On Testing Systems of Linear Inequalities with Known Coefficients

Yuehao Bai
Department of Economics
University of Michigan–Ann Arbor
yuehaob@umich.edu

Andres Santos
Department of Economics
University of California–Los Angeles
andres@econ.ucla.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

May 10, 2022

Abstract

In this paper, we revisit the problem considered in Fang et al. (2020) of testing whether there exists a non-negative solution to a possibly under-determined system of linear equations with known coefficients. We propose two alternative methods for this testing problem – one based on subsampling and a second which is closely related to the two-step method for testing moment inequalities developed in Romano et al. (2014) – and provide weak conditions under which they control size uniformly over a large class of possible distributions of the data. In contrast to Fang et al. (2020), however, our analysis does not accommodate high-dimensional settings in which the dimension of p and/or d grow with the sample size n.

KEYWORDS: linear programming, linear inequalities, subsampling, bootstrap, uniform inference, partial identification

JEL classification codes: C12, C14

1 Introduction

Let $X_1, \ldots, X_n$ be i.i.d. $\sim P \in \mathbf{P}$ on $\mathbf{R}^k$ and consider testing the null hypothesis

$$H_0 : P \in \mathbf{P}_0 \text{ versus } H_1 : P \in \mathbf{P} \setminus \mathbf{P}_0 \ , \tag{1}$$

where

$$\mathbf{P}_0 := \{P \in \mathbf{P} : \exists\, x \geq 0 \text{ s.t. } Ax = \beta(P)\} \ .$$

Here, $\mathbf{P}$ is a large class of possible distributions for the observed data, $\beta(P) \in \mathbf{R}^d$ is an unknown parameter and A is a known $p \times d$ matrix. This hypothesis testing problem was previously considered in Fang et al. (2020), who additionally provide numerous examples where this testing problem arises naturally. See, in particular, Section 2 of their paper. In this paper, we develop two alternative methods for this testing problem: one based on subsampling and a second which is a closely-related to the two-step method for testing moment inequalities developed in Romano et al. (2014). For each testing procedure, we provide weak conditions under which the test controls size uniformly over $\mathbf{P}$.

Our testing procedures differ not only in how the critical value is constructed, but also rely on a different test statistic that does not employ the alternative geometric characterization of the null hypothesis developed in Fang et al. (2020). Specifically, we consider the "natural" choice of test statistic given by

$$T_n := \inf_{x \geq 0} \sqrt{n} |Ax - \hat{\beta}_n| \ , \tag{2}$$

where $\hat{\beta}_n = \hat{\beta}_n(X_1, \ldots, X_n)$ is a suitable estimator of $\beta(P)$. We emphasize, however, that our results do not accommodate high-dimensional settings in which the dimension of p and/or d grow with the sample size n. We leave the development of such results for future work, but emphasize that both of the tests we propose are computationally very attractive: each requires simply repeatedly solving problems like those on the right-hand side of (2). As noted by Fang et al. (2020), this feature is especially important in many of the applications they describe.

In addition to Fang et al. (2020), the problem of testing (1) has been previously considered by Kitamura and Stoye (2018) in the context of testing the validity of a random utility model. In contrast to our tests below, their test requires certain conditions on A that can be violated in some of the examples described in Fang et al. (2020). Our paper is also broadly related to the literature on sub-vector inference in moment inequality models and shape restrictions. See, for example, Romano and Shaikh (2008), Bugni et al. (2017), Kaido et al. (2019), Gandhi et al. (2019), Chernozhukov et al. (2015), Zhu (2020) and Fang and Seo (2021). These testing procedures are sufficiently general to accommodate testing (1), but do not exploit the linear structure in the null hypothesis and, as a result, are computationally less tractable and/or rely on more demanding assumptions than the ones in our analysis below. Notable exceptions include Andrews et al. (2019) and Cox and Shi (2019), who propose methods for sub-vector inference in certain conditional moment inequality models with some linear structure that can be also be used to test (1). We leave a more detailed comparison of the tests proposed in this paper with these tests as well as with the procedure in Fang et al. (2020) for future work.

The remainder of our paper is organized as follows. In Section 2, we describe our subsampling-based test and provide conditions under which it controls size uniformly over $\mathbf{P}$. In Section 3, we describe our two-step test and likewise provide conditions under which it controls size uniformly over $\mathbf{P}$. Our conditions are formulated in a high-level fashion that accommodates a broad variety of possible applications, but, for each test, we discuss more primitive conditions under which they may be verified in the leading example where $\beta(P)$ is a mean or sufficiently "mean-like."

2 Subsampling

In this section, we describe our subsampling-based test. To this end, define

$$L_n(t) := \frac{1}{N_n} \sum_{1 \le j \le N_n} I\{\inf_{x \ge 0} \sqrt{b}|Ax - \hat{\beta}_{b,j}| \le t\} ,$$

where $N_n = \binom{n}{b}$, j indexes the N_n subsets of $X_1, \ldots, X_n$ of size b, and $\hat{\beta}_{b,j}$ is $\hat{\beta}_b$ evaluated at the jth such subset of data. The subsampling-based test we consider is given by

$$\phi_n^{\text{sub}} := I\{T_n > L_n^{-1}(1-\alpha)\} .$$

In order to state the following theorem concerning the behavior of this test, we require some additional notation. To this end, let $\mathcal{C}$ be the set of all convex subsets of $\mathbf{R}^d$. For $\mathbb{C} \in \mathcal{C}$, define

$$J_n(\mathbb{C}, P) := P\{\sqrt{n}(\hat{\beta}_n - \beta(P)) \in \mathbb{C}\} . \tag{3}$$

With this notation, we have the following theorem:

Theorem 2.1. *Let $b = b_n \to \infty$ and $b/n \to 0$. Suppose*

$$\sup_{P \in \mathbf{P}} \sup_{\mathbb{C} \in \mathcal{C}} |J_n(\mathbb{C}, P) - J_b(\mathbb{C}, P)| \to 0 . \tag{4}$$

Then,

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} E_P[\phi_n^{\text{sub}}] \leq \alpha \ .$$

PROOF: Define $\mathbb{X}_0(P) := \{x \geq 0 : Ax = \beta(P)\}$ and, for $P \in \mathbf{P}_0$, let $x_0(P) \in \mathbb{X}_0(P)$. (Note that the dependence of $x_0(P)$ on P is intended to reflect the fact that it is an element of $\mathbb{X}_0(P)$, which is non-empty for $P \in \mathbf{P}_0$. One could write $x_0(\mathbb{X}_0(P))$ instead.) For $c \in \mathbf{R}^p$ and $t \in \mathbf{R}$, define

$$\mathbb{A}(c,t) := \bigcup_{y\geq c} B_t(Ay) \ ,$$

where $B_t(Ay)$ is the closed ball of radius t with center Ay. Here, it is understood that $B_t(Ay) := \{Ay\}$ for $t < 0$. Note that $\mathbb{A}(c,t) = B_t(\bigcup_{y\geq c} Ay)$ and is therefore convex.

Next, for any $P \in \mathbf{P}_0$, note that

$$\begin{aligned}
\inf_{x\geq 0} \sqrt{n}|Ax - \hat{\beta}_n| &= \inf_{x\geq 0} \sqrt{n}|A(x - x_0(P)) - (\hat{\beta}_n - \beta(P))| \\
&= \inf_{y\geq -\sqrt{n}x_0(P)} |Ay - \sqrt{n}(\hat{\beta}_n - \beta(P))| \\
&\leq \inf_{y\geq -\sqrt{b}x_0(P)} |Ay - \sqrt{n}(\hat{\beta}_n - \beta(P))|
\end{aligned}$$

where in the first equality we use the fact that $Ax_0(P) = \beta(P)$ since $x_0(P) \in \mathbb{X}_0(P)$, the second equality uses the substitution $y = \sqrt{n}(x - x_0(P))$, and the inequality uses the fact that $b \leq n$ and $x_0(P) \geq 0$ imply that $-\sqrt{b}x_0(P) \geq -\sqrt{n}x_0(P)$. Similarly,

$$\begin{aligned}
\inf_{x\geq 0} \sqrt{b}|Ax - \hat{\beta}_b| &= \inf_{x\geq 0} \sqrt{b}|A(x - x_0(P)) - (\hat{\beta}_b - \beta(P))| \\
&= \inf_{y\geq -\sqrt{b}x_0(P)} |Ay - \sqrt{b}(\hat{\beta}_b - \beta(P))| \ .
\end{aligned}$$

It follows that

$$\begin{aligned}
K_n(t,P) &:= P\{T_n \leq t\} \\
&= P\{\inf_{y\geq -\sqrt{n}x_0(P)} |Ay - \sqrt{n}(\hat{\beta}_n - \beta(P))| \leq t\} \\
&\geq P\{\inf_{y\geq -\sqrt{b}x_0(P)} |Ay - \sqrt{n}(\hat{\beta}_n - \beta(P))| \leq t\} \\
&= P\{\sqrt{n}(\hat{\beta}_n - \beta(P)) \in \mathbb{A}(-\sqrt{b}x_0(P), t)\} \ .
\end{aligned}$$

and, likewise, that

$$\begin{aligned} K_b(t,P) &:= P\{T_b \le t\} \\ &= P\{\sqrt{b}(\hat{\beta}_b - \beta(P)) \in \mathbb{A}(-\sqrt{b}x_0(P),t)\} \ . \end{aligned}$$

Hence,

$$\begin{aligned} &\sup_{P\in\mathbf{P}}\sup_{t\in\mathbf{R}}\{K_b(t,P)-K_n(t,P)\} \\ &\quad \le \sup_{P\in\mathbf{P}}\sup_{t\in\mathbf{R}}\{P\{\sqrt{b}(\hat{\beta}_b-\beta(P))\in\mathbb{A}(-\sqrt{b}x_0(P),t)\}-P\{\sqrt{n}(\hat{\beta}_n-\beta(P))\in\mathbb{A}(-\sqrt{b}x_0(P),t)\}\} \\ &\quad \le \sup_{P\in\mathbf{P}}\sup_{\mathbb{C}\in\mathcal{C}}|J_b(\mathbb{C},P)-J_n(\mathbb{C},P)|\to 0 \ , \end{aligned}$$

where the convergence to zero follows by assumption. The desired result thus follows from Lemma A.1 in Romano and Shaikh (2012). ■

Remark 2.1. An inspection of the proof reveals that the requirement (4) is stronger than is required for the theorem: the set $\mathcal{C}$ can be replaced with the smaller class $\{\mathbb{A}(c,t) : c \in \mathbf{R}^p, t \in \mathbf{R}\}$ without changing the argument. ■

Remark 2.2. The requirement (4) can be readily verified in the case where $\beta(P)$ is a mean and $\hat{\beta}_n$ is the corresponding sample average under weak assumptions on $\mathbf{P}$. All that is required is a weak uniform integrability requirement. See, for example, Romano and Shaikh (2008). This requirement can also be verified for the case where $\beta(P)$ is sufficiently "mean-like." We leave the development of such results for future work. ■

Remark 2.3. While our methodology does not require $\hat{\beta}_n$ to be non-degenerate, in some cases, it may be desireable to incorporate additional deterministic constraints into the null hypothsis differently. To this end, consider testing (1) with $\mathbf{P}_0$ replaced with

$$\tilde{\mathbf{P}}_0 := \{P \in \mathbf{P} : \exists x \ge 0 \text{ s.t. } Bx = m \text{ and } Ax = \beta(P)\} \ , \tag{5}$$

where B is a known $r \times d$-dimensional matrix and m is a r-dimensional vector of constants. Consider the test statistic

$$\tilde{T}_n := \inf_{x\ge 0: Bx=m} \sqrt{n}|Ax - \hat{\beta}_n| \ . \tag{6}$$

Define

$$\tilde{L}_n(t) := \frac{1}{N_n}\sum_{1\le j\le N_n} I\{\inf_{x\ge 0:Bx=m}\sqrt{b}|Ax-\hat{\beta}_{b,j}| \le t\} \ .$$

The same argument employed in establishing Theorem 2.1 shows that the test of (5) that rejects when $\tilde{T}_n$ exceeds $\tilde{L}_n^{-1}(1-\alpha)$ controls size uniformly over $\mathbf{P}$ under (4). ■

3 Two-Step Method

The above calculations suggest a way of constructing a "two-step" critical value with which to compare T_n in the spirit of Romano et al. (2014). To this end, recall from the proof of Theorem 2.1 that $\mathbb{X}_0(P) := \{x \geq 0 : Ax = \beta(P)\}$ and, for $P \in \mathbf{P}_0$, $x_0(P) \in \mathbb{X}_0(P)$. Denote by $C_n(1-\gamma)$ a confidence set for $x_0(P)$, by which we mean a random set satisfying

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{x_0(P) \in C_n(1-\gamma)\} \geq 1-\gamma \ . \tag{7}$$

For instance, if $\sqrt{n}(\hat{\beta}_n - \beta(P)) \xrightarrow{d} N(0, \Sigma(P))$ as $n \to \infty$ with $\Sigma(P)$ invertible, then we may define

$$C_n(1-\gamma) := \{x \geq 0 : n(Ax - \hat{\beta}_n)'\hat{\Sigma}_n^{-1}(Ax - \hat{\beta}_n) \leq c\} \ ,$$

where c is the $1-\gamma$ quantile of the χ^2_d distribution and $\hat{\Sigma}_n$ is a suitable estimator of $\Sigma(P)$. For this choice of $C_n(1-\gamma)$, it is straightforward (by computing its projection onto each axis in $\mathbf{R}^d$) to compute a "greatest lower bound" w.r.t. the usual partial order on $\mathbf{R}^d$. Denote such a point in general by $\underline{x}_n$. For $\tilde{x} \in \mathbf{R}^d$, define

$$K_n(t, \tilde{x}, P) := P\{\inf_{y \geq -\tilde{x}} |Ay - \sqrt{n}(\hat{\beta}_n - \beta(P))| \leq t\} \ .$$

Using this notation, our two-step test is given by

$$\phi_n^{\text{two-step}} := I\{T_n > K_n^{-1}(1-\alpha+\gamma, \underline{x}_n, \hat{P}_n)\} \ .$$

The following theorem describes its behavior under (7) and an additional high-level assumption. We discuss these assumptions in the subsequent remarks.

Theorem 3.1. *Suppose that* (7) *holds and*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P\{T_n \leq K_n^{-1}(1-\alpha+\gamma, x_0(P), \hat{P}_n)\} \geq 1-\alpha+\gamma \ . \tag{8}$$

Then,

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} E_P[\phi_n^{\text{two-step}}] \leq \alpha \ .$$

PROOF: Let $P \in \mathbf{P}_0$. Define $E_n := \{x_0(P) \in C_n(1-\gamma)\}$. On the event E_n, we have that $-\underline{x}_n \geq -x_0(P)$, so we have that $K_n(t, x_0(P), P) \geq K_n(t, \underline{x}_n, P)$ for all $t \in \mathbf{R}$ and any P (including

$\hat{P}_n$). Thus, with probability one, we have that

$$K_n^{-1}(1-\alpha+\gamma, x_0(P), \hat{P}_n) \leq K_n^{-1}(1-\alpha+\gamma, \underline{x}_n, \hat{P}_n) \ . \tag{9}$$

Next, for any $P \in \mathbf{P}_0$, note that

$$\begin{aligned}
& P\{T_n > K_n^{-1}(1-\alpha+\gamma, \underline{x}_n, \hat{P}_n)\} \\
= \ & P\{T_n > K_n^{-1}(1-\alpha+\gamma, \underline{x}_n, \hat{P}_n) \cap E_n\} + P\{T_n > K_n^{-1}(1-\alpha+\gamma, \underline{x}_n, \hat{P}_n) \cap E_n^c\} \\
\leq \ & P\{T_n > K_n^{-1}(1-\alpha+\gamma, x_0(P), \hat{P}_n) \cap E_n\} + P\{T_n > K_n^{-1}(1-\alpha+\gamma, x_0(P), \hat{P}_n) \cap E_n^c\} \\
\leq \ & P\{T_n > K_n^{-1}(1-\alpha+\gamma, x_0(P), \hat{P}_n)\} + P\{E_n^c\} \ ,
\end{aligned}$$

where the first equality follows by inspection, the second exploits (9), and the third follows from Bonferroni's inequality. The desired conclusion now follows immediately from (7) and (8). ■

Remark 3.1. The two requirements (7) and (8) can be verified using arguments in, for example, Romano and Shaikh (2012) and Romano et al. (2014). The second of these requirements may look particularly high-level and therefore merits further discussion. It is useful to recall, however, from the proof of Theorem 2.1 that

$$T_n = \inf_{y \geq -\sqrt{n}x_0(P)} |Ay - \sqrt{n}(\hat{\beta}_n - \beta(P))| \ .$$

and thus $P\{T_n \leq t\} = P\{\sqrt{n}(\hat{\beta}_n - \beta(P)) \in \mathbb{A}(-\sqrt{n}x_0(P), t)\}$, where $\mathbb{A}(-\sqrt{n}x_0(P), t)$ is defined in the proof of Theorem 2.1. Hence, the assumption may be viewed as requiring that the bootstrap approximation to the distribution of $\sqrt{n}(\hat{\beta}_n - \beta(P))$ hold uniformly over a rich enough class of sets (e.g., all convex sets) and over $P \in \mathbf{P}$. Indeed, it suffices to assume that for every $\epsilon > 0$

$$\sup_{P \in \mathbf{P}} P\{\sup_{\mathbb{C} \in \mathcal{C}} |J_n(\mathbb{C}, P) - J_n(\mathbb{C}, \hat{P}_n)| > \epsilon\} \to 0 \ ,$$

where, as in Section 2, $\mathcal{C}$ is the set of all convex subsets of $\mathbf{R}^d$ and $J_n(\mathbb{C}, P)$ is defined as in (3). As in the preceding section, this condition may be readily verified in the case where $\beta(P)$ is a mean and $\hat{\beta}_n$ is the corresponding sample average under a weak uniform integrability assumption on $\mathbf{P}$. It can also be verified for the case where $\beta(P)$ is sufficiently "mean-like." We leave the development of such results for future work. ■

Remark 3.2. Different choices of $C_n(1-\gamma)$ may have substantial impact on the power of the test described in this section. While we describe one specific "elliptical" construction, it may be desirable to employ a "rectangular" confidence region as in Romano et al. (2014) and Bai et al.

(2019). This choice may be especially useful if it is desired to accommodate high-dimensional settings. As mentioned previously, our analysis does not permit p and/or d to grow with the sample size n. We expect, however, using arguments like those in Bai et al. (2019) and Fang et al. (2020) to be able to extend our results in this direction. ■

Remark 3.3. As in Remark 2.3, the testing procedure described in this section can be modified to test (5) differently. The key insight is to observe, by arguing as in the proof of Theorem 2.1, that

$$\tilde{T}_n = \inf_{y \geq -\sqrt{n} x_0(P) : By = 0} |Ay - \sqrt{n}(\hat{\beta}_n - \beta(P))| ,$$

where $\tilde{T}_n$ is defined in (6). By analogy with $K_n(t, \tilde{x}, P)$, we therefore introduce

$$\tilde{K}_n(t, \tilde{x}, P) := P\{ \inf_{y \geq -\tilde{x} : By = 0} |Ay - \sqrt{n}(\hat{\beta}_n - \beta(P))| \leq t \} .$$

With this notation, consider the test of (5) that rejects whenever $\tilde{T}_n$ exceeds $K_n^{-1}(1 - \alpha + \gamma, \underline{x}_n, \hat{P}_n)$. By arguing as in the proof of Theorem 3.1, this test can be shown to control size uniformly over $\mathbf{P}$ whenever

$$\liminf_{n \to \infty} \inf_{P \in \tilde{\mathbf{P}}_0} P\{\tilde{T}_n \leq \tilde{K}_n^{-1}(1 - \alpha + \gamma, x_0(P), \hat{P}_n)\} \geq 1 - \alpha + \gamma$$

and (7) hold. ■

Remark 3.4. We have omitted a discussion of Studentization at this time, but, as in similar problems, it may be desirable to do so, especially when different components of $\hat{\beta}_n$ may have considerable different variances.. We leave the development of such results for future work. ■

References

Andrews, I., Roth, J. and Pakes, A. (2019). Inference for linear conditional moment inequalities. Tech. rep., NBER.

Bai, Y., Santos, A. and Shaikh, A. (2019). A practical method for testing many moment inequalities. *Working Paper. University of Chicago.*

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

Chernozhukov, V., Newey, W. K. and Santos, A. (2015). Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311.*

Cox, G. and Shi, X. (2019). Simple adaptive size-exact testing for full-vector and subvector inference in moment inequality models. *arXiv preprint arXiv:1907.06317*.

Fang, Z., Santos, A., Shaikh, A. and Torgovitsky, A. (2020). Inference for large-scale linear systems with known coefficients. *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

Fang, Z. and Seo, J. (2021). A projection framework for testing shape restrictions that form convex cones. *Econometrica*, **89** 2439–2458.

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

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

Kitamura, Y. and Stoye, J. (2018). Nonparametric analysis of random utility models. *Econometrica*, **86** 1883–1909.

Romano, J. P. and Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference – Special Issue in Honor of Ted Anderson*.

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.

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

Zhu, Y. (2020). Inference in non-parametric/semi-parametric moment equality models with shape restrictions. *Quantitative Economics*, **11** 609–636.