Inference in partially identified models with many moment inequalities using Lasso*

Federico A. Bugni
Department of Economics
Duke University

Mehmet Caner
Department of Economics
Ohio State University

Anders Bredahl Kock
Department of Economics & Business
Aarhus University

Soumendra Lahiri
Department of Statistics
North Carolina State University

May 18, 2016

Abstract

This paper considers inference in a partially identified moment (in)equality model with many moment inequalities. We propose a novel two-step inference procedure that combines the methods proposed by Chernozhukov et al. (2014c) (CCK14, hereafter) with a first-step moment inequality selection based on the Lasso. Our method controls size uniformly, both in underlying parameter and data distribution. Also, the power of our method compares favorably with that of the corresponding two-step method in CCK14 for large parts of the parameter space, both in theory and in simulations. Finally, our Lasso-based first step is straightforward to implement.

Keywords and phrases: Many moment inequalities, self-normalizing sum, multiplier bootstrap, empirical bootstrap, Lasso, inequality selection.

JEL classification: C13, C23, C26.

*We thank useful comments and suggestions from the participants in the 2015 World Congress in Montreal, the Second International Workshop in Financial Econometrics, and the seminar at University of Maryland. Bugni acknowledges support by National Institutes of Health under grant no. 40-4153-00-0-85-399. Bredahl Kock acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Lahiri acknowledges support from National Science Foundation under grant no. DMS 130068.

1 Introduction

This paper contributes to the growing literature on inference in partially identified econometric models defined by many unconditional moment (in)equalities, i.e., inequalities and equalities. Consider an economic model with a parameter θ belonging to a parameter space Θ, whose main prediction is that the true value of θ, denoted by θ_0, satisfies a collection of moment (in)equalities. This model is partially identified, i.e., the restrictions of the model do not necessarily restrict θ_0 to a single value, but rather they constrain it to belong to a certain set, called the identified set. The literature on partially identified models discusses several examples of economic models that satisfy this structure, such as selection problems, missing data, or multiplicity of equilibria (see, e.g., Manski (1995) and Tamer (2003)).

The first contributions in the literature of partially identified moment (in)equalities focus on the case in which there is a fixed and finite number of moment (in)equalities, both unconditionally[1] and conditionally[2]. In practice, however, there are many relevant econometric models that produce a large set of moment conditions (even infinitely many). As several references in the literature point out (e.g. Menzel (2009, 2014)), the associated inference problems cannot be properly addressed by an asymptotic framework with a fixed number of moment (in)equalities.[3] To address this issue, Chernozhukov et al. (2014c) (hereafter referred to as CCK14) obtain inference results in a partially identified model with *many* moment (in)equalities.[4] According to this asymptotic framework, the number of moment (in)equalities, denoted by p, is allowed to be larger than the sample size n. In fact, the asymptotic framework allows p to be an increasing function of n and even to grow at certain exponential rates. Furthermore, CCK14 allow their moment (in)equalities to be "unstructured", in the sense that they do not impose restrictions on the correlation structure of the sample moment conditions.[5] For these reasons, CCK14 represents a significant advancement relative to the previous literature on inference in moment (in)equalities.

This paper builds on the inference method proposed in CCK14. Their goal is to test whether a collection of p moment inequalities simultaneously holds or not. In order to implement their test they propose a test

[1] These include Chernozhukov et al. (2007), Andrews et al. (2004), Imbens and Manski (2004), Galichon and Henry (2006, 2013), Beresteanu and Molinari (2008), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Stoye (2009), Andrews and Soares (2010), Bugni (2010, 2015), Canay (2010), Romano and Shaikh (2010), Andrews and Jia-Barwick (2012), Bontemps et al. (2012), Bugni et al. (2012), Romano et al. (2014), and Pakes et al. (2015), among others.

[2] These include Kim (2008), Ponomareva (2010), Armstrong (2014b,a), Chetverikov (2013), Andrews and Shi (2013), and Chernozhukov et al. (2013c), among others.

[3] As pointed out by Chernozhukov et al. (2014c), this is true even for conditional moment (in)equality models (which typically produce an infinite number of unconditional moment (in)equalities). As they explain, the unconditional moment (in)equalities generated by conditional moment (in)equality models inherit the structure from the conditional moment conditions, which limits the underlying econometric model.

[4] See also the related technical contributions in Chernozhukov et al. (2013b,a, 2014a,b).

[5] This feature distinguishes their framework from a standard conditional moment (in)equality model. While conditional moment conditions can generate an uncountable set of unconditional moment (in)equalities, their covariance structure is greatly restricted by the conditioning structure.

statistic based on the maximum of p Studentized statistics and several methods to compute the critical values. Their critical values may include a first stage inequality selection procedure with the objective of detecting slack moment inequalities, thus increasing the statistical power. According to their simulation results, including a first stage can result in significant power gains.

Our contribution is to propose a new inference method based on the combination of two ideas. On the one hand, our test statistic and critical values are based on those proposed by CCK14. On the other hand, we propose a new first stage selection procedure based on the Lasso. The Lasso was first proposed in the seminal contribution by Tibshirani (1996) as a regularization technique in the linear regression model. Since then, this method has found wide use as a dimension reduction technique in large dimensional models with strong theoretical underpinnings.[6] It is precisely these powerful shrinkage properties that serve as motivation to consider the Lasso as a procedure to separate out and select binding moment inequalities from the non-binding ones. Our Lasso first step inequality selection can be combined with any of the second step inference procedures in CCK14: self-normalization, multiplier bootstrap, or empirical bootstrap.

The present paper considers using the Lasso to select moments in a partially identified moment (in)equality model. In the context of point identified problems, there is an existing literature that proposes the Lasso to address estimation and moment selection in GMM settings. In particular, Caner (2009) introduce Lasso type GMM-Bridge estimators to estimate structural parameters in a general model. The problem of selection of moment in GMM is studied in Liao (2013) and Cheng and Liao (2015). In addition, Caner and Zhang (2014) and Caner et al. (2016) find a method to estimate parameters in GMM with diverging number of moments/parameters, and selecting valid moments among many valid or invalid moments respectively. In addition, Fan et al. (2015) consider the problem of inference in high dimensional models with sparse alternatives. Finally, Caner and Fan (2015) propose a hybrid two-step estimation procedure based on Generalized Empirical Likelihood, where instruments are chosen in a first-stage using an adaptive Lasso procedure.

We obtain the following results for our two-step Lasso inference methods. First, we provide conditions under which our methods are uniformly valid, both in the underlying parameter θ and the distribution of the data. According to the literature in moment (in)equalities, obtaining uniformly valid asymptotic results is important to guarantee that the asymptotic analysis provides an accurate approximation to finite sample

[6]For excellent reviews of this method see, e.g., Belloni and Chernozhukov (2011), Bühlmann and van de Geer (2011), Fan et al. (2011), and Hastie et al. (2015).

results.[7] Second, by virtue of results in CCK14, all of our proposed tests are asymptotically optimal in a minimax sense. Third, we compare the power of our methods to the corresponding one in CCK14, both in theory and in simulations. Since our two-step procedure and the corresponding one in CCK14 share the second step, our power comparison is a comparison of the Lasso-based first-step vis-à-vis the ones in CCK14. On the theory front, we obtain a region of underlying parameters under which the power of our method dominates that of CCK14. We also conduct extensive simulations to explore the practical consequences of our theoretical findings. Our simulations indicate that a Lasso-based first step is usually as powerful as the one in CCK14, and can sometimes be more powerful. Fourth, we show that our Lasso-based first step is straightforward to implement.

The remainder of the paper is organized as follows. Section 2 describes the inference problem and introduces our assumptions. Section 3 introduces the Lasso as a method to distinguish binding moment inequalities from non-binding ones and Section 4 considers inference methods that use the Lasso as a first step. Section 5 compares the power properties of inference methods based on the Lasso with the ones in the literature. Section 6 provides evidence of the finite sample performance using Monte Carlo simulations. Section 7 concludes. Proofs of the main results and several intermediate results are reported in the appendix.

Throughout the paper, we use the following notation. For any set S, $|S|$ denotes its cardinality, and for any vector $x \in \mathbb{R}^d$, $||x||_1 \equiv \sum_{i=1}^{d} |x_i|$.

2 Setup

For each $\theta \in \Theta$, let $X(\theta) : \Omega \to \mathbb{R}^k$ be a k-dimensional random variable with distribution $P(\theta)$ and mean $\mu(\theta) \equiv E_{P(\theta)}[X(\theta)] \in \mathbb{R}^k$. Let $\mu_j(\theta)$ denote the jth component of $\mu(\theta)$ so that $\mu(\theta) = \{\mu_j(\theta)\}_{j \le k}$. The main tenet of the econometric model is that the true parameter value θ_0 satisfies the following collection of p moment inequalities and $v \equiv k - p$ moment equalities:

$$\begin{aligned} \mu_j(\theta_0) &\le 0 \text{ for } j = 1, \ldots, p, \\ \mu_j(\theta_0) &= 0 \text{ for } j = p+1, \ldots, k. \end{aligned} \tag{2.1}$$

As in CCK14, we are implicitly allowing the collection P of distributions of $X(\theta)$ and the number of moment (in)equalities, $k = p + v$ to depend on n. In particular, we are primarily interested in the case in which

[7] In these models, the limiting distribution of the test statistic is discontinuous in the slackness of the moment inequalities, while its finite sample distribution does not exhibit such discontinuities. In consequence, asymptotic results obtained for any fixed distribution (i.e. pointwise asymptotics) can be grossly misleading, and possibly producing confidence sets that undercover (even asymptotically). See Imbens and Manski (2004), Andrews and Guggenberger (2009), Andrews and Soares (2010), and Andrews and Shi (2013) (Section 5.1).

$p = p_n \to \infty$ and $v = v_n \to \infty$ as $n \to \infty$, but the subscripts will be omitted to keep the notation simple. In particular, p and v can be much larger than the sample size and increase at rates made precise in Section 2.1. We allow the econometric model to be partially identified, i.e., the moment (in)equalities in Eq. (2.1) do not necessarily restrict θ_0 to a single value, but rather they constrain it to belong to the identified set, denoted by $\Theta_I(P)$. By definition, the identified set is as follows:

$$\Theta_I(P) \;\equiv\; \left\{\theta \in \Theta : \left\{ \begin{array}{l} \mu_j(\theta) \le 0 \text{ for } j = 1, \ldots, p, \\ \mu_j(\theta) = 0 \text{ for } j = p+1, \ldots, k. \end{array} \right\}\right\}. \tag{2.2}$$

Our goal is to test whether a particular parameter value $\theta \in \Theta$ is a possible candidate for the true parameter value $\theta_0 \in \Theta_I(P)$. In other words, we are interested in testing:

$$H_0 : \theta_0 = \theta \quad \text{vs.} \quad H_1 : \theta_0 \ne \theta. \tag{2.3}$$

By definition, the identified set is composed of all parameters that are observationally equivalent to the true parameter value θ_0, i.e., every parameter value in $\Theta_I(P)$ is a candidate for θ_0. In this sense, $\theta = \theta_0$ is observationally equivalent to $\theta \in \Theta_I(P)$ and so the hypothesis test in Eq. (2.3) can be equivalently reexpressed as:

$$\begin{gathered} H_0 : \theta \in \Theta_I(P) \quad \text{vs.} \quad H_1 : \theta \notin \Theta_I(P), \\ \text{i.e.,} \quad H_0 : \left\{ \begin{array}{l} \mu_j(\theta) \le 0 \text{ for all } j = 1, \ldots, p, \\ \mu_j(\theta) = 0 \text{ for all } j = p+1, \ldots, k. \end{array} \right\} \quad \text{vs.} \quad H_1 : \text{“not } H_0\text{”}. \end{gathered} \tag{2.4}$$

In this paper, we propose a procedure to implement the hypothesis test in Eq. (2.3) (or, equivalently, Eq. (2.4)) with a given significance level $\alpha \in (0, 1)$ based on a random sample of $X(\theta) \sim P(\theta)$, denoted by $X^n(\theta) \equiv \{X_i(\theta)\}_{i \le n}$. The inference procedure will reject the null hypothesis whenever a certain test statistic $T_n(\theta)$ exceeds a critical value $c_n(\alpha, \theta)$, i.e.,

$$\phi_n(\alpha, \theta) \;\equiv\; 1[T_n(\theta) > c_n(\alpha, \theta)], \tag{2.5}$$

where $1[\cdot]$ denotes the indicator function. By the duality between hypothesis tests and confidence sets, a confidence set for θ_0 can be constructed by collecting all parameter values for which the inference procedure is not rejected, i.e.,

$$C_n(1-\alpha) \equiv \{\theta \in \Theta : T_n(\theta) \leq c_n(\alpha,\theta)\}. \tag{2.6}$$

Our formal results will have the following structure. Let $\mathcal{P}$ denote a set of probability distributions. We will show that for all $P \in \mathcal{P}$ and under H_0,

$$P\left(T_n(\theta) > c_n(\alpha,\theta)\right) \leq \alpha + o(1). \tag{2.7}$$

Moreover, the convergence in Eq. (2.7) will be shown to occur uniformly over both $P \in \mathcal{P}$ and $\theta \in \Theta$. This uniform size control result in Eq. (2.7) has important consequences regarding our inference problem. First, this result immediately implies that the hypothesis test procedure in Eq. (2.5) uniformly controls asymptotic size i.e., for all $\theta \in \Theta$ and under $H_0 : \theta_0 = \theta$,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} E[\phi_n(\alpha,\theta)] \leq \alpha. \tag{2.8}$$

Second, the result also implies that the confidence set in Eq. (2.6) is asymptotically uniformly valid, i.e.,

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P\left(\theta \in C_n(1-\alpha)\right) \geq 1-\alpha. \tag{2.9}$$

The rest of the section is organized as follows. Section 2.1 specifies the assumptions on the probability space $\mathcal{P}$ that are required for our analysis. All the inference methods described in this paper share the test statistic $T_n(\theta)$ and differ only in the critical value $c_n(\alpha,\theta)$. The common test statistic is introduced and described in Section 2.2.

2.1 Assumptions

The collection of distributions $P \equiv \{P(\theta) : \theta \in \Theta\}$ are assumed to satisfy the following assumptions.

Assumption A.1. For every $\theta \in \Theta$, let $X^n(\theta) \equiv \{X_i(\theta)\}_{i\leq n}$ be i.i.d. k-dimensional random vectors distributed according to $P(\theta)$. Further, let $E_{P(\theta)}[X_{1j}(\theta)] \equiv \mu_j(\theta)$ and $Var_{P(\theta)}[X_{1j}(\theta)] \equiv \sigma_j^2(\theta) > 0$, where $X_{ij}(\theta)$ denotes the j component of $X_i(\theta)$.

Assumption A.2. For some $\delta \in (0,1]$, $\max_{j=1,\ldots,k} \sup_{\theta\in\Theta}(E_{P(\theta)}[|X_{1j}(\theta)|^{2+\delta}])^{1/(2+\delta)} \equiv M_{n,2+\delta} < \infty$ and $M_{n,2+\delta}^{2+\delta}(\ln(2k-p))^{(2+\delta)/2} n^{-\delta/2} \to 0$.

Assumption A.3. For some $c \in (0,1)$, $(n^{-(1-c)/2}\ln(2k-p) + n^{-3/2}(\ln(2k-p))^2)B_n^2 \to 0$, where $\sup_{\theta\in\Theta}(E_{P(\theta)}[\max_{j=1,\ldots,k}|Z_{1j}(\theta)|^4])^{1/4} \equiv B_n < \infty$ and $Z_{ij}(\theta) \equiv (X_{ij}(\theta) - \mu_j(\theta))/\sigma_j(\theta)$.

Assumption A.4. For some $c \in (0,1/2)$ and $C > 0$, $\max\{M_{n,3}^3, M_{n,4}^2, B_n\}^2 \ln((2k-p)n)^{7/2} \leq Cn^{1/2-c}$, where $M_{n,2+\delta}$ and B_n are as in Assumptions A.2-A.3.

We now briefly describe these assumptions. Assumption A.1 is standard in microeconometric applications. Assumption A.2 has two parts. The first part requires that $X_{ij}(\theta)$ has finite $(2+\delta)$-moments for all $j = 1,\ldots,k$. The second part limits the rate of growth of $M_{n,2+\delta}$ and the number of moment (in)equalities. Notice that $M_{n,2+\delta}$ is a function of the sample size because $\max_{j=1,\ldots,k}\sup_{\theta\in\Theta}(E_{P(\theta)}[|X_{1j}(\theta)|^{2+\delta}])^{1/(2+\delta)}$ is function of P and $k = v + p$, both of which could depend on n. Also, notice that $2k - p = 2v + p$, i.e., the total number of moment inequalities p plus twice the number of moment equalities v, all of which could depend on n. Assumption A.3 could be interpreted in a similar fashion as Assumption A.2, except that it refers to the standardized random variable $Z_{ij}(\theta) \equiv (X_{ij}(\theta) - \mu_j(\theta))/\sigma_j(\theta)$. Assumption A.4 is a technical assumption that is used to control the size of the bootstrap test in CCK14.[8]

2.2 Test statistic

Throughout the paper, we consider the following test statistic:

$$T_n(\theta) \equiv \max\left\{\max_{j=1,\ldots,p}\frac{\sqrt{n}\hat{\mu}_j(\theta)}{\hat{\sigma}_j(\theta)}, \max_{s=p+1,\ldots,k}\frac{\sqrt{n}\left|\hat{\mu}_s(\theta)\right|}{\hat{\sigma}_s(\theta)}\right\}, \tag{2.10}$$

where, for $j = 1,\ldots,k$, $\hat{\mu}_j(\theta) \equiv \frac{1}{n}\sum_{i=1}^n X_{ij}(\theta)$ and $\hat{\sigma}_j^2(\theta) \equiv \frac{1}{n}\sum_{i=1}^n \left(X_{ij}(\theta) - \hat{\mu}_j(\theta)\right)^2$. Note that Eq. (2.10) is not properly defined if $\hat{\sigma}_j^2(\theta) = 0$ for some $j = 1,\ldots,k$ and, in such cases, we use the convention that $C/0 \equiv \infty \times 1[C > 0] - \infty \times 1[C < 0]$.

The test statistic is identical to that in CCK14 with the exception that we allow for the presence of moment equalities. By definition, large values of $T_n(\theta)$ are an indication that $H_0 : \theta = \theta_0$ is likely to be violated, leading to the hypothesis test in Eq. (2.5). The remainder of the paper considers several procedures to construct critical values that can be associated to this test statistic.

[8] We point out that Assumptions A.1-A.4 are tailored for the construction of confidence sets in Eq. (2.6) in the sense that all the relevant constants are defined uniformly in $\theta \in \Theta$. If we were only interested in the hypothesis testing problem for a particular value of θ, then the previous assumptions could be replaced by their "pointwise" versions at the parameter value of interest.

3 Lasso as a first step moment selection procedure

In order to propose a critical value for our test statistic $T_n(\theta)$, we need to approximate its distribution under the null hypothesis. According to the econometric model in Eq. (2.1), the true parameter satisfies p moment inequalities and v moment equalities. By definition, the moment equalities are always binding under the null hypothesis. On the other hand, the moment inequalities may or may not be binding, and a successful approximation of the asymptotic distribution depends on being able to distinguish between these two cases. Incorporating this information into the hypothesis testing problem is one of the key issues in the literature on inference in partially identified moment (in)equality models.

In their seminal contribution, CCK14 is the first paper in the literature to conduct inference in a partially identified model with many unstructured moment inequalities. Their paper proposes several procedures to select binding moment inequalities from non-binding based on three approximation methods: self-normalization (SN), multiplier bootstrap (MB), and empirical bootstrap (EB). Our relative contribution is to propose a novel approximation method based on the Lasso. By definition, the Lasso penalizes parameters values by their ℓ_1-norm, with the ability of producing parameter estimates that are exactly equal to zero. This powerful shrinkage property is precisely what motivates us to consider the Lasso as a first step moment selection procedure in a model with many moment (in)equalities. As we will soon show, the Lasso is an excellent method to detect binding moment inequalities from non-binding ones, and this information can be successfully incorporated into an inference procedure for many moment (in)equalities.

For every $\theta \in \Theta$, let $J(\theta)$ denote the true set of binding moment inequalities, i.e., $J(\theta) \equiv \{j = 1, \ldots, p \; : \; \mu_j(\theta) \geq 0\}$. Let $\mu_I(\theta) \equiv \{\mu_j(\theta)\}_{j=1}^p$ denote the moment vector for the moment inequalities and let $\hat{\mu}_I(\theta) \equiv \{\hat{\mu}_j(\theta)\}_{j=1}^p$ denote its sample analogue. In order to detect binding moment inequalities, we consider the weighted Lasso estimator of $\mu_I(\theta)$, given by:

$$\hat{\mu}_L(\theta) \;\equiv\; \underset{t \in \mathbb{R}^p}{\arg\min} \left\{ \left(\hat{\mu}_I(\theta) - t\right)' \hat{W}(\theta) \left(\hat{\mu}_I(\theta) - t\right) + \lambda_n \left\| \hat{W}(\theta)^{1/2} t \right\|_1 \right\}, \tag{3.1}$$

where λ_n is a positive penalization sequence that controls the amount of regularization and $\hat{W}(\theta)$ is a positive definite weighting matrix. To simplify the computation of the Lasso estimator, we impose $\hat{W}(\theta) \equiv diag\{1/\hat{\sigma}_j(\theta)^2\}_{j=1}^p$. As a consequence, Eq. (3.1) becomes:

$$\hat{\mu}_L(\theta) \;=\; \left\{ \underset{m \in \mathbb{R}}{\arg\min} \left\{ \left(\hat{\mu}_j(\theta) - m\right)^2 + \lambda_n \hat{\sigma}_j(\theta)|m| \right\} \right\}_{j=1}^p. \tag{3.2}$$

Notice that instead of using the Lasso in one p-dimensional model we instead use it in p one-dimensional models. As we shall see later, $\hat{\mu}_L(\theta)$ in Eq. (3.2) is closely linked to the soft-thresholded least squares estimator, which implies that its computation is straightforward. The Lasso estimator $\hat{\mu}_L(\theta)$ implies a Lasso-based estimator of $J(\theta)$, given by:

$$\hat{J}_L(\theta) \;\equiv\; \{j = 1, \ldots, p : \hat{\mu}_{j,L}(\theta)/\hat{\sigma}_j(\theta) \;\geq\; -\lambda_n\}. \tag{3.3}$$

In order to implement this procedure, we need to choose the sequence λ_n, which determines the degree of regularization imposed by the Lasso. A higher value of λ_n will produce a larger number of moment inequalities considered to be binding, resulting in a lower rejection rate. In consequence, this is a critical choice for our inference methodology. According to our theoretical results, a suitable choice of λ_n is given by:

$$\lambda_n = (4/3 + \varepsilon) n^{-1/2} \left(M_{n,2+\delta}^2 n^{-\delta/(2+\delta)} - n^{-1} \right)^{-1/2} \tag{3.4}$$

for any arbitrary $\varepsilon > 0$. Assumption A.2 implies that λ_n in Eq. (3.4) satisfies $\lambda_n \to 0$. Notice that Eq. (3.4) is infeasible as it depends on the unknown expression $M_{n,2+\delta}$. In practice, one can replace this unknown expression with its sample analogue:

$$\hat{M}_{n,2+\delta}^2 = \max_{j=1,\ldots,k} \sup_{\theta\in\Theta} \left(n^{-1} \sum\nolimits_{i=1}^{n} |X_{ij}(\theta)|^{2+\delta} \right)^{2/(2+\delta)}.$$

In principle, a more rigorous choice of λ_n can be implemented via a modified BIC method designed for divergent number of parameters as in Wang et al. (2009) or Caner et al. (2016).[9]

As explained earlier, our Lasso procedure is used as a first step in order to detect binding moment inequalities from non-binding ones. The following result formally establishes that our Lasso procedure includes all binding ones with a probability that approaches one, uniformly.

Lemma 3.1. *Assume Assumptions A.1-A.3, and let λ_n be as in Eq. (3.4). Then,*

$$P[J(\theta) \subseteq \hat{J}_L(\theta)] \;\geq\; 1 - 2p \exp\left(- \frac{n^{\delta/(2+\delta)}}{2M_{n,2+\delta}^2} \right) \left[1 + K \left(\frac{M_{n,2+\delta}}{n^{\delta/(2(2+\delta))}} + 1 \right)^{2+\delta} \right] + \tilde{K} n^{-c} \;=\; 1 + o(1),$$

where $K, \tilde{K}$ are universal constants and the convergence is uniform in all parameters $\theta \in \Theta$ and distributions P that satisfy the assumptions in the statement.

Thus far, our Lasso estimator of the binding constrains in Eq. (3.3) has been defined in terms of the

[9] Nevertheless, it is unclear whether the asymptotic properties of this method carry over to our partially identified moment (in)equality model. We consider that a rigorous handling of these issues is beyond the scope of this paper.

solution of the p-dimensional minimization problem in Eq. (3.2). We conclude the subsection by providing an equivalent closed form solution for this set.

Lemma 3.2. *Eq.* (3.3) *can be equivalently reexpressed as follows:*

$$\hat{J}_L(\theta) \;=\; \{j = 1, \ldots, p : \hat{\mu}_j(\theta)/\hat{\sigma}_j(\theta) \geq -3\lambda_n/2\}. \tag{3.5}$$

Lemma 3.2 is a very important computational aspect of our methodology. This result reveals that $\hat{J}_L(\theta)$ can be computed by comparing standardized sample averages with a modified threshold of $-3\lambda_n/2$. In other words, our Lasso-based first stage can be implemented without the need of solving the p-dimensional minimization problem in Eq. (3.2).

4 Inference methods with Lasso first step

In the remainder of the paper we show how to conduct inference in our partially identified many moment (in)equality model by combining the Lasso-based first step in Section 3 with a second step based on the inference methods proposed by CCK14. In particular, Section 4.1 combines our Lasso-based first step with their self-normalization approximation, while Section 4.2 combines it with their bootstrap approximations.

4.1 Self-normalization approximation

Before describing our self-normalization (SN) approximation with Lasso first stage, we first describe the "plain vanilla" SN approximation without first stage moment selection. Our treatment extends the SN method proposed by CCK14 to the presence of moment equalities.

As a preliminary step, we now define the SN approximation to the $(1-\alpha)$-quantile of $T_n(\theta)$ in a hypothetical moment (in)equality model composed of $|J|$ moment inequalities and $k-p$ moment equalities, given by:

$$c_n^{SN}(|J|, \alpha) \equiv \begin{cases} 0 & \text{if } 2(k-p)+|J| = 0, \\ \dfrac{\Phi^{-1}\big(1-\alpha/(2(k-p)+|J|)\big)}{\sqrt{1-\Big(\Phi^{-1}\big(1-\alpha/(2(k-p)+|J|)\big)\Big)^2/n}} & \text{if } 2(k-p)+|J| > 0. \end{cases} \tag{4.1}$$

Lemma A.4 in the appendix shows that $c_n^{SN}(|J|, \alpha)$ provides asymptotic uniform size control in a hypothetical moment (in)equality model with $|J|$ moment inequalities and $k-p$ moment equalities under Assumptions A.1-A.2. The main difference between this result and CCK14 (Theorem 4.1) is that we allow for the presence of moment equalities. Since our moment (in)equality model has $|J| = p$ moment inequalities and $k-p$ moment equalities, we can define the regular (i.e. one-step) SN approximation method by using $|J| = p$ in Eq. (4.1), i.e.,

$$c_n^{SN,1S}(\alpha) \equiv c_n^{SN}(p,\alpha) = \frac{\Phi^{-1}\big(1-\alpha/(2k-p)\big)}{\sqrt{1-\Big(\Phi^{-1}\big(1-\alpha/(2k-p)\big)\Big)^2/n}}.$$

The following result is a corollary of Lemma A.4.

Theorem 4.1 (One-step SN approximation)**.** *Assume Assumptions A.1-A.2, $\alpha \in (0, 0.5)$, and that H_0 holds. Then,*

$$P\left(T_n(\theta) > c_n^{SN,1S}(\alpha)\right) \;\leq\; \alpha + \alpha K n^{-\delta/2} M_{n,2+\delta}^{2+\delta}\Big(1 + \Phi^{-1}\left(1-\alpha/(2k-p)\right)\Big)^{2+\delta} = \alpha + o(1),$$

where K is a universal constant and the convergence is uniform in all parameters $\theta \in \Theta$ and distributions P that satisfy the assumptions in the statement.

By definition, this SN approximation considers all moment inequalities in the model as binding. A more powerful test can be constructed by using the data to reveal which moment inequalities are slack. In particular, CCK14 propose a two-step SN procedure which combines a first step moment inequality based on SN methods and the second step SN critical value in Theorem 4.1. If we adapt their procedure to the presence of moment equalities, this would be given by:

$$c^{SN,2S}(\theta,\alpha) \;\equiv\; c_n^{SN}(|\hat{J}_{SN}(\theta)|, \alpha - 2\beta_n) \tag{4.2}$$

with:

$$\hat{J}_{SN}(\theta) \;\equiv\; \Big\{ j \in \{1,\ldots,p\} \;:\; \sqrt{n}\hat{\mu}_j(\theta)/\hat{\sigma}_j(\theta) > -2c^{SN,1S}(\beta_n) \Big\},$$

where $\{\beta_n\}_{n\geq 1}$ is an arbitrary sequence of constants in $(0, \alpha/3)$. By extending arguments in CCK14 to include moment equalities, one can show that inference based on the critical value $c^{SN,2S}(\theta,\alpha)$ in Eq. (4.2) is asymptotically valid in a uniform sense.

In this paper, we propose an alternative SN procedure by using our Lasso-based first step. In particular, we define the following two-step Lasso SN critical value:

$$c_n^{SN,L}(\theta,\alpha) \;\equiv\; c_n^{SN}(|\hat{J}_L(\theta)|, \alpha), \tag{4.3}$$

where $\hat{J}_L(\theta)$ is as in Eq. (3.5). The following result shows that an inference method based on our two-step

Lasso SN critical value is asymptotically valid in a uniform sense.

Theorem 4.2 (Two-step Lasso SN approximation). *Assume Assumptions A.1-A.3, $\alpha \in (0, 0.5)$, and that H_0 holds, and let λ_n be as in Eq. (3.4). Then,*

$$
\begin{aligned}
&P\left(T_n(\theta) > c_n^{SN,L}(\theta,\alpha)\right) \\
&\leq \alpha + \left[\begin{array}{c} \alpha K n^{-\delta/2} M_{n,2+\delta}^{2+\delta}(1+\Phi^{-1}\left(1-\alpha/(2k-p)\right))^{2+\delta} + \\ 4p\exp\left(-2^{-1}n^{\delta/(2+\delta)}M_{n,2+\delta}^{-2}\right)\left[1+K\left(n^{-\delta/(2(2+\delta))}M_{n,2+\delta}+1\right)^{2+\delta}\right] + 2\tilde{K}n^{-c} \end{array}\right] \\
&= \alpha + o(1),
\end{aligned}
$$

where $K, \tilde{K}$ are universal constants and the convergence is uniform in all parameters $\theta \in \Theta$ and distributions P that satisfy the assumptions in the statement.

We now compare our two-step SN Lasso method with the SN methods in CCK14. Since all inference methods share the test statistic, the only difference lies in the critical values. While the one-step SN critical values considers all p moment inequalities as binding, our two-step SN Lasso critical value considers only $|\hat{J}_L(\theta)|$ moment inequalities as binding. Since $|\hat{J}_L(\theta)| \leq p$ and $c_n^{SN}(\alpha, |J|)$ is weakly increasing in $|J|$ (see Lemma A.3 in the appendix), then our two-step SN method results in a weakly larger rejection probability for all sample sizes. In contrast, the comparison between $c_n^{SN,L}(\theta,\alpha)$ and $c_n^{SN,2S}(\theta,\alpha)$ is not straightforward as these differ in two aspects. First, the set of binding constrains $\hat{J}_{SN}(\theta)$ according to SN differs from the set of binding constrains $\hat{J}_L(\theta)$ according to the Lasso. Second, the quantile of the critical values are different: the two-step SN method in Eq. (4.2) considers the $\alpha - 2\beta_n$ quantile while the Lasso-based method considers the usual α quantile. As a result of these differences, the comparison of these critical values is ambiguous and so is the resulting power comparison. This topic will be discussed in further detail in Section 5.

4.2 Bootstrap methods

CCK14 also propose two bootstrap approximation methods: multiplier bootstrap (MB) and empirical bootstrap (EB). Relative to the SN approximation, bootstrap methods have the advantage of taking into account the dependence between the coordinates of $\{\sqrt{n}\hat{\mu}_j(\theta)/\hat{\sigma}_j(\theta)\}_{j=1}^p$ involved in the definition of the test statistic $T_n(\theta)$.

As in the previous subsection, we first define the bootstrap approximation to the $(1-\alpha)$-quantile of $T_n(\theta)$ in a hypothetical moment (in)equality model composed of moment inequalities indexed by the set J and the $k-p$ moment equalities. The corresponding MB and EB approximations are denoted by $c_n^{MB}(\theta, J, \alpha)$ and

$c_n^{EB}(\theta, J, \alpha)$, respectively, and are computed as follows.

Algorithm 4.1. **Multiplier bootstrap (MB)**

1. Generate i.i.d. standard normal random variables $\{\epsilon_i\}_{i=1}^n$, and independent of the data $X^n(\theta)$.

2. Construct the multiplier bootstrap test statistic:

$$W_n^{MB}(\theta, J) = \max\left\{\max_{j\in J}\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i(X_{ij}(\theta)-\hat{\mu}_j(\theta))}{\hat{\sigma}_j(\theta)}, \max_{s=p+1,\ldots,k}\frac{\frac{1}{\sqrt{n}}|\sum_{i=1}^n \epsilon_i(X_{is}(\theta)-\hat{\mu}_s(\theta))|}{\hat{\sigma}_s(\theta)}\right\}.$$

3. Calculate $c_n^{MB}(\theta, J, \alpha)$ as the conditional $(1-\alpha)$-quantile of $W_n^{MB}(\theta, J)$ (given $X^n(\theta)$).

Algorithm 4.2. **Empirical bootstrap (EB)**

1. Generate a bootstrap sample $\{X_i^*(\theta)\}_{i=1}^n$ from the data, i.e., an i.i.d. draw from the empirical distribution of $X^n(\theta)$.

2. Construct the empirical bootstrap test statistic:

$$W_n^{EB}(\theta, J) = \max\left\{\max_{j\in J}\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_{ij}^*(\theta)-\hat{\mu}_j(\theta))}{\hat{\sigma}_j(\theta)}, \max_{s=p+1,\ldots,k}\frac{\frac{1}{\sqrt{n}}|\sum_{i=1}^n (X_{is}^*(\theta)-\hat{\mu}_s(\theta))|}{\hat{\sigma}_s(\theta)}\right\}.$$

3. Calculate $c_n^{EB}(\theta, J, \alpha)$ as the conditional $(1-\alpha)$-quantile of $W_n^{EB}(\theta, J)$ (given $X^n(\theta)$).

All the results in the remainder of the section will apply to both versions of the bootstrap, and under the same assumptions. For this reason, we can use $c_n^B(\theta, J, \alpha)$ to denote the bootstrap critical value where $B \in \{MB, EB\}$ represents either MB or EB. Lemma A.5 in the appendix shows that $c_n^B(\theta, J, \alpha)$ for $B \in \{MB, EB\}$ provides asymptotic uniform size control in a hypothetical moment (in)equality model composed of moment inequalities indexed by the set J and the $k-p$ moment equalities under Assumptions A.1 and A.4. As in Section 4.1, the main difference between this result and CCK14 (Theorem 4.3) is that we allow for the presence of the moment equalities. Since our moment (in)equality model has $|J| = p$ moment inequalities and $k-p$ moment equalities, we can define the regular (i.e. one-step) MB or EB approximation method by using $|J| = p$ in Algorithm 4.1 or 4.2, respectively, i.e.,

$$c_n^{B,1S}(\theta, \alpha) \equiv c_n^B(\theta, \{1, \ldots, p\}, \alpha),$$

where $c_n^B(\theta, J, \alpha)$ is as in Algorithm 4.1 if $B = MB$ or Algorithm 4.2 if $B = EB$. The following result is a corollary of Lemma A.5.

Theorem 4.3 (One-step bootstrap approximation). *Assume Assumptions A.1, A.4, $\alpha \in (0, 0.5)$, and that H_0 holds. Then,*

$$P\left(T_n(\theta) > c_n^{B,1S}(\theta,\alpha)\right) \;\leq\; \alpha + \tilde{C}n^{-\tilde{c}},$$

where $\tilde{c}, \tilde{C} > 0$ are positive constants that only depend on the constants c, C in Assumption A.4. Furthermore, if $\mu_j(\theta) = 0$ for all $j = 1, \ldots, p$, then

$$|P\left(T_n(\theta) > c_n^{B,1S}(\theta,\alpha)\right) - \alpha| \leq \tilde{C}n^{-\tilde{c}}.$$

Finally, the proposed bounds are uniform in all parameters $\theta \in \Theta$ and distributions P that satisfy the assumptions in the statement.

As in the SN approximation method, the regular (one-step) bootstrap approximation considers all moment inequalities in the model as binding. A more powerful bootstrap-based test can be constructed using the data to reveal which moment inequalities are slack. However, unlike in the SN approximation method, Theorem 4.3 shows that the size of the test using the bootstrap critical values converges to α when all the moment inequalities are binding. This difference comes from the fact that the bootstrap can better approximate the correlation structure in the moment inequalities, which is not taken into account by the SN approximation. As we will see in simulations, this translates into power gains in favor of the bootstrap.

CCK14 propose a two-step bootstrap procedure, combining a first step moment inequality based on the bootstrap with the second step bootstrap critical value in Theorem 4.3.[10] If we adapt their procedure to the presence of moment equalities, this would be given by:

$$c^{B,2S}(\theta,\alpha) \;\equiv\; c_n^B(\theta, \hat{J}_B(\theta), \alpha - 2\beta_n) \tag{4.4}$$

with:

$$\hat{J}_B(\theta) \;\equiv\; \{j \in \{1,\ldots,p\} \;:\; \sqrt{n}\hat{\mu}_j(\theta)/\hat{\sigma}_j(\theta) > -2c^{B,1S}(\alpha,\beta_n)\},$$

where $\{\beta_n\}_{n\geq 1}$ is an arbitrary sequence of constants in $(0, \alpha/2)$. Again, by extending arguments in CCK14 to the presence of moment equalities, one can show that an inference method based on the critical value $c^{B,2S}(\theta,\alpha)$ in Eq. (4.4) is asymptotically valid in a uniform sense.

[10] They also consider the so-called "hybrid" procedures in which the first step can be based on one approximation method (e.g. SN approximation) and the second step could be based on another approximation method (e.g. bootstrap). While these are not explicitly addressed in this section they are included in the Monte Carlo section.

This paper proposes an alternative bootstrap procedure by using our Lasso-based first step. For $B \in \{MB, EB\}$, define the following two-step Lasso bootstrap critical value:

$$c_n^{B,L}(\theta,\alpha) \;\equiv\; c_n^B(\theta,\hat{J}_L(\theta),\alpha), \tag{4.5}$$

where $\hat{J}_L(\theta)$ is as in Eq. (3.5), and $c_n^B(\theta,J,\alpha)$ is as in Algorithm 4.1 if $B = MB$ or Algorithm 4.2 if $B = EB$. The following result shows that an inference method based on our two-step Lasso bootstrap critical value is asymptotically valid in a uniform sense.

Theorem 4.4 (Two-step Lasso bootstrap approximation)**.** *Assume Assumptions A.1, A.2, A.3, A.4, $\alpha \in (0, 0.5)$, and that H_0 holds, and let λ_n be as in Eq. (3.4). Then, for $B \in \{MB, EB\}$,*

$$\begin{aligned}
&P\left(T_n(\theta) > c_n^{B,L}(\theta,\alpha)\right)\\
&\leq \alpha + \tilde{C}n^{-\tilde{c}} + Cn^{-c} + 2\tilde{K}n^{-c} + 4p\exp\left(2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2\right)\left[1 + K(M_{n,2+\delta}/n^{\delta/(2(2+\delta)} + 1)^{2+\delta}\right]\\
&= \alpha + o(1),
\end{aligned}$$

where $\tilde{c}, \tilde{C} > 0$ are positive constants that only depend on the constants c, C in Assumption A.4, $K, \tilde{K}$ are universal constants, and the convergence is uniform in all parameters $\theta \in \Theta$ and distributions P that satisfy the assumptions in the statement. Furthermore, if $\mu_j(\theta) = 0$ for all $1 \leq j \leq p$ and

$$\tilde{K}n^{-c} + 2p\exp\left(2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2\right)\left[1 + K(M_{n,2+\delta}/n^{\delta/(2(2+\delta)} + 1)^{2+\delta}\right] \leq \tilde{C}n^{-\tilde{c}}, \tag{4.6}$$

then,

$$|P\left(T_n(\theta) > c_n^{B,L}(\theta,\alpha)\right) - \alpha| \leq 3\tilde{C}n^{-\tilde{c}} + Cn^{-c} = o(1),$$

where all constants are as defined earlier and the convergence is uniform in all parameters $\theta \in \Theta$ and distributions P that satisfy the assumptions in the statement.

By repeating arguments at the end of Section 4.1, it follows that our two-step bootstrap method results in a larger rejection probability than the one-step bootstrap method for all sample sizes.[11] Also, the comparison between $c_n^{B,L}(\theta,\alpha)$ and $c_n^{B,2S}(\theta,\alpha)$ is not straightforward as these differ in the same two aspects described Section 4.1. This comparison will be the topic of the next section.

[11] To establish this result, we now use Lemma A.6 instead of Lemma A.3.

5 Power comparison

CCK14 show that all of their inference methods satisfy uniform asymptotic size control under appropriate assumptions. Theorems 4.2 and 4.4 show that our Lasso-based two-step inference methods also satisfy uniform asymptotic size control under similar assumptions. Given these results, the natural next step is to compare these inference methods in terms of criteria related to power.

One possible such criterion is minimax optimality, i.e., the ability that a test has of rejecting departures from the null hypothesis at the fastest possible rate (without loosing uniform size control). CCK14 show that all their proposed inference methods are asymptotically optimal in a minimax sense, even in the absence of any inequality selection (i.e. defined as in Theorems 4.1 and 4.3 in the presence of moment equalities). Since our Lasso-based inequality selection can only reduce the number of binding moment inequalities (thus increasing rejection), we can also conclude that all of our two-step Lasso-based inference methods (SN, MB, and EB) are also asymptotically optimal in a minimax sense. In other words, minimax optimality is a desirable property that is satisfied by all tests under consideration and, thus, cannot be used as a criterion to distinguish between them.

Thus, we proceed to compare our Lasso-based inference procedures with those proposed by CCK14 in terms of rejection rates. Since all inference methods share the test statistic $T_n(\theta)$, the power comparison depends exclusively on the critical values.

5.1 Comparison with one-step methods

As pointed out in previous sections, our Lasso-based two-step inference methods will always be more powerful than the corresponding one-step analogue, i.e.,

$$\begin{aligned} P\left(T_n(\theta) > c_n^{SN,L}(\theta,\alpha)\right) &\geq P\left(T_n(\theta) > c_n^{SN,1S}(\alpha)\right) \\ P\left(T_n(\theta) > c_n^{B,L}(\theta,\alpha)\right) &\geq P\left(T_n(\theta) > c_n^{B,1S}(\theta,\alpha)\right) \quad \forall B \in \{MB, EB\}, \end{aligned}$$

for all $\theta \in \Theta$ and $n \in \mathbb{N}$. This is a direct consequence of the fact that one-step critical values are based on considering all moment inequalities as binding, while the Lasso-based first-step will restrict attention to the subset of them that are sufficiently close to binding, i.e., $\hat{J}_L(\theta) \subseteq \{1, \ldots, p\}$.

5.2 Comparison with two-step methods

The comparison between our two-step Lasso procedures and the two-step methods in CCK14 is not straightforward for two reasons. First, the set of binding inequalities according to the Lasso might be different from the other methods. Second, our Lasso-based methods considers the usual α quantile while the other two-step methods consider the $\alpha - 2\beta_n$ quantile for a sequence of positive constants $\{\beta_n\}_{n\geq 1}$.

To simplify the discussion, we focus exclusively on the case where the moment (in)equality model is only composed of inequalities, i.e., $k = p$, which is precisely the setup in CCK14. This is done for simplicity of exposition, the introduction of moment equalities would not qualitatively change the conclusions that follow.

We begin by comparing the two-step SN method with the two-step Lasso SN method. For all $\theta \in \Theta$ and $n \in \mathbb{N}$, our two-step Lasso SN method will have more power than the two-step SN method if and only if $c_n^{SN,L}(\theta,\alpha) \;\leq\; c_n^{SN,2S}(\alpha)$. By inspecting the formulas in CCK14, this occurs if and only if:

$$|\hat{J}_L(\theta)| \;\leq\; \frac{\alpha}{\alpha - 2\beta_n}|\hat{J}_{SN}(\theta)|, \tag{5.1}$$

where, by definition, $\{\beta_n\}_{n\geq 1}$ satisfies $\beta_n \leq \alpha/3$. We provide sufficient conditions for Eq. (5.1) in the following result.

Theorem 5.1. *For all $\theta \in \Theta$ and $n \in \mathbb{N}$,*

$$\hat{J}_L(\theta) \quad\subseteq\quad \hat{J}_{SN}(\theta) \tag{5.2}$$

implies

$$P\left(T_n(\theta) > c_n^{SN,L}(\theta,\alpha)\right) \quad\geq\quad P\left(T_n(\theta) > c_n^{SN,2S}(\alpha)\right). \tag{5.3}$$

In turn, Eq. (5.2) occurs under any of the following circumstances:

$$\frac{4}{3}c_n^{SN}(\beta_n) \geq \sqrt{n}\lambda_n, \quad \textit{or}, \tag{5.4}$$

$$\beta_n \leq 0.1, \quad M_{n,2+\delta}^2 n^{2/(2+\delta)} \geq 2, \quad \textit{and} \quad \ln\left(\frac{p}{2\beta_n\sqrt{2\pi}}\right) \geq \frac{9}{8}(\frac{4}{3}+\varepsilon)^2 n^{\delta/(2+\delta)} M_{n,2+\delta}^{-2}, \tag{5.5}$$

where $\varepsilon > 0$ is as in Eq. (3.4).

Theorem 5.1 provides two sufficient conditions under which our two-step Lasso SN method will have greater or equal power than the two-step SN method in CCK14. The power difference is a direct consequence of Eq. (5.2), i.e., our Lasso-based first step inequality selection procedure chooses a subset of the inequalities in the SN-based first step. The first sufficient condition, Eq. (5.4), is sharper than the second one, Eq. (5.5), but the second one is of lower level and, thus, easier to interpret and understand. Eq. (5.5) is composed of three statements and only the third one could be considered restrictive. The first one, $\beta \leq 10\%$, is non-restrictive as CCK14 require that $\beta_n \leq \alpha/3$ and the significance level α is typically less than 30%. The second, $M_{n,2+\delta}^2 n^{2/(2+\delta)} \geq 2$, is also non-restrictive since $M_{n,2+\delta}^2$ is a non-decreasing sequence of positive constants and $n^{2/(2+\delta)} \to \infty$.

In principle, Theorem 5.1 allows for the possibility of the inequality in Eq. (5.3) being an equality. However, in cases in which the Lasso-based first step selects a strict subset of the moment inequalities chosen by the SN method (i.e. the inclusion in Eq. (5.2) is strict), the inequality in Eq. (5.3) can be strict. In fact, the inequality in Eq. (5.3) can be strict even in cases in which the Lasso-based and SN-based first step agree on the set of binding moment inequalities. The intuition for this is that our Lasso-based method considers the usual α-quantile while the other two-step methods consider the $(\alpha - 2\beta_n)$-quantile for the sequence of positive constants $\{\beta_n\}_{n\geq 1}$. This slight difference always plays in favor of the Lasso-based first step having more power.[12]

The relevance of Theorem 5.1 depends on the generality of the sufficient conditions in Eq. (5.4) and (5.5). Figure 1 provides heat maps that indicate combinations of values of $M_{n,2+\delta}$ and p under which Eqs. (5.4) and (5.5) are satisfied. The graphs clearly show these conditions are satisfied for a large portion of the parameter space. In fact, the region in which Eq. (5.4) fails to hold is barely visible. In addition, the graph also confirms that Eq. (5.4) applies more generally than Eq. (5.5).

Remark 5.1. Notice that the power comparison in Theorem 5.1 is a finite sample result. In other words, under any of the sufficient conditions Theorem in 5.1, the rejection of the null hypothesis by an inference method with SN-based first step implies the same outcome for the corresponding inference method with Lasso-based first step. Expressed in terms of confidence sets, the confidence set with our Lasso first step will be a subset of the corresponding confidence set with a SN first step.

To conclude the section, we now compare the power of the two-step bootstrap procedures.

Theorem 5.2. *Assume Assumption A.4 and let $B \in \{MB, EB\}$.*

Part 1: For all $\theta \in \Theta$ and $n \in \mathbb{N}$,

$$\hat{J}_L(\theta) \;\subseteq\; \hat{J}_B(\theta) \tag{5.6}$$

[12] This is clearly shown in Designs 5-6 of our Monte Carlos. In these cases, both first-step methods to agree on the correct set of binding moment inequalities (i.e. $\hat{J}_L(\theta) = \hat{J}_{SN}(\theta)$). Nevertheless, the slight difference in quantiles produces small but positive power advantage in favor of methods that use the Lasso in a first stage.

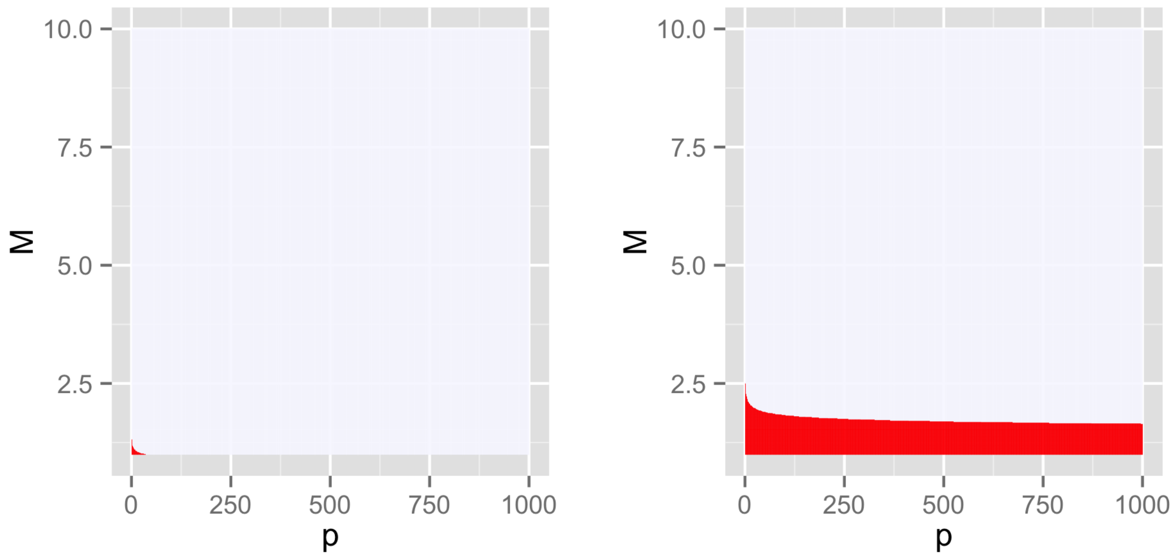

Figure 1: Consider a moment inequality model with $n = 400$, $\beta_n = 0.1\%$, $C = 2$, $M = M_{n,2+\delta} \in [0, 10]$, and $k = p \in \{1, \ldots, 1000\}$. The left (right) panel shows in red the configurations (p, M) that do not satisfy Eq. (5.4) (Eq. (5.5), respectively).

implies

$$P(T_n(\theta) > c_n^{B,2S}(\alpha)) \ \le \ P(T_n(\theta) > c_n^{B,L}(\theta, \alpha)). \tag{5.7}$$

Part 2: *Eq.* (5.6) *occurs with probability approaching one, i.e.,*

$$P\left(\hat{J}_L(\theta) \subseteq \hat{J}_B(\theta)\right) \ \ge \ 1 - Cn^{-c} \tag{5.8}$$

under the following sufficient conditions: $M^2_{n,2+\delta} n^{2/(2+\delta)} \ge 2$, $\beta_n \ge Cn^{-c}$ *for some* $C, c > 0$, *and any one of the following conditions:*

$$1 - \Phi\left(\frac{3}{2^{3/2}}(\frac{4}{3} + \varepsilon) n^{\delta/(2(2+\delta))} M^{-1}_{n,2+\delta}\right) \ge 3\beta_n, \qquad or, \tag{5.9}$$

$$\sqrt{(1 - \rho(\theta)) \log(p)/2} - \sqrt{2 \log(1/[1 - 3\beta_n])} \ge \frac{3}{2^{3/2}}(\frac{4}{3} + \varepsilon) n^{\delta/(2(2+\delta))} M^{-1}_{n,2+\delta}, \tag{5.10}$$

where $\rho(\theta) \equiv \max_{j_1 \ne j_2} corr[X_{j_1}(\theta), X_{j_2}(\theta)]$.

Part 3: *Under any of the sufficient conditions in part 2,*

$$P\left(T_n(\theta) > c_n^{B,2S}(\alpha)\right) \;\leq\; P\left(T_n(\theta) > c_n^{B,L}(\theta,\alpha)\right) + Cn^{-c} \tag{5.11}$$

Theorem 5.2 provides sufficient conditions under which any power advantage of the two-step bootstrap method in CCK14 relative to our two-step bootstrap Lasso vanishes as the sample size diverges to infinity. Specifically, Eq. (5.11) indicates that, under any of the sufficient conditions, this power advantage does not exceed $\tilde{C}n^{-\tilde{c}}$. As in the SN approximation, this relative power difference is a direct consequence of Eq. (5.6), i.e., our Lasso-based first step inequality selection procedure chooses a subset of the inequalities in the bootstrap-based first step.

The relevance of the result in Theorem 5.2 depends on the generality of the sufficient condition. This condition has three parts. The first part, i.e., $M_{n,2+\delta}^2 n^{2/(2+\delta)} \geq 2$, was already argued to be non-restrictive since $M_{n,2+\delta}^2$ is a non-decreasing sequence of positive constants and $n^{2/(2+\delta)} \to \infty$. The second part, i.e., $\beta_n \geq Cn^{-c}$ is also considered mild as $\{\beta_n\}_{n\geq 1}$ is a sequence of positive constants and Cn^{-c} converges to zero. The third part is Eq. (5.9) or (5.10) and we deem it to be the more restrictive condition of the three. In the case of the latter, this condition can be understood as imposing an upper bound on the maximal pairwise correlation within the moment inequalities of the model.

6 Monte Carlo simulations

We now use Monte Carlo simulations to investigate the finite sample properties of our tests and to compare them to those proposed by CCK14. Our simulation setup follows closely the moment inequality model considered in their Monte Carlo simulation section. For a hypothetical fixed parameter value $\theta \in \Theta$, we generate data according to the following equation:

$$X_i(\theta) \;=\; \mu(\theta) + A'\epsilon_i \qquad i = 1, \ldots, n = 400,$$

where $\Sigma(\theta) = A'A$, $\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,p})$, and $p \in \{200, 500, 1000\}$. We simulate $\{\epsilon_i\}_{i=1}^n$ to be i.i.d. with $E[\epsilon_i] = \mathbf{0}_p$ and $Var[\epsilon_i] = \mathbf{I}_{p\times p}$, and so $\{X_i(\theta)\}_{i=1}^n$ are i.i.d. with $E[X_i(\theta)] = \mu(\theta)$ and $Var[X_i(\theta)] = \Sigma(\theta)$. This model satisfies the moment (in)equality model in Eq. (2.1) if and only if $\mu(\theta) \leq \mathbf{0}_p$. In this context, we are interested in implementing the hypothesis test in Eqs. (2.3) (or, equivalently, Eq. (2.4)) with a significance level of $\alpha = 5\%$.

We simulate $\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,p})$ to be i.i.d. according to two distributions: (i) $\epsilon_{i,j}$ follows a t-distribution with four degrees of freedom divided by $\sqrt{2}$, i.e., $\epsilon_{i,j} \sim t_4/\sqrt{2}$ and (ii) $\epsilon_{i,j} \sim U(-\sqrt{3}, \sqrt{3})$. Note that both of these choices satisfy $E[\epsilon_i] = \mathbf{0}_p$ and $Var[\epsilon_i] = \mathbf{I}_{p\times p}$. Since $(\epsilon_{i,1}, \ldots, \epsilon_{i,p})$ are i.i.d., the correlation structure across moment inequalities depends entirely on $\Sigma(\theta)$, for which we consider two possibilities: (i) $\Sigma(\theta)_{[j,k]} = 1[j = k] + \rho \cdot 1[j \neq k]$ and (ii) a Toeplitz structure, i.e., $\Sigma(\theta)_{[j,k]} = \rho^{|j-k|}$ with $\rho \in \{0, 0.5, 0.9\}$. We repeat all experiments $2,000$ times.

The description of the model is completed by specifying $\mu(\theta)$, given in Table 1. We consider ten different specifications of $\mu(\theta)$ which, in combination with the rest of the parameters, results in fourteen simulation designs. Our first eight simulation designs correspond exactly to those in CCK14, half of which satisfy the null hypothesis and half of which do not. We complement these simulations with six designs that do not satisfy the null hypothesis. The additional designs are constructed so that the moment inequalities that agree with the null hypothesis are only slightly or moderately negative.[13] As the slackness of these inequalities decreases, it becomes harder for two-step inference methods to correctly classify the non-binding moment conditions as such. As a consequence, these new designs will help us understand which two-step inference procedures have better ability in detecting slack moment inequalities.

Design no.	$\{\mu_j(\theta) : j \in \{1, \ldots, p\}\}$	$\Sigma(\theta)$	Hypothesis	CCK14 Design no.
1	$-0.8 \cdot 1[j > 0.1p]$	Equicorrelated	H_0	2
2	$-0.8 \cdot 1[j > 0.1p]$	Toeplitz	H_0	4
3	0	Equicorrelated	H_0	1
4	0	Toeplitz	H_0	3
5	0.05	Equicorrelated	H_1	5
6	0.05	Toeplitz	H_1	7
7	$-0.75 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$	Equicorrelated	H_1	6
8	$-0.75 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$	Toeplitz	H_1	8
9	$-0.6 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$	Toeplitz	H_1	New
10	$-0.5 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$	Toeplitz	H_1	New
11	$-0.4 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$	Toeplitz	H_1	New
12	$-0.3 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$	Toeplitz	H_1	New
13	$-0.2 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$	Toeplitz	H_1	New
14	$-0.1 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$	Toeplitz	H_1	New

Table 1: Parameter choices in our simulations.

We implement all the inference methods described in Table 2. These include all of the procedures described in previous sections some additional "hybrid" methods (i.e. MB-H and EB-H). The bootstrap based methods are implemented with $B = 1,000$ bootstrap replications. Finally, for our Lasso-based first

[13] For reasons of brevity, these additional designs only consider $\Sigma(\theta)$ with a Toeplitz structure. We carried out the same designs with equicorrelated $\Sigma(\theta)$ and obtained qualitatively similar results. These are available from the authors, upon request.

Method	No. of steps	First step	Second step	Parameters
SN Lasso	Two	Lasso	SN	$C \in \{2, 4, 6\}$ in Eq. (6.1)
MB Lasso	Two	Lasso	MB	$C \in \{2, 4, 6\}$ in Eq. (6.1)
EB Lasso	Two	Lasso	EB	$C \in \{2, 4, 6\}$ in Eq. (6.1)
SN-1S	One	None	SN	None
SN-2S	Two	SN	SN	None
MB-1S	One	None	MB	None
MB-H	Two	SN	MB	$\beta_n \in \{0.01\%, 0.1\%, 1\%\}$
MB-2S	Two	MB	MB	$\beta_n \in \{0.01\%, 0.1\%, 1\%\}$
EB-1S	One	None	EB	None
EB-H	Two	SN	EB	$\beta_n \in \{0.01\%, 0.1\%, 1\%\}$
EB-2S	Two	EB	EB	$\beta_n \in \{0.01\%, 0.1\%, 1\%\}$

Table 2: Inference methods.

step, we use:

$$\lambda_n \ = \ C \cdot n^{-1/2} \left(\hat{M}_{n,3}^2 n^{-1/3} - n^{-1} \right)^{-1}, \tag{6.1}$$

with $C \in \{2, 4, 6\}$ and $\hat{M}_{n,3} \equiv \max_{j=1,\ldots,p} (n^{-1} \sum_{i=1}^n |X_{ij}(\theta)|^3)^{1/3}$. This corresponds to the empirical analogue of Eq. (3.4) when $\delta = 1$ and $\varepsilon \in \{2/3, 8/3, 14/3\}$.

We shall begin by considering the simulation designs in CCK14 as reported in Tables 3-10. The first four tables are concerned with the finite sample size control. The general finding is that all tests under consideration are very rarely over-sized. The maximal size observed for our procedures is 7.15 (e.g. EB Lasso in Designs 3-4, $p = 1{,}000$, $\rho = 0$, and uniform errors) while the corresponding number for CCK14 is 7.25 (e.g. EB-1S in Designs 3-4, $p = 1{,}000$, $\rho = 0$, and uniform errors). Some procedures, such as SN-1S, can be heavily under-sized. Our simulations reveal that in order to achieve empirical rejection rates close to $\alpha = 5\%$ under the null hypothesis, one requires using a two-step inference procedure with a bootstrap-based second step (either MB or EB).

Before turning to the individual setups for power comparison, let us remark that a first step based on our Lasso procedure compares favorably with a first step based on SN. For example, SN-Lasso with $C = 2$ has more or equal power than SN-2S with $\beta_n = 0.1\%$. While the differences may often be small, this finding is in line with the power comparison in Section 5.

Tables 7-10 contain the designs used by CCK14 to gauge the power of their tests. Tables 7 and 8 consider the case where all moment inequalities are violated. Since none of the moment conditions are slack, there is no room for power gains based on a first-step inequality selection procedure. In this sense, it is not surprising that the first step choice makes no difference in these designs. For example, the power of SN-Lasso is identical to the one of SN-1S while the power of SN-2S is also close to the one of SN-1S. However, the SN-2S has lower power than SN-1S for some values of β_n while the power of SN Lasso appears to be invariant to the choice of C. The latter is in accordance with our previous findings. The bootstrap still improves power for high values of ρ.

Next, we consider Tables 9 and 10. In this setting, 90% of the moment conditions have $\mu_j(\theta) = -0.75$ and our results seem to suggest that this value is relative far away from being binding. We deduce this from the fact that all first-step selection methods agree on the set of binding moment conditions, producing very similar power results. Table 17 shows the percentage of moment inequalities retained by each of the first-step procedures in Design 8. When the error terms are t-distributed, all first-step procedures retain around 10% of the inequalities which is also the fraction that are truly binding (and, in this case, violated). Thus, all two-step inference procedures are reasonably powerful. When the error terms are uniformly distributed, all first-step procedures have an equal tendency to aggressively remove slack inequalities. However, we have seen from the size comparisons that this does not seem to result in oversized tests. Finally, we notice that the power of our procedures hardly varies with the choice of C.

The overall message of the simulation results in Designs 1-8 is that our Lasso-based procedures are comparable in terms of size and power to the ones proposed by CCK14.

Tables 11-16 present simulations results for Designs 9-14. These correspond to modifications of the setup in Design 8 in which progressively decrease the degree of slackness of the non-binding moment inequalities from -0.75 to values between -0.6 and -0.1.

Tables 11-12 shows results for Designs 9 and 10. As in the case of Design 8, the degree of slackness of the non-binding moment inequalities is still large enough so that it can be correctly detected by all first first-step selection methods.

As Table 13 shows, this pattern changes in Design 11. In this case, the MB Lasso with $C = 2$ has a rejection rate that is at least 20 percentage points higher than the most powerful procedure in CCK14. For example, with t-distributed errors, $p = 1{,}000$, and $\rho = 0$, our MB Lasso with $C = 2$ has a rejection rate of 71.40% whereas the MB-2S with $\beta_n = 0.01\%$ has a rejection rate of 20.55%. Table 18 holds the key to these power differences. Ideally, a powerful procedure should retain only the 10% of the moment inequalities that are binding (in this case, violated). The Lasso-based selection indeed often retains close to 10% of the inequalities for $C \in \{2, 4\}$. On the other hand, SN-based selection can sometimes retain more than 90% of the inequalities (e.g. see t-distributed errors, $p = 1{,}000$, and $\rho = 0$).

The power advantage in favor of the Lasso-based first step is also present in Design 12 as shown in Table 14. In this case, the MB Lasso with $C = 2$ has a rejection rate which is at least 15 percentage points higher than the most powerful procedure in CCK14. For t-distributed errors, the MB Lasso always has a rejection rate that is at least 20 percentage points higher than its competitors and sometimes more than 50 percentage points (e.g. $p = 1,000$ and $\rho = 0$). As in the previous design, this power gain mainly comes from the Lasso being better at removing the slack moment conditions.

Table 15 shows the results for Design 13. For t-distributed errors, the MB Lasso with $C = 2$ has a higher rejection rate than the most powerful procedure of CCK14 (which is often MB-1S) by at least 5 percentage points. Sometimes the difference is larger than 45 percentage points (e.g. see $p = 1,000$ and $\rho = 0$). For uniformly distributed errors, there seems to be no significant difference between our procedures and the ones in CCK14; all of them have relatively low power.

Design 14 is our last experiment and it is shown in Table 16. In this case, the degree of slackness of the non-binding moment inequalities is so small that it cannot be detected by any of the first-step selection methods. As a consequence, there are very little differences among the various inference procedures and all of them exhibit relatively low power.

The overall message from Tables 11-16 is that our Lasso-based inference procedures can have higher power than those in CCK14 when the slack moment inequalities are difficult to distinguish from zero.

7 Conclusions

This paper considers the problem of inference in a partially identified moment (in)equality model with possibly many moment inequalities. Our contribution is to propose a novel two-step inference method based on the combination of two ideas. On the one hand, our test statistic and critical values are based on those proposed by CCK14. On the other hand, we propose a new first step selection procedure based on the Lasso. Our two-step inference method can be used to conduct hypothesis tests and to construct confidence sets for the true parameter value.

Our inference method has very desirable properties. First, under reasonable conditions, it is uniformly valid, both in underlying parameter θ and distribution of the data. Second, by virtue of results in CCK14, our test is asymptotically optimal in a minimax sense. Third, the power of our method compares favorably with that of the corresponding two-step method in CCK14, both in theory and in simulations. On the theory front, we provide sufficient conditions under which the power of our method dominates. These can sometimes represent a significant part of the parameter space. Our simulations indicate that our inference methods are usually as powerful as the corresponding ones in CCK14, and can sometimes be more powerful. Fourth, our Lasso-based first step is straightforward to implement.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	4.10	4.00	4.00	5.20	5.00	5.00	4.55	4.50	4.50	0.40	3.95	3.85	2.30	0.55	4.80	4.50	2.50	4.85	4.50	2.50	0.50	4.35	4.40	2.65	4.75	4.55	2.70
		0.5	3.05	3.05	3.05	5.10	5.10	5.10	4.85	4.85	4.85	0.25	3.05	2.95	1.85	0.85	5.05	4.80	2.80	5.05	4.80	2.85	0.85	4.85	4.60	3.10	4.75	4.65	2.85
		0.9	0.90	0.90	0.90	5.00	5.00	5.00	5.15	5.15	5.15	0.05	0.90	0.85	0.45	2.80	4.95	4.80	3.10	4.95	4.80	3.10	2.80	5.15	4.85	2.95	5.10	4.95	3.10
	500	0	4.45	4.35	4.35	5.65	5.45	5.45	5.00	4.85	4.85	0.50	4.30	4.30	2.85	0.65	5.25	5.20	3.40	5.30	5.20	3.40	0.45	4.55	4.55	3.10	4.85	4.75	3.15
		0.5	2.75	2.75	2.75	5.70	5.70	5.70	5.60	5.60	5.60	0.35	2.70	2.70	1.60	1.30	5.65	5.55	3.50	5.65	5.55	3.50	1.20	5.45	5.30	3.25	5.60	5.35	3.35
		0.9	0.55	0.55	0.55	5.80	5.80	5.80	5.60	5.60	5.60	0.15	0.55	0.55	0.40	2.75	5.80	5.60	3.10	5.80	5.60	3.10	2.90	5.60	5.40	3.15	5.65	5.40	3.00
	1000	0	5.05	4.85	4.80	5.85	5.80	5.80	5.55	5.35	5.35	0.45	4.65	4.65	3.00	0.55	5.55	5.45	3.60	5.60	5.40	3.60	0.45	4.85	4.90	3.40	4.95	4.80	3.15
		0.5	2.10	2.10	2.10	4.80	4.80	4.80	4.70	4.70	4.70	0.25	2.10	2.10	1.40	1.15	4.80	4.65	2.95	4.80	4.65	2.95	1.20	4.50	4.50	2.90	4.60	4.45	2.60
		0.9	0.55	0.55	0.55	4.60	4.60	4.60	4.65	4.65	4.65	0.05	0.55	0.55	0.35	2.75	4.55	4.40	2.70	4.60	4.40	2.70	2.65	4.60	4.40	2.65	4.70	4.60	2.85
$U(-\sqrt{3},\sqrt{3})$	200	0	5.00	5.00	5.00	5.60	5.60	5.60	5.80	5.80	5.80	0.45	4.95	4.85	3.20	0.55	5.55	5.45	3.65	5.55	5.45	3.65	0.55	5.90	5.65	3.70	5.70	5.45	3.65
		0.5	3.10	3.10	3.10	5.20	5.20	5.20	5.05	5.05	5.05	0.40	3.10	3.05	2.00	1.15	5.20	4.90	3.15	5.20	4.90	3.15	1.15	5.15	5.00	3.10	5.15	5.00	3.15
		0.9	1.10	1.10	1.05	4.85	4.85	4.85	5.05	5.05	5.05	0.25	1.10	1.10	0.65	2.80	4.85	4.70	3.00	4.85	4.70	3.00	2.75	5.05	4.95	3.00	4.85	4.75	3.05
	500	0	5.05	5.05	5.05	5.70	5.70	5.70	6.05	6.05	6.05	0.80	5.05	5.05	3.10	1.10	5.65	5.45	3.70	5.65	5.45	3.70	1.05	6.10	5.90	3.85	5.70	5.50	3.85
		0.5	2.60	2.60	2.60	5.35	5.35	5.35	5.40	5.40	5.40	0.40	2.55	2.45	1.60	1.10	5.30	5.15	3.25	5.30	5.15	3.25	1.05	5.45	5.15	3.15	5.20	5.10	3.00
		0.9	0.75	0.75	0.75	5.05	5.05	5.05	4.70	4.70	4.70	0.05	0.75	0.75	0.45	2.95	5.05	4.75	3.05	5.05	4.75	3.05	2.90	4.70	4.50	3.00	4.85	4.75	3.00
	1000	0	4.50	4.50	4.50	5.85	5.85	5.85	5.55	5.55	5.55	0.55	4.50	4.40	3.00	0.70	5.80	5.45	3.65	5.80	5.45	3.65	0.80	5.60	5.40	3.65	5.30	5.20	3.75
		0.5	2.15	2.15	2.15	4.95	4.95	4.95	5.10	5.10	5.10	0.40	2.15	2.00	1.50	1.40	4.95	4.85	2.75	4.95	4.85	2.75	1.35	5.15	5.00	2.95	5.25	4.95	2.75
		0.9	0.45	0.45	0.45	5.20	5.20	5.20	5.20	5.20	5.20	0.00	0.45	0.40	0.30	2.75	5.20	4.80	2.70	5.20	4.80	2.70	2.80	5.20	5.05	2.75	5.15	4.95	3.05

Table 3: Simulation results in Design 1: $\mu_j(\theta) = -0.8 \cdot 1[j > 0.1p]$, $\Sigma(\theta)$ equicorrelated

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	4.10	4.00	4.00	5.20	5.00	5.00	4.55	4.50	4.50	0.40	3.95	3.85	2.30	0.55	4.80	4.50	2.50	4.85	4.50	2.50	0.50	4.35	4.40	2.65	4.75	4.55	2.70
		0.5	4.40	4.40	4.40	5.30	5.20	5.20	5.05	5.00	5.00	0.35	4.40	4.20	2.70	0.50	5.05	5.10	3.30	5.10	5.05	3.30	0.45	4.90	4.90	3.05	5.15	4.95	3.20
		0.9	2.30	2.30	2.30	5.20	5.20	5.20	5.00	4.95	4.95	0.35	2.30	2.30	1.55	0.70	5.10	4.85	3.35	5.15	4.85	3.35	0.65	4.85	4.80	3.35	5.30	5.20	3.10
	500	0	4.45	4.35	4.35	5.65	5.45	5.45	5.00	4.85	4.85	0.50	4.30	4.30	2.85	0.65	5.25	5.20	3.40	5.30	5.20	3.40	0.45	4.55	4.55	3.10	4.85	4.75	3.15
		0.5	4.35	4.25	4.25	5.60	5.45	5.45	5.15	5.10	5.10	0.40	4.20	4.15	2.70	0.45	5.35	5.25	3.25	5.40	5.25	3.25	0.45	4.85	4.70	3.00	4.75	4.55	3.05
		0.9	1.75	1.70	1.70	5.20	5.20	5.20	5.30	5.30	5.30	0.10	1.65	1.60	1.15	0.35	5.10	4.95	2.70	5.15	4.95	2.70	0.35	5.00	5.15	2.60	5.25	4.90	2.55
	1000	0	5.05	4.85	4.80	5.85	5.80	5.80	5.55	5.35	5.35	0.45	4.65	4.65	3.00	0.55	5.55	5.45	3.60	5.60	5.40	3.60	0.45	4.85	4.90	3.40	4.95	4.80	3.15
		0.5	4.55	4.15	4.10	5.80	5.60	5.60	5.50	5.40	5.35	0.45	4.00	3.90	2.35	0.65	5.45	5.25	3.35	5.55	5.30	3.35	0.60	4.95	5.00	3.05	4.75	4.60	2.85
		0.9	1.90	1.85	1.85	5.25	5.20	5.20	4.95	4.95	4.95	0.20	1.85	1.70	1.15	0.35	5.15	4.90	2.90	5.15	4.90	2.90	0.25	4.85	4.75	2.85	4.95	4.85	2.90
$U(-\sqrt{3},\sqrt{3})$	200	0	5.00	5.00	5.00	5.60	5.60	5.60	5.80	5.80	5.80	0.45	4.95	4.85	3.20	0.55	5.55	5.45	3.65	5.55	5.45	3.65	0.55	5.90	5.65	3.70	5.70	5.45	3.65
		0.5	4.65	4.65	4.65	5.55	5.55	5.55	5.45	5.45	5.45	0.55	4.65	4.50	3.00	0.65	5.50	5.30	3.80	5.50	5.30	3.80	0.80	5.45	5.20	3.85	5.95	5.85	3.40
		0.9	2.25	2.25	2.25	5.25	5.25	5.25	5.15	5.15	5.15	0.55	2.25	2.15	1.50	0.85	5.15	4.90	3.40	5.15	4.90	3.40	0.80	5.20	5.15	3.50	5.15	5.05	3.45
	500	0	5.05	5.05	5.05	5.70	5.70	5.70	6.05	6.05	6.05	0.80	5.05	5.05	3.10	1.10	5.65	5.45	3.70	5.65	5.45	3.70	1.05	6.10	5.90	3.85	5.70	5.50	3.85
		0.5	4.55	4.55	4.55	5.70	5.70	5.70	5.30	5.30	5.30	0.55	4.50	4.35	2.80	0.65	5.65	5.50	3.50	5.65	5.50	3.50	0.60	5.30	5.25	3.35	5.95	5.75	3.15
		0.9	1.65	1.65	1.65	5.20	5.20	5.20	5.00	5.00	5.00	0.15	1.65	1.65	1.00	0.35	5.15	4.90	2.60	5.15	4.90	2.60	0.35	5.00	5.00	3.00	4.80	4.65	2.90
	1000	0	4.50	4.50	4.50	5.85	5.85	5.85	5.55	5.55	5.55	0.55	4.50	4.40	3.00	0.70	5.80	5.45	3.65	5.80	5.45	3.65	0.80	5.60	5.40	3.65	5.30	5.20	3.75
		0.5	4.70	4.70	4.70	6.15	6.15	6.15	5.90	5.90	5.90	0.80	4.70	4.50	2.60	0.90	6.10	5.90	3.35	6.10	5.90	3.35	0.95	5.90	5.70	3.60	6.05	5.65	3.50
		0.9	2.40	2.40	2.40	5.55	5.55	5.55	5.30	5.30	5.30	0.20	2.40	2.40	1.60	0.50	5.50	5.35	3.50	5.50	5.35	3.50	0.40	5.30	5.25	3.50	5.35	5.10	3.55

Table 4: Simulation results in Design 2: $\mu_j(\theta) = -0.8 \cdot 1[j > 0.1p]$, $\Sigma(\theta)$ Toeplitz

Density	p	ρ	Our methods									CCK14's methods																		
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	3.20	3.05	3.05	4.35	4.30	4.30	3.65	3.55	3.55	3.05	3.05	2.95	1.55	4.30	4.25	4.10	2.15	4.25	4.10	2.15	3.60	3.60	3.50	2.00	3.45	3.35	1.95
		0.5	1.45	1.45	1.45	4.60	4.60	4.60	4.15	4.15	4.15	1.45	1.45	1.40	1.00	4.60	4.60	4.40	2.60	4.60	4.40	2.60	4.20	4.20	4.05	2.55	4.35	4.15	2.45
		0.9	0.25	0.25	0.25	5.20	5.20	5.20	5.30	5.30	5.30	0.25	0.25	0.20	0.15	5.20	5.15	4.90	3.10	5.15	4.90	3.10	5.30	5.30	4.80	3.00	5.15	4.95	2.95
	500	0	3.00	2.95	2.95	4.45	4.25	4.25	3.50	3.50	3.50	2.95	2.95	2.90	1.55	4.25	4.20	3.95	2.20	4.20	3.95	2.20	3.55	3.55	3.50	1.90	3.70	3.50	2.20
		0.5	1.45	1.45	1.45	5.65	5.65	5.65	5.05	5.05	5.05	1.45	1.45	1.35	0.85	5.65	5.65	5.65	3.45	5.65	5.65	3.45	5.10	5.10	5.00	3.15	5.10	5.00	3.15
		0.9	0.20	0.20	0.20	5.45	5.45	5.45	5.55	5.55	5.55	0.20	0.20	0.20	0.15	5.45	5.45	5.25	3.50	5.45	5.25	3.50	5.55	5.55	5.35	2.95	5.35	5.25	3.00
	1000	0	3.40	3.20	3.20	4.65	4.40	4.40	3.75	3.70	3.70	3.20	3.15	3.05	1.80	4.40	4.40	4.25	2.75	4.40	4.25	2.75	3.70	3.70	3.60	1.95	3.90	3.85	2.25
		0.5	1.40	1.40	1.40	4.25	4.25	4.25	3.85	3.85	3.85	1.40	1.40	1.35	0.90	4.25	4.25	4.10	2.90	4.25	4.10	2.90	3.85	3.85	3.75	2.55	4.20	4.05	2.75
		0.9	0.15	0.15	0.15	4.40	4.40	4.40	4.15	4.15	4.15	0.15	0.15	0.15	0.05	4.40	4.40	4.15	2.50	4.40	4.15	2.50	4.15	4.15	4.00	2.50	4.35	4.10	2.60
$U(-\sqrt{3},\sqrt{3})$	200	0	3.85	3.85	3.85	5.20	5.20	5.20	4.90	4.90	4.90	3.85	3.85	3.75	2.00	5.20	5.15	4.90	2.85	5.15	4.90	2.85	4.95	4.90	4.70	2.95	5.30	5.10	2.80
		0.5	1.95	1.95	1.95	4.65	4.65	4.65	4.55	4.55	4.55	1.95	1.95	1.85	1.45	4.65	4.65	4.65	2.80	4.65	4.65	2.80	4.55	4.55	4.45	2.75	4.70	4.45	2.75
		0.9	0.40	0.40	0.40	4.95	4.95	4.95	5.20	5.20	5.20	0.40	0.40	0.40	0.30	4.95	4.95	4.85	2.85	4.95	4.85	2.85	5.20	5.20	5.05	2.90	4.85	4.65	2.90
	500	0	4.20	4.20	4.20	5.60	5.60	5.60	5.65	5.65	5.65	4.20	4.20	4.00	2.50	5.60	5.45	5.20	3.35	5.45	5.20	3.35	5.75	5.70	5.55	3.60	5.75	5.55	3.35
		0.5	1.50	1.50	1.50	5.05	5.05	5.05	4.70	4.70	4.70	1.50	1.50	1.50	1.10	5.05	5.05	4.95	2.80	5.05	4.95	2.80	4.75	4.70	4.65	2.45	5.00	4.75	2.65
		0.9	0.10	0.10	0.10	5.10	5.10	5.10	4.50	4.50	4.50	0.10	0.05	0.05	0.05	5.10	5.10	4.85	2.85	5.10	4.85	2.85	4.60	4.55	4.35	2.80	4.60	4.50	2.65
	1000	0	5.05	5.05	5.05	6.95	6.95	6.95	7.15	7.15	7.15	5.05	5.00	4.95	3.05	6.95	6.95	6.85	4.30	6.95	6.85	4.30	7.25	7.20	7.15	4.20	7.20	6.95	4.50
		0.5	1.40	1.40	1.40	5.90	5.90	5.90	5.55	5.55	5.55	1.40	1.40	1.40	1.05	5.90	5.85	5.65	3.70	5.85	5.65	3.70	5.55	5.55	5.45	3.75	5.85	5.65	3.90
		0.9	0.05	0.05	0.05	5.10	5.10	5.10	5.05	5.05	5.05	0.05	0.05	0.05	0.05	5.10	5.10	4.95	3.30	5.10	4.95	3.30	5.05	5.05	4.85	3.40	5.60	5.30	3.45

Table 5: Simulation results in Design 3: $\mu_j(\theta) = 0$ for all $j = 1, \ldots, p$, $\Sigma(\theta)$ equicorrelated.

Density	p	ρ	Our methods									CCK14's methods																		
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	3.20	3.05	3.05	4.35	4.30	4.30	3.65	3.55	3.55	3.05	3.05	2.95	1.55	4.30	4.25	4.10	2.15	4.25	4.10	2.15	3.60	3.60	3.50	2.00	3.45	3.35	1.95
		0.5	3.85	3.80	3.80	4.90	4.85	4.85	4.50	4.45	4.45	3.80	3.80	3.80	2.45	4.85	4.85	4.55	3.05	4.85	4.55	3.05	4.55	4.55	4.20	2.85	4.55	4.45	2.95
		0.9	2.40	2.40	2.40	5.90	5.90	5.90	5.55	5.55	5.55	2.40	2.40	2.35	1.75	5.90	5.90	5.80	3.60	5.90	5.80	3.60	5.55	5.55	5.45	3.50	5.65	5.40	3.20
	500	0	3.00	2.95	2.95	4.45	4.25	4.25	3.50	3.50	3.50	2.95	2.95	2.90	1.55	4.25	4.20	3.95	2.20	4.20	3.95	2.20	3.55	3.55	3.50	1.90	3.70	3.50	2.20
		0.5	3.90	3.75	3.75	5.70	5.45	5.45	5.00	4.85	4.85	3.75	3.75	3.45	1.95	5.45	5.45	5.30	3.05	5.45	5.30	3.05	4.95	4.95	4.55	2.45	4.50	4.40	2.55
		0.9	2.35	2.35	2.35	5.20	5.20	5.20	4.60	4.55	4.55	2.35	2.35	2.20	1.45	5.20	5.15	5.00	2.85	5.15	5.00	2.85	4.55	4.55	4.45	2.90	4.70	4.65	2.65
	1000	0	3.40	3.20	3.20	4.65	4.40	4.40	3.75	3.70	3.70	3.20	3.15	3.05	1.80	4.40	4.40	4.25	2.75	4.40	4.25	2.75	3.70	3.70	3.60	1.95	3.90	3.85	2.25
		0.5	3.70	3.60	3.60	5.30	5.20	5.20	4.70	4.65	4.65	3.60	3.60	3.40	1.95	5.15	5.15	4.95	3.15	5.15	4.95	3.15	4.70	4.70	4.55	2.80	4.35	4.25	2.80
		0.9	2.30	2.30	2.30	5.20	5.15	5.15	5.00	4.85	4.85	2.30	2.30	2.30	1.55	5.15	5.10	5.05	3.20	5.10	5.05	3.20	4.95	4.95	4.65	2.90	4.90	4.80	2.80
$U(-\sqrt{3},\sqrt{3})$	200	0	3.85	3.85	3.85	5.20	5.20	5.20	4.90	4.90	4.90	3.85	3.85	3.75	2.00	5.20	5.15	4.90	2.85	5.15	4.90	2.85	4.95	4.90	4.70	2.95	5.30	5.10	2.80
		0.5	4.35	4.35	4.35	5.20	5.20	5.20	5.25	5.25	5.25	4.35	4.35	4.05	2.95	5.20	5.15	5.05	3.35	5.15	5.05	3.35	5.40	5.35	5.10	3.50	5.30	5.10	3.45
		0.9	2.50	2.50	2.50	5.40	5.40	5.40	5.30	5.30	5.30	2.50	2.50	2.40	1.95	5.40	5.40	5.15	3.85	5.40	5.15	3.85	5.40	5.35	5.25	3.70	5.35	5.20	3.65
	500	0	4.20	4.20	4.20	5.60	5.60	5.60	5.65	5.65	5.65	4.20	4.20	4.00	2.50	5.60	5.45	5.20	3.35	5.45	5.20	3.35	5.75	5.70	5.55	3.60	5.75	5.55	3.35
		0.5	3.85	3.85	3.85	5.25	5.25	5.25	5.15	5.15	5.15	3.85	3.85	3.75	2.55	5.25	5.15	5.05	3.60	5.15	5.05	3.60	5.20	5.20	5.05	3.35	5.20	5.10	3.25
		0.9	2.75	2.75	2.75	5.35	5.35	5.35	5.55	5.55	5.55	2.75	2.75	2.70	1.50	5.35	5.35	5.25	3.30	5.35	5.25	3.30	5.70	5.70	5.50	3.50	5.30	5.25	3.45
	1000	0	5.05	5.05	5.05	6.95	6.95	6.95	7.15	7.15	7.15	5.05	5.00	4.95	3.05	6.95	6.95	6.85	4.30	6.95	6.85	4.30	7.25	7.20	7.15	4.20	7.20	6.95	4.50
		0.5	5.00	5.00	5.00	6.75	6.75	6.75	6.90	6.90	6.90	5.00	5.00	4.90	3.05	6.75	6.75	6.70	4.55	6.75	6.70	4.55	6.95	6.90	6.75	4.50	6.90	6.85	4.55
		0.9	2.95	2.95	2.95	6.40	6.40	6.40	6.30	6.30	6.30	2.95	2.95	2.95	1.85	6.40	6.40	6.20	3.95	6.40	6.20	3.95	6.35	6.35	6.25	4.15	6.40	6.20	4.25

Table 6: Simulation results in Design 4: $\mu_j(\theta) = 0$ for all $j = 1, \ldots, p$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	70.30	70.25	70.25	75.70	75.70	75.70	73.15	73.15	73.15	70.25	70.10	68.95	53.85	75.70	75.60	74.55	61.55	75.60	74.55	61.55	73.55	73.50	72.50	58.05	73.10	71.90	58.60
		0.5	22.15	22.15	22.15	37.75	37.75	37.75	36.75	36.75	36.75	22.15	22.10	21.60	16.55	37.75	37.70	37.25	28.50	37.70	37.25	28.50	37.00	36.80	36.10	28.15	37.00	36.25	27.95
		0.9	4.40	4.40	4.40	27.80	27.80	27.80	27.20	27.20	27.20	4.40	4.40	4.25	2.80	27.80	27.75	27.00	20.20	27.75	27.00	20.20	27.30	27.25	26.55	20.05	27.00	26.35	20.20
	500	0	77.50	77.45	77.45	84.10	84.05	84.05	80.60	80.55	80.55	77.45	77.40	76.10	62.00	84.05	83.90	83.30	70.90	83.90	83.30	70.90	80.75	80.65	79.85	66.05	80.60	79.90	65.90
		0.5	20.95	20.95	20.95	38.35	38.35	38.35	37.05	37.05	37.05	20.95	20.90	20.75	15.70	38.35	38.25	37.85	30.05	38.25	37.85	30.05	37.30	37.25	36.50	28.45	36.95	36.70	28.90
		0.9	3.35	3.35	3.35	28.30	28.30	28.30	28.10	28.10	28.10	3.35	3.35	3.25	2.10	28.30	28.30	27.45	21.00	28.30	27.45	21.00	28.15	28.15	27.30	20.55	28.15	27.60	20.20
	1000	0	81.70	81.65	81.65	89.15	89.10	89.10	84.65	84.50	84.50	81.65	81.55	80.45	66.30	89.10	89.00	88.05	77.75	89.00	88.05	77.75	84.95	84.95	84.05	70.60	84.95	84.10	70.75
		0.5	17.50	17.50	17.50	36.65	36.65	36.65	34.85	34.85	34.85	17.50	17.45	17.20	13.90	36.65	36.55	35.70	28.30	36.55	35.70	28.30	35.05	34.95	34.40	26.55	35.30	34.45	26.75
		0.9	2.05	2.05	2.05	26.20	26.20	26.20	26.00	26.00	26.00	2.05	2.05	2.05	1.65	26.20	26.15	25.70	19.05	26.15	25.70	19.05	26.20	26.15	25.15	18.45	25.70	24.90	18.15
$U(-\sqrt{3},\sqrt{3})$	200	0	70.00	70.00	70.00	76.50	76.50	76.50	76.30	76.30	76.30	70.00	69.80	68.90	55.90	76.50	76.40	75.80	62.10	76.40	75.80	62.10	76.65	76.60	75.60	63.55	76.35	75.50	62.90
		0.5	22.50	22.50	22.50	38.20	38.20	38.20	37.95	37.95	37.95	22.50	22.50	21.75	17.30	38.20	38.05	37.40	29.00	38.05	37.40	29.00	38.20	38.10	37.25	29.25	38.15	37.35	29.25
		0.9	4.20	4.20	4.20	27.30	27.30	27.30	27.80	27.80	27.80	4.20	4.20	4.05	2.90	27.30	27.15	26.65	20.25	27.15	26.65	20.25	27.80	27.75	27.30	20.15	27.65	27.20	19.95
	500	0	77.30	77.30	77.30	83.70	83.70	83.70	84.45	84.45	84.45	77.30	77.25	76.25	61.90	83.70	83.65	82.80	71.25	83.65	82.80	71.25	84.75	84.75	84.00	72.30	84.25	83.40	72.00
		0.5	20.50	20.50	20.50	40.20	40.20	40.20	39.70	39.70	39.70	20.50	20.45	19.90	16.00	40.20	40.10	39.45	29.95	40.10	39.45	29.95	40.00	39.75	39.20	30.05	39.55	38.65	30.15
		0.9	2.70	2.70	2.70	28.70	28.70	28.70	27.95	27.95	27.95	2.70	2.70	2.70	2.15	28.70	28.70	28.05	20.65	28.70	28.05	20.65	28.10	28.05	27.40	20.55	28.00	27.50	20.70
	1000	0	83.40	83.40	83.40	90.00	90.00	90.00	89.95	89.95	89.95	83.40	83.35	82.05	68.50	90.00	89.85	89.20	79.20	89.85	89.20	79.20	90.35	90.25	89.85	79.70	90.05	89.35	80.05
		0.5	20.55	20.55	20.55	40.65	40.65	40.65	40.10	40.10	40.10	20.55	20.55	20.45	16.60	40.65	40.55	39.70	30.85	40.55	39.70	30.85	40.15	40.10	39.50	31.65	39.80	39.40	31.45
		0.9	2.30	2.30	2.30	28.25	28.25	28.25	28.00	28.00	28.00	2.30	2.30	2.20	1.55	28.25	28.25	28.00	20.80	28.25	28.00	20.80	28.15	28.10	27.65	20.45	28.55	27.75	20.40

Table 7: Simulation results in Design 5: $\mu_j(\theta) = 0.05$ for all $j = 1, \ldots p$, $\Sigma(\theta)$ equicorrelated.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	70.30	70.25	70.25	75.70	75.70	75.70	73.15	73.15	73.15	70.25	70.10	68.95	53.85	75.70	75.60	74.55	61.55	75.60	74.55	61.55	73.55	73.50	72.50	58.05	73.10	71.90	58.60
		0.5	64.10	64.05	64.05	71.15	71.15	71.15	69.05	69.00	69.00	64.05	64.00	62.60	49.40	71.15	71.00	70.10	57.85	71.00	70.10	57.85	69.30	69.20	68.35	54.00	69.20	68.10	54.85
		0.9	33.70	33.70	33.70	53.05	53.05	53.05	52.35	52.35	52.35	33.70	33.65	33.00	25.15	53.05	53.05	52.35	41.35	53.05	52.35	41.35	52.60	52.40	51.40	40.70	52.35	51.35	40.00
	500	0	77.50	77.45	77.45	84.10	84.05	84.05	80.60	80.55	80.55	77.45	77.40	76.10	62.00	84.05	83.90	83.30	70.90	83.90	83.30	70.90	80.75	80.65	79.85	66.05	80.60	79.90	65.90
		0.5	72.90	72.80	72.80	81.65	81.55	81.55	78.35	78.30	78.30	72.75	72.75	71.75	58.20	81.55	81.45	80.55	67.40	81.45	80.55	67.40	78.60	78.50	77.60	64.80	78.35	77.25	64.25
		0.9	43.50	43.50	43.50	65.40	65.40	65.40	64.50	64.50	64.50	43.50	43.45	42.10	31.60	65.40	65.30	64.15	52.10	65.30	64.15	52.10	64.75	64.65	63.70	51.25	64.75	63.45	50.95
	1000	0	81.70	81.65	81.65	89.15	89.10	89.10	84.65	84.50	84.50	81.65	81.55	80.45	66.30	89.10	89.00	88.05	77.75	89.00	88.05	77.75	84.95	84.95	84.05	70.60	84.95	84.10	70.75
		0.5	79.45	79.40	79.40	87.75	87.75	87.75	83.65	83.65	83.65	79.40	79.35	78.55	65.15	87.75	87.65	86.50	76.15	87.65	86.50	76.15	84.15	84.05	83.40	71.30	84.75	84.00	71.35
		0.9	51.25	51.25	51.25	73.30	73.25	73.25	71.55	71.50	71.50	51.25	51.25	50.25	39.00	73.25	73.20	72.50	59.35	73.20	72.50	59.35	72.10	72.00	71.05	58.00	71.95	70.70	58.35
$U(-\sqrt{3},\sqrt{3})$	200	0	70.00	70.00	70.00	76.50	76.50	76.50	76.30	76.30	76.30	70.00	69.80	68.90	55.90	76.50	76.40	75.80	62.10	76.40	75.80	62.10	76.65	76.60	75.60	63.55	76.35	75.50	62.90
		0.5	65.25	65.25	65.25	72.65	72.65	72.65	73.00	73.00	73.00	65.25	65.15	63.50	51.85	72.65	72.50	71.55	59.20	72.50	71.55	59.20	73.50	73.35	72.20	59.95	73.10	72.05	59.65
		0.9	34.10	34.10	34.10	53.45	53.45	53.45	53.90	53.90	53.90	34.10	34.10	33.30	25.70	53.45	53.20	52.00	41.85	53.20	52.00	41.85	54.25	54.10	53.15	41.85	53.80	52.40	41.35
	500	0	77.30	77.30	77.30	83.70	83.70	83.70	84.45	84.45	84.45	77.30	77.25	76.25	61.90	83.70	83.65	82.80	71.25	83.65	82.80	71.25	84.75	84.75	84.00	72.30	84.25	83.40	72.00
		0.5	74.60	74.60	74.60	83.55	83.55	83.55	83.10	83.10	83.10	74.60	74.55	73.60	59.85	83.55	83.45	83.10	69.75	83.45	83.10	69.75	83.45	83.40	82.55	70.85	83.60	82.60	70.40
		0.9	42.85	42.85	42.85	64.80	64.80	64.80	65.70	65.70	65.70	42.85	42.80	41.70	32.90	64.80	64.80	64.15	53.10	64.80	64.15	53.10	66.05	66.05	65.25	53.30	64.90	64.20	53.40
	1000	0	83.40	83.40	83.40	90.00	90.00	90.00	89.95	89.95	89.95	83.40	83.35	82.05	68.50	90.00	89.85	89.20	79.20	89.85	89.20	79.20	90.35	90.25	89.85	79.70	90.05	89.35	80.05
		0.5	80.50	80.50	80.50	88.70	88.70	88.70	89.15	89.15	89.15	80.50	80.35	79.55	66.55	88.70	88.70	87.90	77.00	88.70	87.90	77.00	89.25	89.20	88.30	77.25	88.85	88.05	78.15
		0.9	53.30	53.30	53.30	75.35	75.35	75.35	74.65	74.65	74.65	53.30	53.20	52.40	42.30	75.35	75.10	74.25	62.50	75.10	74.25	62.50	75.30	75.20	74.05	62.65	75.95	74.90	62.75

Table 8: Simulation results in Design 6: $\mu_j(\theta) = 0.05$ for all $j = 1, \ldots p$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	52.15	52.15	52.15	55.45	55.40	55.40	54.20	54.20	54.20	11.80	51.15	50.55	39.50	13.75	54.65	54.10	42.85	55.20	54.15	42.85	12.80	52.90	52.70	41.55	53.00	52.05	42.05
		0.5	27.15	27.15	27.15	35.35	35.35	35.35	34.55	34.55	34.55	7.10	26.90	26.35	20.90	15.15	35.15	34.10	27.20	35.25	34.10	27.20	14.80	34.50	34.10	26.65	34.85	34.30	26.45
		0.9	10.80	10.80	10.80	28.25	28.25	28.25	28.25	28.25	28.25	2.20	10.75	10.60	7.85	19.15	28.25	27.75	20.85	28.25	27.75	20.85	19.20	28.35	27.60	20.70	28.10	27.80	20.55
	500	0	59.40	59.20	59.20	64.70	64.65	64.65	62.70	62.60	62.60	14.15	58.05	57.85	45.20	17.95	63.30	62.90	50.50	63.70	62.95	50.50	15.70	59.80	60.35	48.15	60.85	60.00	47.40
		0.5	25.35	25.35	25.35	36.65	36.65	36.65	36.05	36.05	36.05	7.70	25.05	24.75	19.75	16.00	36.15	35.65	28.35	36.30	35.60	28.35	15.05	35.45	34.95	27.50	35.80	35.30	27.55
		0.9	7.75	7.75	7.75	28.30	28.30	28.30	28.20	28.20	28.20	1.85	7.75	7.50	6.05	20.20	28.25	27.75	21.10	28.25	27.75	21.10	19.60	28.10	27.65	20.70	28.15	27.65	21.05
	1000	0	66.20	66.10	66.10	71.40	71.25	71.25	69.10	69.00	68.95	16.15	64.00	63.80	50.90	19.70	69.40	69.35	56.25	70.30	69.45	56.25	17.95	64.05	65.20	53.00	65.25	64.75	52.50
		0.5	22.75	22.75	22.75	36.05	36.05	36.05	35.45	35.45	35.45	6.25	22.50	22.35	18.00	17.00	35.70	34.95	27.60	35.90	35.00	27.60	15.55	34.30	34.15	26.50	34.65	34.05	26.55
		0.9	5.00	5.00	5.00	26.90	26.90	26.90	26.35	26.35	26.35	1.40	5.00	4.95	3.60	19.05	26.85	26.15	19.40	26.85	26.15	19.40	18.30	26.35	25.65	19.40	26.25	25.80	19.30
$U(-\sqrt{3},\sqrt{3})$	200	0	50.75	50.75	50.75	54.05	54.05	53.90	53.30	53.30	53.30	11.85	50.70	49.85	38.80	13.80	54.05	52.75	41.80	54.05	52.75	41.80	14.20	53.55	52.65	42.20	54.05	53.10	41.85
		0.5	26.75	26.75	26.75	35.15	35.15	35.10	35.25	35.25	35.25	7.10	26.70	26.25	20.90	14.35	35.15	34.65	27.25	35.15	34.65	27.25	15.05	35.30	34.50	26.70	34.95	34.40	27.25
		0.9	9.60	9.60	9.60	27.85	27.85	27.85	27.80	27.80	27.80	2.30	9.60	9.30	6.95	19.20	27.70	27.25	20.40	27.70	27.25	20.40	18.70	27.80	27.30	20.10	27.95	27.40	20.10
	500	0	59.10	59.10	59.05	63.25	63.25	63.25	64.45	64.45	64.35	14.00	59.05	58.00	45.80	16.80	63.05	62.25	50.50	63.05	62.25	50.50	17.50	64.50	63.00	50.15	64.15	63.40	50.40
		0.5	26.15	26.15	26.10	36.80	36.80	36.80	36.55	36.55	36.55	7.25	26.00	25.40	19.60	16.55	36.80	36.05	29.40	36.80	36.05	29.40	16.45	36.70	35.90	28.80	36.65	35.65	28.55
		0.9	7.10	7.10	7.10	28.00	28.00	28.00	27.75	27.75	27.75	1.80	7.05	6.95	5.00	20.35	27.90	27.25	21.15	27.90	27.25	21.15	20.40	27.80	27.15	20.75	28.25	27.40	20.75
	1000	0	64.55	64.55	64.55	70.15	70.15	70.10	69.65	69.65	69.60	15.75	64.45	63.45	48.40	20.05	70.05	68.95	54.15	70.05	68.95	54.15	20.40	70.25	69.00	55.00	69.85	69.15	55.40
		0.5	24.90	24.90	24.85	37.50	37.50	37.50	37.40	37.40	37.40	6.75	24.75	24.35	19.20	17.45	37.45	36.70	29.40	37.45	36.70	29.40	17.70	37.40	36.90	29.60	37.10	36.25	29.80
		0.9	5.75	5.75	5.70	27.65	27.65	27.65	27.55	27.55	27.55	1.15	5.75	5.70	3.65	20.60	27.60	26.95	20.95	27.60	26.95	20.95	20.20	27.65	26.85	20.00	27.60	27.00	20.35

Table 9: Simulation results in Design 7: $\mu_j(\theta) = -0.75 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \le 0.1p]$, $\Sigma(\theta)$ equicorrelated.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	52.15	52.15	52.15	55.45	55.40	55.40	54.20	54.20	54.20	11.80	51.15	50.55	39.50	13.75	54.65	54.10	42.85	55.20	54.15	42.85	12.80	52.90	52.70	41.55	53.00	52.05	42.05
		0.5	43.60	43.55	43.55	48.40	48.40	48.40	47.25	47.20	47.20	10.15	42.95	42.15	32.60	12.50	48.10	47.20	35.80	48.10	47.30	35.85	11.50	46.25	46.00	35.90	47.35	46.55	35.50
		0.9	18.85	18.85	18.85	33.95	33.95	33.90	34.15	34.15	34.15	4.85	18.70	18.50	14.25	8.45	33.50	32.90	24.45	33.65	32.90	24.45	8.60	33.75	33.10	24.40	32.95	32.45	24.30
	500	0	59.40	59.20	59.20	64.70	64.65	64.65	62.70	62.60	62.60	14.15	58.05	57.85	45.20	17.95	63.30	62.90	50.50	63.70	62.95	50.50	15.70	59.80	60.35	48.15	60.85	60.00	47.40
		0.5	53.45	53.45	53.45	59.40	59.30	59.30	57.55	57.45	57.45	12.45	52.65	51.95	41.55	15.55	58.05	57.80	46.60	58.65	57.75	46.65	14.60	55.65	55.40	44.45	56.10	55.40	44.55
		0.9	25.75	25.75	25.75	41.60	41.60	41.60	41.25	41.25	41.25	6.00	25.70	25.25	18.50	11.20	41.00	40.80	31.50	41.25	40.80	31.55	10.75	40.45	40.15	31.10	41.50	40.80	31.05
	1000	0	66.20	66.10	66.10	71.40	71.25	71.25	69.10	69.00	68.95	16.15	64.00	63.80	50.90	19.70	69.40	69.35	56.25	70.30	69.45	56.25	17.95	64.05	65.20	53.00	65.25	64.75	52.50
		0.5	59.60	59.35	59.35	65.75	65.65	65.60	64.10	64.05	64.00	15.40	57.65	57.45	46.45	19.80	64.40	64.25	52.40	65.10	64.40	52.55	18.00	59.80	60.55	49.45	61.80	61.05	49.90
		0.9	30.80	30.75	30.75	46.85	46.85	46.85	46.25	46.25	46.25	6.90	30.50	30.00	23.55	13.60	46.25	45.75	36.75	46.70	45.90	36.75	13.50	45.75	45.40	37.05	45.20	45.05	36.30
$U(-\sqrt{3},\sqrt{3})$	200	0	50.75	50.75	50.75	54.05	54.05	53.90	53.30	53.30	53.30	11.85	50.70	49.85	38.80	13.80	54.05	52.75	41.80	54.05	52.75	41.80	14.20	53.55	52.65	42.20	54.05	53.10	41.85
		0.5	44.80	44.80	44.70	49.75	49.75	49.70	49.65	49.65	49.65	9.95	44.75	43.80	33.65	11.95	49.70	48.70	37.75	49.70	48.70	37.75	12.25	49.75	49.15	38.25	49.65	49.00	37.70
		0.9	19.50	19.50	19.45	34.65	34.65	34.60	34.20	34.20	34.20	4.75	19.50	19.25	14.90	8.85	34.50	34.00	24.55	34.50	34.00	24.55	9.15	34.35	33.75	25.15	34.20	33.45	25.50
	500	0	59.10	59.10	59.05	63.25	63.25	63.25	64.45	64.45	64.35	14.00	59.05	58.00	45.80	16.80	63.05	62.25	50.50	63.05	62.25	50.50	17.50	64.50	63.00	50.15	64.15	63.40	50.40
		0.5	52.75	52.75	52.70	58.00	58.00	57.95	57.50	57.50	57.45	13.40	52.70	51.60	40.15	16.40	57.85	57.05	44.85	57.85	57.05	44.85	16.65	57.80	56.80	45.15	57.80	57.15	45.70
		0.9	25.70	25.70	25.70	40.65	40.65	40.60	39.90	39.90	39.90	5.65	25.65	25.00	19.25	11.05	40.60	40.15	31.75	40.60	40.15	31.75	11.55	39.90	39.30	32.05	40.25	39.95	31.95
	1000	0	64.55	64.55	64.55	70.15	70.15	70.10	69.65	69.65	69.60	15.75	64.45	63.45	48.40	20.05	70.05	68.95	54.15	70.05	68.95	54.15	20.40	70.25	69.00	55.00	69.85	69.15	55.40
		0.5	58.65	58.65	58.65	65.75	65.75	65.70	65.30	65.30	65.30	14.75	58.55	57.45	44.25	18.55	65.65	64.70	51.55	65.65	64.70	51.55	19.20	65.60	64.05	51.35	65.40	64.20	52.45
		0.9	30.65	30.65	30.65	46.55	46.55	46.55	45.85	45.85	45.75	6.80	30.60	29.95	23.35	13.40	46.35	45.55	36.55	46.35	45.55	36.55	13.05	46.00	45.05	36.40	46.55	45.90	35.90

Table 10: Simulation results in Design 8: $\mu_j(\theta) = -0.75 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \le 0.1p]$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods									CCK14's methods																		
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	52.15	52.15	50.65	55.45	55.40	53.90	54.20	54.20	52.45	11.80	40.50	47.80	38.85	13.75	43.20	50.90	46.55	51.10	51.15	42.60	12.80	40.50	48.05	39.90	47.05	48.15	40.45
		0.5	43.60	43.55	41.75	48.40	48.40	46.75	47.25	47.20	44.80	10.15	33.55	39.15	32.30	12.50	37.50	44.50	40.85	44.55	44.75	35.55	11.50	35.95	42.80	35.20	42.55	43.00	34.65
		0.9	18.85	18.85	14.95	33.95	33.95	26.40	34.15	34.15	25.95	4.85	16.15	17.90	14.20	8.45	25.70	31.75	29.35	32.10	32.10	24.45	8.60	24.95	31.25	24.15	30.45	31.10	24.15
	500	0	59.40	59.20	58.80	64.70	64.65	64.40	62.70	62.60	61.90	14.15	40.65	52.00	44.35	17.95	46.65	56.65	51.10	56.90	57.85	49.60	15.70	40.30	51.45	44.55	48.80	50.20	44.65
		0.5	53.45	53.45	53.10	59.40	59.30	58.80	57.55	57.45	56.70	12.45	38.15	47.15	40.35	15.55	43.10	52.80	48.45	52.75	53.50	46.05	14.60	38.85	48.80	42.05	47.20	48.45	42.50
		0.9	25.75	25.75	23.80	41.60	41.60	38.45	41.25	41.25	37.50	6.00	18.05	23.55	18.40	11.20	28.15	36.40	35.20	38.05	38.45	31.45	10.75	27.30	35.40	30.20	36.60	36.95	30.45
	1000	0	66.20	66.10	65.90	71.40	71.25	71.15	69.10	69.00	68.30	16.15	41.80	54.35	48.25	19.70	48.15	59.90	54.10	61.80	62.70	54.90	17.95	40.25	52.70	48.05	49.15	51.45	47.35
		0.5	59.60	59.35	59.20	65.75	65.65	65.50	64.10	64.05	63.45	15.40	38.00	49.95	44.70	19.80	44.85	57.20	51.15	58.55	59.35	51.40	18.00	38.65	49.75	45.55	49.00	50.45	46.05
		0.9	30.80	30.75	30.40	46.85	46.85	45.20	46.25	46.25	44.45	6.90	20.30	27.30	23.30	13.60	30.25	39.95	39.35	41.25	41.90	36.35	13.50	28.65	38.40	35.50	39.50	40.15	35.10
$U(-\sqrt{3},\sqrt{3})$	200	0	50.75	50.75	15.45	54.05	54.00	17.85	53.30	53.30	18.05	11.85	45.35	49.10	38.80	13.80	48.45	52.30	48.25	53.20	52.45	41.80	14.20	48.60	52.50	42.20	53.40	52.90	41.85
		0.5	44.80	44.80	14.30	49.75	49.75	16.65	49.65	49.65	16.95	9.95	39.05	43.50	33.65	11.95	43.60	48.15	44.15	48.40	48.35	37.75	12.25	43.80	48.70	38.25	48.65	48.75	37.70
		0.9	19.50	19.50	7.10	34.65	34.65	12.50	34.20	34.20	12.55	4.75	17.35	19.20	14.90	8.85	28.40	33.40	30.30	33.60	33.65	24.55	9.15	28.70	33.20	25.15	33.20	33.10	25.50
	500	0	59.10	59.10	18.90	63.25	63.25	23.10	64.45	64.45	23.55	14.00	46.40	56.60	45.75	16.80	51.05	60.85	58.25	61.00	61.30	50.50	17.50	51.65	61.90	50.15	62.10	62.35	50.40
		0.5	52.75	52.75	18.75	58.00	57.95	21.80	57.50	57.50	22.25	13.40	40.65	50.05	40.15	16.40	45.45	55.90	53.10	55.70	55.90	44.85	16.65	46.00	55.90	45.10	55.95	56.30	45.70
		0.9	25.70	25.70	9.25	40.65	40.65	15.55	39.90	39.90	15.50	5.65	19.15	24.45	19.25	11.05	29.80	38.65	36.85	38.75	38.95	31.75	11.55	30.15	37.95	32.05	37.95	39.05	31.95
	1000	0	64.55	64.55	21.95	70.15	70.15	26.70	69.65	69.60	27.55	15.75	43.60	58.95	48.25	20.05	49.60	65.40	64.00	65.25	66.40	54.00	20.40	50.50	65.80	54.95	65.45	66.85	55.35
		0.5	58.65	58.65	21.55	65.75	65.70	26.25	65.30	65.30	26.80	14.75	39.90	54.00	44.05	18.55	46.25	60.80	58.85	61.55	62.15	51.55	19.20	46.70	60.60	51.25	61.45	62.05	52.45
		0.9	30.65	30.65	11.90	46.55	46.55	19.85	45.85	45.85	19.60	6.80	20.85	28.40	23.25	13.40	31.60	41.60	41.25	43.25	43.55	36.40	13.05	31.85	41.30	36.15	42.50	43.30	35.80

Table 11: Simulation results in Design 9: $\mu_j(\theta) = -0.6 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods									CCK14's methods																		
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	52.15	51.95	42.45	55.45	55.25	45.45	54.20	53.90	42.20	11.80	17.75	27.65	31.40	13.75	20.80	31.05	34.25	32.50	34.40	36.50	12.80	18.40	28.50	31.85	27.85	29.05	32.70
		0.5	43.60	43.45	32.85	48.40	48.30	36.90	47.25	47.10	34.55	10.15	15.15	23.95	27.45	12.50	18.40	27.10	31.10	28.30	29.90	31.35	11.50	17.15	24.95	29.00	24.60	25.80	28.95
		0.9	18.85	18.60	8.10	33.95	32.80	14.45	34.15	32.80	13.85	4.85	6.70	10.60	12.55	8.45	11.55	17.15	20.85	19.40	21.05	22.65	8.60	11.45	16.40	19.65	18.85	20.15	21.85
	500	0	59.40	59.10	56.85	64.70	64.60	61.70	62.70	62.50	57.60	14.15	19.50	27.05	31.55	17.95	22.95	30.95	32.40	33.40	35.20	39.05	15.70	20.35	27.40	30.40	26.30	27.05	30.65
		0.5	53.45	53.45	49.35	59.40	59.30	54.70	57.55	57.25	51.20	12.45	16.65	24.90	29.60	15.55	20.80	29.35	31.30	31.60	33.30	36.75	14.60	18.40	26.15	29.70	25.30	26.20	30.75
		0.9	25.75	25.65	15.75	41.60	41.50	25.85	41.25	40.75	24.65	6.00	7.50	11.55	13.90	11.20	13.25	18.05	22.00	20.80	21.90	26.50	10.75	12.40	17.15	21.55	19.20	20.70	24.65
	1000	0	66.20	66.05	64.85	71.40	71.20	70.15	69.10	68.85	65.40	16.15	19.40	25.85	30.40	19.70	23.80	31.80	30.80	34.80	36.85	40.45	17.95	20.35	26.10	28.50	25.50	26.50	29.75
		0.5	59.60	59.35	57.65	65.75	65.55	64.15	64.10	63.85	60.30	15.40	19.05	25.45	30.15	19.80	23.60	30.90	31.15	32.60	33.90	37.55	18.00	21.30	26.50	29.75	26.25	27.10	30.75
		0.9	30.80	30.75	23.75	46.85	46.80	36.80	46.25	45.95	35.50	6.90	8.55	12.50	15.40	13.60	15.40	20.40	23.35	22.95	23.80	28.60	13.50	15.00	19.30	23.10	21.40	22.45	26.50
$U(-\sqrt{3},\sqrt{3})$	200	0	50.75	37.65	11.85	54.05	41.10	13.85	53.30	41.00	14.05	11.85	16.85	29.40	35.35	13.80	19.30	32.20	41.85	33.45	35.85	39.20	14.20	19.85	32.55	38.85	34.20	36.80	39.45
		0.5	44.80	33.60	10.15	49.75	37.50	12.00	49.65	38.00	12.30	9.95	14.85	25.25	30.75	11.95	17.40	27.90	37.20	29.80	32.10	35.00	12.25	17.90	29.35	34.75	30.90	32.50	35.65
		0.9	19.50	15.90	4.80	34.65	25.50	8.90	34.20	25.60	9.05	4.75	7.00	11.40	13.70	8.85	12.00	18.10	24.10	20.85	22.15	23.35	9.15	12.30	18.05	22.35	21.20	23.05	24.35
	500	0	59.10	44.90	14.00	63.25	49.70	16.85	64.45	50.30	17.30	14.00	17.20	26.60	35.75	16.80	21.50	31.50	41.15	33.85	35.95	42.95	17.50	22.05	32.30	40.15	34.75	37.10	42.85
		0.5	52.75	41.10	13.40	58.00	46.10	16.50	57.50	46.85	16.55	13.40	16.90	23.55	31.75	16.40	20.05	28.25	37.90	30.85	33.10	38.65	16.65	20.35	28.55	37.20	31.70	33.45	39.20
		0.9	25.70	20.45	5.80	40.65	32.15	11.10	39.90	31.85	11.50	5.65	7.60	11.75	15.15	11.05	12.90	18.65	23.70	22.05	23.55	27.70	11.55	13.15	18.70	23.35	21.55	23.05	27.70
	1000	0	64.55	49.05	15.75	70.15	55.40	20.15	69.65	55.70	19.85	15.75	18.65	25.60	33.20	20.05	23.30	30.50	38.90	33.25	35.10	42.70	20.40	23.75	31.15	39.25	35.10	36.55	42.95
		0.5	58.65	46.95	14.80	65.75	54.00	18.60	65.30	53.70	18.95	14.75	17.50	23.70	31.10	18.55	21.50	28.45	36.65	31.80	32.30	39.60	19.20	22.40	29.50	36.55	32.30	34.25	40.20
		0.9	30.65	25.85	6.90	46.55	38.40	13.50	45.85	38.15	12.95	6.80	8.15	11.95	15.10	13.40	15.00	19.80	25.25	23.00	24.20	29.50	13.05	14.95	19.65	25.35	23.50	24.50	28.90

Table 12: Simulation results in Design 10: $\mu_j(\theta) = -0.5 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	52.10	48.25	26.70	55.45	51.55	30.10	54.20	49.10	27.05	11.80	12.10	12.90	13.15	13.75	14.05	15.20	14.65	15.90	16.30	16.50	12.80	13.10	13.95	13.70	14.50	14.35	14.10
		0.5	43.60	38.80	19.15	48.40	43.30	22.65	47.25	40.85	21.00	10.15	10.40	11.20	11.45	12.50	12.65	13.55	13.30	14.30	14.35	14.70	11.50	11.90	12.60	12.30	13.20	13.10	12.90
		0.9	18.85	12.35	5.75	33.95	20.65	10.25	34.15	19.60	10.20	4.85	4.85	5.00	5.00	8.45	8.45	8.90	9.20	9.30	9.30	10.65	8.60	8.60	9.00	8.15	9.30	9.25	9.80
	500	0	59.40	58.35	42.65	64.70	63.55	47.30	62.70	60.75	42.05	14.15	14.35	14.85	13.85	17.95	17.90	18.70	16.40	19.70	19.55	17.80	15.70	16.05	16.70	14.40	16.75	16.55	14.70
		0.5	53.45	51.90	34.95	59.40	57.70	40.00	57.55	55.05	36.40	12.45	12.50	12.90	11.75	15.55	15.75	16.10	14.90	16.70	16.55	15.75	14.60	14.60	14.90	12.95	15.05	15.05	13.25
		0.9	25.75	20.60	9.60	41.60	33.35	16.85	41.25	32.10	15.90	6.00	6.00	6.05	5.40	11.20	11.20	11.30	10.70	11.90	11.90	11.30	10.75	10.75	10.85	9.30	11.40	11.40	10.25
	1000	0	66.20	65.60	56.10	71.40	70.75	61.15	69.10	67.20	55.30	16.15	16.20	16.50	13.40	19.70	19.80	19.85	17.20	20.55	20.35	18.60	17.95	18.05	18.05	14.40	18.00	18.05	14.45
		0.5	59.60	58.70	48.25	65.75	65.25	54.30	64.10	62.45	49.50	15.40	15.45	15.65	13.80	19.80	19.90	20.20	17.10	20.50	20.65	18.35	18.00	18.05	18.10	15.15	18.50	18.50	15.50
		0.9	30.80	28.50	14.45	46.85	42.25	23.55	46.25	41.80	22.95	6.90	6.95	6.90	6.15	13.60	13.60	13.55	12.20	14.10	14.05	12.70	13.50	13.55	13.50	10.15	13.60	13.35	11.40
$U(-\sqrt{3},\sqrt{3})$	200	0	50.70	13.40	11.85	53.95	15.25	13.80	53.25	15.65	14.00	11.85	11.85	12.75	13.25	13.80	13.95	14.95	16.20	15.45	15.55	16.00	14.20	14.30	15.10	15.40	15.30	15.40	16.15
		0.5	44.75	11.40	9.95	49.75	13.70	11.95	49.65	13.80	12.20	9.95	10.20	10.80	11.15	11.95	12.05	12.95	14.35	13.80	13.65	14.35	12.25	12.35	13.30	13.35	13.90	14.20	14.55
		0.9	19.45	5.75	4.75	34.65	10.20	8.85	34.20	10.45	9.05	4.75	4.80	5.15	5.25	8.85	8.95	9.35	10.15	10.00	10.05	11.10	9.15	9.35	9.70	9.30	10.50	10.40	11.35
	500	0	59.10	15.60	14.00	63.25	19.10	16.80	64.40	19.50	17.30	14.00	14.00	14.05	12.60	16.80	16.90	17.05	17.60	18.25	17.90	16.45	17.50	17.45	17.90	15.70	19.15	18.95	17.05
		0.5	52.75	15.55	13.40	57.95	18.80	16.40	57.50	19.00	16.40	13.40	13.40	13.65	12.15	16.40	16.50	16.60	16.95	17.80	17.60	16.50	16.65	16.70	17.15	15.05	17.85	17.85	16.65
		0.9	25.70	7.30	5.65	40.65	12.85	11.05	39.80	12.80	11.45	5.65	5.75	5.95	5.15	11.05	11.00	11.10	11.05	11.60	11.60	11.30	11.55	11.55	11.55	9.75	11.65	11.75	11.30
	1000	0	64.55	18.10	15.75	70.05	22.60	20.05	69.55	22.75	19.85	15.75	15.70	15.60	12.80	20.05	20.10	20.05	19.50	20.65	20.35	18.10	20.40	20.35	20.50	17.50	21.50	21.00	18.30
		0.5	58.65	17.50	14.75	65.70	21.60	18.55	65.30	22.25	18.95	14.75	14.75	14.75	12.40	18.55	18.55	18.70	18.50	19.45	19.25	16.80	19.20	19.20	19.30	16.45	20.05	19.70	17.05
		0.9	30.65	8.65	6.80	46.55	15.95	13.40	45.75	15.65	12.90	6.80	6.80	6.80	5.80	13.40	13.35	13.30	12.50	13.85	13.85	12.40	13.05	13.10	12.80	10.80	13.70	13.60	11.95

Table 13: Simulation results in Design 11: $\mu_j(\theta) = -0.4 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	51.60	30.90	15.75	55.15	33.95	17.70	53.50	31.50	16.45	11.80	11.75	11.70	8.15	13.75	13.65	13.30	11.20	13.70	13.35	10.30	12.80	12.75	12.45	9.05	12.90	12.50	9.15
		0.5	42.85	23.15	12.85	48.05	26.60	15.60	46.75	24.70	14.40	10.15	10.15	9.80	6.75	12.50	12.50	12.10	10.05	12.55	12.15	8.95	11.50	11.55	11.05	8.15	11.55	11.00	7.95
		0.9	18.25	6.25	5.10	31.75	11.05	9.00	31.65	11.10	9.05	4.85	4.85	4.70	3.60	8.45	8.45	8.35	7.45	8.45	8.35	6.50	8.60	8.55	8.20	6.20	8.25	8.20	6.15
	500	0	58.90	45.85	23.75	64.60	51.40	28.40	62.35	45.70	24.05	14.15	14.10	14.00	9.85	17.95	17.75	17.25	13.85	17.80	17.25	12.50	15.70	15.65	15.30	11.20	15.45	15.10	10.95
		0.5	53.30	38.85	20.15	59.35	44.25	24.70	57.05	40.20	22.35	12.45	12.45	12.05	8.40	15.55	15.55	15.20	11.95	15.55	15.20	11.10	14.60	14.60	14.25	10.25	14.50	13.85	10.00
		0.9	25.30	11.55	7.00	40.80	18.95	12.15	39.60	17.60	11.55	6.00	6.00	5.90	3.85	11.20	11.20	10.80	9.15	11.20	10.80	7.75	10.75	10.75	10.45	7.30	10.60	10.45	7.15
	1000	0	66.00	58.00	36.15	71.30	63.10	41.45	68.70	58.00	35.50	16.15	16.15	15.65	10.90	19.70	19.65	19.25	15.55	19.65	19.25	14.20	17.95	17.95	17.50	12.05	17.55	17.25	12.10
		0.5	59.55	50.40	28.95	65.65	56.85	35.20	63.75	51.75	30.70	15.40	15.40	15.20	11.45	19.80	19.75	19.50	15.95	19.75	19.50	14.65	18.00	17.95	17.50	13.10	17.95	17.80	13.05
		0.9	30.60	17.05	9.10	46.20	26.95	16.70	45.40	26.15	16.25	6.90	6.90	6.75	4.70	13.60	13.60	13.50	11.05	13.60	13.50	9.05	13.50	13.50	13.10	8.75	13.05	12.50	8.50
$U(-\sqrt{3},\sqrt{3})$	200	0	37.30	11.85	11.85	40.30	13.80	13.80	39.95	14.00	14.00	11.85	11.65	11.50	8.30	13.80	13.80	13.60	12.20	13.85	13.60	10.60	14.20	14.20	13.70	10.45	13.55	13.35	10.10
		0.5	32.90	10.10	9.95	36.85	11.95	11.95	37.45	12.20	12.20	9.95	9.95	9.75	7.65	11.95	11.90	11.70	10.80	11.90	11.70	9.15	12.25	12.25	11.85	9.15	12.10	11.75	9.25
		0.9	15.75	4.75	4.75	24.90	8.85	8.85	25.05	9.05	9.05	4.75	4.75	4.65	3.90	8.85	8.75	8.70	7.70	8.75	8.70	6.30	9.15	9.15	8.95	6.35	8.95	8.80	6.25
	500	0	43.85	14.00	14.00	48.15	16.80	16.80	48.95	17.30	17.30	14.00	14.00	13.55	8.95	16.80	16.80	16.30	15.10	16.85	16.30	11.85	17.50	17.45	17.00	12.45	17.30	17.10	11.95
		0.5	39.85	13.40	13.40	44.50	16.45	16.40	45.35	16.50	16.40	13.40	13.40	13.15	9.40	16.40	16.35	16.20	14.60	16.40	16.20	12.25	16.65	16.55	16.25	12.20	16.40	16.00	12.45
		0.9	19.85	5.75	5.65	30.85	11.10	11.05	30.90	11.45	11.45	5.65	5.65	5.40	3.90	11.05	11.00	10.65	9.50	11.00	10.70	7.90	11.55	11.50	11.20	7.65	10.90	10.55	7.65
	1000	0	46.55	15.75	15.75	53.25	20.10	20.05	53.65	19.85	19.85	15.75	15.65	15.25	10.60	20.05	20.00	19.50	17.80	20.00	19.50	14.30	20.40	20.30	19.75	14.55	20.65	19.90	14.75
		0.5	45.00	14.75	14.75	52.50	18.55	18.55	51.55	18.95	18.95	14.75	14.75	14.40	10.90	18.55	18.50	18.15	17.10	18.50	18.15	13.90	19.20	19.15	18.75	14.10	19.50	18.85	13.95
		0.9	24.35	6.80	6.80	37.00	13.45	13.40	36.60	12.90	12.90	6.80	6.80	6.70	5.15	13.40	13.30	13.05	11.35	13.30	13.05	8.80	13.05	13.05	12.70	8.90	13.25	12.75	9.00

Table 14: Simulation results in Design 12: $\mu_j(\theta) = -0.3 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	39.05	15.50	12.40	42.00	17.70	14.30	38.85	16.10	13.25	11.80	11.75	11.70	7.80	13.75	13.65	13.30	9.75	13.65	13.30	9.75	12.80	12.75	12.45	8.80	12.90	12.40	8.80
		0.5	30.85	12.80	10.50	34.10	15.25	12.95	32.35	14.00	11.95	10.15	10.15	9.80	6.45	12.50	12.50	12.05	8.80	12.50	12.05	8.80	11.50	11.50	11.05	7.85	11.50	11.00	7.60
		0.9	9.45	5.10	4.90	15.40	8.95	8.55	14.90	8.95	8.60	4.85	4.85	4.65	3.45	8.45	8.45	8.35	6.15	8.45	8.35	6.15	8.60	8.55	8.20	6.10	8.25	8.20	5.80
	500	0	51.75	22.85	16.75	56.55	27.20	20.25	52.20	23.25	17.70	14.15	14.10	14.00	9.75	17.95	17.75	17.25	12.25	17.75	17.25	12.25	15.70	15.65	15.30	11.20	15.45	15.10	10.90
		0.5	44.25	19.25	13.85	50.10	23.90	17.30	46.40	20.75	15.75	12.45	12.45	12.05	8.20	15.55	15.55	15.15	10.90	15.55	15.15	10.90	14.60	14.60	14.25	10.15	14.50	13.85	9.90
		0.9	15.40	7.05	6.10	24.80	12.05	11.30	22.90	11.50	10.80	6.00	6.00	5.90	3.85	11.20	11.20	10.80	7.65	11.20	10.80	7.65	10.75	10.75	10.45	7.30	10.60	10.45	7.00
	1000	0	59.95	32.25	21.05	65.55	38.60	24.40	60.70	32.05	21.75	16.15	16.15	15.65	10.90	19.70	19.65	19.25	14.15	19.65	19.25	14.15	17.95	17.95	17.50	12.00	17.55	17.25	12.05
		0.5	53.50	27.00	18.45	59.70	32.90	23.10	54.95	28.60	20.50	15.40	15.40	15.20	11.35	19.80	19.75	19.50	14.50	19.75	19.50	14.50	18.00	17.95	17.45	13.10	17.95	17.80	12.95
		0.9	22.25	9.00	7.00	33.70	16.25	14.00	32.50	15.90	13.85	6.90	6.90	6.75	4.70	13.60	13.60	13.50	9.05	13.60	13.50	9.05	13.50	13.50	13.10	8.75	13.05	12.50	8.50
$U(-\sqrt{3},\sqrt{3})$	200	0	13.70	11.85	11.85	15.75	13.80	13.80	15.80	14.00	14.00	11.85	11.65	11.50	8.20	13.80	13.80	13.60	10.35	13.80	13.60	10.35	14.20	14.20	13.70	10.30	13.50	13.35	9.85
		0.5	11.80	9.95	9.95	14.15	11.95	11.95	14.30	12.20	12.20	9.95	9.95	9.75	7.55	11.95	11.90	11.70	9.05	11.90	11.70	9.05	12.25	12.25	11.85	9.00	12.05	11.75	9.00
		0.9	5.85	4.75	4.75	10.40	8.85	8.85	10.55	9.05	9.05	4.75	4.75	4.65	3.80	8.85	8.75	8.70	6.00	8.75	8.70	6.00	9.15	9.15	8.95	6.25	8.95	8.80	5.85
	500	0	16.05	14.00	14.00	19.85	16.80	16.80	20.20	17.30	17.30	14.00	14.00	13.55	8.95	16.80	16.80	16.30	11.80	16.80	16.30	11.80	17.50	17.45	17.00	12.35	17.30	17.10	11.75
		0.5	15.90	13.40	13.40	19.30	16.40	16.40	19.30	16.40	16.40	13.40	13.40	13.15	9.40	16.40	16.35	16.20	12.15	16.35	16.20	12.15	16.65	16.55	16.25	12.15	16.40	16.00	12.30
		0.9	7.40	5.65	5.65	12.85	11.05	11.05	12.90	11.45	11.45	5.65	5.65	5.40	3.75	11.05	11.00	10.65	7.90	11.00	10.65	7.90	11.55	11.50	11.15	7.60	10.90	10.55	7.60
	1000	0	18.60	15.75	15.75	23.30	20.05	20.05	23.25	19.85	19.85	15.75	15.65	15.25	10.60	20.05	20.00	19.50	14.15	20.00	19.50	14.15	20.40	20.30	19.75	14.55	20.65	19.85	14.75
		0.5	17.80	14.75	14.75	22.30	18.55	18.55	22.60	18.95	18.95	14.75	14.75	14.40	10.90	18.55	18.50	18.15	13.85	18.50	18.15	13.85	19.20	19.15	18.75	14.10	19.50	18.85	13.85
		0.9	8.65	6.80	6.80	15.95	13.40	13.40	15.60	12.90	12.90	6.80	6.80	6.70	5.15	13.40	13.30	13.05	8.65	13.30	13.05	8.65	13.05	13.05	12.70	8.90	13.25	12.75	8.95

Table 15: Simulation results in Design 13: $\mu_j(\theta) = -0.2 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods									CCK14's methods																	
			SN Lasso			MB Lasso			EB Lasso			SN-1S	SN-2S			MB-1S	MB-H			MB-2S			EB-1S	EB-H			EB-2S		
			C=2	C=4	C=6	C=2	C=4	C=6	C=2	C=4	C=6		0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%		0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	15.50	12.05	11.85	17.80	13.95	13.75	16.15	12.95	12.75	11.80	11.75	11.70	7.80	13.75	13.65	13.30	9.75	13.65	13.30	9.75	12.80	12.75	12.45	8.80	12.90	12.40	8.80
		0.5	12.80	10.20	10.15	15.10	12.55	12.50	14.00	11.60	11.55	10.15	10.15	9.80	6.45	12.50	12.50	12.05	8.80	12.50	12.05	8.80	11.50	11.50	11.05	7.85	11.50	11.00	7.60
		0.9	5.10	4.85	4.85	8.95	8.50	8.45	8.95	8.50	8.50	4.85	4.85	4.65	3.45	8.45	8.45	8.35	6.15	8.45	8.35	6.15	8.60	8.55	8.20	6.10	8.25	8.20	5.80
	500	0	22.00	15.05	14.30	26.10	18.80	18.05	22.20	16.25	15.75	14.15	14.10	14.00	9.75	17.95	17.75	17.25	12.25	17.75	17.25	12.25	15.70	15.65	15.30	11.20	15.45	15.10	10.90
		0.5	18.45	12.80	12.55	22.60	15.90	15.60	19.75	14.85	14.65	12.45	12.45	12.05	8.20	15.55	15.55	15.15	10.90	15.55	15.15	10.90	14.60	14.60	14.25	10.15	14.50	13.85	9.90
		0.9	7.30	6.05	6.00	12.65	11.25	11.20	11.90	10.70	10.70	6.00	6.00	5.90	3.85	11.20	11.20	10.80	7.65	11.20	10.80	7.65	10.75	10.75	10.45	7.30	10.60	10.45	7.00
	1000	0	27.30	18.00	16.60	33.45	21.90	20.25	27.05	19.45	18.20	16.15	16.15	15.65	10.90	19.70	19.65	19.25	14.15	19.65	19.25	14.15	17.95	17.95	17.50	12.00	17.55	17.25	12.05
		0.5	24.55	16.85	15.80	29.35	21.40	20.00	25.70	19.00	18.05	15.40	15.40	15.20	11.35	19.80	19.75	19.50	14.50	19.75	19.50	14.50	18.00	17.95	17.45	13.10	17.95	17.80	12.95
		0.9	9.10	6.95	6.90	16.35	13.70	13.60	15.65	13.60	13.45	6.90	6.90	6.75	4.70	13.60	13.60	13.50	9.05	13.60	13.50	9.05	13.50	13.50	13.10	8.75	13.05	12.50	8.50
$U(-\sqrt{3},\sqrt{3})$	200	0	11.85	11.85	11.85	13.85	13.80	13.80	14.00	14.00	14.00	11.85	11.65	11.50	8.20	13.80	13.80	13.60	10.35	13.80	13.60	10.35	14.20	14.20	13.70	10.30	13.50	13.35	9.85
		0.5	10.15	9.95	9.95	11.95	11.95	11.95	12.30	12.20	12.20	9.95	9.95	9.75	7.55	11.95	11.90	11.70	9.05	11.90	11.70	9.05	12.25	12.25	11.85	9.00	12.05	11.75	9.00
		0.9	4.80	4.75	4.75	8.85	8.85	8.85	9.05	9.05	9.05	4.75	4.75	4.65	3.80	8.85	8.75	8.70	6.00	8.75	8.70	6.00	9.15	9.15	8.95	6.25	8.95	8.80	5.85
	500	0	14.00	14.00	14.00	16.85	16.80	16.80	17.30	17.30	17.30	14.00	14.00	13.55	8.95	16.80	16.80	16.30	11.80	16.80	16.30	11.80	17.50	17.45	17.00	12.35	17.30	17.10	11.75
		0.5	13.40	13.40	13.40	16.45	16.40	16.40	16.50	16.40	16.40	13.40	13.40	13.15	9.40	16.40	16.35	16.20	12.15	16.35	16.20	12.15	16.65	16.55	16.25	12.15	16.40	16.00	12.30
		0.9	5.75	5.65	5.65	11.10	11.05	11.05	11.45	11.45	11.45	5.65	5.65	5.40	3.75	11.05	11.00	10.65	7.90	11.00	10.65	7.90	11.55	11.50	11.15	7.60	10.90	10.55	7.60
	1000	0	15.75	15.75	15.75	20.15	20.05	20.05	19.85	19.85	19.85	15.75	15.65	15.25	10.60	20.05	20.00	19.50	14.15	20.00	19.50	14.15	20.40	20.30	19.75	14.55	20.65	19.85	14.75
		0.5	14.75	14.75	14.75	18.60	18.55	18.55	18.95	18.95	18.95	14.75	14.75	14.40	10.90	18.55	18.50	18.15	13.85	18.50	18.15	13.85	19.20	19.15	18.75	14.10	19.50	18.85	13.85
		0.9	6.80	6.80	6.80	13.45	13.40	13.40	12.90	12.90	12.90	6.80	6.80	6.70	5.15	13.40	13.30	13.05	8.65	13.30	13.05	8.65	13.05	13.05	12.70	8.90	13.25	12.75	8.95

Table 16: Simulation results in Design 14: $\mu_j(\theta) = -0.1 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods			CCK14's methods								
			Lasso selection			SN selection			MB selection			EB Selection		
			C=2	C=4	C=6	0.01%	0.10%	1.00%	0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	9.99	10.00	10.00	10.35	10.15	10.06	10.16	10.13	10.05	10.25	10.17	10.06
		0.5	9.99	10.00	10.00	10.26	10.11	10.04	10.12	10.10	10.04	10.17	10.11	10.04
		0.9	10.00	10.00	10.01	10.06	10.02	10.01	10.02	10.02	10.01	10.03	10.02	10.01
	500	0	9.97	10.00	10.00	10.49	10.21	10.10	10.23	10.19	10.09	10.37	10.25	10.10
		0.5	9.98	10.00	10.00	10.37	10.17	10.08	10.18	10.14	10.07	10.27	10.18	10.07
		0.9	9.99	10.00	10.00	10.11	10.05	10.02	10.04	10.04	10.02	10.06	10.04	10.02
	1000	0	9.94	10.00	10.00	10.64	10.27	10.12	10.29	10.23	10.10	10.49	10.32	10.12
		0.5	9.95	10.00	10.00	10.49	10.20	10.09	10.22	10.17	10.08	10.35	10.23	10.09
		0.9	9.99	10.00	10.00	10.14	10.05	10.02	10.05	10.04	10.02	10.07	10.05	10.02
$U(-\sqrt{3},\sqrt{3})$	200	0	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
		0.5	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
		0.9	2.00	2.00	2.01	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
	500	0	5.00	5.00	5.01	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
		0.5	5.00	5.00	5.01	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
		0.9	5.00	5.00	5.01	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	1000	0	10.00	10.00	10.02	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
		0.5	10.00	10.00	10.01	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
		0.9	10.00	10.00	10.02	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00

Table 17: Percentage of moment inequalities retained by first step selection procedures in Design 8: $\mu_j(\theta) = -0.75 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$, $\Sigma(\theta)$ Toeplitz.

Density	p	ρ	Our methods			CCK14's methods								
			Lasso selection			SN selection			MB selection			EB Selection		
			C=2	C=4	C=6	0.01%	0.10%	1.00%	0.01%	0.10%	1.00%	0.01%	0.10%	1.00%
$t_4/\sqrt{2}$	200	0	9.99	12.01	46.98	95.88	81.21	49.76	78.87	74.71	44.79	82.96	79.09	47.60
		0.5	9.99	12.93	55.22	96.32	82.13	50.27	79.80	75.76	45.03	82.90	78.99	47.15
		0.9	10.00	26.27	86.88	97.22	84.36	51.58	78.14	73.37	37.67	78.91	74.25	38.46
	500	0	9.97	10.41	25.12	98.06	89.18	63.79	86.63	83.75	57.92	90.44	87.96	61.93
		0.5	9.98	10.62	31.38	98.33	89.97	64.65	87.31	84.48	58.51	90.38	87.87	61.67
		0.9	10.00	16.18	70.57	98.89	91.81	66.98	86.58	83.40	52.79	88.21	85.11	53.90
	1000	0	9.95	10.14	16.32	98.94	93.19	73.30	90.72	88.60	67.23	94.05	92.39	71.51
		0.5	9.96	10.20	19.76	99.11	93.79	74.24	91.45	89.38	67.98	94.05	92.41	71.49
		0.9	9.99	12.54	55.08	99.42	95.12	76.52	91.05	88.81	63.52	92.20	90.07	65.04
$U(-\sqrt{3},\sqrt{3})$	200	0	2.00	17.04	20.00	19.60	17.27	10.54	16.73	15.95	9.38	16.66	15.88	9.33
		0.5	2.00	16.40	20.00	19.58	17.22	10.49	16.71	15.95	9.32	16.63	15.84	9.20
		0.9	2.00	15.35	20.00	19.55	17.10	10.43	15.83	14.88	7.57	15.68	14.75	7.49
	500	0	5.01	41.99	50.00	49.62	46.61	34.35	45.25	44.09	30.97	45.20	43.98	30.77
		0.5	5.01	40.28	50.00	49.59	46.55	34.26	45.21	43.99	30.85	45.21	43.96	30.72
		0.9	5.00	37.28	50.00	49.56	46.40	34.17	43.90	42.36	26.88	43.84	42.26	26.82
	1000	0	10.02	83.05	100.00	99.63	96.22	78.71	94.10	92.53	72.32	93.88	92.26	71.79
		0.5	10.01	79.39	100.00	99.62	96.12	78.53	94.00	92.39	71.98	93.82	92.14	71.51
		0.9	10.01	72.46	100.00	99.58	95.92	78.06	91.86	89.74	64.80	91.92	89.76	64.87

Table 18: Percentage of moment inequalities retained by first step selection procedures in Design 11: $\mu_j(\theta) = -0.4 \cdot 1[j > 0.1p] + 0.05 \cdot 1[j \leq 0.1p]$, $\Sigma(\theta)$ Toeplitz.

A Appendix

Throughout the appendix, we omit the dependence of all expressions on θ. Furthermore, LHS and RHS abbreviate "left hand side" and "right hand side", respectively.

A.1 Auxiliary results

Lemma A.1. *Assume Assumptions A.1-A.2. Then, for any γ s.t. $\sqrt{n}\gamma/\sqrt{1+\gamma^2} \in [0, n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1}]$,*

$$P\left(\max_{j=1,\dots,p} |\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > \gamma\right) \leq 2p(1-\Phi(\sqrt{n}\gamma/\sqrt{1+\gamma^2}))[1 + Kn^{-\delta/2}M_{n,2+\delta}^{2+\delta}(1+\sqrt{n}\gamma/\sqrt{1+\gamma^2})^{2+\delta}], \quad \text{(A.1)}$$

where K is a universal constant.

Proof. For any $i = 1,\dots,n$ and $j = 1,\dots,p$, let $Z_{ij} \equiv (X_{ij} - \mu_j)/\sigma_j$ and $U_j \equiv \sqrt{n}\sum_{i=1}^n (Z_{ij}/n)/\sqrt{\sum_{i=1}^n (Z_{ij}^2/n)}$. We divide the rest of the proof into three steps.

Step 1. By definition, $\sqrt{n}(\hat{\mu}_j - \mu_j)/\hat{\sigma}_j = U_j/\sqrt{1 - U_j^2/n}$ and so

$$\sqrt{n}|\hat{\mu}_j - \mu_j|/\hat{\sigma}_j = |U_j|/\sqrt{1 - |U_j|^2/n}. \quad \text{(A.2)}$$

Since the RHS of Eq. (A.2) is increasing in $|U_j|$, it follows that:

$$\left\{\max_{j=1,\dots,p} |\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > \gamma\right\} = \left\{\max_{1\leq j\leq p} |U_j|/\sqrt{1-|U_j|^2/n} > \sqrt{n}\gamma\right\} \subseteq \left\{\max_{1\leq j\leq p} |U_j| \geq \sqrt{n}\gamma/\sqrt{1+\gamma^2}\right\}. \quad \text{(A.3)}$$

Step 2. For every $j = 1,\dots,p$, $\{Z_{ij}\}_{i=1}^n$ is a sequence of independent random variables with $E[Z_{ij}] = 0$, $E[Z_{ij}^2] = 1$, and $E[|Z_{ij}|^{2+\delta}] \leq M_{n,2+\delta}^{2+\delta} < \infty$. If we let $S_{nj} = \sum_{i=1}^n Z_{ij}$, $V_{nj}^2 = \sum_{i=1}^n Z_{ij}^2$, and $0 < D_{nj} = [n^{-1}\sum_{i=1}^n E[|Z_{ij}|^{2+\delta}]]^{1/(2+\delta)} \leq M_{n,2+\delta} < \infty$, then CCK14 (Lemma A.1) implies that for all $t \in [0, n^{\delta/(2(2+\delta))}D_{nj}^{-1}]$,

$$\left|\frac{P(S_{nj}/V_{nj} \geq t)}{1-\Phi(t)} - 1\right| \leq Kn^{-\delta/2}D_{nj}^{2+\delta}(1+t)^{2+\delta}, \quad \text{(A.4)}$$

where K is a universal constant.

By using that $U_j = S_{nj}/V_{nj}$, $D_{nj} \leq M_{n,2+\delta}$, and applying Eq. (A.4) to $t = \sqrt{n}\gamma/\sqrt{1+\gamma^2}$, it follows that for any γ s.t. $\sqrt{n}\gamma/\sqrt{1+\gamma^2} \in [0, n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1}]$,

$$\left|P\left(U_j \geq \sqrt{n}\gamma/\sqrt{1+\gamma^2}\right) - (1-\Phi(\sqrt{n}\gamma/\sqrt{1+\gamma^2}))\right| \leq Kn^{-\delta/2}D_{nj}^{2+\delta}\left(1-\Phi(\sqrt{n}\gamma/\sqrt{1+\gamma^2})\right)\left(1+\sqrt{n}\gamma/\sqrt{1+\gamma^2}\right)^{2+\delta}.$$

Thus, for any γ s.t. $\sqrt{n}\gamma/\sqrt{1+\gamma^2} \in [0, n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1}]$,

$$\sum_{j=1}^p P\left(U_j \geq \sqrt{n}\gamma/\sqrt{1+\gamma^2}\right) \leq p\left(1-\Phi(\sqrt{n}\gamma/\sqrt{1+\gamma^2})\right)\left[1 + Kn^{-\delta/2}M_{n,2+\delta}^{2+\delta}(1+\sqrt{n}\gamma/\sqrt{1+\gamma^2})^{2+\delta}\right]. \quad \text{(A.5)}$$

By applying the same argument for $-Z_{ij}$ instead of Z_{ij}, it follows that for any γ s.t. $\sqrt{n}\gamma/\sqrt{1+\gamma^2} \in [0, n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1}]$,

$$\sum_{j=1}^{p} P\left(-U_j \geq \sqrt{n}\gamma/\sqrt{1+\gamma^2}\right) \leq p\left(1-\Phi(\sqrt{n}\gamma/\sqrt{1+\gamma^2})\right)\left[1+Kn^{-\delta/2}M_{n,2+\delta}^{2+\delta}(1+\sqrt{n}\gamma/\sqrt{1+\gamma^2})^{2+\delta}\right]. \quad \text{(A.6)}$$

Step 3. Consider the following argument.

$$\begin{aligned}
P\left(\max_{j=1,\dots,p}|\hat{\mu}_j-\mu_j|/\hat{\sigma}_j > \gamma\right) &\leq P\left(\max_{1\leq j\leq p}|U_j| \geq \sqrt{n}\gamma/\sqrt{1+\gamma^2}\right)\\
&\leq \sum_{j=1}^{p} P\left(|U_j| \geq \sqrt{n}\gamma/\sqrt{1+\gamma^2}\right)\\
&\leq \sum_{j=1}^{p} P\left(U_j \geq \sqrt{n}\gamma/\sqrt{1+\gamma^2}\right) + \sum_{j=1}^{p} P\left[-U_j \geq \sqrt{n}\gamma/\sqrt{1+\gamma^2}\right]\\
&\leq 2p\left(1-\Phi(\sqrt{n}\gamma/\sqrt{1+\gamma^2})\right)\left[1+Kn^{-\delta/2}M_{n,2+\delta}^{2+\delta}\left(1+\sqrt{n}\gamma/\sqrt{1+\gamma^2}\right)^{2+\delta}\right],
\end{aligned}$$

where the first inequality follows from Eq. (A.3), the second inequality follows from Bonferroni bound, and the fourth inequality follows from Eqs. (A.5) and (A.6). ☐

Lemma A.2. *Assume Assumptions A.1-A.2 and let $\{\gamma_n\}_{n\geq 1} \subseteq \mathbb{R}$ satisfy $\gamma_n \geq \gamma_n^*$ for all n sufficiently large, where*

$$\gamma_n^* \equiv n^{-1/2}(M_{n,2+\delta}^2 n^{-\delta/(2+\delta)} - n^{-1})^{-1/2} = (nM_{n,2+\delta}^{2+\delta})^{-1/(2+\delta)}(1-(nM_{n,2+\delta}^{2+\delta})^{-2/(2+\delta)})^{-1/2} \to 0. \quad \text{(A.7)}$$

Then,

$$P\left(\max_{j=1,\dots,p}|\hat{\mu}_j-\mu_j|/\hat{\sigma}_j > \gamma_n\right) \leq 2p\exp\left(-2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2\right)\left[1+K(M_{n,2+\delta}/n^{\delta/(2(2+\delta))}+1)^{2+\delta}\right] \to 0. \quad \text{(A.8)}$$

Proof. First, note that the convergence to zero in Eq. (A.7) follows from $nM_{n,2+\delta}^{2+\delta} \to \infty$. Since $\gamma_n \geq \gamma_n^*$, Eq. (A.8) holds if we show:

$$P\left(\max_{j=1,\dots,p}|\hat{\mu}_j-\mu_j|/\hat{\sigma}_j > \gamma_n^*\right) \leq 2p\exp\left(-2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2\right)\left[1+K(M_{n,2+\delta}/n^{\delta/(2(2+\delta))}+1)^{2+\delta}\right] \to 0. \quad \text{(A.9)}$$

As we show next, Eq. (A.9) follows from using Lemma A.1 with $\gamma = \gamma_n^*$. This choice of γ implies $\sqrt{n}\gamma_n^*/\sqrt{1+(\gamma_n^*)^2} = n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1}$ making $\gamma = \gamma_n^*$ a valid choice in Lemma A.1. Then, Lemma A.1 with $\gamma = \gamma_n^*$ implies that:

$$\begin{aligned}
P\left(\max_{j=1,\dots,p}|\hat{\mu}_j-\mu_j|/\hat{\sigma}_j > \gamma_n^*\right) &\leq 2p\left(1-\Phi(n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1})\right)\left[1+Kn^{-\delta/2}M_{n,2+\delta}^{2+\delta}(1+n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1})^{2+\delta}\right]\\
&\leq 2p\exp\left(-2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2\right)\left[1+K(n^{-\delta/(2(2+\delta))}M_{n,2+\delta}+1)^{2+\delta}\right],
\end{aligned}$$

where we have used that $1-\Phi(t) \leq e^{-t^2/2}$. We now show that the RHS of the above display converges to zero by Assumption A.2. First, notice that $M_{n,2+\delta}^{(2+\delta)}(\ln(2k-p))^{(2+\delta)/2}n^{-\delta/2} \to 0$. Next, $(2k-p) > 1$ implies that

$M_{n,2+\delta}^{(2+\delta)} n^{-\delta/2} \to 0$ and, in turn, this implies that $n^{-\delta/(2(2+\delta))} M_{n,2+\delta} \to 0$. Furthermore, notice that $M_{n,2+\delta}^{(2+\delta)}(\ln(2k - p))^{(2+\delta)/2} n^{-\delta/2} \to 0$, $M_{n,2+\delta}^{(2+\delta)}(\ln(2k-p))^{(2+\delta)/2} n^{-\delta/2} > 0$, and $(2k-p) \geq p$ implies that $n^{\delta/(2+\delta)}(M_{n,2+\delta}^2 \ln p)^{-1} \to \infty$. This implies that:

$$p \exp\left(-2^{-1} n^{\delta/(2+\delta)} / M_{n,2+\delta}^2]\right) = \exp\left(\ln p \left[1 - 2^{-1}[n^{\delta/(2+\delta)}(M_{n,2+\delta}^2 \ln p)^{-1}]\right]\right) \to 0,$$

completing the proof. □

Proof of Lemma 3.1. By definition, $J \subseteq J_I$ where J_I is as defined in the proof of Theorem 4.2. Then, the result is a corollary of Step 2 in the proof of Theorem 4.2. □

Proof of Lemma 3.2. Fix $j = 1, \ldots, p$ arbitrarily. Bühlmann and van de Geer (2011, Eq. (2.5)) implies that the Lasso estimator in Eq. (3.2) satisfies:

$$\hat{\mu}_{L,j} = \text{sign}(\hat{\mu}_j) \times \max\{|\hat{\mu}_j| - \hat{\sigma}_j \lambda_n / 2, 0\} \quad \forall j = 1, \ldots, p. \tag{A.10}$$

To complete the proof, it suffices to show that:

$$\{\hat{\mu}_{L,j} \geq -\hat{\sigma}_j \lambda_n\} \quad = \quad \{\hat{\mu}_j \geq -3\hat{\sigma}_j \lambda_n / 2\}. \tag{A.11}$$

We divide the verification into four cases. First, consider that $\hat{\sigma}_j = 0$. If so, $-\hat{\sigma}_j \lambda_n = -3\hat{\sigma}_j \lambda_n/2 = 0$ and $\hat{\mu}_{L,j} = sign(\hat{\mu}_j) \times \max\{|\hat{\mu}_j|, 0\} = \hat{\mu}_j$, and so Eq. (A.11) holds. Second, consider that $\hat{\sigma}_j > 0$ and $\hat{\mu}_j \geq 0$. If so, $\hat{\mu}_j \geq 0 \geq -3\hat{\sigma}_j \lambda_n/2$ and so the RHS condition in Eq. (A.11) is satisfied. In addition, Eq. (A.10) implies that $\hat{\mu}_{L,j} \geq 0 \geq -\hat{\sigma}_j \lambda_n$ and so the LHS of condition in Eq. (A.11) is also satisfied. Thus, Eq. (A.11) holds. Third, consider that $\hat{\sigma}_j > 0$ and $\hat{\mu}_j \in [-\hat{\sigma}_j \lambda_n/2, 0)$. If so, $\hat{\mu}_j \geq -\hat{\sigma}_j \lambda_n/2 \geq -3\hat{\sigma}_j \lambda_n/2$ and so the RHS condition in Eq. (A.11) is satisfied. In addition, Eq. (A.10) implies that $\hat{\mu}_{L,j} = 0 \geq -\hat{\sigma}_j \lambda_n$ and so the LHS of condition in Eq. (A.11) is also satisfied. Thus, Eq. (A.11) holds. Fourth and finally, consider that $\hat{\sigma}_j > 0$ and $\hat{\mu}_j < -\hat{\sigma}_j \lambda_n/2$. Then, Eq. (A.10) implies that $\hat{\mu}_{L,j} = \hat{\mu}_j + \hat{\sigma}_j \lambda_n/2$ and so Eq. (A.11) holds. □

A.2 Results for the self-normalization approximation

Lemma A.3. *For any $\pi \in (0, 0.5]$, $n \in \mathbb{N}$, and $d \in \{0, 1 \ldots, 2k - p\}$, define the function:*

$$CV(d) \equiv \begin{cases} 0 & \text{if } d = 0, \\ \frac{\Phi^{-1}(1-\pi/d)}{\sqrt{1-(\Phi^{-1}(1-\pi/d))^2/n}} & \text{if } d > 0. \end{cases}$$

Then, $CV : \{0, 1 \ldots, 2k - p\} \to \mathbb{R}_+$ is weakly increasing for n sufficiently large.

Proof. First, we show that $CV(d) \leq CV(d+1)$ for $d = 0$. To see this, use that $\pi \leq 0.5$ such that $\Phi^{-1}(1-\pi) \geq 0$, implying that $CV(1) \geq 0 = CV(0)$.

Second, we show that $CV(d) \leq CV(d+1)$ for any $d > 0$. To see this, notice that $CV(d)$ and $CV(d+1)$ are both the result of the composition $g_1(g_2(\cdot)) : \{1\ldots, 2k-p\} \to \mathbb{R}$ where:

$$\begin{aligned} g_1(y) &\equiv y/\sqrt{1-y^2/n} : [0, \sqrt{n}) \to \mathbb{R}_+ \\ g_2(d) &\equiv \Phi^{-1}(1-\pi/d) : \{1\ldots, 2k-p\} \to \mathbb{R}. \end{aligned}$$

We first show that $g_1(g_2(\cdot))$ is properly defined by verifying that the range of g_2 is included in support of g_1. Notice that g_2 is an increasing function and so $g_2(d) \in [g_2(1), g_2(2(k-p)+p)] = [\Phi^{-1}(1-\pi), \Phi^{-1}(1-\pi/(2k-p))]$. For the lower bound, $\pi \leq 0.5$ implies that $\Phi^{-1}(1-\pi) \geq 0$. For the upper bound, consider the following argument. On the one hand, $(1-\Phi(\sqrt{n})) \leq \exp(-n/2)/2$ holds for all n large enough. On the other hand, Assumption A.2 implies that $\exp(-n/2)/2 \leq \pi/(2k-p)$. By combining these two, we conclude that $\Phi^{-1}(1-\pi/(2k-p)) \leq \sqrt{n}$ for all n large enough, as desired. From here, the monotonicity of $CV(d)$ follows from the fact that g_1 and g_2 are both weakly increasing functions and so $CV(d) = g_1(g_2(d)) \leq g_1(g_2(d+1)) = CV(d+1)$. ☐

Lemma A.4. *Assume Assumptions A.1-A.2, $\alpha \in (0, 0.5)$, and that H_0 holds. For any non-stochastic set $L \subseteq \{1, \ldots, p\}$, define:*

$$\begin{aligned} T_n(L) &\equiv \max\left\{\max_{j\in L} \sqrt{n}\hat{\mu}_j/\hat{\sigma}_j, \max_{s=p+1,\ldots,k} \sqrt{n}|\hat{\mu}_s|/\hat{\sigma}_s\right\} \\ c_n^{SN}(|L|, \alpha) &\equiv \frac{\Phi^{-1}(1-\alpha/(2(k-p)+|L|))}{\sqrt{1-(\Phi^{-1}(1-\alpha/(2(k-p)+|L|)))^2/n}}. \end{aligned}$$

Then,

$$P\left(T_n(L) > c_n^{SN}(|L|, \alpha)\right) \leq \alpha + R_n,$$

where $R_n \equiv \alpha K n^{-\delta/2} M_{n,2+\delta}^{2+\delta}(1+\Phi^{-1}(1-\alpha/(2k-p)))^{2+\delta} \to 0$ and K is a universal constant.

Proof. Under H_0, $\sqrt{n}\hat{\mu}_j/\hat{\sigma}_j \leq \sqrt{n}(\hat{\mu}_j - \mu_j)/\hat{\sigma}_j$ for all $j \in L$ and $\sqrt{n}|\hat{\mu}_s|/\hat{\sigma}_s = \sqrt{n}|\hat{\mu}_s - \mu_s|/\hat{\sigma}_s$ for $s = p+1, \ldots, k$. From this, we deduce that:

$$\begin{aligned} T_n(L) &= \max\left\{\max_{j\in L} \sqrt{n}\hat{\mu}_j/\hat{\sigma}_j, \max_{s=p+1,\ldots,k} \sqrt{n}|\hat{\mu}_s|/\hat{\sigma}_s\right\} \\ &\leq \max\left\{\max_{j\in L} \sqrt{n}(\hat{\mu}_j-\mu_j)/\hat{\sigma}_j, \max_{s=p+1,\ldots,k} \sqrt{n}|\hat{\mu}_s-\mu_s|/\hat{\sigma}_s\right\} = T_n^*(L). \end{aligned}$$

For any $i = 1, \ldots, n$ and $j = 1, \ldots, k$, let $Z_{ij} \equiv (X_{ij} - \mu_j)/\sigma_j$ and $U_j \equiv \sqrt{n}\sum_{i=1}^n (Z_{ij}/n)/\sqrt{\sum_{i=1}^n (Z_{ij}^2/n)}$. It then follows that $\sqrt{n}[\hat{\mu}_j - \mu_j]/\hat{\sigma}_j = U_j/\sqrt{1-U_j^2/n}$ and so,

$$\begin{aligned}\sqrt{n}(\hat{\mu}_j - \mu_j)/\hat{\sigma}_j &= U_j/\sqrt{1-|U_j|^2/n}\\ \sqrt{n}|\hat{\mu}_j - \mu_j|/\hat{\sigma}_j &= |U_j|/\sqrt{1-|U_j|^2/n}.\end{aligned}$$

Notice that the expressions on the RHS are increasing in U_j and $|U_j|$, respectively. Therefore, for any $c \geq 0$,

$$\begin{aligned}\{T_n^*(L) > c\} &= \left\{\max_{j\in L}\sqrt{n}(\hat{\mu}_j - \mu_j)/\hat{\sigma}_j > c\right\} \cup \left\{\max_{s=p+1,\ldots,k}\sqrt{n}|\hat{\mu}_s - \mu_s|/\hat{\sigma}_s > c\right\}\\ &= \left\{\max_{j\in L} U_j/\sqrt{1-|U_j|^2/n} > c\right\} \cup \left\{\max_{s=p+1,\ldots,k}|U_s|/\sqrt{1-|U_s|^2/n} > c\right\}\\ &= \left\{\max_{j\in L} U_j > c/\sqrt{1+c^2/n}\right\} \cup \left\{\max_{s=p+1,\ldots,k}|U_s| > c/\sqrt{1+c^2/n}\right\}.\end{aligned}$$

From here, we conclude that for all $c \geq 0$ such that: $c/\sqrt{1+c^2/n} \in [0, n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1}]$,

$$\begin{aligned}P(T_n(L) > c) &\leq P(T_n^*(L) > c)\\ &\leq P\left(\left\{\max_{j\in L} U_j > c/\sqrt{1+c^2/n}\right\} \cup \left\{\max_{s=p+1,\ldots,k}|U_s| > c/\sqrt{1+c^2/n}\right\}\right)\\ &\leq \sum_{j\in L} P\left(U_j > c/\sqrt{1+c^2/n}\right) + \sum_{s=p+1}^{k} P\left(|U_s| > c/\sqrt{1+c^2/n}\right)\\ &\leq \sum_{j\in L} P\left(U_j > c/\sqrt{1+c^2/n}\right) + \sum_{s=p+1}^{k} P\left(U_s > c/\sqrt{1+c^2/n}\right) + \sum_{g=p+1}^{k} P\left(-U_g > c/\sqrt{1+c^2/n}\right)\\ &\leq (2(k-p)+|L|)\left(1-\Phi(c/\sqrt{1+c^2/n})\right)\left[1+Kn^{-\delta/2}M_{n,2+\delta}^{2+\delta}(1+c/\sqrt{1+c^2/n})^{2+\delta}\right], \qquad \text{(A.12)}\end{aligned}$$

where the first inequality follows from $T_n(L) \leq T_n^*(L)$, the third inequality is based on a Bonferroni bound, the last inequality follows from Eqs. (A.4)-(A.5) in Lemma A.1 upon choosing $\gamma = c/\sqrt{n}$ in that result.

We are interested in applying Eq. (A.12) with $c = c_n^{SN}(|L|,\alpha)$ which satisfies:

$$(2(k-p)+|L|)\left(1-\Phi(c_n^{SN}(|L|,\alpha)/\sqrt{1+c_n^{SN}(|L|,\alpha)^2/n})\right) = \alpha. \qquad \text{(A.13)}$$

Before doing this, we need to verify that this is a valid choice, i.e., we need to verify that, for all sufficiently large n,

$$c_n^{SN}(|L|,\alpha)/\sqrt{1+c_n^{SN}(|L|,\alpha)^2/n} \quad \in \quad [0, n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1}]. \qquad \text{(A.14)}$$

On the one hand, note that $c_n^{SN}(|L|,\alpha) \geq 0$ implies that $c_n^{SN}(|L|,\alpha)/\sqrt{1+c_n^{SN}(\alpha,|L|)^2/n} \geq 0$. On the other hand, note that, by definition, $c_n^{SN}(|L|,\alpha)/\sqrt{1+c_n^{SN}(|L|,\alpha)^2/n} = \Phi^{-1}(1-\alpha/(2(k-p)+|L|))$ and so it suffices to show that $\Phi^{-1}(1-\alpha/(|L|+2(k-p)))M_{n,2+\delta}n^{-\delta/(2(2+\delta))} \to 0$. To show this, note that $\Phi^{-1}(1-\alpha/(2(k-p)+|L|)) \leq \sqrt{2\ln((|L|+2(k-p))/\alpha)} \leq \sqrt{2\ln((2k-p)/\alpha)}$, where the first inequality uses that $1-\Phi(t) \leq \exp(-t^2/2)$ for any $t > 0$ and the second inequality follows from $|L| \leq p$. These inequalities and $\ln((2k-p)/\alpha)M_{n,2+\delta}^2 n^{-\delta/(2+\delta)} \to 0$ (by

Assumption A.2) complete the verification.

Therefore, by Eq. (A.12) with $c = c_n^{SN}(|L|, \alpha)$ we conclude that:

$$P(T_n > c_n^{SN}(|L|, \alpha)) \;\leq\; \alpha + \alpha K n^{-\delta/2} M_{n,2+\delta}^{2+\delta} (1 + \Phi^{-1}(1 - \alpha/(2(k-p) + |L|)))^{2+\delta} \;\leq\; \alpha + R_n,$$

where the first inequality uses Eq. (A.13) and the second inequality follows from the definition R_n and $f(x) \equiv \Phi^{-1}(1 - \alpha/(2(k-p) + x))$ being increasing and $|L| \leq p$. To conclude the proof, it suffices to show that $R_n \to 0$. To this end, consider the following argument:

$$\begin{aligned} R_n &\equiv \alpha K n^{-\delta/2} M_{n,2+\delta}^{2+\delta} (1 + \Phi^{-1}(1 - \alpha/(2k-p)))^{2+\delta} \\ &\leq \alpha 2^{1+\delta} K n^{-\delta/2} M_{n,2+\delta}^{2+\delta} (1 + |\Phi^{-1}(1 - \alpha/(2k-p))|^{2+\delta}) \\ &\leq \alpha 2^{1+\delta} K n^{-\delta/2} M_{n,2+\delta}^{2+\delta} (1 + 2^{1/2} (\ln((2k-p)/\alpha))^{(2+\delta)/2})) = o(1), \end{aligned}$$

where the first inequality uses the convexity of $f(x) = x^{2+\delta}$ and $\delta > 0$ and Jensen's Inequality to show $(1+a)^{2+\delta} \leq 2^{1+\delta}(1 + a^{2+\delta})$ for any $a > 0$, the second inequality follows from $1 - \Phi(t) \leq \exp(-t^2/2)$ for any $t > 0$ and so $\Phi^{-1}(1 - \alpha/(2k-p)) \leq \sqrt{2 \ln((2k-p)/\alpha)}$, and the convergence to zero is based on $n^{-\delta/2} M_{n,2+\delta}^{2+\delta} (\ln(2k-p))^{(2+\delta)/2} \to 0$ (by Assumption A.2) which for $2k - p > 1$ implies that $n^{-\delta/2} M_{n,2+\delta}^{2+\delta} \to 0$. □

Proof of Theorem 4.1. This result follows from Lemma A.4 with $L = \{1, \ldots, p\}$. □

Proof of Theorem 4.2. This proof follows similar steps than CCK14 (Proof of Theorem 4.2). Let us define the sequence of sets:

$$J_I \;\equiv\; \{j = 1, \ldots, p : \mu_j/\sigma_j \geq -3\lambda_n/4\}$$

We divide the proof into three steps.

<u>Step 1.</u> We show that $\hat{\mu}_j \leq 0$ for all $j \in J_I^c$ with high probability, i.e., for any $c \in (0,1)$,

$$P\left(\cup_{j \in J_I^c} \{\hat{\mu}_j > 0\}\right) \leq 2p \exp\left(-2^{-1} n^{\delta/(2+\delta)} / M_{n,2+\delta}^2\right) \left[1 + K(M_{n,2+\delta}/n^{\delta/(2(2+\delta))} + 1)^{2+\delta}\right] + \tilde{K} n^{-c} \to 0,$$

where K and $\tilde{K}$ are universal constants.

First, we show that for any $r \in (0,1)$,

$$\left\{\cup_{j \in J_I^c} \{\hat{\mu}_j > 0\}\right\} \cap \left\{\sup_{j=1,\ldots,p} |\hat{\sigma}_j/\sigma_j - 1| \leq r/(1+r)\right\} \subseteq \left\{\sup_{j=1,\ldots,p} |\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > (1-r)\lambda_n 3/4\right\}.$$

To see this, suppose that there is an index $j = 1, \ldots, p$ s.t. $\mu_j/\sigma_j < -\lambda_n 3/4$ and $\hat{\mu}_j > 0$. Then, $|\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > \lambda_n(3/4)(\sigma_j/\hat{\sigma}_j)$. In turn, $\sup_{j=1,\ldots,p} |1 - \hat{\sigma}_j/\sigma_j| \leq r/(1+r)$ implies that $|1 - \sigma_j/\hat{\sigma}_j| \leq r$ and so $(\sigma_j/\hat{\sigma}_j)\lambda_n 3/4 \geq (1-r)\lambda_n 3/4$. By combining these, we conclude that $\sup_{j=1,\ldots,p} |\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > (1-r)\lambda_n(3/4)$.

Based on this, consider the following derivation for any $r \in (0,1)$,

$$P(\cup_{j\in J_I^c}\{\hat{\mu}_j > 0\}) \;=\; \left\{ \begin{array}{c} P(\cup_{j\in J_I^c}\{\hat{\mu}_j > 0\} \cap \sup_{j=1,\ldots,p} |\hat{\sigma}_j/\sigma_j - 1| \leq r/(1+r)) + \\ P(\cup_{j\in J_I^c}\{\hat{\mu}_j > 0\} \cap \sup_{j=1,\ldots,p} |\hat{\sigma}_j/\sigma_j - 1| > r/(1+r)) \end{array} \right\}$$
$$\leq\; P\left(\sup_{j=1,\ldots,p} |\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > (1-r)\lambda_n 3/4\right) + P\left(\sup_{j=1,\ldots,p} |\hat{\sigma}_j/\sigma_j - 1| > r/(1+r)\right). \tag{A.15}$$

By evaluating Eq. (A.15) with $r = r_n = (((n^{-(1-c)/2}\ln p + n^{-3/2}(\ln p)^2)B_n^2)^{-1} - 1)^{-1} \to 0$ (by Assumption A.3), we deduce that:

$$P(\cup_{j\in J_I^c}\{\hat{\mu}_j > 0\}) \;\leq\; 2p\exp(-2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2)[1 + K(M_{n,2+\delta}/n^{\delta/(2(2+\delta))} + 1)^{2+\delta}] + \tilde{K}n^{-c},$$

where the first term is a consequence of Lemma A.2, $r_n \to 0$, and $(1-r_n)\lambda_n 3/4 \geq n^{-1/2}(M_{n,2+\delta}^2 n^{-\delta/(2+\delta)} - n^{-1})^{-1/2}$ for all n sufficiently large, and the second term is a consequence of CCK14 (Lemma A.5) and $r_n/(1+r_n) = [n^{-(1-c)/2}\ln p + n^{-3/2}(\ln p)^2]B_n^2 \to 0$.

Step 2. We show that $J_I \subseteq \hat{J}_L$ with high probability, i.e.,

$$P(J_I \subseteq \hat{J}_L) \;\geq\; 1 - 2p\exp(-2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2)[1 + K(M_{n,2+\delta}/n^{\delta/(2(2+\delta))} + 1)^{2+\delta}] + \tilde{K}n^{-c},$$

where $K, \tilde{K}$ are uniform constants.

First, we show that for any $r \in (0,1)$,

$$\left\{\{J_I \not\subseteq \hat{J}_L\} \cap \left\{\sup_{j=1,\ldots,p} |\hat{\sigma}_j/\sigma_j - 1| \leq r/(1+r)\right\}\right\} \;\subseteq\; \left\{\sup_{j=1,\ldots,p} |\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > \lambda_n(1-r)3/4\right\}.$$

To see this, consider the following argument. Suppose that $j \in J_I$ and $j \notin \hat{J}_L$, i.e., $\mu_j/\sigma_j \geq -\lambda_n 3/4$ and $\hat{\mu}_{L,j}/\hat{\sigma}_j < -\lambda_n$ or, equivalently by Eq. (A.11), $\hat{\mu}_j/\hat{\sigma}_j < -\lambda_n 3/2$. Then, $|\mu_j - \hat{\mu}_j|/\hat{\sigma}_j > \lambda_n[\frac{3}{2} - \frac{3}{4}(\sigma_j/\hat{\sigma}_j)]$. In turn, $\sup_{j=1,\ldots,p}|1 - \hat{\sigma}_j/\sigma_j| \leq r/(1+r)$ implies that $|\sigma_j/\hat{\sigma}_j - 1| \leq r$ and so $\lambda_n[\frac{3}{2} - \frac{3}{4}(\sigma_j/\hat{\sigma}_j)] \geq \lambda_n(1-r)3/4$. By combining these, we conclude that $\sup_{j=1,\ldots,p}|\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > \lambda_n(1-r)3/4$, as desired.

Based on this, consider the following derivation for any $r \in (0,1)$,

$$P(J_I \not\subseteq \hat{J}_L) \;=\; \left\{ \begin{array}{c} P\left(\{J_I \not\subseteq \hat{J}_L\} \cap \{\sup_{j=1,\ldots,p} |\hat{\sigma}_j/\sigma_j - 1| \leq r/(1+r)\}\right) \\ +P\left(\{J_I \not\subseteq \hat{J}_L\} \cap \{\sup_{j=1,\ldots,p} |\hat{\sigma}_j/\sigma_j - 1| > r/(1+r)\}\right) \end{array} \right\}$$
$$\leq\; P\left(\sup_{j=1,\ldots,p} |\hat{\mu}_j - \mu_j|/\hat{\sigma}_j > \lambda_n(1-r)3/4\right) + P\left(\sup_{j=1,\ldots,p} |\hat{\sigma}_j/\sigma_j - 1| > r/(1+r)\right).$$

Notice that the expression on the RHS is exactly the RHS of Eq. (A.15). Consequently, by evaluating this equation in $r = r_n$ and repeating arguments used in step 1, the desired result follows.

Step 3. We now complete the argument. Consider the following derivation:

$$\begin{aligned}\Big\{\{T_n > c_n^{SN,L}(\alpha)\} \cap \{J_I \subseteq \hat{J}_L\} \cap \{\cap_{j\in J_I^c}\{\hat{\mu}_j \leq 0\}\}\Big\} &\subseteq \Big\{\{T_n > c_n^{SN}(|J_I|,\alpha)\} \cap \{\cap_{j\in J_I^c}\{\hat{\mu}_j \leq 0\}\}\Big\} \\ &\subseteq \left\{\max\left\{\max_{j\in J_I}\frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j}, \max_{s=p+1,\ldots,k}\frac{\sqrt{n}|\hat{\mu}_s|}{\hat{\sigma}_s}\right\} > c_n^{SN}(\alpha,|J_I|)\right\},\end{aligned}$$

where we have used $c_n^{SN,L}(\alpha) = c_n^{SN}(\alpha,|\hat{J}_L|)$, Lemma A.3 (in that $c_n^{SN}(\alpha,d)$ is a non-negative increasing function of $d \in \{0,1\ldots,2k-p\}$), and we take $\max_{j\in J_I}\sqrt{n}\hat{\mu}_j/\hat{\sigma}_j = -\infty$ if $J_I = \emptyset$. Thus,

$$\begin{aligned}P(T_n > c_n^{SN,L}(\alpha)) &= \left\{\begin{array}{c} P(\{T_n > c_n^{SN,L}(\alpha)\} \cap \{\{J_I \subseteq \hat{J}_L\} \cap \{\cap_{j\in J_I^c}\{\hat{\mu}_j \leq 0\}\}\})+ \\ P(\{T_n > c_n^{SN,L}(\alpha)\} \cap \{\{J_I \not\subseteq \hat{J}_L\} \cup \{\cup_{j\in J_I^c}\{\hat{\mu}_j > 0\}\}\}) \end{array}\right\} \\ &\leq P\left(\max\left\{\max_{j\in J_I}\frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j}, \max_{s=p+1,\ldots,k}\frac{\sqrt{n}|\hat{\mu}_s|}{\hat{\sigma}_s}\right\} > c_n^{SN}(\alpha,|J_I|)\right) + P(J_I \not\subseteq \hat{J}_L) + P(\cup_{j\in J_I^c}\{\hat{\mu}_j > 0\}) \\ &\leq \alpha + \left\{\begin{array}{c} \alpha K n^{-\delta/2} M_{n,2+\delta}^{2+\delta}(1+\Phi^{-1}(1-\alpha/(2k-p)))^{2+\delta}+ \\ 4p\exp(-2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2)[1+K(M_{n,2+\delta}/n^{\delta/(2(2+\delta))}+1)^{2+\delta}] + 2\tilde{K}n^{-c} \end{array}\right\} \\ &\leq \alpha + o(1), \qquad\qquad (A.16)\end{aligned}$$

where the third line uses Lemma A.4 and steps 1 and 2, and the convergence in the last line holds uniformly in the manner required by the result. □

A.3 Results for the bootstrap approximation

Lemma A.5. *Assume Assumptions A.1, A.4, $\alpha \in (0, 0.5)$, and that H_0 holds. For any non-stochastic set $L \subseteq \{1,\ldots,p\}$, define:*

$$T_n(L) \;\equiv\; \max\left\{\max_{j\in L}\sqrt{n}\hat{\mu}_j/\hat{\sigma}_j, \max_{s=p+1,\ldots,k}\sqrt{n}|\hat{\mu}_s|/\hat{\sigma}_s\right\},$$

and let $c_n^B(L,\alpha)$ with $B \in \{MB, EB\}$ denote the conditional $(1-\alpha)$-quantile based on the bootstrap. Then,

$$P(T_n(L) > c_n^B(L,\alpha)) \leq \alpha + \tilde{C}n^{-\tilde{c}},$$

where $\tilde{c}, \tilde{C} > 0$ are positive constants that only depend on the constants c, C in Assumption A.4. Furthermore, if $\mu_j = 0$ for all $j \in L$ then:

$$|P(T_n(L) > c_n^B(L,\alpha)) - \alpha| \leq \tilde{C}n^{-\tilde{c}}.$$

Finally, since $\tilde{c}, \tilde{C}$ depend only on the constants c, C in Assumption A.4, the proposed bounds are uniform in all parameters $\theta \in \Theta$ and distributions P that satisfy the assumptions in the statement.

Proof. In the absence of moment equalities equalities, this results follow from replacing $\{1,\ldots,p\}$ with L in CCK14 (proof of Theorem 4.3). As we show next, our proof can be completed by simply redefining the set of moment inequalities by adding the moment equalities as two sets of inequalities with reversed sign.

Define $A = A(L) \equiv L \cup \{p+1, \ldots, k\} \cup \{k+1, \ldots, 2k-p\}$ with $|A| = |L| + 2(k-p)$ and for any $i = 1, \ldots, n$, define the following $|A|$-dimensional auxiliary data vector:

$$X_i^E \equiv \left\{ \{X_{ij}\}'_{j \in L}, \{X_{is}\}'_{s=p+1,\ldots,k}, \{-X_{is}\}'_{s=p+1,\ldots,k} \right\}'.$$

Based on these definitions, we modify all expressions analogously, e.g.,

$$\begin{aligned} \mu^E &= \{\{\mu_j\}'_{j \in L}, \{\mu_s\}'_{s=p+1,\ldots,k}, \{-\mu_s\}'_{s=p+1,\ldots,k}\}', \\ \sigma^E &= \{\{\sigma_j\}'_{j \in L}, \{\sigma_s\}'_{s=p+1,\ldots,k}, \{\sigma_s\}'_{s=p+1,\ldots,k}\}', \end{aligned}$$

and notice that H_0 is equivalently re-written as $\mu^E \leq \mathbf{0}_{|A|}$.

In the new notation, the test statistic is re-written as $T_n(L) = \max_{j \in A} \sqrt{n} \hat{\mu}_j^E / \hat{\sigma}_j^E$, and the critical values can re-written analogously. In particular, the MB and EB test statistics are respectively defined as follows:

$$\begin{aligned} W_n^{MB}(L) &= \max_{j \in A} \sqrt{n} \sum_{i=1}^n \epsilon_i (X_{ij}^E - \hat{\mu}_j^E) / \hat{\sigma}_j^E, \\ W_n^{EB}(L) &= \max_{j \in A} \sqrt{n} \sum_{i=1}^n (X_{ij}^{*,E} - \hat{\mu}_j^E) / \hat{\sigma}_j^E. \end{aligned}$$

Given this setup, the result follows immediately from CCK14 (Theorem 4.3). □

Proof of Theorem 4.3. This result follows from Lemma A.5 with $|L| = \{1, \ldots, p\}$. □

Lemma A.6. *For any $\alpha \in (0, 0.5)$, $n \in \mathbb{N}$, $B \in \{MB, EB\}$, and $L_1 \subseteq L_2 \subseteq \{1, \ldots, p\}$,*

$$c_n^B(L_1, \alpha) \;\leq\; c_n^B(L_2, \alpha).$$

Furthermore, under the above assumptions, $P\left(c_n^B(L_1, \alpha) \geq 0\right) \geq 1 - Cn^{-c}$, where c, C are universal constants.

Proof. By definition, $L_1 \subseteq L_2$ implies that $W_n^B(L_1) \leq W_n^B(L_2)$ which, in turn, implies $c_n^B(L_1, \alpha) \leq c_n^B(L_2, \alpha)$.

We now turn to the second result. If the model has at least one moment equality, then $W_n^B(L_1) \geq 0$ and so $c_n^B(\alpha, L_1) \geq 0$. If the model has no moment equalities, then we consider consider a different argument depending on the type of bootstrap procedure being implemented.

First, consider MB. Conditionally on the sample, $W_n^{MB}(L_1) = \max_{j \in L} (1/\sqrt{n}) \sum_{i=1}^n \epsilon_i \, (X_{ij} - \hat{\mu}_j)/\hat{\sigma}_j$ is the maximum of L_1 zero mean Gaussian random variables. Thus, $\alpha \in (0, 0.5)$ implies that $c_n^{MB}(\alpha, L_1) \geq 0$.

Second, consider EM. Let $c_0(L_1, \alpha)$ denote the $(1-\alpha)$-quantile of $\max_{j \in L_1} Y_j$ with $\{Y_j\}_{j \in L_1} \sim N(\mathbf{0}, E[\tilde{Z}\tilde{Z}'])$ with $\tilde{Z} = \{Z_j\}_{j \in L_1}$ and Z as in Assumption A.3. At this point, we apply CCK14 (Eq. (66)) to our hypothetical model with the moment inequalities indexed by L_1. Applied to this model, their Eq. (66) yields:

$$P\left(c_n^{EB}(L_1, \alpha) \geq c_0(L_1, \alpha + \gamma_n)\right) \;\geq\; 1 - Cn^{-c}, \tag{A.17}$$

where $\gamma_n \equiv \zeta_{n2} + \nu_n + 8\zeta_{n1}\sqrt{\log p} \in (0, 2Cn^{-c})$, for sequences $\{(\zeta_{n1}, \zeta_{n2}, \nu_n)\}_{n\geq 1}$ and universal positive constants (c, C), all specified in CCK14. Since $\alpha < 0.5$ and $\gamma_n < 2Cn^{-c}$, it follows that for all n sufficiently large, $\alpha + \gamma_n < 0.5$ and so $c_0(\alpha + \gamma_n, L_1) > 0$. The desired result follows from combining this with Eq. (A.17). $\square$

Proof of Theorem 4.4. This proof follows similar steps than CCK14 (Proof of Theorem 4.4). Let us define the sequence of sets:

$$J_I \ \equiv\ \{j = 1, \ldots, p : \mu_j/\sigma_j \geq -3\lambda_n/4\}$$

We divide the proof into three steps. Steps 1-2 are exactly as in the proof of Theorem 4.2 so they are omitted.

Step 3. Defining $T_n(J_I)$ as in Lemma A.5 and consider the following derivation:

$$\begin{aligned}
&\left\{T_n > c_n^B(\hat{J}_L, \alpha)\right\} \cap \left\{J_I \subseteq \hat{J}_L\right\} \cap \left\{\cap_{j\in J_I^c}\{\hat{\mu}_j \leq 0\}\right\} \cap \left\{c_n^B(J_I, \alpha) \geq 0\right\}\\
&\subseteq \left\{T_n > c_n^B(J_I, \alpha)\right\} \cap \left\{\cap_{j\in J_I^c}\{\hat{\mu}_j \leq 0\}\right\} \cap \left\{c_n^B(\alpha, J_I) \geq 0\right\}\\
&\subseteq \left\{T_n(J_I) > c_n^B(J_I, \alpha)\right\},
\end{aligned}$$

where the first inclusion follows from Lemma A.6, and the second inclusion follows from noticing that $\cap_{j\in J_I^c}\{\hat{\mu}_j \leq 0\}$ and $\{T_n > c_n^B(\alpha, J_I) \geq 0\}$ implies that $\{T_n(J_I) > c_n^B(\alpha, J_I)\}$. Thus,

$$\begin{aligned}
P\left(T_n > c_n^{B,L}(\alpha)\right) &= P\left(T_n > c_n^B(\hat{J}_L, \alpha)\right)\\
&= \left\{\begin{array}{c} P\left(\{T_n > c_n^B(\hat{J}_L, \alpha)\} \cap \{\{J_I \subseteq \hat{J}_L\} \cap \{\cap_{j\in J_I^c}\{\hat{\mu}_j \leq 0\}\} \cap \{c_n^B(\alpha, J_I) \geq 0\}\}\right) + \\ P\left(\{T_n > c_n^B(\hat{J}_L, \alpha)\} \cap \{\{J_I \not\subseteq \hat{J}_L\} \cup \{\cup_{j\in J_I^c}\{\hat{\mu}_j > 0\}\} \cup \{c_n^B(\alpha, J_I) < 0\}\}\right) \end{array}\right\}\\
&\leq P(T_n(J_I) > c_n^B(J_I, \alpha)) + P(J_I \not\subseteq \hat{J}_L) + P(\cup_{j\in J_I^c}\{\hat{\mu}_j > 0\}) + P(c_n^B(\alpha, J_I) < 0)\\
&\leq \alpha + Cn^{-c} + \tilde{C}n^{-\tilde{c}} + 4p\exp(-2^{-1}n^{\delta/(2+\delta)}M_{n,2+\delta}^{-2})[1 + K(n^{-\delta/(2(2+\delta))}M_{n,2+\delta} + 1)^{2+\delta}] + 2\tilde{K}n^{-c}\\
&\leq \alpha + o(1), \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (A.18)
\end{aligned}$$

where the convergence in the last line is uniform in the manner required by the result. The third line of Eq. (A.18) uses Lemmas A.5 and A.6 as well as steps 1 and 2.

We next turn to the second part of the result. By the case under consideration, $\mu = \mathbf{0}_p$ and so $J_I = \{1, \ldots, p\}$. Thus, in this case, $\{J_I \subseteq \hat{J}_L\} = \{\hat{J}_L = J_I = \{1, \ldots, p\}\}$. By this and step 2 of Theorem 4.2, it follows that:

$$P\left(\hat{J}_L = J_I = \{1, \ldots, p\}\right) \ \geq\ 1 - 2p\exp(-2^{-1}n^{\delta/(2+\delta)}/M_{n,2+\delta}^2)[1 + K(M_{n,2+\delta}/n^{\delta/(2(2+\delta))} + 1)^{2+\delta}] + \tilde{K}n^{-c}, \quad (A.19)$$

where $K, \tilde{K}$ are uniform constants. In turn, notice that $\{\hat{J}_L = J_I = \{1, \ldots, p\}\}$ implies that $c_n^{B,1S}(\alpha) = c_n^B(J_I, \alpha) =$

$c_n^B(\hat{J}_L,\alpha)=c_n^{B,L}(\alpha)$. Thus,

$$
\begin{aligned}
P(T_n>c_n^{B,L}(\alpha)) &= P(\{T_n>c_n^{B,L}(\alpha)\}\cap\{\hat{J}_L=J_I=\{1,\ldots,p\}\})+P(\{T_n>c_n^{B,L}(\alpha)\}\cap\{\hat{J}_L=J_I=\{1,\ldots,p\}\}^c)\\
&\geq P(\{T_n>c_n^{B,1S}(\alpha)\}\cap\{\hat{J}_L=J_I=\{1,\ldots,p\}\})\\
&\geq P(T_n>c_n^{B,1S}(\alpha))-P(\{\hat{J}_L=J_I=\{1,\ldots,p\}\}^c)\\
&\geq \alpha-2\tilde{C}n^{-\tilde{c}}, \qquad (A.20)
\end{aligned}
$$

where the last inequality uses the second result in Theorem 4.3, Eq. (4.6), and Eq. (A.19). If we combine this with Eq. (A.18), the result follows. □

A.4 Results for power comparison

Proof of Theorem 5.1. The arguments in the main text show that Eq. (5.2) implies Eq. (5.3). To complete the proof, it suffices to show that the two sufficient conditions imply Eq. (5.2). By definition and Lemma 3.2,

$$
\begin{aligned}
\hat{J}_{SN} &= \{j=1,\ldots,p:\hat{\mu}_j/\hat{\sigma}_j\geq -2c_n^{SN,1S}(\beta_n)/\sqrt{n}\},\\
\hat{J}_L &= \{j=1,\ldots,p:\hat{\mu}_{j,L}/\hat{\sigma}_j\geq-\lambda_n\}=\{j=1,\ldots,p:\hat{\mu}_j/\hat{\sigma}_j\geq-\lambda_n 3/2\},
\end{aligned}
$$

Condition 1. We show this by contradiction, i.e., suppose that Eq. (5.4) and $\hat{J}_L\not\subseteq\hat{J}_{SN}$ hold. By the latter, $\exists j=1,\ldots,p$ s.t. $j\in\hat{J}_L\cap\hat{J}_{SN}^c$, i.e., $-2c_n^{SN,1S}(\beta_n)/\sqrt{n}>\hat{\mu}_j/\hat{\sigma}_j\geq-\lambda_n3/2$, which implies that $c_n^{SN,1S}(\beta_n)4/3<\sqrt{n}\lambda_n$, contradicting Eq. (5.4).

Condition 2. By definition, $c_n^{SN,1S}(\beta_n)4/3\geq\sqrt{n}\lambda_n$ is equivalent to

$$
\left(\Phi^{-1}(1-\beta_n/p)\right)^2 \;\geq\; n\lambda_n^2\frac{9}{16}. \qquad (A.21)
$$

The remainder of the proof shows that Eq. (A.21) holds under Eq. (5.5).

First, we establish a lower bound for the LHS of Eq. (A.21). For any $x\geq1$, consider the following inequalities:

$$
1-\Phi(x)\;\geq\;\frac{1}{x+1/x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\;\geq\;\frac{1}{2x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\;\geq\;\frac{1}{2\sqrt{2\pi}}e^{-x^2},
$$

where the first inequality holds for all $x>0$ by Gordon (1941, Eq. (10)), the second inequality holds by $x\geq1$ and so $x>1/x$, and the third inequality holds by $e^{-x^2/2}\leq1/x$ for all $x>0$. Note that for $\beta_n\leq10\%$ and $p\geq1$, $\Phi^{-1}(1-\beta_n/p)\geq1$. Evaluating the previous display at $x=\Phi^{-1}(1-\beta_n/p)$ yields:

$$
\left(\Phi^{-1}(1-\beta_n/p)\right)^2\;\geq\;\ln\left(\frac{p}{2\sqrt{2\pi}\beta_n}\right). \qquad (A.22)
$$

Second, we establish an upper bound for the RHS of Eq. (A.21). By Eq. (3.4),

$$n\lambda_n^2 \;=\; (4/3+\varepsilon)^2\frac{n}{n^{2/(2+\delta)}M_{n,2+\delta}^2-1} \;\leq\; 2(4/3+\varepsilon)^2 n^{\delta/(2+\delta)}M_{n,2+\delta}^{-2}, \tag{A.23}$$

where the last inequality used that $1/(x-1)\leq 2/x$ for $x\geq 2$ and that $n^{2/(2+\delta)}M_{n,2+\delta}^2\geq 2$. Thus,

$$\frac{9}{16}n\lambda_n^2 \;\leq\; \frac{18}{16}(4/3+\varepsilon)^2 n^{\delta/(2+\delta)}M_{n,2+\delta}^{-2} \;=\; \frac{9}{8}(4/3+\varepsilon)^2 n^{\delta/(2+\delta)}M_{n,2+\delta}^{-2}. \tag{A.24}$$

To conclude the proof, notice that Eq. (A.21) follows directly from combining Eqs. (5.5), (A.22), and (A.24). □

Proof of Theorem 5.2. This result has several parts.

Part 1: The same arguments used for SN method imply that Eq. (5.6) implies Eq. (5.7).

Part 2: By definition and Lemma 3.2,

$$\begin{aligned}\hat{J}_B &= \{j=1,\ldots,p:\hat{\mu}_j/\hat{\sigma}_j\geq -2c_n^{B,1S}(\beta_n)/\sqrt{n}\},\\ \hat{J}_L &= \{j=1,\ldots,p:\hat{\mu}_{j,L}/\hat{\sigma}_j\geq -\lambda_n\}=\{j=1,\ldots,p:\hat{\mu}_j/\hat{\sigma}_j\geq -\lambda_n 3/2\},\end{aligned}$$

Suppose that $\hat{J}_L\subseteq\hat{J}_B$ does not occur, i.e., $\exists j\in\hat{J}_L\cap\hat{J}_B^c$ s.t. $-2c_n^B(\beta_n)/\sqrt{n}>\hat{\mu}_j/\hat{\sigma}_j\geq-\lambda_n 3/2$. From this, we conclude that:

$$\{c_n^B(\beta_n)4/3\geq\lambda_n\sqrt{n}\}\;\subseteq\;\{\hat{J}_L\subseteq\hat{J}_B\}.$$

Let $c_0(3\beta_n)$ denote the $(1-3\beta_n)$-quantile of $\max_{1\leq j\leq p}Y_j$ with $(Y_1,\ldots,Y_p)\sim N(\mathbf{0},E[ZZ'])$ with Z as in Assumption A.3. In the remainder of this step, we consider two strategies to establish the following result:

$$c_0(3\beta_n)4/3\;\geq\;\lambda_n\sqrt{n}. \tag{A.25}$$

Under Eq. (A.25), we can conclude that:

$$\{c_n^B(\beta_n)\geq c_0(3\beta_n)\}\;\subseteq\;\{c_n^B(\beta_n)4/3\geq\lambda_n\sqrt{n}\}\;\subseteq\;\{\hat{J}_L\subseteq\hat{J}_B\}.$$

From this and since $c_0(\cdot)$ is decreasing, we conclude that for any $\mu_n\leq 3\beta_n$,

$$P\left(\hat{J}_L\subseteq\hat{J}_B\right)\;\geq\;P\left(c_n^B(\beta_n)\;\geq\;c_0(3\beta_n)\right)\;\geq\;P\left(c_n^B(\beta_n)\geq c_0(\mu_n)\right). \tag{A.26}$$

To complete the proof, it suffices to provide a uniformly high lower bound for the RHS of Eq. (A.26). To this end, we consider CCK14 (Eq. (66)) at the following values: $\alpha=\beta_n$, $\nu_n=Cn^{-c}$, and (ζ_{n2},ζ_{n1}) s.t. $\zeta_{n2}+8\zeta_{n1}\sqrt{\ln p}\leq Cn^{-c}$. Under our assumptions, these choices yield $\mu_n\equiv\beta_n+\zeta_{n2}+\upsilon_n+8\zeta_{n1}\sqrt{\log p}\leq\beta_n+2Cn^{-c}\leq 3\beta_n$. By plugging these on CCK14 (Eq. (66)), the RHS of Eq. (A.26) exceeds $1-Cn^{-c}$, as desired.

To complete the proof the step, we now describe the two strategies that can be used to show Eq. (A.25). The first strategy relies on Eq. (5.9) and the second strategy relies on Eq. (5.10).

Strategy 1. By definition,

$$c_0(3\beta_n) \;\geq\; \Phi^{-1}(1-3\beta_n), \tag{A.27}$$

By combining Eqs. (5.9), (A.23), and (A.27), it follows that:

$$c_0(3\beta_n)4/3 \;\geq\; \Phi^{-1}(1-3\beta_n)4/3 \;\geq\; \sqrt{2}(4/3+\varepsilon)n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1} \geq \sqrt{n}\lambda_n.$$

Strategy 2. First, the Borell-Cirelson-Sudakov inequality (see, e.g., Boucheron et al. (2013, Theorem 5.8)), implies that for $x \geq 0$,

$$P\left(\max_{1\leq j\leq p} Y_j \;\leq\; E[\max_{1\leq j\leq p} Y_j] - x\right) \;\leq\; e^{-x^2/2}, \tag{A.28}$$

where we used that the diagonal $E[ZZ']$ is a vector of ones. Equating the RHS of Eq. (A.28) to $(1-3\beta_n)$ yields $x = \sqrt{2\log(1/[1-3\beta_n])}$ such that:

$$c_0(3\beta_n) \;\geq\; E[\max_{1\leq j\leq p} Y_j] - \sqrt{2\log(1/[1-3\beta_n])}. \tag{A.29}$$

We now provide a lower bound for the first term on the RHS of Eq. (A.29). Consider the following derivation:

$$E[\max_{1\leq j\leq p} Y_j] \;\geq\; \min_{i\neq j}\sqrt{E(Y_i-Y_j)^2\log(p)/2} \;\geq\; \sqrt{2(1-\rho)\log(p)/2}, \tag{A.30}$$

where the first inequality follows from Sudakov's minorization inequality (see, e.g., Boucheron et al. (2013, Theorem 13.4)) and the second inequality follows from $E[ZZ']$ having a diagonal elements equal to one and the maximal absolute correlation less that ρ. Eqs. (A.29)-(A.30) imply that:

$$c_0(3\beta_n) \;\geq\; \sqrt{(1-\rho)\log(p)/2} - \sqrt{2\log(1/[1-3\beta_n])}. \tag{A.31}$$

By combining Eqs. (5.10), (A.23), and (A.31), it follows that:

$$c_0(3\beta_n)4/3 \geq 4/3(\sqrt{(1-\rho)\log(p)/2} - \sqrt{2\log(1/[1-3\beta_n])}) \geq \sqrt{2}(4/3+\varepsilon)n^{\delta/(2(2+\delta))}M_{n,2+\delta}^{-1} \geq \sqrt{n}\lambda_n.$$

Part 3: Consider the following argument.

$$\begin{aligned}
P(T_n \geq c_n^{B,2S}(\alpha)) &= P(T_n \geq c_n^{B,2S}(\alpha) \cap \hat{J}_L \subseteq \hat{J}_B) + P(T_n \geq c_n^{B,2S}(\alpha) \cap \hat{J}_L \not\subseteq \hat{J}_B)\\
&\leq P(T_n \geq c_n^{B,L}(\alpha)) + P(\hat{J}_L \not\subseteq \hat{J}_B)\\
&\leq P(T_n \geq c_n^{B,L}(\alpha)) + Cn^{-c},
\end{aligned}$$

where the first inequality uses part 1, and the second inequality uses that the sufficient conditions imply Eq. (5.8). ☐

References

Andrews, D. W. K., S. Berry, and P. Jia-Barwick (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria with an Application to Discount Chain Store Location," Mimeo: Yale University and M.I.T.

Andrews, D. W. K. and P. Guggenberger (2009): "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709.

Andrews, D. W. K. and P. Jia-Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805–2826.

Andrews, D. W. K. and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Armstrong, T. B. (2014a): "Asymptotically Exact Inference in Conditional Moment Inequality Models," *Journal of Econometrics*, 186, 51–65.

——— (2014b): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 181, 92–116.

Belloni, A. and V. Chernozhukov (2011): *High dimensional sparse econometric models: An introduction*, Springer.

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

Bontemps, C., T. Magnac, and E. Maurin (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1155.

Boucheron, S., G. Lugosi, and P. Massart (2013): *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press.

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

——— (2015): "A comparison of inferential methods in partially identified models in terms of error in coverage probability (Formerly circulated as "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Elements of the Identified Set")," *Econometric Theory*, FirstView, 1–56.

Bugni, F. A., I. A. Canay, and P. Guggenberger (2012): "Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models," *Econometrica*, 80, 1741–1768.

Bühlmann, P. and S. van de Geer (2011): *Statistics for high-dimensional data: methods, theory and applications*, Springer Science & Business Media.

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

Caner, M. (2009): "Lasso Type GMM Estimator," *Econometric Theory*, 25, 270–290.

Caner, M. and Q. M. Fan (2015): "Hybrid GEL Estimators: Instrument Selection with Adaptive Lasso," *Journal of Econometrics*, 187, 256–274.

Caner, M., X. Han, and Y. Lee (2016): "Adaptive elastic net GMM estimation with many Invalid Moment Conditions: Simultaneous Model and Moment Selection," Forthcoming: Journal of Business and Economics Statistics.

Caner, M. and H. Zhang (2014): "Adaptive Elastic Net GMM with Diverging Number of Moments," *Journal of Business and Economics Statistics*, 32, 30–47.

Cheng, X. and Z. Liao (2015): "Select the Valid and Relevant Moment Conditions: A One-step Procedure for GMM with Many Moments," *Journal of Econometrics*, 186, 443–464.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013a): "Comparison and anti-concentration bounds for maxima of Gaussian random vectors," Working paper. Forthcoming in Probability Theory Related Fields.

——— (2013b): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *Annals of Statistics*, 41, 2786–2819.

——— (2014a): "Anti-concentration and honest, adaptive confidence bands," *Annals of Statistics*, 42, 1787–1818.

——— (2014b): "Central Limit Theorems and Bootstrap in High Dimensions," Working paper.

——— (2014c): "Testing Many Moment Inequalities," Working paper.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013c): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Chetverikov, D. (2013): "Adaptive Test of Conditional Moment Inequalities," Unpublished manuscript.

Fan, J., Y. Liao, and J. Yao (2015): "Power Enhancement in High-Dimensional Cross-Sectional Tests," *Econometrica*, 84, 1496–1541.

Fan, J., J. Lv, and L. Qi (2011): "Sparse high dimensional models in economics," *Annual review of economics*, 3, 291–317.

Galichon, A. and M. Henry (2006): "Inference in Incomplete Models," Mimeo: Ecole Polytechnique, Paris - Department of Economic Sciences and Pennsylvania State University.

——— (2013): "Dilation Bootstrap: A methodology for constructing confidence regions with partially identified models," *Journal of Econometrics*, 177, 109–115.

Gordon, R. D. (1941): "Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument," *The Annals of Mathematical Statistics*, 12, 364–366.

Hastie, T., R. Tibshirani, and M. Wainwright (2015): *Statistical Learning with Sparsity: The Lasso and Generalizations*, CRC Press.

Imbens, G. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Kim, K. (2008): "Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities," Mimeo: Michigan State University.

Liao, Z. (2013): "Adaptive GMM Shrinkage Estimation with Consistent Moment Selection," *Econometric Theory*, 29, 1–48.

Manski, C. F. (1995): *Identification Problems in the Social Sciences*, Harvard University Press.

Menzel, K. (2009): "Essays of set estimation and inference with moment inequalities," Ph.D. Thesis dissertation, MIT.

——— (2014): "Consistent estimation with many moment inequalities," *Journal of Econometrics*, 182, 329350.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Application," *Econometrica*, 83, 315–334.

Ponomareva, M. (2010): "Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates," Mimeo: Northern Illinois University.

Romano, J. P. and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138, 2786–2807.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014): "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 82, 1979–2002.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117.

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 299–1315.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–165.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B (Methodological)*, 58, 267–288.

Wang, H., B. Li, and C. Leng (2009): "Shrinkage tuning parameter selection with a diverging number of parameters," *J. R. Statist. Soc. B*, 71, 671–683.