x_q \to x_0$.

Before proving the lemma, we note that for any subsequence, n_q such that $A_{n_q} \to A_0$ and $g_{n_q} \to g_0$, and for any $B_q \to B_0$ and $h_q \to h_0$, condition (d)(i) is satisfied. Specifically, let x_q denote a sequence that belongs to $\text{poly}(A_{n_q}, g_{n_q}) \cap \text{poly}(B_q, h_q)$ for all q, and such that $x_q \to x_0$. Then

$$a'_{j,0} x_0 = \lim_{q\to\infty} a'_{j,n_q} x_q \leq \lim_{q\to\infty} g_{j,n_q} = g_{j,0}. \tag{178}$$

Also, by the convergence of h_q, we have that

$$b'_{j,0} x_0 = \lim_{q\to\infty} b'_{j,q} x_q \leq \lim_{q\to\infty} h_{j,q} = h_{j,0}. \tag{179}$$

Therefore, $x_0 \in \text{poly}(A_0, g_0) \cap \text{poly}(B_0, h_0)$.

We also note that for any q, B_q, and h_q satisfying (b), condition (c) must also be satisfied. If not, then there exists a q, an $x \in \text{poly}(A_{n_q}, g_{n_q})$ and a $j' \in J(x; B_q, h_q)$ such that $b_{j',q}$ cannot be written as a linear combination of a_{j,n_q} for $j \in J(x; A_{n_q}, g_{n_q})$. This implies that there exists a v such that $b'_{j',q} v > 0$ and $v \perp a_{j,n_q}$ for all $j \in J(x; A_{n_q}, g_{n_q})$. But then, $x + \alpha v \in \text{poly}(A_{n_q}, g_{n_q})$ for sufficiently small α, at the same time that $b'_{j',q}(x + \alpha v) > h_q$. This contradicts the fact that $\text{poly}(A_{n_q}, g_{n_q}) \subseteq \text{poly}(B_q, h_q)$. Therefore, (c) holds.

We now prove the lemma by finding a subsequence, n_q, and sequences $\{B_q\}$ and $\{h_q\}$ that satisfy conditions (a), (b), and (d)(ii). We first consider A_n and g_n. By the compactness of the unit circle, let n_q be a subsequence so that A_{n_q} converges to some A_0. Also suppose g_{n_q} converges along the subsequence to some vector $g_0 \in \mathbb{R}^{d_A}_{+,\infty}$.

Let J_A^+ denote the subset of $\{1, ..., d_A\}$ for which $g_{j,0} > 0$, and let J_A^0 denote the subset for which $g_{j,0} = 0$. Consider $A_{J_A^0,0}$, which defines a cone in $\mathbb{R}^{d_m}$: $\text{poly}(A_{J_A^0,0}, \mathbf{0}) = \{x \in \mathbb{R}^{d_m} : A_{J_A^0,0} x \leq \mathbf{0}\}$. Let S denote the smallest linear subspace of $\mathbb{R}^{d_m}$ that contains this cone. Let the dimension of S be d_S. Let J_A^S be the subset of J_A^0 for which $a_{j,0} \perp S$ for all $j \in J_A^S$. Let $J_A^N = \{1, ..., d_A\}/J_A^S$.

Next, we define sequences B_q and h_q that satisfy conditions (a), (b), and (d)(ii) by induction on the dimension of S. If $d_S = 0$, then no B_q or h_q is required. Condition (a) is satisfied by the above choice of the subsequence. Condition (b) is satisfied because $\text{poly}(B_q, h_q) = \mathbb{R}^{d_m}$ for all q. Condition (d)(ii) is satisfied because $\text{poly}(A_0, g_0) = \{\mathbf{0}\}$, and then we can take $x_q = 0$ for all q, which belongs to $\text{poly}(A_{n_q}, g_{n_q})$ and converges to $x_0 = 0 \in \text{poly}(A_0, g_0)$.

If $d_S > 0$, then suppose the conclusion of Lemma 7 holds for all values of the dimension of S less than d_S.

Let $C_q = \text{poly}(A_{J_A^S,n_q}, g_{J_A^S,n_q})$. Let C_q^S be the projection of C_q onto S. That is, $C_q^S = \{P_S x : x \in C_q\}$, where P_S denotes the projection onto S and $M_S = I - P_S$. The fact that C_q is a polyhedral set (defined by finitely many affine inequalities) implies by Theorem 19.3 in Rockafellar (1970) that C_q^S is also a polyhedral set. Therefore, there exists a $d_{B_1} \times d_m$ matrix of unit vectors in S, B_q^1 and a vector h_q^1 such that $C_q^S = \{y \in S : B_q^1 y \leq h_q^1\}$. We note that C_q^S contains zero, so $h_q^1 \geq 0$. In the special case of $d_S = d_m$, $C_q^S = C_q$ and we let B_q^1 be the matrix composed of all the non-zero rows of A_{n_q} and let h_q^1 be the corresponding elements of g_{n_q}.

Let n_q be a further subsequence so that $B_q^1 \to B_0^1$ and $h_q^1 \to h_0^1$, where some of the elements of h_0^1 may be $+\infty$, in which case the convergence holds elementwise. We note that this construction satisfies conditions (a) and (b) because $\text{poly}(A_{n_q}, g_{n_q}) \subseteq C_q \subseteq \text{poly}(B_q^1, h_q^1)$ for all q, where the second subset holds because $B_q^1 x = B_q^1 M_S x + B_q^1 P_S x = B_q^1 P_S x \leq h_q^1$ for all $x \in C_q$ because the rows of B_q^1 belong to S and $P_S x \in C_q^S$.

Let J_B^+ denote the set of $j \in \{1, ..., d_{B_1}\}$ for which $h_{j,0}^1 > 0$, and let J_B^0 denote the set for which $h_{j,0}^1 = 0$, where $h_{j,0}^1$ is the jth element of h_0^1. Consider $B_{J_B^0,0}^1$ and $A_{J_A^0,0}$, which together define a cone in S: $\{x \in S : B_{J_B^0,0}^1 x \leq 0 \text{ and } A_{J_A^0,0} x \leq 0\}$. As before, let $S^\dagger$ denote the smallest linear subspace of S that contains this cone. Let $J_B^{S^\dagger}$ denote the set of all $j \in J_B^0$ for which $b_{j,0}^1 \perp S^\dagger$. Also let $J_A^{S^\dagger}$ denote the set of all $j \in J_A^0$ for which $a_{j,0} \perp S^\dagger$. Let the dimension of $S^\dagger$ be $d_{S^\dagger}$.

If $d_{S^\dagger} < d_S$, then the result follows by the induction assumption. In particular, if we let

$$\tilde{A}_q = \begin{bmatrix} A_{n_q} \\ B_q^1 \end{bmatrix} \text{ and } \tilde{g}_q = \begin{bmatrix} g_{n_q} \\ h_q^1 \end{bmatrix},$$

then the subspace, $\tilde{S}$, defined to be the smallest linear subspace containing $\text{poly}(\tilde{A}_0, \tilde{g}_0)$, is equal to $S^\dagger$. Therefore, there exists a further subsequence, n_q, and another matrix of inequalities, B_q^2 and h_q^2 such that: (a) $B_q^2 \to B_0^2$ and $h_q^2 \to h_0^2$, (b) $\text{poly}(\tilde{A}_q, \tilde{g}_q) \subseteq \text{poly}(B_q^2, h_q^2)$ for all q along the subsequence, and (d)(ii) $\text{poly}(\tilde{A}_q, \tilde{g}_q) \cap \text{poly}(B_q^2, h_q^2) \to \text{poly}(\tilde{A}_0, \tilde{g}_0) \cap \text{poly}(B_0^2, h_0^2)$ pointwise. It is easy to see that these conditions imply conditions (a), (b), and (d)(ii) for the original A_n and g_n along this subsequence, with

$$B_q = \begin{bmatrix} B_q^1 \\ B_q^2 \end{bmatrix} \text{ and } h_q = \begin{bmatrix} h_q^1 \\ h_q^2 \end{bmatrix},$$

using the fact that $\text{poly}(\tilde{A}_q, \tilde{g}_q) = \text{poly}(A_{n_q}, g_{n_q}) \cap \text{poly}(B_q^1, h_q^1)$.

Therefore, we only need to show condition (d)(ii) in the case that $d_{S^\dagger} = d_S$. In this case, $S = S^\dagger$, and so $J_B^{S^\dagger} = \emptyset$ and $J_A^{S^\dagger} = J_A^S$. Fix $x_0 \in \text{poly}(A_0, g_0) \cap \text{poly}(B_0^1, h_0^1)$. We show that for every $\epsilon > 0$ there exists a Q such that for all $q \geq Q$ there exists a $y_q \in \text{poly}(A_{n_q}, g_{n_q}) \cap \text{poly}(B_q^1, h_q^1)$ such that $\|y_q - x_0\| \leq 2\epsilon$. If true, then this can be used to construct a sequence satisfying $y_q \to x_0$, establishing condition (d)(ii).

Fix $\epsilon > 0$. By Lemma 9, there exists a point, $\tilde{x}$, in S that satisfies ${b_{j,0}^1}'\tilde{x} < h_{j,0}^1$ for all $j \in \{1, ..., d_{B_1}\}$, and $a_{j,0}'\tilde{x} < g_{j,0}$ for all $j \in J_A^N$. There exists a $\lambda \in (0, 1)$ small enough that $x^\dagger = \lambda\tilde{x} + (1-\lambda)x_0 \in \bar{B}(x_0, \epsilon)$, where $\bar{B}(x_0, \epsilon)$ denotes the closed ball of radius ϵ around x_0. Note that $x^\dagger$ satisfies $a_{j,0}'x^\dagger < g_{j,0}$ for all $j \in J_A^N$ and ${b_{j,0}^1}'x^\dagger < h_{j,0}$ for all $j \in \{1, ..., d_{B_1}\}$. Therefore, there exists a $\delta \in (0, \epsilon)$ and a Q such that for all $q \geq Q$, and for all $x \in \bar{B}(x^\dagger, \delta)$, ${b_{j,q}^1}'x < h_{j,q}^1$ for all $j \in \{1, ..., d_{B_1}\}$, and $a_{j,n_q}'x < g_{j,n_q}$ for all $j \in J_A^N$. Notice that, for all $q \geq Q$, $x^\dagger \in C_q^S = \{y \in S : B_q^1 y \leq h_q^1\}$, which means that there exists a $y_q \in C_q$ such that $x^\dagger = P_S y_q$. By Lemma 10 applied to $K = \{x^\dagger\}$ (where the condition is satisfied because, by Lemma 11, $S = \{x \in \mathbb{R}^{d_m} : A_{J^S,0}x \leq 0\}$), there exists a larger Q such that for all $q \geq Q$, $y_q \in \bar{B}(x^\dagger, \delta)$. Therefore, $\|y_q - x_0\| \leq 2\epsilon$. □

Proof of Lemma 8. Fix $x \in \mathbb{R}^{d_m}$. The fact that $\text{poly}(A, g) \subseteq \text{poly}(B, h)$ implies that $\mu^*(x; A, g) = \mu^*(x; [A; B], [g, h])$. Denote the common value by μ^*.

If there does not exist a $\bar{j} \in J(x; A, g)$ such that $a_{\bar{j}} \neq \mathbf{0}$, then $x = \mu^*$ and $a_j'x < g_j$ for all $j \notin J(x; A, g)$. Suppose, to reach a contradiction, that there does exist a $\bar{j} \in J(x; [A; B], [g; h])$ such that $b_{\bar{j}-d_A} \neq 0$. Then, there would exist a point, y, very close to x (say, $y = x + \epsilon b_{\bar{j}-d_A}$ for some $\epsilon > 0$) such that $y \notin \text{poly}(B, h)$ but $y \in \text{poly}(A, g)$. This contradicts the assumption that $\text{poly}(A,g) \subseteq \text{poly}(B,h)$. Therefore, there does not exist a $\bar{j} \in J(x;[A;B],[g;h])$ such that $b_{\bar{j}-d_A} \neq 0$. This implies that, in this case, $\tau_j(x,A,g) = 0$ for all $j \in \{1,...,d_A\}$ and $\tau_j(x;[A;B],[g;h]) = 0$ for all $j \in \{1,...,d_A+d_B\}$. Therefore, $\tau(x;A,g) = \tau(x;[A;B],[g;h])$.

Suppose there does exist a $\bar{j} \in J(x;A,g)$ such that $a_{\bar{j}} \neq \mathbf{0}$. Then, the same $\bar{j}$ can be used to define $\tau_j(x;[A;B],[g;h])$ because $J(x;A,g) \subseteq J(x;[A;B],[g;h])$.

We show that for every $j = 1,...,d_B$, $\tau_{j+d_A}(x;[A;B],[g;h]) \geq \tau(x;A,g)$. The result holds trivially if $\|b_j\|\|a_{\bar{j}}\| = b_j'a_{\bar{j}}$ because then $\tau_{j+d_A}(x;[A;B],[g;h]) = \infty$. Suppose, to reach a contradiction, that $\tau_{j+d_A}(x;[A;B],[g;h]) < \tau(x;A,g)$. Let $\tau^* = \tau_{j+d_A}(x;[A;B],[g;h])$, and consider two cases.

(i) If $\tau^* = 0$, then for some $\epsilon > 0$, the point $t^* = \mu^* + \epsilon(I_{d_m} - a_{\bar{j}}a_{\bar{j}}'\|a_{\bar{j}}\|^{-2})b_j$ belongs to $\text{poly}(A,g)$ but not $\text{poly}(B,h)$. To see that $t^* \in \text{poly}(A,g)$, note that for all $\ell \in J(x;A,g)$, the fact that $\tau(x;A,g) > 0$ implies that a_ℓ is collinear with $a_{\bar{j}}$. Then, $a_\ell' t^* = a_\ell'\mu^* = g_\ell$. For all $\ell \notin J(x;A,g)$, $a_\ell'\mu^* < g_\ell$, so ϵ can be chosen small enough that $a_\ell't^* < g_\ell$ for all $\ell \notin J(x;A,g)$. To see that $t^* \notin \text{poly}(B,h)$, note that

$$b_j't^* = b_j'\mu^* + \epsilon\|b_j\|^2 - \epsilon(b_j'a_{\bar{j}})\|a_{\bar{j}}\|^{-2} = h_j + \epsilon\|b_j\|^2 - \epsilon(b_j'a_{\bar{j}})^2\|a_{\bar{j}}\|^{-2} > h_j, \tag{180}$$

where the second equality follows because $\tau^* = 0$ and b_j is not collinear with $a_{\bar{j}}$ (so $b_j'\mu^* = h_j$), and the inequality follows because $(b_j'a_{\bar{j}})^2 < \|a_{\bar{j}}\|^2\|b_j\|^2$. This contradicts the assumption that $\text{poly}(A,g) \subseteq \text{poly}(B,h)$. Therefore, in this case, $\tau^* \geq \tau(x;A,g)$.

(ii) If $\tau^* > 0$, then let $t^* = \mu^* + \tau^*\left(\frac{b_j}{\|b_j\|} - \frac{a_{\bar{j}}}{\|a_{\bar{j}}\|}\right)$. We show that t^* belongs to the interior of $\text{poly}(A,g)$ but is on the boundary of $\text{poly}(B,h)$. Note that for every $\ell \in \{1,...,d_A\}$,

$$a_\ell't^* = a_\ell'\mu^* + \tau^*\left(\frac{a_\ell'b_j}{\|b_j\|} - \frac{a_\ell'a_{\bar{j}}}{\|a_{\bar{j}}\|}\right) = a_\ell'\mu^* + \frac{\|a_{\bar{j}}\|(h_j - b_j'\mu^*)}{\|a_{\bar{j}}\|\|b_j\| - b_j'a_{\bar{j}}}\left(\frac{a_\ell'b_j}{\|b_j\|} - \frac{a_\ell'a_{\bar{j}}}{\|a_{\bar{j}}\|}\right). \tag{181}$$

When a_ℓ is collinear with $a_{\bar{j}}$, the right hand side of (181) is less than g_ℓ because $a_\ell'\mu^* \leq g_\ell$, $h_j > b_j'\mu^*$ (because $\tau^* > 0$), and $\|a_{\bar{j}}\|\|b_j\| > a_{\bar{j}}'b_j$ (because b_j is not collinear with $a_{\bar{j}}$). When a_ℓ is not collinear with $a_{\bar{j}}$, the right hand side of (181) is less than g_ℓ because

$$\begin{aligned}
&\|a_{\bar{j}}\|(h_j - b_j'\mu^*)\left(\frac{a_\ell'b_j}{\|b_j\|} - \frac{a_\ell'a_{\bar{j}}}{\|a_{\bar{j}}\|}\right) - (g_\ell - a_\ell'\mu^*)(\|a_{\bar{j}}\|\|b_j\| - b_j'a_{\bar{j}}) \\
<&(h_j - b_j'\mu^*)\left(\|a_{\bar{j}}\|\left(\frac{a_\ell'b_j}{\|b_j\|} - \frac{a_\ell'a_{\bar{j}}}{\|a_{\bar{j}}\|}\right) - (\|a_{\bar{j}}\|\|a_\ell\| - a_\ell'a_{\bar{j}})\right) \\
\leq&0,
\end{aligned} \tag{182}$$

where the first inequality follows because

$$(g_\ell - a_\ell'\mu^*)(\|a_{\bar{j}}\|\|b_j\| - b_j'a_{\bar{j}}) > (h_j - b_j'\mu^*)(\|a_{\bar{j}}\|\|a_\ell\| - a_\ell'a_{\bar{j}}) \tag{183}$$

(by the assumption that $\tau^* < \tau(x; A, g) \leq \tau_\ell(x, A, g)$), and the second inequality follows because $b_j'\mu^* \leq h_j$ and $\|a_\ell\|\|b_j\| \geq a_\ell' b_j$. This shows that t^* is on the interior of poly(A, g). We also show that t^* is on the boundary of poly(B, h) by calculating that $b_j't^* = h_j$. By a similar calculation to above, we see that

$$\begin{aligned}&(b_j't^* - h_j)(\|b_j\|\|a_{\bar{j}}\| - b_j'a_{\bar{j}})\\ =&\|a_{\bar{j}}\|(h_j - b_j'\mu^*)\left(\|b_j\| - \frac{b_j'a_{\bar{j}}}{\|a_{\bar{j}}\|}\right) + (b_j'\mu^* - h_j)(\|b_j\|\|a_{\bar{j}}\| - b_j'a_{\bar{j}}) = 0.\end{aligned} \tag{184}$$

This implies that there exists a point, y, very close to t^* (say $y = t^* + \epsilon b_j$ for some $\epsilon > 0$) such that $y \notin$ poly(B, h) but $y \in$ poly(A, g). This contradicts the assumption that poly$(A, g) \subseteq$ poly(B, h). Therefore, $\tau_{j+d_A}(x; [A; B], [g; h]) \geq \tau(x; A, g)$ for all $j = 1, ..., d_B$. □

Lemma 9. *Let A be a $d_A \times d_m$ matrix. Let g be nonnegative. Let J^+ denote the subset of $\{1, ..., d_A\}$ such that $g_j > 0$, and let J^0 denote the subset of $\{1, ..., d_A\}$ such that $g_j = 0$. Let S denote the smallest linear subspace containing* poly$(A_{J^0}, 0) = \{x \in \mathbb{R}^{d_m} : A_{J^0}x \leq 0\}$. *Let J^S be the subset of J^0 for which $A_{J^S} \perp S$. Let $J^N = \{1, ..., d_A\}/J^S$. There exists a $\tilde{x} \in S$ such that $a_j'\tilde{x} < g_j$ for all $j \in J^N$.*

Proof of Lemma 9. First, let $M > \max_{j\in J^+}\|a_j\|$, and let $\epsilon \in (0, \min_{j\in J^+}\{g_j\}/M)$. Then, for all $\tilde{x} \in \bar{B}(0, \epsilon)$, $a_j'\tilde{x} < g_j$ for all $j \in J^+$, where $\bar{B}(x, \epsilon)$ denotes the closed ball of radius ϵ around x. Also, for every $j \in J^N \cap J^0$, $\{x \in S : a_j'x = 0\}$ defines a subspace of S. We note that for all $j \in J^N \cap J^0$, $\{x \in S : a_j'x = 0\}$ is a proper subset of S, because otherwise j would belong to J^S. By the definition of S, $S \cap$ poly$(A_{J^N\cap J^0}, 0)$ is not contained within any of these subspaces. In particular, for each $j \in J^N \cap J^0$, we can find a $\tilde{x}_j$ and a neighborhood, N_j, (relatively open in S) that belongs to $S \cap$ poly$(A_{J^N\cap J^0,0}, 0)/\{x \in S : a_j'x = 0\}$. Indeed, we can consider $j \in J^N \cap J^0$ sequentially, and define each neighborhood to be a subset of the previous one. Therefore, the final $\tilde{x}_j$ must belong to $S \cap$ poly$(A_{J^N\cap J^0,0}, \mathbf{0})$ and satisfy $a_j'\tilde{x} < 0$ for all $j \in J^N \cap J^0$. Take $\tilde{x} = \lambda\tilde{x}_j$, where $\lambda > 0$ is small enough that $\tilde{x} \in \bar{B}(\mathbf{0}, \epsilon)$. Then, $\tilde{x}$ satisfies $a_j'\tilde{x} < g_j$ for all $j \in J^N$. □

Lemma 10. *Let $A_n \to A_0$ and $g_n \to 0$, where $g_n \geq 0$ for all n. Suppose $S = \{x \in \mathbb{R}^{d_m} : A_0x \leq 0\}$ is a linear subspace of $\mathbb{R}^{d_m}$. Let $S^\perp$ denote the orthogonal subspace to S in $\mathbb{R}^{d_m}$. Let P_Sx denote the projection of $x \in \mathbb{R}^{d_m}$ onto S and let M_Sx denote $x - P_Sx$. Then, for* *every $K \subseteq S$, compact, and for every $\varepsilon > 0$, we have*

$$\{x \in \text{poly}(A_n, g_n) : P_S x \in K, \|M_S x\| \geq \varepsilon\} = \emptyset \tag{185}$$

eventually as $n \to \infty$.

Proof of Lemma 10. Suppose that the conclusion of the lemma is not true. Then there exists a sequence $\{x_n \in \text{poly}(A_n, g_n)\}$ and a subsequence n_m such that $P_S x_{n_m} \in K$ and $\|M_S x_{n_m}\| \geq \varepsilon$ for all $m \geq 1$. Define the unit vector $x_{n_m}^{\perp} = M_S x_{n_m} / \|M_S x_{n_m}\|$. Then, by the compactness of K and the unit circle, there exists a further subsequence n_q such that $P_S x_{n_q} \to x^S$ and $x_{n_q}^{\perp} \to x^{\perp}$ for some $x^S \in S$ and $x^{\perp} \in S^{\perp}$ as $q \to \infty$.

Because $x^{\perp} \in S^{\perp}$ and $x^{\perp} \neq 0$, we know that $x^{\perp} \notin S = \{x \in \mathbb{R}^{d_m} : A_0 x \leq 0\}$, and therefore there exists a j such that

$$a_{j,0}' x^{\perp} > 0. \tag{186}$$

Also, since $x^S \in S$, $a_{j,0}' x^S \leq 0$. Since S is a linear subspace, we have $a_{j,0}'(-x^S) \leq 0$ as well. This shows that $a_{j,0}' x^S = 0$ (and more generally, $S = \{x \in \mathbb{R}^{d_m} : A_0 x = \mathbf{0}\}$).

Now consider

$$\begin{aligned} a_{j,n_q}' x_{n_q} - g_{j,n_q} &= a_{j,n_q}' P_S x_{n_q} + a_{j,n_q}' M_S x_{n_q} - g_{j,n_q} \\ &= o(1) + a_{j,0}' x^S + \|M_S x_{n_q}\|(o(1) + a_{j,0}' x^{\perp}) - o(1) \\ &= o(1) + \|M_S x_{n_q}\|(o(1) + a_{j,0}' x^{\perp}). \end{aligned} \tag{187}$$

By (186), $o(1) + a_{j,0}' x^{\perp} > 0$ eventually. This, combined with $\|M_S x_{n_q}\| \geq \epsilon$ implies that

$$a_{j,n_q}' x_{n_q} - g_{j,n_q} > 0 \tag{188}$$

eventually. This contradicts the definition of the sequence x_n which requires that $x_n \in \text{poly}(A_n, g_n)$ for all n. □

Lemma 11. *Let A be a matrix. Let S be the smallest linear subspace containing $C = \text{poly}(A, \mathbf{0})$. Let $J = \{j : a_j \perp S\}$. Then, $S = \text{poly}(A_J, \mathbf{0})$.*

Proof of Lemma 11. First, notice that if $x \in S$, then $x \perp a_j$ for all $j \in J$, and therefore, $A_J x = 0$, so $x \in \text{poly}(A_J, \mathbf{0})$.

To go the other way, let $x \in \text{poly}(A_J, \mathbf{0})$. Lemma 9 implies that there exists an $\tilde{x} \in S$ such that $a_j' \tilde{x} < 0$ for all $j \in J^c$, where $J^c = \{1, ..., d_A\}/J$. Consider $y = x + M\tilde{x}$ for M large. We note that $A_J y = A_J x + M A_J \tilde{x} \leq 0$ since $x \in \text{poly}(A_J, \mathbf{0})$ and $\tilde{x} \in S \subseteq \text{poly}(A_J, \mathbf{0})$. We also note that for every $j \in J^c$, $a_j'y = a_j'x + Ma_j'\tilde{x} \to -\infty$ as M diverges. Thus, there exists an M large enough that $y \in \text{poly}(A, \mathbf{0})$. This implies that $y \in S$ because $\text{poly}(A, \mathbf{0}) \subseteq S$. This also implies that $x = y - M\tilde{x} \in S$ because S is a linear subspace. □

C Supporting Materials for Section 3.2

C.1 Lemmas 12-13 and Their Proofs

Lemma 12. *Let B and C be conformable matrices and d be a conformable vector. There exists a matrix $A = A(B, C)$ and a vector $b = b(C, d)$ such that*

$$\{\delta : C\delta \geq B\mu - d\} \neq \emptyset \;\Leftrightarrow\; A\mu \leq b.$$

Furthermore, $A(B, C) = H(C)B$ and $b(C, d) = H(C)d$, where $H(C)$ is the matrix with rows formed by the vertices of the polyhedron $\{h \in \mathbb{R}^k : h \geq \mathbf{0}, C'h = \mathbf{0}, \mathbf{1}'h = 1\}$.

Proof of Lemma 12. By Theorem 2.7 in Gale (1960), $\{\delta : C\delta \geq B\mu - d\} \neq \emptyset$ is equivalent to

$$h'(B\mu - d) \leq 0 \text{ for all } h \geq 0, C'h = \mathbf{0}. \tag{189}$$

The equivalence is not affected by adding the scale normalization: $\mathbf{1}'h = 1$ to (189). Thus, $\{\delta : C\delta \geq B\mu - d\} \neq \emptyset$ is equivalent to

$$h'(B\mu - d) \leq 0 \text{ for all } h \in \mathcal{H} := \{h \geq 0 : C'h = 0, \mathbf{1}'h = 1\}. \tag{190}$$

Since the rows of $H(C)$ are vertices, and thus elements of $\mathcal{H}$, we have

$$\{\delta : C\delta \geq B\mu - d\} \neq \emptyset \Rightarrow H(C)(B\mu - d) \leq 0.$$

Then by the definition of A and b, we have

$$\{\delta : C\delta \geq B\mu - d\} \neq \emptyset \Rightarrow A\mu \leq b.$$

Conversely, since the rows of $H(C)$ are vertices of $\mathcal{H}$, for any $h \in \mathcal{H}$, there exists $\mathbb{R}^k$-vector $c \geq 0$ such that $H(C)'c = h$. Thus, if $A\mu \leq b$, we must have $h'(B\mu - d) = c'H(C)(B\mu - d) = c'(A\mu - b) \leq 0$. That this holds for all $h \in \mathcal{H}$ implies that $\{\delta : C\delta \geq B\mu - d\} \neq \emptyset$. Thus the lemma is proved. □

Lemma 13. $\hat{r} = \text{rk}(B_Z'\mathcal{H}_0)$.

Proof of Lemma 13. Denote B_Z, C_Z, and d_Z by B, C, and d. Let $h_1', \ldots, h_{m_1}'$ be all the rows of $H(C)$ orthogonal to $B\hat{\mu} - d$. Then by definition, $A_{\hat{\mathcal{J}}} = [B'h_1, \ldots, B'h_{m_1}]'$, and thus $\text{rk}(A_{\hat{\mathcal{J}}}) = \text{rk}(B'h_1, \ldots, B'h_{m_1})$. Since $h_1, \ldots, h_{m_1} \in \{h \geq \mathbf{0} : h'C = \mathbf{0}, h'(B\hat{\mu} - d) = 0\}$, we have $B'h_1, \ldots, B'h_{m_1} \in \{B'h : h \geq \mathbf{0}, h'C = \mathbf{0}, h'(B\hat{\mu} - d) = 0\}$. This implies that $\text{rk}(A_{\hat{\mathcal{J}}}) \leq \text{rk}(\{B'h : h \geq \mathbf{0}, h'C = \mathbf{0}, h'(B\hat{\mu} - d) = 0\}) = \text{rk}(B'\mathcal{H}_0)$.

Next, suppose that $\tilde{h}_1, \ldots, \tilde{h}_{m_2} \in \{h \geq \mathbf{0} : h'C = \mathbf{0}, h'(B\hat{\mu} - d) = 0\}$ such that $\text{rk}(B'\mathcal{H}_0) = \text{rk}(B'\tilde{h}_1, \ldots, B\tilde{h}_{m_2})$. By the definition of $H(C)$, $\tilde{h}_1, \ldots, \tilde{h}_{m_2}$ must all be convex combinations of the rows of $H(C)$. In fact, they must all be convex combinations of $h_1, \ldots, h_{m_1}$ defined in the first part of the proof because any other row (say, h_*) of $H(C)$ must satisfy the strict inequality $h_*'(B\hat{\mu} - d) > 0$ (since they correspond to the inactive inequalities). Consequently, $B'\tilde{h}_1, \ldots, B'\tilde{h}_{m_2}$ must be convex combinations of $B'h_1, \ldots, B'h_{m_1}$. This implies that $\text{rk}(B'\mathcal{H}_0) \leq \text{rk}(A_{\hat{\mathcal{J}}})$. Therefore, the lemma is proved. □

C.2 General Algorithm for Calculating rk$(B_Z'\mathcal{H}_0)$

We suppress the subscript Z in B_Z, C_Z, and d_Z for notational ease. As discussed in Section 3.2, the rank of the polyhedron $B'\mathcal{H}_0$ is the dimension of the smallest linear subspace containing it. That is, its linear span, denoted by $\text{span}(B'\mathcal{H}_0)$. Let $G := \begin{pmatrix} I_{J_0}' & C & B\hat{\mu} - d \end{pmatrix}$, where J_0 is defined in Section 3.2. From the discussions in Section 3.2, we know that

$$\text{span}(B'\mathcal{H}_0) = \{B'h : h \in \mathbb{R}^k, G'h = \mathbf{0}\}. \tag{191}$$

The number of restrictions in $G'h = \mathbf{0}$ is given by the number of columns of G. However, some of those restrictions may be redundant or contained in the null space of B'. We do some linear algebra to calculate $\text{span}(B'\mathcal{H}_0)$.

Let $G = (g_1, g_2, ..., g_k)'$, where each g_j' is the jth row of G. Suppose the first $\text{rk}(G)$ rows of G have rank $\text{rk}(G)$. This can be achieved by rearranging the rows of G together with the rows of B and the elements of h accordingly. Partition G as $[G_1', G_2']'$, where $G_1 = (g_1, ..., g_{\text{rk}(G)})'$ and $G_2 = (g_{\text{rk}(G)+1}, ..., g_k)'$. We can then solve the equations, $G'h = \mathbf{0}$, for the first $\text{rk}(G)$ elements of h as a function of the other elements. Specifically, let $h = (h_1', h_2')'$ be partitioned conformably with G. Then, we can solve for $h_1 = \Gamma h_2$, where $\Gamma = -(G_1G_1')^{-1}G_1G_2'$.

Then $\text{span}(\mathcal{H}_0)$ can be written as

$$\text{span}(\mathcal{H}_0) = \{(h_1', h_2')' \in \mathbb{R}^k : h_1 = \Gamma h_2\}. \tag{192}$$

Accordingly,

$$\text{span}(B'\mathcal{H}_0) = B'\text{span}(\mathcal{H}_0) = \{(B_1'\Gamma + B_2')h_2 : h_2 \in \mathbb{R}^{k-\text{rk}(G)}\}, \tag{193}$$

where $B = [B_1', B_2']'$ is partitioned conformably with h. This implies that

$$\text{rk}(B'\mathcal{H}_0) = \text{rk}(\Gamma' B_1 + B_2).$$

To end this section, we provide the pseudo-code to calculate $\text{rk}(B'\mathcal{H}_0)$ in Algorithm 3. This can be plugged in Algorithm 2 in Section 3.2, replacing line 12, to compute the sCC and the sRCC tests.

Algorithm 3: Pseudo-code for calculating $\text{rk}(B'\mathcal{H}_0)$ when $\text{rk}(B) < k$.

%$\mathcal{H}_0 = \{h \in \mathbb{R}^k : G'h = 0\}$**, where** $G = (g_1, \ldots, g_k)'$ **and** g_j' **is the** j**th row of** G**.**

if $\text{rk}(G) = k$ **then**
  $\text{rk}(B'\mathcal{H}_0) := 0$
else
  if $\text{rk}(g_1, \ldots, g_{\text{rk}(G)}) < \text{rk}(G)$ **then**
    Rearrange rows of G so that $\text{rk}(g_1, \ldots, g_{\text{rk}(G)}) = \text{rk}(G)$ holds. Rearrange the elements of h and the rows of B accordingly.
  end if
  $G_1 := (g_1, \ldots, g_{\text{rk}(G)})'$
  $G_2 := (g_{\text{rk}(G)+1}, \ldots, g_k)'$
  $\Gamma \ := -(G_1 G_1')^{-1} G_1 G_2'$
  $\text{rk}(B'\mathcal{H}_0) := \text{rk}((\Gamma', I_{k-\text{rk}(G)})B)$.
end if

D Asymptotic Validity of the Subvector Tests

D.1 General Conditions for Asymptotic Validity

We fix a realization of $\{Z_i\}_{i=1}^n$ and denote it by z. Let $\mathcal{F}_z$ be a collection of distributions F_z. The following high-level assumption is sufficient for the uniform asymptotic validity of the sCC and the sRCC tests. This assumption is the conditional version of Assumption 2.

Assumption 3. *The given sequence* $\{(F_{z,n}, \theta_n) : F_{z,n} \in \mathcal{F}_z, \theta_n \in \Theta_0(F_{z,n})\}_{n=1}^\infty$ *satisfies, for every subsequence,* n_m*, there exists a further subsequence,* n_q*, and there exists a sequence of positive definite* $d_m \times d_m$ *matrices,* $\{D_q\}$*, such that:*

(a) *Under the sequence* $\{F_{z,n_q}\}_{q=1}^{\infty}$,

$$\sqrt{n_q}D_q^{-1/2}(\overline{m}_{n_q}(\theta_{n_q}) - \mathbb{E}_{F_{z,n_q}}\overline{m}_{n_q}(\theta_{n_q})) \to_d N(\mathbf{0}, \Omega), \tag{194}$$

for a positive definite correlation matrix Ω, *and*

$$\|D_q^{-1/2}\widehat{\Sigma}_{n_q}(\theta_{n_q})D_q^{-1/2} - \Omega\| \to_p 0. \tag{195}$$

(b) *Let* $A(\theta)$ *and* $b(\theta)$ *be defined in Lemma 12.* $\Lambda_q A(\theta_{n_q})D_q \to \bar{A}_0$ *for some* $d_A \times d_m$ *matrix* $\bar{A}_0$, *and for every* $J \subseteq \{1, ..., d_A\}$, $\text{rk}(I_J A(\theta_{n_q})D_q) = \text{rk}(I_J\bar{A}_0)$, *where* Λ_q *is the diagonal* $d_A \times d_A$ *matrix whose* j*th diagonal entry is one if* $e_j' A(\theta_{n_j}) = 0$ *and* $\|e_j' A(\theta_{n_q})D_q\|^{-1}$ *otherwise.*

The following corollary of Theorem 3 shows the asymptotic properties of the sRCC test.

Corollary 2. (a) *Suppose Assumption* 3(a) *holds for all sequences* $\{(F_{z,n}, \theta_n) : F_{z,n} \in \mathcal{F}_z, \theta_n \in \Theta_0(F_{z,n})\}_{n=1}^{\infty}$. *Then,*

$$\limsup_{n\to\infty} \sup_{F_z \in \mathcal{F}_z} \sup_{\theta \in \Theta_0(F_z)} \mathbb{E}_{F_z}(\phi_n^{\text{sRCC}}(\theta, \alpha)|z) \leq \alpha.$$

Next consider a sequence $\{(F_{z,n}, \theta_n) : F_{z,n} \in \mathcal{F}_z, \theta_n \in \Theta_0(F_{z,n})\}_{n=1}^{\infty}$ *satisfying Assumption* 3(a,b).

(b) *If, along any further subsequence, for all* $j = 1, ..., d_A$, $\sqrt{n_q}e_j'\Lambda_q(A(\theta_{n_q})\mathbb{E}_{F_{z,n_q}}\overline{m}_{n_q}(\theta_{n_q}) - b(\theta_{n_q})) \to 0$, *and if* $\bar{A}_0 \neq \mathbf{0}_{d_A \times d_m}$, *then*

$$\lim_{n\to\infty} \mathbb{E}_{F_{z,n}}\phi_n^{\text{sRCC}}(\theta_n, \alpha) = \alpha.$$

(c) *If, for* $J \subseteq \{1, ..., d_A\}$, *along any further subsequence,* $\sqrt{n_q}e_j'\Lambda_q(A(\theta_{n_q})\mathbb{E}_{F_{z,n_q}}\overline{m}_{n_q}(\theta_{n_q}) - b(\theta_{n_q})) \to -\infty$ *as* $q \to \infty$, *for all* $j \notin J$, *then*

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

Remark. Corollary 2 follows from Theorem 3 because $\Theta_0(F_{z,n})$ has the equivalent representation

$$\Theta_0(F_{z,n}) = \{\theta \in \Theta : A(\theta)\mathbb{E}_{F_{z,n}}[\overline{m}_n(\theta)|z] \leq b(\theta)\}, \tag{196}$$

by Lemma 12. There is one subtle point: $A(\theta) = H(C_z(\theta))B_z(\theta)$ might change dimension because $H(C_z)$ might change dimension with C_z, and C_z might change with the sample size.

But this does not cause a problem because the dimension of $H(C_z)$ and thus that of $A(\theta)$ is bounded by a function of k and p which does not change with the sample size.[40] Due to this boundedness, for any subsequence of $\{n\}$ we can always find a further subsequence along which the dimension of $A(\theta)$ does not change. Then the problem falls into the framework of Theorem 3.

D.2 Primitive Conditions under i.i.d. Sampling

Now we assume that $\{W_i\}_{i=1}^n$ is an i.i.d. sample unconditionally and derive primitive conditions for Assumption 3(a). Let the conditional distribution of W_i given $Z_i = z_i$ be represented by the mapping: $F_| : z_i \mapsto F_{|z_i}$. Let $\mathcal{F}_|$ denote a collection of $F_|$ and let $\mathcal{F}_z = \{\times_{i=1}^n F_{|z_i} : F_| \in \mathcal{F}_|\}$, where $\times_{i=1}^n F_{|z_i}$ denotes the joint distribution whose marginal distributions are independent $F_{|z_i}$. The following assumption is sufficient for (194) in Assumption 3. In the assumption $\sigma_{j|z}^2(\theta) := n^{-1}\sum_{i=1}^n \mathrm{Var}_{F_{|z_i}}(m_j(W_i,\theta)|z_i)$ and

$$D_{|z}(\theta) = \mathrm{diag}(\sigma_{1|z}^2(\theta), \ldots, \sigma_{d_m|z}^2(\theta)). \tag{197}$$

Let $\mathrm{eig}_{\min}(V)$ denote the minimum eigenvalue of a matrix V.

Assumption 4. *There exists an $M_0 < \infty$ and an $\epsilon_0 > 0$ such that for all $F_| \in \mathcal{F}_|$, the following hold.*

(a) $\sigma_{j|z}^2(\theta) > 0$ *for all* $j = 1, \ldots, d_m$, $\theta \in \Theta$, *and for all* n.

(b) $n^{-1}\sum_{i=1}^n \mathbb{E}_{F_{|z_i}}((m_j(W_i,\theta)/\sigma_{j|z}(\theta))^4|z_i) < M_0$ *for all* j, *all* $\theta \in \Theta$, *and for all* n.

(c) $\mathrm{eig}_{\min}(n^{-1}\sum_{i=1}^n[\mathrm{Var}_{F_{|z_i}}(D_{|z}^{-1/2}(\theta)m(W_i,\theta)|z_i)]) > \epsilon_0$ *for all* $\theta \in \Theta$ *and for all* n.

Remark. Part (b) requires $m(W_i,\theta)$ to have finite 4th moment conditional on $Z_i = z_i$. This is used both to derive the asymptotic normality of $\overline{m}_n(\theta)$ using the Lindeberg-Feller central limit theorem under the sequence of $F_{z,n}$, and to show the consistency of the average conditional variance estimator $\widehat{\Sigma}_n(\theta)$. Part (c) requires that the average conditional variance of $m(W_i,\theta)$ be invertible uniformly over θ and $F_| \in \mathcal{F}_|$. This is required since we use the quasi-likelihood ratio statistic which involves inverting an estimator of the average conditional variance.

When the nearest neighbor matching variance estimator in (39) is used, the following additional assumption is used for consistency.

Assumption 5. (a) *$\{z_i\}_{i=1}^\infty$ is a bounded sequence of distinct values.*

[40] This fact is know as the McMullen's upper bound theorem. See e.g. Section 8.4 of Ziegler (1995).

(b) $\Sigma_{Z,n} \to \Sigma_Z$ *where* Σ_Z *is finite positive definite matrix.*

(c) *There exist* $M_g > 0$ *and* $M_V > 0$ *such that for all* $\theta \in \Theta$ *and* $F_| \in \mathcal{F}_|$ *the conditional mean and variance,* $\mathbb{E}_{F_{|z_i}}[D_{|z}(\theta)^{-1/2}m(W_i,\theta)|z_i]$ *and* $Var_{F_{|z_i}}(D_{|z}(\theta)^{-1/2}m(W_i,\theta)|z_i)$*, are Lipschitz continuous in* z_i *with Lipschitz constants* M_g *and* M_V*, respectively.*

Remark. The boundedness part of part (a) is used to show that z_i and its nearest neighbor get close to each other on average as $n \to \infty$. This can be guaranteed by pre-normalizing Z_i before applying the matching estimator. For example, if the raw conditioning variable $\tilde{Z}_i$ is supported in $(0,\infty)$, one can let $Z_i = \Phi(\tilde{Z}_i)$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function. This and the Lipschitz continuity in part (c) together ensure that the nearest neighbor provides the correct information about the conditional variance in the limit. The distinct values required by part (a) ensures that each point can be the nearest neighbor to at most a uniformly bounded number of other points. This holds with probability one if Z_i has no probability mass on any single point. It can be made to hold by adding a tiny continuous noise to Z_i when Z_i has repeated values. The noise should be set small enough to be a tie breaker only in the nearest neighbor calculation. Part (b) of the assumption can be established for a probability-one set of $\{Z_i\}$ values by the strong law of large numbers.

The following theorem verifies Assumption 3(a).

Theorem 4. (a) *Assumption* 4 *implies* (194) *in Assumption* 3 *for all sequences* $\{(F_{z,n},\theta_n) : F_{z,n} \in \mathcal{F}_z, \theta_n \in \Theta_0(F_{z,n})\}_{n=1}^{\infty}$.

(b) *If* $\{z_i\}_{i=1}^n$ *contains at least two instances of each value eventually as* $n \to \infty$*, and Assumption* 4 *holds, then* (195) *holds for* $\widehat{\Sigma}_n(\theta)$ *defined in* (37)*, for all sequences* $\{(F_{z,n},\theta_n) : F_{z,n} \in \mathcal{F}_z, \theta_n \in \Theta_0(F_{z,n})\}_{n=1}^{\infty}$.

(c) *If Assumptions* 4 *and* 5 *hold, then* (195) *holds for* $\widehat{\Sigma}_n(\theta)$ *defined in* (39)*, for all sequences* $\{(F_{z,n},\theta_n) : F_{z,n} \in \mathcal{F}_z, \theta_n \in \Theta_0(F_{z,n})\}_{n=1}^{\infty}$.

D.3 Proof of Theorem 4

(a) Let $\{(F_{z,n},\theta_n) : F_{z,n} \in \mathcal{F}_z, \theta_n \in \Theta_0(F_{z,n})\}$ be an arbitrary sequence. Let $F_{|z_i,n}$ denote the conditional distribution of W_i given $Z_i = z_i$ implied by $F_{z,n}$. Let $\sigma^2_{j|z,n}(\theta)$ and $D_{|z,n}(\theta)$ be defined just like $\sigma^2_{j|z}(\theta)$ and $D_{|z}(\theta)$ except with $F_{|z_i}$ replaced by $F_{|z_i,n}$. Let $D_n = D_{|z,n}(\theta_n)$. Then D_n is positive definite for every n by Assumption 4(a).

Let $\Omega_n = D_n^{-1/2} n^{-1} \sum_{i=1}^n \mathrm{Var}_{F_{|z_i,n}}(m(W_i,\theta_n)|z_i) D_n^{-1/2}$. Algebra shows that the square of the (j, ℓ)th element of Ω_n is bounded by

$$2n^{-1}\sum_{i=1}^{n}\mathbb{E}_{F_{|z_i,n}}\left[\left(\frac{m_j(W_i,\theta_n)}{\sigma_{j|z,n}(\theta_n)}\right)^4|z_i\right]+2n^{-1}\sum_{i=1}^{n}\mathbb{E}_{F_{|z_i,n}}\left[\left(\frac{m_\ell(W_i,\theta_n)}{\sigma_{j|z,n}(\theta_n)}\right)^4|z_i\right], \tag{198}$$

which is bounded by $4M_0$ by Assumption 4(a). Thus $\text{vec}(\Omega_n) \in [0, 4M_0]^{d_m^2}$, which is a compact set. This implies that a subsequence n_q can be found for any subsequence of $\{n\}$ such that $\Omega_{n_q} \to \Omega_\infty$. Furthermore, Assumption 4(c) implies that Ω_∞ is positive definite.

It remains to verify the Lindeberg condition for the Lindeberg-Feller central limit theorem (CLT) along the subsequence $\{n_q\}$. Let a be an arbitrary real vector on the unit sphere in $\mathbb{R}^{d_m}$. Let

$$\hat{m}_{n,i}(\theta) = a'D_n^{-1/2}(m(W_i,\theta) - \mathbb{E}_{F_{|z_i,n}}[m(W_i,\theta)|z_i]). \tag{199}$$

Let

$$s_q^2 = n_q{}^{-1}\sum_{i=1}^{n_q}\mathbb{E}_{F_{|z_i,n_q}}\left[\hat{m}_{n_q,i}(\theta_{n_q})^2|z_i\right]. \tag{200}$$

For an arbitrary $\varepsilon > 0$, consider the derivation,

$$\begin{aligned}
&\sum_{i=1}^{n_q} n_q^{-1}s_q^{-2}\mathbb{E}_{F_{|z_i,n_q}}[\hat{m}_{n_q,i}(\theta_{n_q})^2 1\{n_q^{-1}s_q^{-2}\hat{m}_{n_q,i}(\theta_{n_q})^2 > \varepsilon\}|z_i]\\
&\le n_q^{-2}s_q^{-4}\varepsilon^{-1}\sum_{i=1}^{n_q}\mathbb{E}_{F_{|z_i,n_q}}[\hat{m}_{n_q,i}(\theta_{n_q})^4|z_i]\\
&\le 16n_q^{-2}s_q^{-4}\varepsilon^{-1}\sum_{i=1}^{n_q}\mathbb{E}_{F_{|z_i,n_q}}[(a'D_{n_q}^{-1/2}m(W_i,\theta_{n_q}))^4|z_i]\\
&\le 16n_q^{-2}s_q^{-4}\varepsilon^{-1}\sum_{i=1}^{n_q}\mathbb{E}_{F_{|z_i,n_q}}[\|D_{n_q}^{-1/2}m(W_i,\theta_{n_q})\|^4|z_i]\\
&= O(n_q^{-1}s_q^{-4}\varepsilon^{-1})\\
&\to 0, \text{ as } q \to \infty,
\end{aligned} \tag{201}$$

where the first inequality holds because $1(x > \varepsilon) \le \frac{x}{\varepsilon}$ for any $x \ge 0$, the second inequality holds because $E[(X - E(X))^4] \le 16E[X^4]$, the third inequality holds by the Cauchy-Schwarz inequality and $\|a\|= 1$, the equality holds by Assumption 4(b), and the convergence holds because $s_q^2 \to a'\Omega_\infty a$ by the definition of the subsequence $\{n_q\}$. Therefore, the Lindeberg condition holds and the CLT applies, proving part (a).

(b) Note that $\widehat{\Sigma}_n(\theta)$ is the weighted average of the standard sample variance estimator within subsamples with same z_i values. Thus, by standard argument, we have

$$\mathbb{E}_{F_{z,n}}[\widehat{\Sigma}_n(\theta_n)|z] = \sum_{\ell\in\mathcal{Z}} \frac{n_\ell}{n}\mathrm{Var}_{F_{|\ell,n}}(m(W_i,\theta_n)|\ell) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{F_{|z_i,n}}(m(W_i,\theta_n)|z_i), \tag{202}$$

where the second equality holds by rearranging terms. Thus,

$$\mathbb{E}_{F_{z,n}}[D_n^{-1/2}\widehat{\Sigma}_n(\theta_n)D_n^{-1/2}|z] = \Omega_n. \tag{203}$$

Also by standard calculation, the (j,j') element of $D_n^{-1/2}\widehat{\Sigma}_n(\theta_n)D_n^{-1/2}$ has a conditional variance given z:

$$\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}_{F_{|z_i,n}}\left(\frac{m_j(W_i,\theta_n)m_{j'}(W_i,\theta_n)}{\sigma_{j|z,n}(\theta_n)\sigma_{j'|z,n}(\theta_n)}|z_i\right) + \frac{1}{n^2}\sum_{i=1}^{n}\frac{\omega^2_{j|z_i,n}(\theta_n)\omega^2_{j'|z_i,n}(\theta)+\omega_{jj'|z_i,n}(\theta_n)^2}{n_{z_i}-1}, \tag{204}$$

where $\omega_{j|z_i,n}(\theta) = \mathrm{Var}_{F_{|z_i,n}}\left(\frac{m_j(W_i,\theta)}{\sigma_{j|z,n}(\theta)}|z_i\right)$ and $\omega_{jj'|z_i,n}(\theta) = \mathrm{Cov}_{F_{|z_i,n}}\left(\frac{m_j(W_i,\theta)}{\sigma_{j|z,n}(\theta)}, \frac{m_{j'}(W_i,\theta)}{\sigma_{j'|z,n}(\theta)}|z_i\right)$. By standard algebraic manipulation, we have

$$\begin{aligned}\mathrm{Var}_{F_{|z_i,n}}\left(\frac{m_j(W_i,\theta_n)m_{j'}(W_i,\theta_n)}{\sigma_{j|z,n}(\theta_n)\sigma_{j'|z,n}(\theta_n)}|z_i\right) &\le \frac{1}{2}(M_{ji}+M_{j'i}), \text{ and}\\ \omega^2_{j|z_i,n}(\theta_n)\omega^2_{j'|z_i,n}(\theta)+\omega_{jj'|z_i,n}(\theta_n)^2 &\le M_{ji}+M_{j'i}\end{aligned} \tag{205}$$

where $M_{ji} = \mathbb{E}_{F_{|z_i,n}}\left[\left(\frac{m_j(W_i,\theta_n)}{\sigma_{j|z,n}(\theta_n)}\right)^4|z_i\right]$. Therefore, by Assumption 4 and the additional assumption that $n_{z_i} \ge 2$ for all i, we have that the expression in (204) is bounded by $\frac{1}{n}(M_0+2M_0)$, which converges to zero as $n\to\infty$. This proves part (b).

(c) First, we prove that

$$n^{-1}\sum_{i=1}^{n}\|z_i - z_{\ell_Z(i)}\|^2 \to 0. \tag{206}$$

To begin, define $\tilde{z}_i = \Sigma_{Z,n}^{-1/2}z_i$. By Assumption 5(b), $\Sigma_{Z,n}^{-1/2}\to\Sigma_Z^{-1/2}$ as $n\to\infty$ and this limit is finite. Thus, $\Sigma_{Z,n}^{-1/2}$ is uniformly bounded over all large enough n. This and Assumption 5(a) together imply that the elements of the array $\{\tilde{z}_1,\ldots,\tilde{z}_n\}_{n\ge1}$ are chosen from a bounded set. Then Lemma 1 of Abadie and Imbens (2008) applies directly and implies that

$$n^{-1}\sum_{i=1}^{n}\|\tilde{z}_i - \tilde{z}_{\ell_Z(i)}\|^2 \to 0. \tag{207}$$

Consider the derivation

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}\|z_i - z_{\ell_Z(i)}\|^2 &= n^{-1}\sum_{i=1}^{n}(\tilde{z}_i - \tilde{z}_{\ell_Z(i)})'\Sigma_{Z,n}(\tilde{z}_i - \tilde{z}_{\ell_Z(i)}) \\
&\leq n^{-1}\sum_{i=1}^{n}\|\tilde{z}_i - \tilde{z}_{\ell_Z(i)}\|^2 \text{eig}_{\max}(\Sigma_{Z,n}) \\
&\to 0, \qquad (208)
\end{aligned}$$

where $\text{eig}_{\max}(\cdot)$ stands for maximum eigenvalue and the convergence holds by (207) and Assumption 5(b). This proves (206).

Next consider an arbitrary unit vector a in $\mathbb{R}^{d_m}$, let

$$s_{n,i}^2(\theta) = a'D_n^{-1/2}(m(W_i,\theta) - m(W_{\ell_Z(i)},\theta))(m(W_i,\theta) - m(W_{\ell_Z(i)},\theta))'D_n^{-1/2}a. \qquad (209)$$

Then $a'D_n^{-1/2}\widehat{\Sigma}_n(\theta)D_n^{-1/2}a = \frac{1}{2n}\sum_{i=1}^{n} s_{n,i}^2(\theta_n)$. Since a is arbitrary, it suffices to show that for any subsequence of $\{n\}$ there exists a further subsequence $\{n_q\}$ such that

$$\frac{1}{2n_q}\sum_{i=1}^{n_q} s_{n_q,i}^2(\theta_{n_q}) \to_p a'\Omega_\infty a. \qquad (210)$$

as $q \to \infty$.

Let $\hat{m}_{n,i}(\theta)$ be defined in the proof of part (a). Then

$$\begin{aligned}
&\mathbb{E}_{F_{z,n}}[s_{n,i}^2(\theta_n)|z] \\
&= \mathbb{E}_{F_{z,n}}[a'(\hat{m}_{n,i}(\theta_n) - \hat{m}_{n,\ell_Z(i)}(\theta_n) + \Delta_{ni})(\hat{m}_{n,i}(\theta_n) - \hat{m}_{n,\ell_Z(i)}(\theta_n) + \Delta_{ni})'a|z] \\
&= a'\mathbb{E}_{F_{|z_i,n}}[\hat{m}_{n,i}(\theta_n)\hat{m}_{n,i}(\theta_n)'|z_i]a + a'\mathbb{E}_{F_{|z_{\ell_Z(i)},n}}[\hat{m}_{n,\ell_Z(i)}(\theta_n)\hat{m}_{n,\ell_Z(i)}(\theta_n)'|z_{\ell_Z(i)}]a + a'\Delta_{ni}\Delta_{ni}'a \\
&= 2a'D_n^{-1/2}\text{Var}_{F_{|z_i,n}}[m(W_i,\theta_n)|z_i]D_n^{-1/2}a + a'\Delta_{ni}^V a + a'\Delta_{ni}\Delta_{ni}'a, \qquad (211)
\end{aligned}$$

where $\Delta_{ni} = \mathbb{E}_{F_{|z_i,n}}[D_n^{-1/2}m(W_i,\theta_n)|z_i] - \mathbb{E}_{F_{|z_{\ell_Z(i)},n}}[D_n^{-1/2}m(W_{\ell_Z(i)},\theta_n)|z_{\ell_Z(i)}]$, and

$$\Delta_{ni}^V = \text{Var}_{F_{|z_{\ell_Z(i)},n}}[D_n^{-1/2}m(W_{\ell_Z(i)},\theta_n)|z_{\ell_Z(i)}] - \text{Var}_{F_{|z_i,n}}[D_n^{-1/2}m(W_i,\theta_n)|z_i].$$

By Assumption 5(c) we have,

$$\|\Delta_{ni}\| \leq M_g\|z_i - z_{\ell_Z(i)}\| \text{ and } \|\Delta_{ni}^V\| \leq M_V\|z_i - z_{\ell_Z(i)}\|. \qquad (212)$$

Thus,

$$n^{-1}\sum_{i=1}^{n}a'\Delta_{ni}\Delta_{ni}'a \leq n^{-1}\sum_{i=1}^{n}\|\Delta_{ni}\|^2 \leq n^{-1}\sum_{i=1}^{n}M_g^2\|z_i - z_{\ell_Z(i)}\|^2 \to 0, \tag{213}$$

and

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}a'\Delta_{ni}^{V}a \leq n^{-1}\sum_{i=1}^{n}\|\Delta_{ni}^{V}\| \leq n^{-1}\sum_{i=1}^{n}M_V\|z_i - z_{\ell_Z(i)}\| \\ \leq M_V\sqrt{n^{-1}\sum_{i=1}^{n}\|z_i - z_{\ell_Z(i)}\|^2} \to 0. \end{aligned} \tag{214}$$

For an arbitrary subsequence of n, consider a further subsequence $\{n_q\}$ such that $\Omega_n \to \Omega_\infty$. Such a further subsequence always exists by the proof of part (a). Then as $q \to \infty$,

$$n_q^{-1}\sum_{i=1}^{n_q}2a'\text{Var}_{F_{|z_i,n_q}}[D_{n_q}^{-1/2}m(W_i,\theta_{n_q})|z_i]a \to 2a'\Omega_\infty a. \tag{215}$$

Combining (211), (213), (214), and (215), we have

$$\mathbb{E}_{F_{z,n_q}}[a'D_{n_q}^{-1/2}\widehat{\Sigma}_{n_q}(\theta_{n_q})D_{n_q}^{-1/2}a|z] = \frac{1}{2n_q}\sum_{i=1}^{n_q}\mathbb{E}_{F_{z,n_q}}[s_{n_q,i}^2(\theta_n)|z] \to a'\Omega_\infty a. \tag{216}$$

Now it suffices to show that

$$\mathbb{E}_{F_{z,n}}\left[\left(n^{-1}\sum_{i=1}^{n}\left(s_{n,i}^2(\theta_n) - \mathbb{E}_{F_{z,n}}[s_{n,i}^2(\theta_n)|z]\right)\right)^2|z\right] \to 0. \tag{217}$$

Let $\varepsilon_i(\theta) = a'\hat{m}_{n,i}(\theta)$ and $\sigma_i^2(\theta) = a'\text{Var}_{F_{|z_i,n}}(D_n^{-1/2}m(W_i,\theta|z_i))a = \mathbb{E}_{F_{|z_i,n}}[\varepsilon_i^2(\theta)|z_i]$. Consider

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left(s_{n,i}^2(\theta_n) - \mathbb{E}_{F_{z,n}}[s_{n,i}^2(\theta_n)|z]\right) &= n^{-1}\sum_{i=1}^{n}(\varepsilon_i^2(\theta_n) - \sigma_i^2(\theta_n)) \\ &+ n^{-1}\sum_{i=1}^{n}(\varepsilon_{\ell_Z(i)}^2(\theta_n) - \sigma_{\ell_Z(i)}^2(\theta_n)) \\ &+ 2n^{-1}\sum_{i=1}^{n}(a'\Delta_{ni})\varepsilon_i(\theta_n) \\ &- 2n^{-1}\sum_{i=1}^{n}(a'\Delta_{ni})\varepsilon_{\ell_Z(i)}(\theta_n) \end{aligned}$$

$$+ 2n^{-1} \sum_{i=1}^{n} \varepsilon_i(\theta_n)\varepsilon_{\ell_Z(i)}(\theta_n). \tag{218}$$

Clearly, all the summands on the right-hand side have conditional expectation zero. Now we show that the conditional variance (which then is the conditional second moment) of each of them converges to zero.

For the first summand on the right-hand side of (218), we have

$$\begin{aligned}
\mathbb{E}_{F_{z,n}}\left[\left(n^{-1}\sum_{i=1}^{n}(\varepsilon_i^2(\theta_n)-\sigma_i^2(\theta_n))\right)^2 |z\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}_{F_{|z_i,n}}(\varepsilon_i^2(\theta_n)|z_i) \\
&\le \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{|z_i,n}}[\varepsilon_i^4(\theta_n)|z_i] \\
&\le \frac{16}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{|z_i,n}}[(a'D_n^{-1/2}m(W_i,\theta_n))^4|z_i] \\
&\le \frac{16}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{|z_i,n}}[\|D_n^{-1/2}m(W_i,\theta_n)\|^4|z_i] \\
&\to 0, \qquad (219)
\end{aligned}$$

where the convergence holds by Assumption 4(b). For the second summand on the right-hand side of (218), we have

$$\begin{aligned}
&\mathbb{E}_{F_{z,n}}\left[\left(n^{-1}\sum_{i=1}^{n}(\varepsilon_{\ell_Z(i)}^2(\theta_n)-\sigma_{\ell_Z(i)}^2(\theta_n))\right)^2 |z\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{z,n}}[\left(\varepsilon_{\ell_Z(i)}^2(\theta_n)-\sigma_{\ell_Z(i)}^2(\theta_n)\right)^2|z] \\
&+ \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}_{F_{z,n}}[\left(\varepsilon_{\ell_Z(i)}^2(\theta_n)-\sigma_{\ell_Z(i)}^2(\theta_n)\right)\left(\varepsilon_{\ell_Z(j)}^2(\theta_n)-\sigma_{\ell_Z(j)}^2(\theta_n)\right)|z] \\
&\le \frac{\overline{L}+2\overline{L}^2}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{|z_i,n}}[\left(\varepsilon_i^2(\theta_n)-\sigma_i^2(\theta_n)\right)^2|z_i] \to 0, \qquad (220)
\end{aligned}$$

where $\overline{L}$ is the maximum number of times a j is $\ell_Z(i)$ for some i. This number is bounded by $3^{d_z}-1$, which does not depend on n (see e.g. Zeger and Gersho (1994)). The convergence holds by (219).

For the third summand in (218), we have

$$
\begin{aligned}
\mathbb{E}_{F_{z,n}}\left[\left(n^{-1}\sum_{i=1}^{n}a'\Delta_{ni}\varepsilon_i(\theta_n)\right)^2|z\right] &= \frac{1}{n^2}\sum_{i=1}^{n}(a'\Delta_{ni})^2\mathbb{E}_{F_{|z_i,n}}[\varepsilon_i^2(\theta_n)|z_i] \\
&\leq \frac{M_g\overline{B}}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{|z_i,n}}[\varepsilon_i^2(\theta_n)|z_i] \\
&\leq \frac{M_g\overline{B}}{n^2}\sum_{i=1}^{n}(1+\mathbb{E}_{F_{|z_i,n}}[\varepsilon_i^4(\theta_n)|z_i]) \\
&\to 0, \qquad (221)
\end{aligned}
$$

where $\overline{B}$ is the maximum distance of two points in the sequence $\{z_i\}_{i=1}^n$, which is bounded by Assumption 5(a), the first inequality holds by Assumption 5(c), the second inequality holds by $x^2 \leq (\max(1,|x|))^2 \leq \max\{1,x^4\} \leq 1+x^4$ and the convergence holds by (219).

For the fourth summand in (218), we have

$$
\begin{aligned}
\mathbb{E}_{F_{z,n}}\left[\left(n^{-1}\sum_{i=1}^{n}a'\Delta_{ni}\varepsilon_{\ell_Z(i)}(\theta_n)\right)^2|z\right] &= \frac{1}{n^2}\sum_{i=1}^{n}(a'\Delta_i(\theta_n))^2\mathbb{E}_{F_{|z_{\ell_Z(i)},n}}[\varepsilon_{\ell_Z(i)}^2(\theta_n)|z_{\ell_Z(i)}] \\
&\leq \frac{M_g\overline{B}}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{z_{\ell_Z(i)},n}}[\varepsilon_{\ell_Z(i)}^2(\theta_n)|z_{\ell_Z(i)}] \\
&\leq \frac{M_g\overline{B}}{n^2}\sum_{i=1}^{n}(1+\mathbb{E}_{F_{|z_{\ell_Z(i)},n}}[\varepsilon_{\ell_Z(i)}^4(\theta_n)|z_{\ell_Z(i)}]) \\
&\leq \frac{M_g\overline{LB}}{n^2}\sum_{i=1}^{n}(1+\mathbb{E}_{F_{|z_i,n}}[\varepsilon_i^4(\theta_n)|z_i]) \\
&\to 0, \qquad (222)
\end{aligned}
$$

where $\overline{L}$ is number discussed below (220).

For the fifth summand on the right-hand side of (218), we have

$$
\begin{aligned}
&\mathbb{E}_{F_{z,n}}\left[\left(n^{-1}\sum_{i=1}^{n}\varepsilon_i(\theta_n)\varepsilon_{\ell_Z(i)}(\theta_n)\right)^2|z\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{z,n}}[\varepsilon_i^2(\theta_n)\varepsilon_{\ell_Z(i)}^2(\theta_n)|z] \\
&+ \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}_{F_{z,n}}[\varepsilon_i(\theta_n)\varepsilon_{\ell_Z(i)}(\theta_n)\varepsilon_j(\theta_n)\varepsilon_{\ell_Z(j)}(\theta_n)|z]
\end{aligned}
$$

$$\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{z,n}}[\varepsilon_i^2(\theta_n)\varepsilon_{\ell_Z(i)}^2(\theta_n)|z] + \frac{\overline{L}}{n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{z,n}}[\varepsilon_i^2(\theta_n)\varepsilon_{\ell_Z(i)}^2(\theta_n)|z]$$
$$\leq \frac{1+\overline{L}}{2n^2}\sum_{i=1}^{n}\mathbb{E}_{F_{z,n}}[\varepsilon_i^4(\theta_n)+\varepsilon_{\ell_Z(i)}^4(\theta_n)|z]$$
$$\to 0, \tag{223}$$

where the first inequality holds because $\mathbb{E}_{F_{z,n}}[\varepsilon_i(\theta_n)\varepsilon_{\ell_Z(i)}(\theta_n)\varepsilon_j(\theta_n)\varepsilon_{\ell_Z(j)}(\theta_n)|z]$ is nonzero only when $j=\ell_Z(i)$ and $\ell_Z(j)=i$ and this occurs at most $\overline{L}$ times for each i, the second inequality holds by $2xy \leq x^2+y^2$, and the convergence holds by (219) and the last two lines of (222).

Combining (218)-(223), we have that (217) holds, which then proves part (c). □

E Numerical Details for Section 5.2

E.1 Calculation of the Identified Set

Let Y_U denote $\log(s_{N,i}+2/N)-\log(1-s_{N,i}+\underline{s})$ and let Y_L denote $\log(s_{N,i}+\underline{s})-\log(1-s_{N,i}+2/N)$. For every θ_0 in the identified set, there exists a $\delta=(\delta_1,\delta_2)'\in\mathbb{R}^2$ such that for all $z=(z_c,z_e)'\in\{0,1\}^2$,

$$\mathbb{E}[Y_U|z]-\mathbb{E}[X|z]\theta_0 \geq \delta_1+\delta_2 z_c \geq \mathbb{E}[Y_L|z]-\mathbb{E}[X|z]\theta_0. \tag{224}$$

The identified set for θ_0 can be solved via two linear programming problems once $\mathbb{E}[Y_U|z]$, $\mathbb{E}[Y_L|z]$, and $\mathbb{E}[X|z]$ are calculated. Note that

$$\mathbb{E}[X|z]=\mathbb{E}[1\{2z_e+\varepsilon\geq 0\}]=\Phi(2z_e). \tag{225}$$

We also need to calculate $\mathbb{E}[Y_U|z]$ and $\mathbb{E}[Y_L|z]$. Let

$$\ell(s,N,c)=\sum_{i=0}^{N}\binom{N}{i}s^i(1-s)^{N-i}\log(i+c). \tag{226}$$

Then

$$\mathbb{E}[Y_U|z]=\mathbb{E}[\ell(s^*,N,2)-\ell(1-s^*,N,N\underline{s})|z]$$
$$\mathbb{E}[Y_L|z]=\mathbb{E}[\ell(s^*,N,N\underline{s})-\ell(1-s^*,N,2)|z], \tag{227}$$

where $s^* = \frac{\exp(-1\{2z_e+\varepsilon>0\}-z_c+\varepsilon)}{1+\exp(-1\{2z_e+\varepsilon>0\}-z_c+\varepsilon)}$. The conditional expectations can then be calculated by simulating a large number of ε draws. After obtaining, $\mathbb{E}[Y_U|z]$ and $\mathbb{E}[Y_L|z]$, we use linear programming based on (224) to calculate the upper and the lower bound for θ_0. We find the bounds to be $[-1.203, -0.757]$.

E.2 Bisection Algorithm for Calculating Confidence Sets

We compute the confidence sets for θ_0 implied by the sRCC, sCC, and the ARP hybrid tests by following the same steps:

1. Find a point in the confidence set. Let this point be denoted $\hat{\theta}_0$. Methods to find such a point are given below.

2. Find a point $\hat{\theta}_L < \hat{\theta}_0$ that is outside the identified set. This is done by checking whether $\hat{\theta}_0 - j$ is rejected for $j = 1, 2, ...$ iteratively until a rejected point is found.

3. Use bisection to find a point between $\hat{\theta}_0$ and $\hat{\theta}_{LB}$ that is a boundary point of the confidence set. Specifically, check whether $\hat{\theta}_M = (\hat{\theta}_0 + \hat{\theta}_{LB})/2$ is rejected or not. If yes, check the midpoint between $\hat{\theta}_M$ and $\hat{\theta}_0$, and if not, check the midpoint between $\hat{\theta}_M$ and $\hat{\theta}_{LB}$. Continue this way until the desired accuracy is reached. The accuracy that we use is 0.00049, yielding CI endpoints that are accurate to the third digit.

4. Let the last rejected point from Step 2 be denoted $\hat{\theta}_L$. This is the computed lower endpoint of the confidence interval.

5. Follow analogous steps to find the upper endpoint $\hat{\theta}_U$.

6. Finally, the confidence interval is $[\hat{\theta}_L, \hat{\theta}_U]$.

Remarks. (1) The bisection algorithm is much more efficient than grid search in finding the endpoints accurately. However there is one caveat: the confidence set from inverting either the sRCC, the sCC, or the ARP hybrid test is not guaranteed to be an interval. The algorithm above always yields an interval. This interval sometimes is the confidence set itself, sometimes is the convex hull, and sometimes is a sub-interval of the confidence set. In our experience, however, the difference between the bisection confidence interval and the test-inversion confidence set is rarely big. The coverage probability of the bisection confidence interval is in fact slightly *higher* than the test-inversion confidence set (computed via grid-search).

(2) The initial point in the confidence set for the sRCC and sCC tests are obtained using iterated generalized method of moments (GMM). To describe the iterated GMM procedure,

first note that the moment inequality model of the subvector simulation example is of the form:

$$Y_Z - X_Z\theta_0 - C_Z\delta \leq \mathbf{0}, \tag{228}$$

where

$$Y_Z = \begin{pmatrix} \mathbb{E}[-n^{-1}\sum_{i=1}^n Y_{U,i}\mathcal{I}(Z_i)|\{Z_i\}] \\ \mathbb{E}[n^{-1}\sum_{i=1}^n Y_{L,i}\mathcal{I}(Z_i)|\{Z_i\}] \end{pmatrix},$$
$$X_Z = \begin{pmatrix} \mathbb{E}[-n^{-1}\sum_{i=1}^n X_i\mathcal{I}(Z_i)|\{Z_i\}] \\ \mathbb{E}[n^{-1}\sum_{i=1}^n X_i\mathcal{I}(Z_i)|\{Z_i\}] \end{pmatrix}, \text{ and}$$
$$C_Z = \begin{pmatrix} -n^{-1}\sum_{i=1}^n \mathcal{I}(Z_i)Z_{ci}'] \\ n^{-1}\sum_{i=1}^n \mathcal{I}(Z_i)Z_{ci}'] \end{pmatrix}.$$

Let $\hat{Y}_Z = \begin{pmatrix} -n^{-1}\sum_{i=1}^n Y_{U,i}\mathcal{I}(Z_i) \\ n^{-1}\sum_{i=1}^n Y_{L,i}\mathcal{I}(Z_i)] \end{pmatrix}$ and $\hat{X}_Z = \begin{pmatrix} -n^{-1}\sum_{i=1}^n X_i\mathcal{I}(Z_i) \\ n^{-1}\sum_{i=1}^n X_i\mathcal{I}(Z_i) \end{pmatrix}$. Let $\widehat{\Sigma}_Z(\theta)$ be an estimator of the conditional variance of $\sqrt{n}(Y_Z - X_Z\theta)$ given $\{Z_i\}_{i=1}^n$ as a function of θ.

Let $\widehat{\Sigma}_Z = \widehat{\Sigma}_Z(0)$, and let

$$(\hat{\theta}_1, \hat{\delta}', \hat{\mu}')' = \arg\min_{\theta,\delta',\mu':\mu\leq 0}(Y_Z - X_Z\theta - C_Z\delta - \mu)'\widehat{\Sigma}_Z^{-1}(Y_Z - X_Z\theta - C_Z\delta - \mu)'. \tag{229}$$

If $\hat{\theta}_1$ is not rejected by the sCC (sRCC) test, let $\hat{\theta}_0 = \hat{\theta}_1$. Otherwise, update $\widehat{\Sigma}_Z$ with $\widehat{\Sigma}_Z(\hat{\theta}_1)$ and repeat (229). Iterate until either (1) a point that is not rejected is found, or (2) the update to the GMM estimator of θ is small, or (3) a maximum number of iteration is reached. If the iteration ended in either (2) or (3), let the confidence interval be the singleton that is the last point checked. In our simulations, (2) and (3) never occurred. In the majority of the times, $\hat{\theta}_1$ is already not rejected and no iteration is needed.

(3) The initial point in the confidence set for the ARP hybrid test is obtained in a similar iterative approach, except that instead of using GMM, we minimize the maximum violation of the inequalities. That is, (229) is replaced by

$$\begin{aligned}(\hat{\theta}_1, \hat{\delta}', \tau)' = \arg\min_{\theta,\delta',\tau:\tau\geq 0} \tau \text{ s.t.}& \\ \widehat{D}_Z^{-1/2}(Y_Z - X_Z\theta_0 + C_Z\delta) \leq \tau\mathbf{1},&\end{aligned} \tag{230}$$

where $\widehat{D}_Z$ is the diagonal matrix sharing diagonal elements with $\widehat{\Sigma}_Z$. □

References

Abadie, A. and Imbens, G. W. (2008). Estimation of the conditional variance in paired experiments. *Annales d'Économie et de Statistique*, No. 91/92, Econometric Evaluation of Public Policies: Methods and Applications (July-December 2008):175–187.

Andrews, D., Cheng, X., and Guggenberger, P. (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics*.

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.

Aubin, J.-P. and Frankowska, H. (1990). *Set-Valued Analysis*. Birkhäuser.

Gale, D. (1960). *The Theory of Linear Economic Models*. McGraw-Hill Book Company, 1 edition.

Luenberger, D. (1969). *Optimization by Vector Space Methods*. John Wiley & Sons, Inc.

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.

Rockafellar, R. (1970). *Convex Analysis*. Princeton University Press.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer.

Zeger, K. and Gersho, A. (1994). Number of nearest neighbors in a euclidean code. *IEEE Transactions on Information Theory*, 40:1647–1649.

Ziegler, G. (1995). *Lectures on Polytopes, Graduate Texts in mathematics, vol. 152*. Springer, Berlin.