Inference in Moment Inequality Models with Combined Data Sources *

Yi Sun†

Feb 15, 2017

Abstract

In this paper, we study the estimation and inference problems for parameters when the data set is obtained by combining data from different sources. In the case where we want to estimate the means of random variables in our data set, we consider the set of estimators that are linear combinations of certain sample averages, and find that one estimator, defined to be the Adjusted estimator, that achieves the smallest possible asymptotic variance among all estimators in this set. In particular, the Adjusted estimator has smaller asymptotic variance than two commonly used estimators, the Short estimator and the Long estimator for the mean. Based on this result, we study inference problems in moment inequality models. We implement GMS procedure with three constructions of the sample moments: the Short, the Long and the Adjusted sample moments. We show that the resulting three GMS tests control asymptotic sizes. Based on local power analysis and simulations, we recommend using GMS with the Adjusted sample moments for better power.

Keywords: Combined Data Sources, Moment Inequalities, Generalized Moment Selection, Asymptotic Size, Local Power.

*I am deeply indebted to Ivan Canay, for his extensive guidance and encouragement. I am also grateful to Elie Tamer, Alex Torgovitsky, and seminar participants at various institutions for comments and feedback.

†School of Economics, The University of Sydney. Email: yi.sun@sydney.edu.au.

1 Introduction

In applied research, it is very common for researchers to construct the data set they use by combining data obtained from different sources, each source containing a random sample of observations on some variables. For example, in the recent paper Banerjee et al. (2013), the authors study how participation in a micro-finance program diffuses through social networks, and their data are obtained from two sources. The first one is a detailed individual survey that was administered to a randomly selected subsample of villagers and their spouses, and it contains information on individual's age, sub-caste, education, language, native home, occupation and social network data. And the second one is a full census that collected data on all households in the villages using a village questionnaire, and it contains demographic information, GPS coordinates and data on a variety of amenities for every household in each village. In general, for a unit i (can be individuals, firms, countries, households, etc) in our study, we may have survey data that provides information on a random variable X, and have another source of data (can be full census, administrative data, aggregate data on a neighborhood, etc) that provides information on a random variable Y. And we refer to this type of data set as having combined data sources in this paper. More examples on combining data sources include Acemoglu et al. (2015), Wollmann (2014) and Dupas et al. (2015) among others.

With combined data sources, it is very likely that data coming from different sources contain different numbers of observations. And as a result, the complied data set is likely to have different numbers of observations for different variables. In this paper, we focus on the situations where the second source contains not only the units present in the first source, but also additional units. In consequence, the available data to the researcher is a sequence of independent and identically distributed (i.i.d.) random variables $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n_1}, Y_{n_1+1}, ..., Y_{n_1+n_2}$. So for the first n_1 units there is information on both X and Y, while for the last n_2 units there is only information on Y.[1] In the example of Banerjee et al. (2013) above, the first source contains data on its variables for about 46% of all households per village, while the second source has data on a different set of variables for all households in each village.

In this situation, the common practice for researchers is to work with the sample $\{(X_i, Y_i) : 1 \leq i \leq n_1\}$ and discard the last n_2 observations on Y, with or without explicitly mentioning it. In this paper, we study whether this is the best thing to do or if there exist better alternatives. In order to answer this question, we consider moment condition models that are defined by moment equalities or moment inequalities, where some of the moments only depend on the random variable Y and do not depend on X.

To fix ideas, suppose we have the following moment inequality model:

$$\begin{cases} \mathrm{E}[m_1(X_i, Y_i, \theta)] \geq 0 \\ \mathrm{E}[m_2(Y_i, \theta)] \geq 0 \end{cases}$$

[1]Since the labeling of the units does not matter, we can let the n_1 units that have both observations on X and Y be the first n_1 units in our data set.

where the second moment inequality depends on Y but not on X. To estimate a moment equality model or to do inference in a moment inequality model, a necessary step is to construct sample analogues as consistent estimators of the moment functions. If a researcher chooses to work with the sample $\{(X_i, Y_i) : 1 \le i \le n_1\}$ instead, she will use $(\bar{m}_{1,1:n_1}, \bar{m}_{2,1:n_1})$ as the sample moments, where $\bar{m}_{j,a:b}$ denotes the average of m_j over observations $i = a, ..., b$, respectively. In other words, she will use n_1 values of m_1 and m_2 to form the sample moments, and we call the resulting sample analogues the Short sample moments. In this situation, one may also realize that there are $n_1 + n_2$ observations for m_2, and consider using $(\bar{m}_{1,1:n_1}, \bar{m}_{2,1:(n_1+n_2)})$ as the sample moments. We call the resulting sample analogues the Long sample moments.

In this paper, we propose using the Adjusted sample moments:

$$\begin{aligned} \bar{m}_1^{\text{adj}} &= \bar{m}_{1,1:n_1} + f(\hat{\Sigma}_n)[\bar{m}_{2,1:n_1} - \bar{m}_{2,(n_1+1):(n_1+n_2)}] \\ \bar{m}_2^{\text{adj}} &= \bar{m}_{2,1:(n_1+n_2)} \end{aligned} \tag{1}$$

where $f(\Sigma)$ is a known function of the variance covariance matrix Σ of (m_1, m_2), and $\hat{\Sigma}_n$ is a consistent estimator of Σ. And we present results that support the Adjusted sample moments in several steps. First, we look at estimating the means of two random variables (X, Y) with normally distributed data, and show that the maximum likelihood estimator (MLE) takes the form in (1). Motivated by the finding in the normal case, we focus on the set of estimators that can be written as linear combinations of X and Y over $i = 1, ..., n_1$ and $i = n_1 + 1, ..., n$ and show that the Adjusted estimator in (1) achieves the lowest asymptotic variance within this set. In particular, the Short, the Long and the Adjusted estimators all belong to this set, so the Adjusted estimator has smaller asymptotic variance compared to the Short and the Long estimators. Next, we apply the three estimators to construct sample moments in a moment inequality model. From our previous results, the Adjusted sample moments have the smallest asymptotic variance in estimating the moment condition among these three estimators. We then show that using the Adjusted estimator leads to inference on θ that, not only controls asymptotic size, but also delivers higher asymptotic power under certain specific alternative hypothesis than inference based on the Short or the Long sample moments. As a special case, the asymptotic power results extend immediately to doing inference on θ in moment equality models.

Following the literature, inference for parameters in moment inequality model considered in this paper is carried out by inverting suitable tests to obtain confidence sets for the parameters. And the tests are implemented with generalized moment selection (GMS) critical values. GMS is introduced by Andrews and Soares (2010), and is shown to deliver confidence sets that have correct coverage probability asymptotically. Moreover, compared to other existing methods (plug-in asymptotic and subsampling) in the literature that also have the correct asymptotic coverage probabilities, GMS tests are shown to have better power properties. In this paper, we construct GMS tests using the Short, the Long, and the Adjusted sample moments, and compare these three inference procedures. We view this paper as an attempt to answer how the asymptotic variance of the sample moments affect the quality of inference with GMS methods.

There are not many papers in the literature that deal with estimation and inference problems with combined data sources. As far as we know, there is one paper, Lynch and Wachter (2008) that studies similar problems to ours, but our paper has some important differences from their paper. Lynch and Wachter (2008) studies the estimation problem in generalized method of moments (GMM) models with samples of unequal length, and also proposes the Adjusted estimator for asymptotic efficiency gain over the Short and the Long estimator. However, we show the optimality of the Adjusted estimator within a certain set of estimators, and thus provide stronger motivations for using the Adjusted estimator. We also explain in more details why the properties of the asymptotic variance of the Adjusted estimator matter, other than improving the efficiency of estimation. Moreover, we study the inference problem in moment inequality models, which is very different from the estimation of GMM models and is not mentioned in their paper.

The remainder of the paper is organized as follows. Section 2 sets up the data generating process we consider in this paper, and discuss the problem of estimating means of variables with combined data sources. MLE in the normal case is derived in Section 2.2, and general results on the asymptotic variances of a set of estimators are presented in Section 2.3. Section 3 studies inference in moment inequality model in general form. The three test statistics we consider are defined in Section 3.2, GMS method is described in Section 3.3, size and power properties are discussed in the rest of Section 3. Section 4 gives an example on moment inequality model, and show how the tests are implemented. Results of numerical experiment on power for the example model are shown in Section 4.3.

2 Estimating Means of Variables with Combined Data Sources

Consider the situation where we have different numbers of observations for different variables in our data set. This can happen, for example, when researchers obtain their data sets by combining data from different sources with a unique identifier. In this section, we study how to estimate the means of the variables in a data set in this situation. In particular, we look at estimators that are consistent and asymptotically normally distributed.

To do this, we first describe the data generating process we consider. Then we look at the maximum likelihood estimator (MLE) for the means when the variables are jointly normally distributed to gain some intuition. Based on the intuition from the normal case, we consider a particular set of estimators in the general case without imposing any distributional assumptions.

2.1 Data Generating Process

Let X and Y be two random variables in our data set. Suppose we are interested in estimating $\mu = (\mathrm{E}[X], \mathrm{E}[Y])'$, and we observe:

$$\begin{aligned} X_i : &\quad i = 1, 2, ..., n_1, \\ Y_i : &\quad i = 1, 2, ..., n_1, n_1 + 1, n_1 + 2, ..., n_1 + n_2, \end{aligned}$$

where subscript i indicates that the observed value is obtained from individual i. For example, X can be individual's income from a survey, and Y can be individual's age from census data. Let $n = n_1 + n_2$, and we make the following assumption on the data generating process:

Assumption 2.1. X_i and Y_i are i.i.d. with a common i, and

(a) $\mathrm{E}[X]$ and $\mathrm{E}[Y]$ are finite,

(b) $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$ are positive and finite.

In the i.i.d. case we consider here, an intuitive way of constructing estimators for μ would be using the sample means. Since we have different numbers of observations on X and on Y, we need to specify which part of the observations we are using when we calculate sample means. In particular, we use the following notations for sample averages:

$$\begin{aligned} \bar{X}_{1:n_1} &= \frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \\ \bar{Y}_{1:n} &= \frac{1}{n}\sum_{i=1}^{n} Y_i, \\ \bar{Y}_{1:n_1} &= \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i, \\ \bar{Y}_{(n_1+1):n} &= \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2} Y_i, \end{aligned}$$

where the subscript $a : b$ indicates that we are using $i = a, ..., b$ to construct the sample average, respectively.

There are two estimators for μ in this setting that are widely used by applied researchers, and both estimators are constructed using sample averages. The first one is $(\bar{X}_{1:n_1}, \bar{Y}_{1:n})'$, where we take the sample averages over all the observations available for each variable. We call this estimator the Long estimator for the sample mean, and denote it as $\hat{\mu}^{\ell}$. One may also discard the part of data set that has no observations for X_i, and only use the observations with $i = 1, 2, ..., n_1$ to calculate the sample averages, and this leads us to the second estimator $(\bar{X}_{1:n_1}, \bar{Y}_{1:n_1})'$. We call this second estimator the Short estimator for the sample mean, and denote it as $\hat{\mu}^{\mathrm{s}}$. Although these two estimators are very popular, their properties have not been fully discussed. In this section, we will explore more on how to estimate μ, and study the asymptotic variances of different estimators.

To start with, we look at a special case where we impose parametric assumptions on F, and study the MLE for μ.

2.2 Intuition from the Normal Case

Starting with Fisher in 1920s, there are many theorems that show that a normal distribution with mean zero and covariance matrix equals to the inverse of the Fisher information matrix is a "best possible" asymptotic distribution that an estimator can have. Important results in this literature include the famous Cramér−Rao bound. This asymptotic distribution is achieved by MLE, and is, indeed, the best possible using certain qualifications. See the two convolution theorems by Hájek for more details. To gain some intuitions on how to find a good estimator, in this subsection only, we assume further that (X, Y) are jointly normally distributed with known variance covariance matrix Σ and unknown mean vector μ, and look at the MLE for μ .

Denote the correlation between X and Y as ρ, and denote the standard deviation of X and Y as σ_X and σ_Y, respectively. When $|\rho| < 1$, Σ is invertible, and the likelihood function evaluated at $\mu = (\mu_X, \mu_Y)'$ can be written as:

$$L(\mu_X, \mu_Y) = \prod_{i=1}^{n_1} \frac{1}{2\pi} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}[(X_i,Y_i)-(\mu_X,\mu_Y)]\Sigma^{-1}[(X_i,Y_i)-(\mu_X,\mu_Y)]'} \prod_{i=n_1+1}^{n_1+n_2} \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-(Y_i-\mu_Y)^2/2\sigma_Y^2}.$$

Maximizing the likelihood function with respect to μ_X and μ_Y is the same problem as maximizing the log likelihood function:

$$l(\mu_X, \mu_Y) = C - \frac{1}{2} \sum_{i=1}^{n_1} [(X_i, Y_i) - (\mu_X, \mu_Y)] \Sigma^{-1} [(X_i, Y_i) - (\mu_X, \mu_Y)]' - \frac{1}{2} \sum_{i=n_1+1}^{n_1+n_2} (Y_i - \mu_Y)^2/\sigma_Y^2,$$

where C is a constant. Taking first order conditions with respect to μ_X and μ_Y,

$$\begin{aligned}
0 = \frac{\partial l(\mu_X, \mu_Y)}{\partial \mu_X} &= \frac{n_1 \sigma_Y}{(1-\rho^2)\sigma_X^2 \sigma_Y} \left[(\bar{X}_{1:n_1} - \mu_X)\sigma_Y - (\bar{Y}_{1:n_1} - \mu_Y)\rho\sigma_X \right] \\
0 = \frac{\partial l(\mu_X, \mu_Y)}{\partial \mu_Y} &= \frac{n_1 \sigma_X}{(1-\rho^2)\sigma_X^2 \sigma_Y^2} \left[(\bar{Y}_{1:n_1} - \mu_Y)\sigma_X - (\bar{X}_{1:n_1} - \mu_X)\rho\sigma_Y + \frac{n_2}{n_1}(1-\rho^2)\sigma_X(\bar{Y}_{(n_1+1):n} - \mu_Y) \right],
\end{aligned}$$

which delivers the following estimators for E$[X]$ and E$[Y]$:

$$\begin{cases}
\hat{\mu}_X^{MLE} &= \bar{X}_{1:n_1} - \frac{\rho\sigma_X}{\sigma_Y}(\bar{Y}_{1:n_1} - \bar{Y}_{1:n}) = \bar{X}_{1:n_1} - \frac{n_2}{n} \times \frac{\rho\sigma_X}{\sigma_Y}(\bar{Y}_{1:n_1} - \bar{Y}_{(n_1+1):n}) \\
\hat{\mu}_Y^{MLE} &= \frac{n_1 \bar{Y}_{1:n_1} + n_2 \bar{Y}_{(n_1+1):n}}{n} = \bar{Y}_{1:n}
\end{cases}.$$

When $|\rho| = 1$, we get the same results for MLE with slightly different derivations.

Notice that the MLE for E$[X]$ is different from the sample average $\bar{X}_{1:n_1}$ in general. More specifically, whenever ρ is not 0, the MLE uses the difference between $\bar{Y}_{1:n_1}$ and $\bar{Y}_{1:n}$ to adjust the estimator $\bar{X}_{1:n_1}$.

To better understand the MLE for E$[X]$, consider the case where $\rho > 0$ and $(\bar{Y}_{1:n_1} - \bar{Y}_{1:n})$ is positive. Since $\bar{Y}_{1:n}$ has smaller variance than $\bar{Y}_{1:n_1}$, $\bar{Y}_{1:n}$ is considered a more accurate estimate for E$[Y]$, and $\bar{Y}_{1:n_1}$ is considered likely to be above its mean when $\bar{Y}_{1:n_1} > \bar{Y}_{1:n}$. Since X and Y are positively correlated, $\bar{X}_{1:n_1}$ is also likely to be above its mean in this case. Thus MLE adjusts downwards relative to $\bar{X}_{1:n_1}$ and

$$\hat{\mu}_X^{MLE} - \bar{X}_{1:n_1} = -\frac{\rho\sigma_X}{\sigma_Y}(\bar{Y}_{1:n_1} - \bar{Y}_{1:n}) < 0.$$

So the difference between $\hat{\mu}_X^{MLE}$ and $\bar{X}_{1:n_1}$ is larger when the correlation ρ or the difference $(\bar{Y}_{1:n_1} - \bar{Y}_{1:n})$ is bigger. Similarly, when $\rho < 0$ and $(\bar{Y}_{1:n_1} - \bar{Y}_{1:n})$ is positive, $\bar{Y}_{1:n_1}$ is again considered likely to be above its mean, and $\bar{X}_{1:n_1}$ is likely to be below its mean as X and Y are negatively correlated. In this case, the MLE for E$[X]$ adjusts upwards and is larger than $\bar{X}_{1:n_1}$.

Intuitively, whenever X and Y are correlated, having more observations on Y can be informative on estimating the mean of X. However, when X and Y are uncorrelated, having data on Y will not help us estimate the mean of X. Indeed, when $\rho = 0$, the MLE for E$[X]$ equals to $\bar{X}_{1:n_1}$, regardless of the value of $(\bar{Y}_{1:n_1} - \bar{Y}_{1:n})$.

Now we compare the variance of $\hat{\mu}^{\mathrm{s}}$, $\hat{\mu}^{\ell}$ and $\hat{\mu}^{MLE}$. Recall that $\hat{\mu}^{\mathrm{s}} = (\bar{X}_{1:n_1}, \bar{Y}_{1:n_1})'$ and $\hat{\mu}^{\ell} = (\bar{X}_{1:n_1}, \bar{Y}_{1:n})'$. By direct calculations

$$\begin{aligned}
\mathrm{Var}[\hat{\mu}^{\mathrm{s}}] &= \begin{pmatrix} \frac{1}{n_1}\sigma_X^2 & \frac{1}{n_1}\rho\sigma_X\sigma_Y \\ \frac{1}{n_1}\rho\sigma_X\sigma_Y & \frac{1}{n_1}\sigma_Y^2 \end{pmatrix} \\
\mathrm{Var}[\hat{\mu}^{\ell}] &= \begin{pmatrix} \frac{1}{n_1}\sigma_X^2 & \frac{1}{n}\rho\sigma_X\sigma_Y \\ \frac{1}{n}\rho\sigma_X\sigma_Y & \frac{1}{n}\sigma_Y^2 \end{pmatrix} \\
\mathrm{Var}[\hat{\mu}^{MLE}] &= \begin{pmatrix} \frac{1}{n_1}\sigma_X^2 - \frac{n_2}{nn_1}\rho^2\sigma_X^2 & \frac{1}{n}\rho\sigma_X\sigma_Y \\ \frac{1}{n}\rho\sigma_X\sigma_Y & \frac{1}{n}\sigma_Y^2 \end{pmatrix}.
\end{aligned}$$

And the differences in the variances are:

$$\begin{aligned}
\mathrm{Var}[\hat{\mu}^{\mathrm{s}}] - \mathrm{Var}[\hat{\mu}^{MLE}] &= \begin{pmatrix} \frac{n_2}{nn_1}\rho^2\sigma_X^2 & \frac{n_2}{nn_1}\rho\sigma_X\sigma_Y \\ \frac{n_2}{nn_1}\rho\sigma_X\sigma_Y & \frac{n_2}{nn_1}\sigma_Y^2 \end{pmatrix}, \\
\mathrm{Var}[\hat{\mu}^{\ell}] - \mathrm{Var}[\hat{\mu}^{MLE}] &= \begin{pmatrix} \frac{n_2}{nn_1}\rho^2\sigma_X^2 & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}$$

Based on the differences in the variances, if we are interested in estimating E$[X]$, then $\hat{\mu}_X^{MLE}$ is a better estimator than $\bar{X}_{1:n_1}$, since the variance of $\hat{\mu}_X^{MLE}$ is no larger than the variance of $\bar{X}_{1:n_1}$ in general and is strictly smaller when $\rho \neq 0$. If we are interested in estimating E$[Y]$, then $\hat{\mu}_Y^{MLE} = \bar{Y}_{1:n}$ is a better estimator than $\bar{Y}_{1:n_1}$ since $\hat{\mu}_Y^{MLE}$ has smaller variance. If we are interested in estimating μ and, possibly, any linear combinations of E$[X]$ and E$[Y]$, then $\hat{\mu}_X^{MLE}$ is a better estimator than $\hat{\mu}^{\mathrm{s}}$ and than $\hat{\mu}^{\ell}$, since the two matrices of the differences in variances above are positive semi-definite. We'll discuss more on this problem later in this section.

2.3 The General Case

Now we remove the distributional assumption and study the general case. In the example above, notice that the MLE is a linear combination of $\bar{X}_{1:n_1}$, $\bar{Y}_{1:n_1}$ and $\bar{Y}_{n_1:n_2}$. So we consider all two dimensional vectors that are linear combinations of these sample averages and look for consistent estimators for μ that can be written as a linear combination of them.

Take any linear combination of $\bar{X}_{1:n_1}$, $\bar{Y}_{1:n_1}$ and $\bar{Y}_{n_1:n_2}$

$$\begin{pmatrix} a_1\bar{X}_{1:n_1} + a_2\bar{Y}_{1:n_1} + a_3\bar{Y}_{n_1:n_2} \\ b_1\bar{X}_{1:n_1} + b_2\bar{Y}_{1:n_1} + b_3\bar{Y}_{n_1:n_2} \end{pmatrix}.$$

By LLN, we have $\bar{X}_{1:n_1} \xrightarrow{p} \mathrm{E}[X]$, $\bar{Y}_{1:n_1} \xrightarrow{p} \mathrm{E}[Y]$, $\bar{Y}_{(n_1+1):n} \xrightarrow{p} \mathrm{E}[Y]$ as $n_1, n_2 \to +\infty$, so the linear combination above is consistent for μ if and only if

$$\begin{cases} a_2 + a_3 = b_1 = 0 \\ a_1 = b_2 + b_3 = 1. \end{cases}$$

Based on this observation, we define the following set

$$\mathcal{M} = \{\hat{\mu}(a,b) | \big(\bar{X}_{1:n_1} + a(\bar{Y}_{1:n_1} - \bar{Y}_{(n_1+1):n}), b\bar{Y}_{1:n_1} + (1-b)\bar{Y}_{(n_1+1):n}\big)', a, b \in R\},$$

i.e. $\mathcal{M}$ is the set of all linear combinations of $\bar{X}_{1:n_1}$, $\bar{Y}_{1:n_1}$ and $\bar{Y}_{n_1:n_2}$ that deliver consistent estimators for μ. Notice that $\hat{\mu}^{\mathrm{s}}$ is in $\mathcal{M}$ for $a = 0$ and $b = 1$, and that $\hat{\mu}^{\ell}$ is also in $\mathcal{M}$ for $a = 0$ and $b = n_1/n$.

Next, we study the variances of the estimators in $\mathcal{M}$ for a fixed n. By direct calculation we have

$$\mathrm{Var}[\bar{X}_{1:n_1} + a(\bar{Y}_{1:n_1} - \bar{Y}_{(n_1+1):n})] = \frac{n}{n_1 n_2}\sigma_Y^2 a^2 + \frac{2\rho\sigma_X\sigma_Y}{n_1} a + \frac{1}{n_1}\sigma_X^2.$$

This variance, when viewed as a function of a, is uniquely minimized at

$$a_n^* = -n_2\rho\sigma_X/(n\sigma_Y),$$

and the corresponding estimator for $\mathrm{E}[X]$ is

$$\bar{X}_{1:n_1} - \frac{n_2\rho\sigma_X}{n\sigma_Y}(\bar{Y}_{1:n_1} - \bar{Y}_{(n_1+1):n}).$$

Similarly, we have

$$\mathrm{Var}[b\bar{Y}_{1:n_1} + (1-b)\bar{Y}_{(n_1+1):n}] = [\frac{n}{n_1 n_2}b^2 - \frac{2b}{n_2} + \frac{1}{n_2}]\sigma_Y^2.$$

And this variance, when viewed as a function of b, is uniquely minimized at

$$b_n^* = n_1/n,$$

and the corresponding estimator for $\mathrm{E}[Y]$ is $\bar{Y}_{1:n}$. The optimal weights a_n^* and b_n^* are infeasible since σ_X, σ_Y and ρ are unknown.

Notice that $\hat{\mu}(a_n^*, b_n^*)$ takes the same form as the $\hat{\mu}^{MLE}$ in the normal case, and our previous discussions on the comparisons of the variances of $\hat{\mu}_S$, $\hat{\mu}_L$ and $\hat{\mu}^{MLE}$ apply, since the calculations of the variances are valid without distributional assumptions.

In the general case without any distributional assumptions, we find that $\hat{\mu}(a_n^*, b_n^*)$ is the element in $\mathcal{M}$ of which each entry achieves the smallest possible variance for given n_1 and n. Next, we study the asymptotic variance covariance matrices of the estimators in $\mathcal{M}$. In order to do this, we need to specify how n_1 and n_2 change as $n \to \infty$, and we make the following assumption.

Assumption 2.2. Assume the sample sizes satisfy $n_1/n \to \lambda \in (0,1)$ as $n \to \infty$, where λ is a constant.

And we have the following theorem on the asymptotic variances.

Theorem 2.1. *Suppose Assumption 2.1, Assumption 2.2 and Assumption 2.3 hold. Let $V(\mathcal{M})$ be the set of asymptotic variance covariance matrices of the estimators in set $\mathcal{M}$:*

$$V(\mathcal{M}) = \left\{V \middle| \exists \hat{\mu}(a,b) \in \mathcal{M}, \sqrt{n}[\hat{\mu}(a,b) - (\mathrm{E}[X], \mathrm{E}[Y])'] \xrightarrow{d} N(0,V)\right\},$$

and let

$$V^* = \begin{bmatrix} \frac{1}{\lambda}\operatorname{Var}[X] - \frac{(1-\lambda)}{\lambda}\frac{\operatorname{Cov}^2(X,Y)}{\operatorname{Var}[Y]} & \operatorname{Cov}(X,Y) \\ \operatorname{Cov}(X,Y) & \operatorname{Var}[Y] \end{bmatrix}.$$

Then for all $V \in V(\mathcal{M})$, it follows that $V - V^*$ is positive semi-definite. Moreover, let $\hat{\mu}(a^*, b^*)$ be the element in $\mathcal{M}$ with

$$a^* = -(1-\lambda)\frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}[Y]} \text{ and } b^* = \lambda,$$

then $\hat{\mu}(a^*, b^*)$ has asymptotic variance covariance matrix equal to V^*.

Proof: By CLT, we have

$$\sqrt{n}\left(\begin{pmatrix} \bar{X}_{1:n_1} \\ \bar{Y}_{1:n_1} \\ \bar{Y}_{(n_1+1):n} \end{pmatrix} - \begin{pmatrix} \mathrm{E}[X] \\ \mathrm{E}[Y] \\ \mathrm{E}[Y] \end{pmatrix}\right) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \frac{1}{\lambda} \times \begin{pmatrix} \operatorname{Var}[X] & \operatorname{Cov}(X,Y) & 0 \\ \operatorname{Cov}(X,Y) & \operatorname{Var}[Y] & 0 \\ 0 & 0 & \frac{\lambda}{1-\lambda}\operatorname{Var}[Y] \end{pmatrix}\right).$$

Take $\hat{\mu}(a,b) \in \mathcal{M}$, i.e.

$$\hat{\mu}(a,b) = \left(\bar{X}_{1:n_1} + a(\bar{Y}_{1:n_1} - \bar{Y}_{(n_1+1):n}), b\bar{Y}_{1:n_1} + (1-b)\bar{Y}_{(n_1+1):n}\right)',$$

then the asymptotic distribution of $\hat{\mu}(a,b)$ is

$$\sqrt{n}\left(\hat{\mu}(a,b) - \begin{pmatrix} \mathrm{E}[X] \\ \mathrm{E}[Y] \end{pmatrix}\right) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, V(a,b)\right),$$

where

$$V(a,b) = \frac{1}{\lambda}\times\begin{pmatrix} a^2\frac{1}{1-\lambda}\operatorname{Var}[Y] + \operatorname{Var}[X] + 2a\operatorname{Cov}(X,Y) & a\big[b-(1-b)\frac{\lambda}{1-\lambda}\big]\operatorname{Var}[Y] + b\operatorname{Cov}(X,Y) \\ a\big[b-(1-b)\frac{\lambda}{1-\lambda}\big]\operatorname{Var}[Y] + b\operatorname{Cov}(X,Y) & \big[b^2+(1-b)^2\frac{\lambda}{1-\lambda}\big]\operatorname{Var}[Y] \end{pmatrix}.$$

The difference between $V(a,b)$ and V^* is

$$V(a,b) - V^* = \frac{1}{\lambda}\times\begin{pmatrix} \frac{1}{1-\lambda}\big[a+(1-\lambda)\frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}[Y]}\big]^2\operatorname{Var}[Y] & (b-\lambda)\big[\frac{1}{1-\lambda}a\operatorname{Var}[Y] + \operatorname{Cov}(X,Y)\big] \\ (b-\lambda)\big[\frac{1}{1-\lambda}a\operatorname{Var}[Y] + \operatorname{Cov}(X,Y)\big] & \frac{1}{1-\lambda}\big[b-\lambda\big]^2\operatorname{Var}[Y] \end{pmatrix}.$$

Notice that

$$det(V(a,b) - V^*) = 0$$

and that the diagonal elements of $V(a,b) - V^*$ are always non-negative. In particular,

- $\big(V(a,b) - V^*\big)_{11} \geq 0$, with equality if and only if $a = -(1-\lambda)\frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}[Y]} = a^*$.
- $\big(V(a,b) - V^*\big)_{22} \geq 0$, with equality if and only if $b = \lambda = b^*$.

So $V(a,b) - V^*$ is positive semi-definite when $(a,b) \neq (a^*,b^*)$. And when $(a,b) = (a^*,b^*)$, we get exactly $\hat{\mu}(a^*,b^*)$ with asymptotic variance V^*.

■

We use the following notion to compare matrices. Let A and B be two symmetric matrices of the same size. We say $A \leq B$ if $B - A$ is positive semi-definite, and $A < B$ if $B - A$ is positive definite, respectively.

Remark 2.1. If we look at the elements in $\mathcal{M}$ with *fixed* weights a and b, then $\hat{\mu}^*$ is the only one that has asymptotic variance equals to V^*. If we look at sequences of elements $\hat{\mu}(a_n,b_n)$ in $\mathcal{M}$ instead, then the asymptotic variance of $\hat{\mu}(a_n,b_n)$ is always greater than or equal to V^*. Moreover, the asymptotic variance of $\hat{\mu}(a_n,b_n)$ equals to V^* if and only if

$$a_n \to a^* \text{ and } b_n \to b^* \text{ as } n \to \infty.$$

In particular, the optimal weightings (a_n^*, b_n^*) that we found for a given sample size n satisfy this condition, i.e.

$$a_n^* \to a^* \text{ and } b_n^* \to b^* \text{ as } n \to \infty.$$

The optimal weights a^*, b^* are infeasible, but they depend on unknown quantities that can be consistently estimated from the data. In particular, if we can find estimators $\hat{a}_n$ and $\hat{b}_n$ such that

$$\begin{cases} \hat{a}_n = a^* + o_p(1) \\ \hat{b}_n = b^* + o_p(1) \end{cases},$$

then replacing (a^*,b^*) with $(\hat{a}_n,\hat{b}_n)$ will lead to an estimator that also has asymptotic variance V^*. There are many choices for $(\hat{a}_n,\hat{b}_n)$ that satisfies the above condition. Based on our previous discussions, an intuitive choice for $(\hat{a}_n, \hat{b}_n)$ would be using a feasible, plug-in counterpart of (a_n^*, b_n^*):

$$\hat{a}_n^* = -\frac{n_2}{n}\frac{\hat{\mathrm{Cov}}(X,Y)}{\hat{\mathrm{Var}}[Y]}$$

$$\hat{b}_n^* = \frac{n_1}{n},$$

where

$$\hat{\mathrm{Cov}}(X,Y) = \frac{1}{n_1}\sum_{i=1}^{n_1}(X_i - \bar{X}_{1:n_1})(Y_i - \bar{Y}_{1:n})$$

$$\hat{\mathrm{Var}}[Y] = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y}_{1:n})^2.$$

One can also get $\hat{a}_n^*$ by regressing X_i on Y_i with a constant

$$X_i = \beta_0 + \beta_1 Y_i + \epsilon_i \quad i = 1, 2, ..., n_1,$$

and use $\hat{a}_n^* = -\frac{n_2}{n}\hat{\beta}_1$. With the regression method, we are using only observations $i = 1, ..., n_1$ to estimate $\hat{\mathrm{Var}}[Y]$. While directly estimating $\hat{\mathrm{Var}}[Y]$ using observations $i = 1, ..., n$ should be more accurate than the regression method, it can be more convenient to implement $\hat{a}_n^*$ the regression method when X and Y are multi-dimensional.

The resulting feasible estimator is given by

$$\hat{\mu}^{\mathrm{adj}} = \big(\bar{X}_{1:n_1} + \hat{a}_n^*(\bar{Y}_{1:n_1} - \bar{Y}_{(n_1+1):n}), \hat{b}_n^*\bar{Y}_{1:n_1} + (1 - \hat{b}_n^*)\bar{Y}_{(n_1+1):n}\big)',$$

and we call it the Adjusted estimator for the means. The intuition behind this estimator is similar to the intuition behind the $\hat{\mu}^{MLE}$ in the previous discussion: whenever X and Y are correlated, we can use the information contained in the extra observations on Y_i to better estimate the mean of X. The asymptotic variance of $\hat{\mu}^{\mathrm{adj}}$ is $V^{\mathrm{adj}} = V^*$.

Remark 2.2. The asymptotic variance of the Short estimator $\hat{\mu}^{\mathrm{s}} = (\bar{X}_{1:n_1}, \bar{Y}_{1:n_1})'$ is

$$V^{\mathrm{s}} = \begin{pmatrix} \frac{1}{\lambda}\mathrm{Var}[X] & \frac{1}{\lambda}\mathrm{Cov}(X,Y) \\ \frac{1}{\lambda}\mathrm{Cov}(X,Y) & \frac{1}{\lambda}\mathrm{Var}[Y] \end{pmatrix}.$$

And the asymptotic variance of the Long estimator $\hat{\mu}^{\ell} = (\bar{X}_{1:n_1}, \bar{Y}_{1:n})'$ is

$$V^{\ell} = \begin{pmatrix} \frac{1}{\lambda}\mathrm{Var}[X] & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}[Y] \end{pmatrix}.$$

From the theorem, we know that $V^{\mathrm{s}} - V^{\mathrm{adj}}$ and $V^{\ell} - V^{\mathrm{adj}}$ are positive semi-definite. However,

$$V^{\mathrm{s}} - V^{\ell} = \begin{pmatrix} 0 & (\frac{1}{\lambda} - 1)\mathrm{Cov}(X,Y) \\ (\frac{1}{\lambda} - 1)\mathrm{Cov}(X,Y) & (\frac{1}{\lambda} - 1)\mathrm{Var}[Y] \end{pmatrix}$$

is *not* positive semi-definite in general. More specifically, when $\rho = 0$,

$$V^{\mathrm{s}} - V^{\ell} = \begin{pmatrix} 0 & 0 \\ 0 & (\frac{1}{\lambda} - 1)\,\mathrm{Var}[Y] \end{pmatrix}$$

is positive semi-definite. But when $\rho \neq 0$,

$$\det(V^{\mathrm{s}} - V^{\ell}) = -(\frac{1}{\lambda} - 1)^2[\mathrm{Cov}(X, Y)]^2 < 0,$$

and $V^{\mathrm{s}} - V^{\ell}$ is not positive semi-definite. This tells us that $\hat{\mu}^{\ell}$ does not necessarily have smaller asymptotic variance than $\hat{\mu}^{\mathrm{s}}$, although it uses more data.

To see why the positive semi-definiteness of the differences in the asymptotic variances matters, consider the case where we are interested in estimating a linear combination of $\mathrm{E}[X]$ and $\mathrm{E}[Y]$. Suppose we use corresponding linear combinations of $\hat{\mu}^{\mathrm{s}}$, $\hat{\mu}^{\ell}$ and $\hat{\mu}^{\mathrm{adj}}$ as our estimators, then the resulting estimator based on $\hat{\mu}^{\mathrm{adj}}$ will always have a asymptotic variance that is smaller or equal to the asymptotic variances of the resulting estimators based on $\hat{\mu}^{\mathrm{s}}$ and $\hat{\mu}^{\ell}$, regardless of what linear combination we take. However, similar result does not hold between $\hat{\mu}^{\mathrm{s}}$ and $\hat{\mu}^{\ell}$. Depending on the linear combination we use, the asymptotic variance of the resulting estimator based on $\hat{\mu}^{\ell}$ may or may not be smaller than the asymptotic variance of the resulting estimator based on $\hat{\mu}^{\mathrm{s}}$. For example, suppose we are interested in estimating $(\mathrm{E}[X] - \mathrm{E}[Y])$, then we can construct the following three estimators:

$$\begin{aligned}
(1, -1)\hat{\mu}^{\mathrm{s}} &= \bar{X}_{1:n_1} - \bar{Y}_{1:n_1} \\
(1, -1)\hat{\mu}^{\ell} &= \bar{X}_{1:n_1} - \bar{Y}_{1:n} \\
(1, -1)\hat{\mu}^{\mathrm{adj}} &= \bar{X}_{1:n_1} + \hat{a}_n^*(\bar{Y}_{1:n_1} - \bar{Y}_{(n_1+1):n}) - [\hat{b}_n^*\bar{Y}_{1:n_1} + (1 - \hat{b}_n^*)\bar{Y}_{(n_1+1):n}].
\end{aligned}$$

And the asymptotic variances of these three estimators are:

$$\begin{aligned}
AsyVar((1, -1)\hat{\mu}^{\mathrm{s}}) = (1, -1)V^{\mathrm{s}}(1, -1)' &= \frac{1}{\lambda}\,\mathrm{Var}[X] - \frac{2}{\lambda}\,\mathrm{Cov}(X, Y) + \frac{1}{\lambda}\,\mathrm{Var}[Y] \\
AsyVar((1, -1)\hat{\mu}^{\ell}) = (1, -1)V^{\ell}(1, -1)' &= \frac{1}{\lambda}\,\mathrm{Var}[X] - 2\,\mathrm{Cov}(X, Y) + \mathrm{Var}[Y] \\
AsyVar((1, -1)\hat{\mu}^{\mathrm{adj}}) = (1, -1)V^{\mathrm{adj}}(1, -1)' &= \frac{1}{\lambda}\,\mathrm{Var}[X] - \frac{1-\lambda}{\lambda}\frac{\mathrm{Cov}^2(X, Y)}{\mathrm{Var}[Y]} - 2\,\mathrm{Cov}(X, Y) + \mathrm{Var}[Y].
\end{aligned}$$

Now we take the differences

$$\begin{aligned}
AsyVar((1, -1)\hat{\mu}^{\mathrm{s}}) - AsyVar((1, -1)\hat{\mu}^{\mathrm{adj}}) &= \frac{1-\lambda}{\lambda}\frac{1}{\mathrm{Var}[Y]}(\mathrm{Cov}(X, Y) - \mathrm{Var}[Y])^2 \geq 0 \\
AsyVar((1, -1)\hat{\mu}^{\ell}) - AsyVar((1, -1)\hat{\mu}^{\mathrm{adj}}) &= \frac{1-\lambda}{\lambda}\frac{\mathrm{Cov}^2(X, Y)}{\mathrm{Var}[Y]} \geq 0
\end{aligned}$$

So $(1, -1)\hat{\mu}^{\mathrm{adj}}$ is a better estimator for $(\mathrm{E}[X] - \mathrm{E}[Y])$ than $(1, -1)\hat{\mu}^{\mathrm{s}}$ and $(1, -1)\hat{\mu}^{\ell}$ in the sense of having a smaller asymptotic variance. However

$$AsyVar((1,-1)\hat{\mu}^{\mathrm{s}}) - AsyVar((1,-1)\hat{\mu}^{\ell}) = \frac{1-\lambda}{\lambda}[\mathrm{Var}[Y] - 2\,\mathrm{Cov}(X,Y)],$$

which can be positive or negative or 0, depending on the values of $\mathrm{Var}[Y]$ and $\mathrm{Cov}(X,Y)$. So $(1,-1)\hat{\mu}^{\ell}$ is not necessarily a better estimator for $(\mathrm{E}[X]-\mathrm{E}[Y])$ than $(1,-1)\hat{\mu}^{\mathrm{s}}$. In particular, when the variance of X is large, the variance of Y is small, and X are highly positively correlated with Y, we can have $\sigma_Y < 2\rho\sigma_X$ and the asymptotic variance of $(1,-1)\hat{\mu}^{\mathrm{s}}$ is smaller than the asymptotic variance of $(1,-1)\hat{\mu}^{\ell}$.

2.4 Implications on Inference for the Mean

Intuitively, for a hypothesis testing problem, the quality of the test will be closely related to the properties of the estimators used in the test. In this subsection, we discuss how the asymptotic variance results above are related to the inference problem for the mean.

Suppose we are interested in the following hypothesis testing problem:

$$\begin{aligned} H_0 &: (\mathrm{E}[X], \mathrm{E}[Y])' = \mu_0 \\ H_1 &: (\mathrm{E}[X], \mathrm{E}[Y])' \neq \mu_0, \end{aligned}$$

where we are interested in testing a joint hypothesis on $\mathrm{E}[X]$ and $\mathrm{E}[Y]$, and our goal is to find a test that controls asymptotic size and has good asymptotic power. For this purpose, we make the following assumption on the set of distributions we consider.

Assumption 2.3. Let $\mathcal{F}_0$ be the family of distributions that satisfies

$$\mathcal{F}_0 = \{F : \mathrm{E}_F|\mathrm{Var}_F[X]^{-1/2}(X - \mathrm{E}_F[X])|^{2+\delta} \leq C \text{ and } \mathrm{E}_F|\mathrm{Var}_F[Y]^{-1/2}(Y - \mathrm{E}_F[Y])|^{2+\delta} \leq C\},$$

where $C < \infty$ and $\delta > 0$ are constants.

To see how asymptotic variance is related to the power of the test, we consider the following two test statistics:

$$\begin{aligned} T^{\ell}(\mu_0) &= n[\hat{\mu}^{\ell} - \mu_0]'(\hat{V}^{\ell})^{-1}[\hat{\mu}^{\ell} - \mu_0], \\ T^{\mathrm{adj}}(\mu_0) &= n[\hat{\mu}^{\mathrm{adj}} - \mu_0]'(\hat{V}^{\mathrm{adj}})^{-1}[\hat{\mu}^{\mathrm{adj}} - \mu_0], \end{aligned}$$

where $\hat{V}^{\ell} = V^{\ell} + o_p(1)$ and $\hat{V}^{\mathrm{adj}} = V^{\mathrm{adj}} + o_p(1)$ are consistent estimators for the variance covariance matrices.

Under the null hypothesis, $T^{\ell}(\mu_0)$ and $T^{\mathrm{adj}}(\mu_0)$ both converge in distribution to χ^2_2 as $n \to \infty$. So to test the hypothesis at the significance level α, we would reject when the test statistic is greater than $1-\alpha$ quantile of χ^2_2 for both tests. Under the assumptions we made, the resulting two tests both have asymptotic sizes equal to α, and both have powers tending to 1 as $n \to \infty$ against any fixed alternatives.

To further compare these two tests, we now study their behavior under the same local alternative sequence. Consider a sequence of distributions $F_n \in \mathcal{F}_0$ with

$$\sqrt{n}[(\mathrm{E}_{F_n}[X], \mathrm{E}_{F_n}[Y])' - \mu_0] \to \eta,$$

where $\eta \neq (0,0)$. So this sequence $\{F_n : n \geq 1\}$ is in the alternative space. Under this sequence, we have

$$\sqrt{n}[\hat{\mu}^{\ell} - \mu_0]) \quad \to^d \quad N(\eta, V^{\ell}),$$

and

$$\sqrt{n}[\hat{\mu}^{\mathrm{adj}} - (\theta_1, \theta_2')]) \quad \to^d \quad N(\eta, V^{\mathrm{adj}}).$$

Then the limiting distribution of the test statistic is given by

$$\begin{aligned} T^{\ell}(\mu_0) \quad &\to^d \quad \chi_2^2\left(\eta'(V^{\ell})^{-1}\eta\right) \\ T^{\mathrm{adj}}(\mu_0) \quad &\to^d \quad \chi_2^2\left(\eta'(V^{\mathrm{adj}})^{-1}\eta\right) \end{aligned}$$

where $\chi_2^2(a)$ is the non-central chi-square distribution with 2 degrees of freedom, and noncentrality parameter equals to a.

Notice that $(V^{\mathrm{adj}})^{-1} - (V^{\ell})^{-1}$ is positive semi-definite, we have

$$\eta'(V^{\mathrm{adj}})^{-1}\eta - \eta'(V^{\ell})^{-1}\eta = \eta'[(V^{\mathrm{adj}})^{-1} - (V^{\ell})^{-1}]\eta \geq 0$$

for any $\eta \in R^2$, with strict inequality for at least some $\eta \in R^2$. So under the local alternative, the limiting distribution of $T^{\mathrm{adj}}(\mu_0)$ first-order stochastically dominates the limiting distribution of $T^{\ell}(\mu_0)$. Since the critical values for both tests are the same, the test using T^{adj} has higher probability in rejecting the null under the local alternative. So the test based on $\hat{\mu}^{\mathrm{adj}}$ has better local power than the test based on $\hat{\mu}^{\ell}$. Similarly, the test based on $\hat{\mu}^{\mathrm{adj}}$ has better local power than the test based on $\hat{\mu}^{\mathrm{s}}$.

Remark 2.3. All the results in this section can be extended to the case where X and Y are random vectors.

3 Inference in Moment Inequality Models with Combined Data Sources

Now we apply the above results to moment inequality models. In particular, we consider the situation where we have different numbers of observations for different random vectors in our data set, and where some of our moment conditions only depend on the random vectors with more observations. In this situation, we have different numbers of observations for different moment conditions, and we can use the Short, the Long and the Adjusted estimators to construct sample averages for the moments. Based on these three different sample averages for moments, we consider three test statistics and use GMS to form critical values for all three tests. We then study the properties of these three inference procedures.

We now introduce the moment inequality model we consider. Assume the true value θ_0 lies in $\Theta \subset R^d$ and satisfies the moment conditions:

$$\begin{cases} \mathrm{E}_{F_0} m_j(X_i, Y_i, \theta_0) \geq 0 & j = 1, ..., p_1 \\ \mathrm{E}_{F_0} m_j(X_i, Y_i, \theta_0) = 0 & j = p_1 + 1, ..., p_1 + v_1 \\ \mathrm{E}_{F_0} m_j(Y_i, \theta_0) \geq 0 & j = p_1 + v_1 + 1, ..., p_1 + v_1 + p_2 \\ \mathrm{E}_{F_0} m_j(Y_i, \theta_0) = 0 & j = p_1 + v_1 + p_2 + 1, ..., p_1 + v_1 + p_2 + v_2 \end{cases}$$

where $m_j(\cdot, \theta) : j = 1, ..., p_1 + v_1 + p_2 + v_2$ are known real-valued moment functions, and $\{(X_i, Y_i)\} : i \geq 1$ are i.i.d. random vectors with joint distribution F_0. Denote $k_1 = p_1 + v_1$, $k_2 = p_2 + v_2$ and $k = k_1 + k_2$. So we consider the case where k_2 of our k moment inequalities depend on random vector Y_i but do not depend on random vector X_i. Denote

$$m_1^\star(X_i, Y_i, \theta_0) = (m_1(X_i, Y_i, \theta_0), ..., m_{k_1}(X_i, Y_i, \theta_0))',$$

$$m_2^\star(Y_i, \theta_0) = (m_{k_1+1}(Y_i, \theta_0), ..., m_{k_1+k_2}(Y_i, \theta_0))',$$

so $m_1^\star(X_i, Y_i, \theta_0)$ is a $k_1 \times 1$ vector that contains the values of the first k_1 moments, and $m_2^\star(Y_i, \theta_0)$ is a $k_2 \times 1$ vector that contains the values of the last k_2 moments.

As before, the observed sample is

$$\begin{aligned} &X_i : \quad i = 1, 2, ..., n_1 \\ &Y_i : \quad i = 1, 2, ..., n_1, n_1 + 1, n_1 + 2, ..., n_1 + n_2, \end{aligned}$$

where X and Y are random vectors, and we maintain the same assumptions on the data generating process and on n_1 and n_2 as in previous section.

Generic values of the parameters are denoted (θ, F), and the parameter space $\mathcal{F}$ for (θ, F) is the set of all (θ, F) that satisfy

$$\begin{array}{ll}
\text{(i)} & \theta \in \Theta, \\
\text{(ii)} & \mathrm{E}_F m_j(X_i, Y_i, \theta) \geq 0 \text{ for } j = 1, ..., p_1, \\
 & \mathrm{E}_F m_j(Y_i, \theta) \geq 0 \text{ for } j = k_1 + 1, ..., k_1 + p_2, \\
\text{(iii)} & \mathrm{E}_F m_j(X_i, Y_i, \theta) = 0 \text{ for } j = p_1 + 1, ..., p_1 + v_1, \\
 & \mathrm{E}_F m_j(Y_i, \theta) \geq 0 \text{ for } j = k_1 + p_2 + 1, ..., k, \\
\text{(iv)} & \{(X_i, Y_i) : i \geq 1\} \text{ are i.i.d. under } F, \\
\text{(v)} & \sigma^2_{F,j}(\theta) = \mathrm{Var}_F(m_j(X_i, Y_i, \theta)) \in (0, \infty) \text{ for } j = 1, ..., k_1, \\
 & \sigma^2_{F,j}(\theta) = \mathrm{Var}_F(m_j(Y_i, \theta)) \in (0, \infty) \text{ for } j = k_1 + 1, ..., k, \\
\text{(vi)} & \mathrm{Corr}_F((m_1^\star(X_i, Y_i, \theta), m_2^\star(Y_i, \theta))') \in \Psi, \\
\text{(vii)} & \mathrm{E}_F |m_j(X_i, Y_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq C \text{ for } j = 1, ..., k_1 \\
 & \mathrm{E}_F |m_j(Y_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq C \text{ for } j = k_1 + 1, ..., k
\end{array}$$

where $M < \infty$ and $\delta > 0$ are constants, and Ψ is a set of $k \times k$ correlation matrices.

We consider a confidence set (CS) for the parameter θ obtained by inverting a test. The test is based on a test statistic $T_n(\theta_0)$ for testing $H_0 : \theta = \theta_0$, and we will discuss some commonly used test statistics later.

The nominal $1 - \alpha$ CS for θ is

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_{1-\alpha}(\theta)\}$$

where $c_{1-\alpha}(\theta)$ is a critical value. Widely used method for obtaining the critical values include the PA, the subsampling and the GMS method. Andrews and Guggenberger (2009) shows that the least favorable asymptotic null distribution of the statistic $T_n(\theta)$ are those for which the moment inequalities hold as equalities, and the PA critical value takes the $1 - \alpha$ quantile of the asymptotic null distribution of $T_n(\theta)$ when the moment inequalities hold as equalities. The subsampling method uses the empirical distribution of $T_{n,b,j}(\theta)$ to approximate the distribution of $T_n(\theta)$, where $T_{n,b,j}(\theta)$ is a subsample statistic defined exactly as $T_n(\theta)$ but calculated based on the jth subsample of size b rather than the full sample. The GMS method uses some selection criterion to detect which moment inequalities are binding, and incorporate this information when approximating the distribution of $T_n(\theta)$. All these three methods yield uniformly asymptotically valid tests, but Andrews and Soares (2010) show that the GMS tests have better local power. In this paper, we use GMS critical values.

The exact confidence size of CS_n is

$$ExCS_n = \inf_{(\theta,F) \in \mathcal{F}} P_F(T_n(\theta) \leq c_{1-\alpha}(\theta)),$$

and the asymptotic confidence size is

$$AsyCS = \liminf_{n \to \infty} ExCS_n.$$

In the definition of the asymptotic confidence size, we take $\inf_{(\theta,F)\in\mathcal{F}}$ before $\lim_{n\to\infty}$, and this builds uniformity over (θ, F) into the definition of the asymptotic confidence size. Uniformity is required for the asymptotic size to give a good approximation to the finite sample size of the confidence sets.

3.1 Forms of the Sample Averages

Now we consider three constructions of the sample moment functions. Similar to the previous sections, $\bar{m}_{j,a:b}(\theta)$ denotes the sample averages of the jth moment m_j over $j = a, ..., b$. We first present the sample analogues and the asymptotic variance of $(m_1^\star(X_i, Y_i, \theta_0)', m_2^\star(X_i, Y_i, \theta_0)')'$, and those notations will be useful in displaying the three sample moment functions.

Depending on which part of the sample we are using, we have the following sample analogues for $(m_1^\star(X_i, Y_i, \theta_0)'$ and $m_2^\star(X_i, Y_i, \theta_0)')'$:

$$
\begin{aligned}
\bar{m}^\star_{1,1:n_1}(\theta) &= (\bar{m}_{1,1:n_1}(\theta), ..., \bar{m}_{k_1,1:n_1}(\theta))', \\
\bar{m}^\star_{2,1:n}(\theta) &= (\bar{m}_{k_1+1,1:n}(\theta), ..., \bar{m}_{k,1:n}(\theta))', \\
\bar{m}^\star_{2,1:n_1}(\theta) &= (\bar{m}_{k_1+1,1:n_1}(\theta), ..., \bar{m}_{k,1:n_1}(\theta))', \\
\bar{m}^\star_{2,(n_1+1):n}(\theta) &= (\bar{m}_{k_1+1,(n_1+1):n}(\theta), ..., \bar{m}_{k,(n_1+1):n}(\theta))'.
\end{aligned}
$$

Let $\Sigma(\theta)$ be the variance covariance matrix of $(m_1^\star(X_i, Y_i, \theta)', m_2^\star(Y_i, \theta_0)')'$, and

$$
\begin{aligned}
\Sigma(\theta) &= \begin{pmatrix} \mathrm{Var}(m_1^\star(X_i, Y_i, \theta)) & \mathrm{Cov}(m_1^\star(X_i, Y_i, \theta), m_2^\star(Y_i, \theta)) \\ \mathrm{Cov}(m_1^\star(X_i, Y_i, \theta), m_2^\star(Y_i, \theta))' & \mathrm{Var}(m_2^\star(Y_i, \theta)) \end{pmatrix} \\
&\equiv \begin{pmatrix} \Sigma_{11}(\theta) & \Sigma_{12}(\theta) \\ \Sigma_{21}(\theta) & \Sigma_{22}(\theta) \end{pmatrix}
\end{aligned}
$$

A consistent estimator $\hat{\Sigma}_n(\theta)$ for $\Sigma(\theta)$ is given by

$$
\begin{pmatrix} \hat{\Sigma}_{n,11}(\theta) & \hat{\Sigma}_{n,12}(\theta) \\ \hat{\Sigma}_{n,21}(\theta) & \hat{\Sigma}_{n,22}(\theta) \end{pmatrix}
$$

where

$$
\hat{\Sigma}_n(\theta) = \sum_{i=1}^{n_1} \begin{pmatrix} m_1^\star(X_i, Y_i, \theta) - \bar{m}^\star_{1,1:n_1}(\theta) \\ m_2^\star(Y_i, \theta) - \bar{m}^\star_{2,1:n}(\theta) \end{pmatrix} \begin{pmatrix} m_1^\star(X_i, Y_i, \theta) - \bar{m}^\star_{1,1:n_1}(\theta) \\ m_2^\star(Y_i, \theta) - \bar{m}^\star_{2,1:n}(\theta) \end{pmatrix}'
$$

Since more data points are available on $m_2^\star(Y_i, \theta)$, one can also use

$$
\frac{1}{n}\sum_{i=1}^{n}[m_2^\star(Y_i, \theta) - \bar{m}^\star_{2,1:n}(\theta)][m_2^\star(Y_i, \theta) - \bar{m}^\star_{2,1:n}(\theta)]'
$$

as an estimator for $\hat{\Sigma}_{n,22}(\theta)$ instead. Denote

$$
D(\theta) = \mathrm{Diag}(\Sigma(\theta))^{-1/2}, \text{ and } \hat{D}_n(\theta) = \mathrm{Diag}(\hat{\Sigma}_n(\theta))^{-1/2}.
$$

So $D(\theta)$ is a diagonal matrix of size k that consists of the inverses of the standard deviations of the moments, and $\hat{D}_n(\theta)$ is a consistent estimator for $D(\theta)$.

The Short, the Long sample moments are defined to be

$$\bar{m}_n^{\mathrm{s}}(\theta) = (\bar{m}_{1,1:n_1}(\theta), ..., \bar{m}_{k,1:n_1}(\theta))' = (\bar{m}_{1,1:n_1}^{\star}(\theta)', \bar{m}_{2,1:n_1}^{\star}(\theta)')',$$

$$\bar{m}_n^{\ell}(\theta) = (\bar{m}_{1,1:n_1}(\theta), ..., \bar{m}_{k_1,1:n_1}(\theta), \bar{m}_{k_1+1,1:n}(\theta)..., \bar{m}_{k,1:n}(\theta))' = (\bar{m}_{1,1:n_1}^{\star}(\theta)', \bar{m}_{2,1:n}^{\star}(\theta)')',$$

respectively, and the Adjusted sample moments are defined to be

$$\bar{m}_n^{\mathrm{adj}}(\theta) = \left(\bar{m}_{1,1:n}^{\mathrm{adj}\star}(\theta)', \bar{m}_{2,1:n_1}^{\star}(\theta)'\right)',$$

where

$$\bar{m}_{1,1:n}^{\mathrm{adj}\star}(\theta) = \bar{m}_{1,1:n_1}^{\star}(\theta) - \frac{n_2}{n}\hat{\Sigma}_{n,12}(\theta)\hat{\Sigma}_{n,22}^{-1}(\theta)[\bar{m}_{2,1:n_1}^{\star}(\theta) - \bar{m}_{2,(n_1+1):n}^{\star}(\theta)].$$

The Asymptotic variance of $\sqrt{n}\bar{m}_n^{\mathrm{s}}(\theta)$, $\sqrt{n}\bar{m}_n^{\ell}(\theta)$ and $\sqrt{n}\bar{m}_n^{\mathrm{adj}}(\theta)$ is denoted as $\Sigma^{\mathrm{s}}(\theta)$, $\Sigma^{\ell}(\theta)$ and $\Sigma^{\mathrm{adj}}(\theta)$, respectively. And we show that

$$\Sigma^{\mathrm{s}}(\theta) = \begin{pmatrix} \frac{1}{\lambda}\Sigma_{11}(\theta) & \frac{1}{\lambda}\Sigma_{12}(\theta) \\ \frac{1}{\lambda}\Sigma_{21}(\theta) & \frac{1}{\lambda}\Sigma_{22}(\theta) \end{pmatrix},$$

$$\Sigma^{\ell}(\theta) = \begin{pmatrix} \frac{1}{\lambda}\Sigma_{11}(\theta) & \Sigma_{12}(\theta) \\ \Sigma_{21}(\theta) & \Sigma_{22}(\theta) \end{pmatrix},$$

$$\Sigma^{\mathrm{adj}}(\theta) = \begin{pmatrix} \frac{1}{\lambda}\Sigma_{11}(\theta) - \frac{1-\lambda}{\lambda}\Sigma_{12}(\theta)\Sigma_{22}^{-1}(\theta)\Sigma_{21}(\theta) & \Sigma_{12}(\theta) \\ \Sigma_{21}(\theta) & \Sigma_{22}(\theta) \end{pmatrix}.$$

Let $\hat{\Sigma}_n^{\mathrm{s}}(\theta)$, $\hat{\Sigma}_n^{\ell}(\theta)$ and $\hat{\Sigma}_n^{\mathrm{adj}}(\theta)$ be a consistent estimator for $\Sigma^{\mathrm{s}}(\theta)$, $\Sigma^{\ell}(\theta)$ and $\Sigma^{\mathrm{adj}}(\theta)$, respectively. And we take

$$\hat{\Sigma}_n^{\mathrm{s}}(\theta) = \frac{n}{n_1}\hat{\Sigma}_n(\theta),$$

$$\hat{\Sigma}_n^{\ell}(\theta) = \begin{pmatrix} \frac{n}{n_1}\hat{\Sigma}_{n,11}(\theta) & \hat{\Sigma}_{n,12}(\theta) \\ \hat{\Sigma}_{n,21}(\theta) & \hat{\Sigma}_{n,22}(\theta) \end{pmatrix},$$

$$\hat{\Sigma}_n^{\mathrm{adj}}(\theta) = \begin{pmatrix} \frac{n}{n_1}\hat{\Sigma}_{n,11}(\theta) - \frac{n_2}{n_1}\hat{\Sigma}_{n,12}(\theta)\hat{\Sigma}_{n,22}^{-1}(\theta)\hat{\Sigma}_{n,21}(\theta) & \hat{\Sigma}_{n,12}(\theta) \\ \hat{\Sigma}_{n,21}(\theta) & \hat{\Sigma}_{n,22}(\theta) \end{pmatrix}.$$

3.2 Test Statistics

In general, the test statistics $T_n(\theta)$ is obtained using a test function $S(m, \Sigma)$. The test function $S(m, \Sigma)$ is a real function that has two parts of entries. The first part m of its entries takes value in $R_{[+\infty]}^{p_1} \times R^{v_1} \times R_{[+\infty]}^{p_2} \times R^{v_2}$, and represents a vector of possible values of the moments. The second part Σ takes value in the space of $k \times k$ variance matrix, and represents the corresponding covariance matrix of the first part m. The value of the test statistics $T_n(\theta)$ is obtained by evaluating S at a consistent estimator for the sample moments and a consistent estimator of its variance covariance matrix.

The test function S is required to satisfy Assumptions T1-T6, and here are two examples of commonly used test functions that do so. The first example of test functions is the modified method of moments (MMM) test function S_1, defined by

$$S_1(m,\Sigma)=\left(\sum_{j=1}^{p_1}+\sum_{j=k_1+1}^{k_1+p_2}\right)[m_j/\sigma_j]_-^2+\left(\sum_{j=p_1+1}^{k_1}+\sum_{j=k_1+p_2+1}^{k}\right)[m_j/\sigma_j]^2,$$

where

$$[x]_-=\begin{cases} x, & \text{if } x<0,\\ 0, & \text{if } x\geq 0,\end{cases}$$

m_j is the jth entry of m, and σ_j^2 is the jth diagonal element of Σ.

The second example of widely used test function is the Gaussian quasi-likelihood ratio (QLR) test function S_2, defined by

$$S_2(m,\Sigma)=\inf_{t=(t_1,0_{v_1},t_2,0_{v_2}):t_1\in R_{+,\infty}^{p_1},t_2\in R_{+,\infty}^{p_2}}(m-t)'\Sigma(m-t).$$

Now we construct test statistic based on the three estimators of sample moments. For $\mathrm{q}\in\{\mathrm{s},\ell,\mathrm{adj}\}$, the corresponding test statistic $T_n^{\mathrm{q}}(\theta)$ is defined to be

$$T_n^{\mathrm{q}}(\theta)=S(\sqrt{n}\bar{m}_n^{\mathrm{q}}(\theta),\hat{\Sigma}_n^{\mathrm{q}}(\theta)).$$

3.3 Generalized Moment Selection

For $\mathrm{q}\in\{\mathrm{s},\ell,\mathrm{adj}\}$, we can write

$$\begin{aligned} T_n^{\mathrm{q}}(\theta) &= S(\sqrt{n}\bar{m}_n^{\mathrm{q}}(\theta),\hat{\Sigma}_n^{\mathrm{q}}(\theta))\\ &= S(\hat{D}_n^{\mathrm{q}}(\theta)\sqrt{n}\bar{m}_n^{\mathrm{q}}(\theta),\hat{\Omega}_n^{\mathrm{q}}(\theta)),\end{aligned}$$

where

$$\hat{D}_n^{\mathrm{q}}(\theta)=\mathrm{Diag}(\hat{\Sigma}_n^{\mathrm{q}}(\theta))^{-1/2} \text{ and } \hat{\Omega}_n^{\mathrm{q}}(\theta)=\hat{D}_n^{\mathrm{q}}(\theta)\hat{\Sigma}_n^{q}(\theta)\hat{D}_n^{q}(\theta).$$

So the test statistics $T_n^q(\theta)$ depends on the corresponding normalized sample moments and sample correlation matrix.

We say a sequence of distributions $\{F_n:n\geq 1\}$ is a sequence of null distributions if $(\theta_0,F_n)\in\mathcal{F}$. Under an appropriate sequence of null distributions $\{F_n:n\geq 1\}$, let $h_{11}\in R_{+,\infty}^{p_1}$ be the limit of

$$(\sqrt{n}\mathrm{E}_{F_n}m_1(W_i,\theta_0)/\sigma_{F_n,1}(\theta_0),...,\sqrt{n}\mathrm{E}_{F_n}m_{p_1}(W_i,\theta_0)/\sigma_{F_n,p_1}(\theta_0)),$$

and let $h_{12} \in R^{p_2}_{+,\infty}$ be the limit of

$$(\sqrt{n}\mathrm{E}_{F_n} m_{k_1+1}(W_i, \theta_0)/\sigma_{F_n,k_1+1}(\theta_0), ..., \sqrt{n}\mathrm{E}_{F_n} m_{k_1+p_2}(W_i, \theta_0)/\sigma_{F_n,k_1+p_2}(\theta_0)).$$

Then the asymptotic null distribution of $T^{\mathrm{q}}_n(\theta_0)$ under this sequence $\{F_n : n \geq 1\}$ is that of

$$S((\Omega^{\mathrm{q}}_0)^{1/2} Z^* + D^{\mathrm{q}}(\theta) D^{-1}(\theta)(h'_{11}, 0'_{v_1}, h'_{12}, 0'_{v_2})', \Omega^{\mathrm{q}}_0)$$

where

$$Z^* \sim N(0_k, I_k), D^{\mathrm{q}}(\theta) = \mathrm{Diag}(\Sigma^{\mathrm{q}}(\theta))^{-1/2},$$

and Ω^{q}_0 is a $k \times k$ correlation matrix.

Notice that the asymptotic null distribution of the test statistic depends on the parameters h_{11} and h_{12}. Since h_{11}, h_{12} can not be consistently estimated, the GMS method considers replacing h_{11}, h_{12} with vectors whose values depend on a measure of the slackness of the moment inequalities. In particular, the degree of slackness of the moment inequalities is measured by

$$\xi^{\mathrm{q}}_n(\theta) = \kappa_n^{-1} \sqrt{n} \hat{D}^{\mathrm{q}}_n(\theta) \bar{m}^{\mathrm{q}}_n(\theta)$$

evaluated at $\theta = \theta_0$, where $\{\kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity at suitable rate as $n \to \infty$. Andrews and Soares (2010) recommend

$$\kappa_n = (\ln n)^{1/2}.$$

The constructions of $\xi^{\mathrm{q}}_n(\theta)$ with $\bar{m}^{\mathrm{s}}_n(\theta)$, $\bar{m}^{l}_n(\theta)$ and $\bar{m}^{\mathrm{adj}}_n(\theta)$ have similar properties, and we can use $\bar{m}^{\mathrm{adj}}_n(\theta)$ to implement GMS for all three tests. However, if a researcher uses the Short estimator to obtain sample moments, and is not aware of the Adjusted sample moments, then she is likely to use the Short estimator to measure the degree of slackness. Thus for each test, we use its corresponding estimator for the sample moments to construct ξ_n.

We replace the vector $(h'_{11}, 0'_{v_1}, h'_{12}, 0'_{v_2})'$ in the limiting distributions above by corresponding values of $\varphi(\xi^{\mathrm{q}}_n(\theta_0), \hat{\Omega}^{\mathrm{q}}_n(\theta_0))' \in R^k_{[+\infty]}$, where φ is a function that has certain properties. In particular, the value of φ_j is large when the corresponding element in $(h'_{11}, 0'_{v_1}, h'_{12}, 0'_{v_2})'$ is large, and this tells us the corresponding moment inequality should be treated as slack. And the value of φ_j is 0 when the corresponding element in $(h'_{11}, 0'_{v_1}, h'_{12}, 0'_{v_2})'$ is small, and this tells us that the corresponding moment inequality is likely to be binding or close to binding, and thus is treated as binding. An example for φ is

$$\varphi_j(\xi, \Omega) = \begin{cases} 0, & \text{if } \xi_j \leq 1 \\ \infty, & \text{if } \xi_j > 1 \end{cases}$$

for $j = 1, ..., k$. And the critical value $\hat{c}^{\mathrm{q}}_n(\theta_0, 1-\alpha)$ for $T^{\mathrm{q}}_n(\theta)$ is the $1-\alpha$ quantile of the distribution of

$$S((\hat{\Omega}^{\mathrm{q}}_n)^{1/2} Z^* + \hat{D}^{\mathrm{q}}_n(\theta) \hat{D}^{-1}_n(\theta) \varphi(\xi^{\mathrm{q}}_n(\theta_0), \hat{\Omega}^{\mathrm{q}}_n(\theta_0))', \hat{\Omega}^{\mathrm{q}}_n),$$

and it can be obtained via simulations.

3.4 Asymptotic Confidence Sizes of the Confidence Sets

So far we defined three test statistics and GMS critical value for each test. Now we study the confidence sizes of the resulting confidence sets.

Theorem 3.1. *Suppose Assumptions T1-T3, GMS1, and GMS2 hold and $0 < \alpha < 1/2$. Then the nominal level $1 - \alpha$ GMS confidence sets based on $T_n^{\mathrm{s}}(\theta)$, $T_n^{\ell}(\theta)$ and $T_n^{\mathrm{adj}}(\theta)$ satisfy the following statements:*

(a) $AsyCS \geq 1 - \alpha$.

(b) $AsyCS = 1 - \alpha$ if Assumptions GMS3, GMS4 and T7 also hold.

Theorem 3.1 shows that the resulting confidence sets are asymptotically valid in a uniform sense, and are not conservative. To compare powers of the three tests under local alternative, we study an example in Section 4. The proof of Theorem 3.1 follows the general proof of the GMS procedure.

4 Inference in Moment Inequality Models with Combined Data Sources: An Example

Although we know the limiting behavior of the three test statistics and their GMS critical values under the $n^{-1/2}$-local alternatives, it is hard to compare analytically the local powers of the three tests in general. In this section, we study a simple example where we have only two linear moment inequalities. We use this example to show how to implement the three tests we discussed, and we also study more on their local powers using numerical experiment and simulations.

Consider the following moment inequality model:

$$\begin{cases} E_{F_0}[X_i - \theta_{0,1}] \geq 0 \\ E_{F_0}[Y_i - \theta_{0,2}] \geq 0. \end{cases}$$

And again, we suppose that we have different numbers of observations for X and Y as in the previous sections. To simplify notations, in this example we also assume that the variances of X and Y are 1. We are interested in confidence sets for the true parameter value θ_0, and our goal is to construct confidence sets for identifiable parameters that are uniformly consistent in levels.

Generic values of the parameters are denoted by (θ, F), where $\theta = (\theta_1, \theta_2) \in R^2$. The parameter space $\mathcal{F}$ in this example is the set of all the (θ, F) that satisfy

$$\begin{array}{rl} \text{(i)} & E_F[X_i] \geq \theta_1, E_F[Y_i] \geq \theta_2 \\ \text{(ii)} & \{(X_i, Y_i) : i \geq 1\} \text{ are i.i.d. under F} \\ \text{(iii)} & \mathrm{Var}_F[X_i] = \mathrm{Var}_F[Y_i] = 1 \\ \text{(v)} & E_F|X_i - \theta_1|^{2+\delta} \leq C, E_F|Y_i - \theta_2|^{2+\delta} \leq C \end{array}$$

We denote $\mathrm{Corr}_F(X_i, Y_i)$ as ρ, and the dependence of ρ on F is suppressed.

4.1 Construction of the Confidence Sets

In this example, we use the MMM test function, i.e. S_1 to construct test statistics.

GMS based on the Short sample moments

The Short sample moments in this example are

$$\bar{m}_n^{\mathrm{s}}(\theta) = (\bar{X}_{1:n_1} - \theta_1, \bar{Y}_{1:n_1} - \theta_2)'.$$

When F is the true distribution that generates the data, by CLT we have

$$\sqrt{n} \begin{pmatrix} \bar{X}_{1:n_1} - \mathrm{E}_F[X] \\ \bar{Y}_{1:n_1} - \mathrm{E}_F[Y] \end{pmatrix} \to^d Z^{\mathrm{s}} \sim N\left(0, \Sigma^{\mathrm{s}}\right),$$

where

$$\Sigma^{\mathrm{s}} = \begin{pmatrix} \frac{1}{\lambda} & \frac{\rho}{\lambda} \\ \frac{\rho}{\lambda} & \frac{1}{\lambda} \end{pmatrix}.$$

The MMM test statistic based on the Short sample moments is

$$T_n^{\mathrm{s}}(\theta) = [(\bar{X}_{1:n_1} - \theta_1)/\hat{\sigma}_1^{\mathrm{s}}]_-^2 + [(\bar{Y}_{1:n_1} - \theta_2)/\hat{\sigma}_2^{\mathrm{s}}]_-^2,$$

where $\hat{\sigma}_1^{\mathrm{s}} = (\hat{\Sigma}_{11}^{\mathrm{s}})^{1/2}$, $\hat{\sigma}_2^{\mathrm{s}} = (\hat{\Sigma}_{22}^{\mathrm{s}})^{1/2}$, and $\hat{\Sigma}^{\mathrm{s}}$ is a consistent estimator for Σ^{s}. Under a sequence of distributions $\{F_n : n \geq 1\}$ in the null such that

$$\sqrt{n}(\mathrm{E}_{F_n}[X] - \theta_1, \mathrm{E}_{F_n}[Y] - \theta_2)' \to h_1 \in R^2_{+,\infty},$$

we have

$$\begin{aligned} \sqrt{n}\bar{m}_n^{\mathrm{s}}(\theta) &= \sqrt{n} \begin{pmatrix} \bar{X}_{1:n_1} - \mathrm{E}_{F_n}[X] \\ \bar{Y}_{1:n_1} - \mathrm{E}_{F_n}[Y] \end{pmatrix} + \sqrt{n} \begin{pmatrix} \mathrm{E}_{F_n}[X] - \theta_1 \\ \mathrm{E}_{F_n}[Y] - \theta_2 \end{pmatrix} \\ &\to^d Z^{\mathrm{s}} + h_1 \\ &\sim N\left(h_1, \Sigma^{\mathrm{s}}\right). \end{aligned}$$

So the asymptotic null distribution of $T_n^{\mathrm{s}}(\theta)$ under this sequence is that of

$$S_1(Z^{\mathrm{s}} + h_1, \Sigma^{\mathrm{s}}) = \lambda[Z_1^{\mathrm{s}} + h_{1,1}]_-^2 + \lambda[Z_2^{\mathrm{s}} + h_{1,2}]_-^2.$$

Since h_1 can not be estimated consistently, the GMS method replaces h_1 by φ, where

$$\varphi_j(\xi) = \begin{cases} 0 & \text{if } \xi_j \le 1 \\ \infty & \text{if } \xi_j > 1. \end{cases}$$

And we use $\bar{m}_n^{\mathrm{s}}$ to construct the degree of slackness of the moment inequalities:

$$\xi^{\mathrm{s}}(\theta) = \kappa_n^{-1}\sqrt{n}\bar{m}_n^{\mathrm{s}}(\theta) = \left(\kappa_n^{-1}\sqrt{n}(\bar{X}_{1:n_1} - \theta_1), \kappa_n^{-1}\sqrt{n}(\bar{Y}_{1:n_1} - \theta_2)\right)',$$

where $\kappa_n = \sqrt{\ln n}$. The critical value $c_n^{\mathrm{s}}(\theta, 1-\alpha)$ is defined to be the $1-\alpha$ quantile of

$$[(Z_1^{S*} + \varphi_1(\xi^{\mathrm{s}}))/\hat{\sigma}_1^{\mathrm{s}}]_-^2 + [(Z_2^{S*} + \varphi_2(\xi^{\mathrm{s}}))/\hat{\sigma}_2^{\mathrm{s}}]_-^2,$$

where $Z^{S*} \sim N(0, \hat{\Sigma}^{\mathrm{s}})$, and is independent of our data set.

GMS based on the Long sample moments

The Long sample moments in this example are

$$\bar{m}_n^{\ell}(\theta) = (\bar{X}_{1:n_1} - \theta_1, \bar{Y}_{1:n} - \theta_2)'.$$

When F is the true distribution that generates the data, by CLT we have

$$\sqrt{n}\begin{pmatrix} \bar{X}_{1:n_1} - \mathrm{E}_F[X] \\ \bar{Y}_{1:n} - \mathrm{E}_F[Y] \end{pmatrix} \to^d Z^{\ell} \sim N\left(0, \Sigma^{\ell}\right),$$

where

$$\Sigma^{\ell} = \begin{pmatrix} \frac{1}{\lambda} & \rho \\ \rho & 1 \end{pmatrix}.$$

The MMM test statistic based on the Long sample moments is

$$T_n^{\ell}(\theta) = [(\bar{X}_{1:n_1} - \theta_1)/\hat{\sigma}_1^{\ell}]_-^2 + [(\bar{Y}_{1:n} - \theta_2)/\hat{\sigma}_2^{\ell}]_-^2,$$

where $\hat{\sigma}_1^{\ell} = (\hat{\Sigma}_{11}^{\ell})^{1/2}$, $\hat{\sigma}_2^{\ell} = (\hat{\Sigma}_{22}^{\ell})^{1/2}$, and $\hat{\Sigma}^{\ell}$ is a consistent estimator for Σ^{ℓ}. Under a sequence of distributions $\{F_n : n \ge 1\}$ in the null such that

$$\sqrt{n}(\mathrm{E}_{F_n}[X] - \theta_1, \mathrm{E}_{F_n}[Y] - \theta_2)' \to h_1 \in R_{+,\infty}^2,$$

we have

$$\begin{aligned}
\sqrt{n}\bar{m}_n^{\ell}(\theta) &= \sqrt{n}\begin{pmatrix}\bar{X}_{1:n_1} - \mathrm{E}_{F_n}[X] \\ \bar{Y}_{1:n} - \mathrm{E}_{F_n}[Y]\end{pmatrix} + \sqrt{n}\begin{pmatrix}\mathrm{E}_{F_n}[X] - \theta_1 \\ \mathrm{E}_{F_n}[Y] - \theta_2\end{pmatrix} \\
&\to^d Z^{\ell} + h_1 \\
&\sim N\left(h_1, \Sigma^{\ell}\right).
\end{aligned}$$

So the asymptotic null distribution of $T_n^{\ell}(\theta)$ under this sequence is that of

$$S_1(Z^{\ell} + h_1, \Sigma^{\ell}) = \lambda[Z_1^{\ell} + h_{1,1}]_-^2 + [Z_2^{\ell} + h_{1,2}]_-^2.$$

With the Long sample moments, we use $\bar{m}_n^{\ell}(\theta)$ to construct the degree of slackness of the moment inequalities, and

$$\xi^{\ell}(\theta) = \kappa_n^{-1}\sqrt{n}\bar{m}_n^{\ell}(\theta) = \left(\kappa_n^{-1}\sqrt{n}(\bar{X}_{1:n_1} - \theta_1), \kappa_n^{-1}\sqrt{n}(\bar{Y}_{1:n} - \theta_2)\right)'.$$

The critical value $c_n^{\ell}(\theta_1, \theta_2, 1-\alpha)$ is defined to be the $1-\alpha$ quantile of

$$[(Z_1^{L*} + \varphi_1(\xi^{\ell}))/\hat{\sigma}_1^{\ell}]_-^2 + [(Z_2^{L*} + \varphi_2(\xi^{\ell}))/\hat{\sigma}_2^{\ell}]_-^2,$$

where $Z^{L*} \sim N(0, \hat{\Sigma}^{\ell})$, and is independent of our data set.

GMS based on the Adjusted sample moments

Now we implement the adjusted sample moments using the regression method. Consider the following linear regression

$$X_i - \theta_1 = \beta_0 + \beta_1(Y_i - \theta_2) + u_i, \quad i \in \{1, 2, ..., n_1\}.$$

The OLS estimator is given by:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n_1}(Y_i - \theta_2)(X_i - \theta_1) - n_1(\bar{Y}_{1:n_1} - \theta_2)(\bar{X}_{1:n_1} - \theta_1)}{\sum_{i=1}^{n_1}(Y_i - \theta_2)^2 - n_1(\bar{Y}_{1:n_1} - \theta_2)^2}.$$

Based on this regression, let

$$\bar{m}_n^{\mathrm{adj}}(\theta) = \begin{pmatrix}(\bar{X}_{1:n_1} - \theta_1) - \frac{n_2}{n}\hat{\beta}_1[\bar{Y}_{1:n_1} - \bar{Y}_{(n_1+1):n}] \\ \bar{Y}_{1:n} - \theta_2\end{pmatrix} = \begin{pmatrix}(\bar{X}_{1:n_1} - \theta_1) + \hat{\beta}_1[\bar{Y}_{1:n} - \bar{Y}_{1:n_1}] \\ \bar{Y}_{1:n} - \theta_2\end{pmatrix}$$

be the Adjusted sample moments. Notice that $\hat{\beta}_1 \to^p \rho$ by LLN, when F is the true distribution that generates the data, by CLT we have

$$
\begin{aligned}
&\sqrt{n}\begin{pmatrix} \bar{X}_{1:n_1} + \hat{\beta}_1[\bar{Y}_{1:n} - \bar{Y}_{1:n_1}] - \mathrm{E}_F[X] \\ \bar{Y}_{1:n} - \mathrm{E}_F[Y] \end{pmatrix} \\
= \quad & \sqrt{n}\begin{pmatrix} (\bar{X}_{1:n_1} - \mathrm{E}_F[X]) + \hat{\beta}_1(\bar{Y}_{1:n} - E_F[Y]) - \hat{\beta}_1(\bar{Y}_{1:n_1} - E_F[Y]) \\ \bar{Y}_{1:n} - \mathrm{E}_F[Y] \end{pmatrix} \\
\to^d \quad & Z^{\text{adj}} \sim N\left(0, \Sigma^{\text{adj}}\right),
\end{aligned}
$$

where

$$
\Sigma^{\text{adj}} = \begin{pmatrix} \frac{1-(1-\lambda)\rho^2}{\lambda} & \rho \\ \rho & 1 \end{pmatrix}.
$$

The MMM test statistic based on the Adjusted sample moments is

$$
T_n^{\text{adj}}(\theta) = [(\bar{m}_{n,1}^{\text{adj}}(\theta_1, \theta_2) - \theta_1)/\hat{\sigma}_1^{\text{adj}}]_-^2 + [(\bar{Y}_{1:n} - \theta_2)/\hat{\sigma}_2^{\text{adj}}]_-^2,
$$

where $\hat{\sigma}_1^{\text{adj}} = (\hat{\Sigma}_{11}^{\text{adj}})^{1/2}$, $\hat{\sigma}_2^{\text{adj}} = (\hat{\Sigma}_{22}^{\text{adj}})^{1/2}$, and $\hat{\Sigma}^{\text{adj}}$ is a consistent estimator for Σ^{adj}. Under a sequence of distributions $\{F_n : n \geq 1\}$ in the null such that

$$
\sqrt{n}(\mathrm{E}_{F_n}[X] - \theta_1, \mathrm{E}_{F_n}[Y] - \theta_2)' \to h_1 \in R_{+,\infty}^2,
$$

we have

$$
\begin{aligned}
\sqrt{n}\bar{m}_n^{\text{adj}}(\theta) \quad &= \quad \sqrt{n}\begin{pmatrix} \bar{X}_{1:n_1} + \hat{\beta}_1[\bar{Y}_{1:n} - \bar{Y}_{1:n_1}] - \mathrm{E}_{F_n}[X] \\ \bar{Y}_{1:n} - \mathrm{E}_{F_n}[Y] \end{pmatrix} + \sqrt{n}\begin{pmatrix} \mathrm{E}_{F_n}[X] - \theta_1 \\ \mathrm{E}_{F_n}[Y] - \theta_2 \end{pmatrix} \\
&\to \quad Z^{\text{adj}} + h_1 \\
&\sim \quad N\left(h_1, \Sigma^{\text{adj}}\right).
\end{aligned}
$$

So the asymptotic null distribution of $T_n^{\text{adj}}(\theta_1, \theta_2)$ under this sequence is that of

$$
S_1(Z^{\text{adj}} + h_1, \Sigma^{\text{adj}}) = \frac{\lambda}{1-(1-\lambda)\rho^2}[Z_1^{\text{adj}} + h_{1,1}]_-^2 + [Z_2^{\text{adj}} + h_{1,2}]_-^2.
$$

Here, we use $\bar{m}_n^{\text{adj}}(\theta)$ to construct the degree of slackness of the moment inequalities, and

$$
\xi^{\text{adj}}(\theta) = \kappa_n^{-1}\sqrt{n}\bar{m}_n^{\text{adj}}(\theta).
$$

The critical value $c_n^{\text{adj}}(\theta_1, \theta_2, 1-\alpha)$ is defined to be the $1-\alpha$ quantile of

$$
[(Z_1^{A*} + \varphi_1(\xi^{\text{adj}}))/\hat{\sigma}_1^{\text{adj}}]_-^2 + [(Z_2^{A*} + \varphi_2(\xi^{\text{adj}}))/\hat{\sigma}_2^{\text{adj}}]_-^2,
$$

where $Z^{A*} \sim N(0, \hat{\Sigma}^{\text{adj}})$, and is independent of our data set.

Notice that these three sample moments are all consistent and asymptotically normal estimators for the moment conditions. Moreover, the asymptotic distributions of the three sample moments are centered at the same h_1, but with different variance covariance matrices Σ^{s}, Σ^{ℓ} and Σ^{adj}, respectively. Moreover, $\Sigma^{\mathrm{s}} - \Sigma^{\mathrm{adj}}$ and $\Sigma^{\ell} - \Sigma^{\mathrm{adj}}$ are positive semi-definite, so the Adjusted sample moments have the smallest asymptotic variance covariance matrix among all three constructions of sample moments.

4.2 Comparison of Power under the Local Alternatives

From the theorems in the previous section, we know that each of the three tests leads to a confidence set that has asymptotic size $1-\alpha$ and is not conservative. In this subsection, we study the powers of the three tests against $n^{-1/2}$-local alternatives. First, we define the $n^{-1/2}$-local alternatives we consider here.

For given $\theta_{n,*}$, we consider the tests of

$$\begin{aligned} &H_0 : \mathrm{E}_{F_n}[(X,Y)'] - \theta_{n,*} \geq 0 \\ &H_1 : H_0 \text{ does not hold} \end{aligned}$$

where F_n denotes the true distribution that generates the data we observe.

Assumption 4.1 (Local Alternative)**.** Suppose the data is generated with true parameters $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ that satisfy

(a) $\sqrt{n}(\mathrm{E}_{F_n}[X] - \theta_{n,1}, \mathrm{E}_{F_n}[Y] - \theta_{n,2})' \to h_1 \in R^2_{+,\infty}$.

(b) $\theta_n = \theta_{n,*} - \eta n^{-1/2}(1+o(1))$ for some $\eta \in R^2$. More over, $\theta_{n,*} \to \theta_0$ and $F_n \to F_0$ for some $(\theta_0, F_0) \in \mathcal{F}$.

(c) $\sup_{n\geq 1} \mathrm{E}_{F_n} |(X_i - \theta_{n,*,1})/\sigma_{F_n,X}(\theta_{n,*})|^{2+\delta} < \infty$, $\sup_{n\geq 1} \mathrm{E}_{F_n} |(Y_i - \theta_{n,*,2})/\sigma_{F_n,Y}(\theta_{n,*})|^{2+\delta} < \infty$

(d) $h_{1,1} - \eta_1 < 0$ or $h_{1,2} - \eta_2 < 0$.

This assumption specifies that the true sequence $\{\theta_n\}$ are local to the null sequence $\{\theta_{n,*}\}$, and the true distribution is in the alternative when n is large. To study the behavior of the GMS tests, we also need the following assumption.

Assumption 4.2. Assume $\kappa_n^{-1} n^{1/2} \mathrm{E}_{F_n}[X_i - \theta_{n,1}] \to \pi_{1,1}$ and $\kappa_n^{-1} n^{1/2} \mathrm{E}_{F_n}[Y_i - \theta_{n,2}] \to \pi_{1,2}$, where $\pi_{1,1}, \pi_{1,2} \in R_{+,\infty}$.

Under the assumptions on $n^{-1/2}$-local alternatives above, the parameters satisfy

$$h_{1,1} - \eta_1 < 0 \text{ or } h_{1,2} - \eta_2 < 0,$$

and the limiting distributions of $T_n^{\mathrm{s}}(\theta_1,\theta_2)$, $T_n^{\ell}(\theta_1,\theta_2)$ and $T_n^{\mathrm{adj}}(\theta_1,\theta_2)$ are those of

$$S_1(Z^{\mathrm{s}} + h_1 - \eta, \Sigma^{\mathrm{s}}), \quad S_1(Z^{\ell} + h_1 - \eta, \Sigma^{\ell}), \text{ and } S_1(Z^{\mathrm{adj}} + h_1 - \eta, \Sigma^{\mathrm{adj}}),$$

respectively. And the GMS critical values $c_n^{\mathrm{s}}(\theta_1, \theta_2, 1-\alpha)$, $c_n^{\ell}(\theta_1, \theta_2, 1-\alpha)$, $c_n^{\mathrm{adj}}(\theta_1, \theta_2, 1-\alpha)$ converges in probability to the $1-\alpha$ quantile of

$$S_1(Z^{\mathrm{s}} + \varphi(\pi_1), \Sigma^{\mathrm{s}}), \quad S_1(Z^{\ell} + \varphi(\pi_1), \Sigma^{\ell}), \text{ and } S_1(Z^{\mathrm{adj}} + \varphi(\pi_1), \Sigma^{\mathrm{adj}}),$$

respectively.[2]

Consider a particular type of sequence in the local alternative space that satisfies

- $h_{1,1} - \eta_1 < 0$ and $\varphi_1 = 0$,
- and $h_{1,2} - \eta_2 = +\infty$ and $\varphi_2 = +\infty$.

This describes the case where

- the first moment inequality is locally violated, and GMS treats it as binding
- the second moment inequality is slack, and GMS correctly detects that it is slack.

Under this sequence, the second moment inequality doesn't contribute to the test statistic or the critical value. In particular, the GMS critical value of the test based on the Short sample converges in probability to the $1-\alpha$ quantile of

$$\begin{aligned} S_1(Z^{\mathrm{s}} + \varphi(\pi_1), \Sigma^{\mathrm{s}}) &= \lambda[Z_1^{\mathrm{s}} + \varphi_1(\pi_1)]_-^2 + \lambda[Z_2^{\mathrm{s}} + \varphi_2(\pi_1)]_-^2 \\ &= [\sqrt{\lambda} Z_1^{\mathrm{s}}]_-^2 \\ &= [Z]_-^2, \end{aligned}$$

where Z has the standard normal distribution $N(0,1)$. Notice that this distribution doesn't depend on the matrix Σ^{s}, as long as Σ_{11}^{s} is positive. We can apply the same steps as above and show that the GMS critical values of the tests based on the Long sample mean and based on the Adjusted sample mean also converge to the $1-\alpha$ quantile of $[Z]_-^2$.

Now we study the asymptotic distributions of our test statistics under this sequence. The asymptotic distribution of $T_n^{\mathrm{s}}(\theta_1, \theta_2)$ is that of

$$\begin{aligned} S_1(Z^{\mathrm{s}} + h_1 - \eta, \Sigma^{\mathrm{s}}) &= \lambda[Z_1^{\mathrm{s}} + h_{1,1} - \eta_1]_-^2 + \lambda[Z_2^{\mathrm{s}} + h_{1,2} - \eta_2]_-^2 \\ &= \lambda[Z_1^{\mathrm{s}} + h_{1,1} - \eta_1]_-^2 \\ &= [\sqrt{\lambda} Z_1^{\mathrm{s}} + \sqrt{\lambda}(h_{1,1} - \eta_1)]_-^2 \\ &= [Z + \sqrt{\lambda}(h_{1,1} - \eta_1)]_-^2 \end{aligned}$$

Notice that this distribution depends on Σ^{s} only through Σ_{11}^{s}. Similarly, the asymptotic distribution of $T_n^{\ell}(\theta_1, \theta_2)$ is that of

$$[Z + \sqrt{\lambda}(h_{1,1} - \eta_1)]_-^2$$

[2]This is true everywhere except when some of the elements of π are 1, which happens with probability 0. We are making a simplification here to give the main idea behind the local power comparisons.

and the asymptotic distribution of $T_n^{\mathrm{adj}}(\theta_1, \theta_2)$ is that of

$$[Z + \frac{\sqrt{\lambda}}{\sqrt{1-(1-\lambda)\rho^2}}(h_{1,1} - \eta_1)]_-^2$$

Since

$$\frac{\sqrt{\lambda}}{\sqrt{1-(1-\lambda)\rho^2}} \geq \sqrt{\lambda} > 0,$$

we have

$$\frac{\sqrt{\lambda}}{\sqrt{1-(1-\lambda)\rho^2}}(h_{1,1} - \eta_1) \leq \sqrt{\lambda}(h_{1,1} - \eta_1) \leq 0$$

So for a fixed value z,

$$z + \frac{\sqrt{\lambda}}{\sqrt{1-(1-\lambda)\rho^2}}(h_{1,1} - \eta_1) \leq z + \sqrt{\lambda}(h_{1,1} - \eta_1)$$

and

$$[z + \frac{\sqrt{\lambda}}{\sqrt{1-(1-\lambda)\rho^2}}(h_{1,1} - \eta_1)]_-^2 \geq [z + \sqrt{\lambda}(h_{1,1} - \eta_1)]_-^2.$$

Thus, the asymptotic distribution of $T_n^{\mathrm{adj}}(\theta_1, \theta_2)$ first order stochastically dominates the asymptotic distributions of $T_n^{\mathrm{s}}(\theta_1, \theta_2)$ and $T_n^{\ell}(\theta_1, \theta_2)$. And the GMS test based on the adjusted sample mean has better power against this particular type of local sequence.

Intuitively, the Adjusted sample moments estimate the first moment condition with a smaller asymptotic variance, so when the test statistics and the critical values do not depend on the second moment condition, the GMS test based on the Adjusted sample moments have better power against the same local alternative.

Similarly, along a particular type of sequence in the alternative that satisfies

- $h_{1,1} - \eta_1 = +\infty$ and $\varphi_1 = +\infty$,
- $h_{1,2} - \eta_2 < 0$ and $\varphi_2 = 0$,

the GMS test based on the Adjusted sample moments have the same power as the GMS test based on the Long sample moments, and both these two tests have better power than the GMS test based on the Short sample moments.

In general, it is hard to compare the local powers of these three tests analytically, and we then consider comparing their local powers using simulation studies.

4.3 Numerical Experiment of the Local Powers

In this numerical study, we simulate the powers of these tests from their limiting distributions. We simulate rejection probabilities of these three tests at nominal level $\alpha = 0.05$ for different combinations of values for ρ and λ with $100,000$ draws from the limiting distributions. For each combination of ρ and λ, we simulate rejection probabilities of these three tests under the following local parameter values:

- $h_{1,2} - \eta_2 = -4, -1, 1, 4, 10000, 10000$, $h_{1,1} - \eta_1$ takes value on a grid on $[-4, 0]$. The critical values are generated with $\pi = (0,0)'$ except the second time $h_{1,2} - \eta_2 = 10000$. The first time $h_{1,2} - \eta_2 = 10000$ corresponds to the case where the second moment condition is slack but GMS faills to detect it, and the second time $h_{1,2} - \eta_2 = 10000$ corresponds to the case where the second moment condition is slack and GMS detects it and sets $\pi_2 = \infty$.

- $h_{1,1} - \eta_1 = -4, -1, 1, 4, 10000, 10000$, $h_{1,2} - \eta_2$ takes value on a grid on $[-4, 0]$. The critical values are generated with $\pi = (0,0)'$ except the second time $h_{1,1} - \eta_1 = 10000$.

Whenever a component of $h_1 - \eta$ is smaller than 0, the corresponding moment condition is violated.

We take $\rho = 0.8, 0.5, 0.2$ and $\lambda = 0.2, 0.8$, and we simulate power for each combination of values for ρ and λ. In general, when $|\rho|$ is larger, X_i and Y_i are more correlated, and having extra observations of Y_i will be more helpful in estimating the mean of X_i, and we expect the test based on the Adjusted sample moments to have more significant power advantage over the other two tests. Also, when λ is smaller, the number of observations of X_i is a smaller fraction of the number of observations of Y_i in the limit, and there will be relatively more information in the extra observations of Y_i, and we expect the test based on the Adjusted sample moments to have more significant power advantage over the other two tests. This is true in our simulation.

In our simulations, we find that the powers of the test based on the Adjusted sample moments and the test based on the Long sample moments are higher than the power of the test based on the Short sample moments in almost all cases. We also find that the power of test based on Adjusted sample moments is higher than the power of the test the test based on the Long sample moments when $|\rho|$ is large, λ is small and $h_{1,1} - \eta_1 < 0$, and are close in other cases.

The power curves corresponding to $\rho = 0.8$ and $\lambda = 0.2$ are plotted in Figure 1 and Figure 2, and the test based on the Adjusted sample moments has significantly higher power than the other two tests in Figure 1. The power curves corresponding to $\rho = 0.5$ and $\lambda = 0.8$ are plotted in Figure 3 and Figure 4, and the test based on the Adjusted sample moments has power that is close to the test based on the Long sample moments, and is better than the other two tests for some combinations of local parameter values. We find similar patterns in power comparisons for the cases in between. For a combination of higher λ and lower ρ, the power advantage is very small.

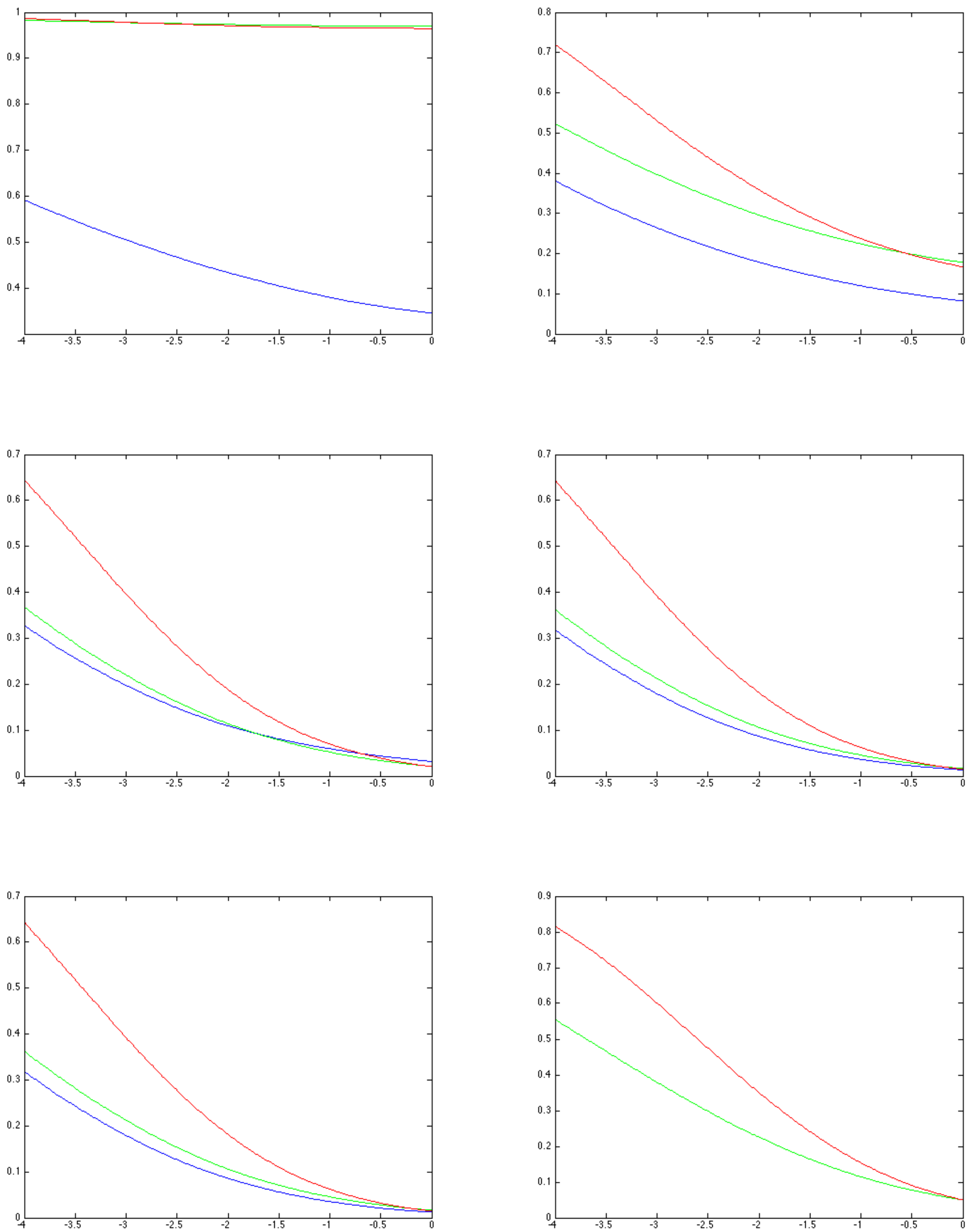

Figure 1: Power curves of the three tests with $\rho = 0.8, \lambda = 0.2, h_{1,2} - \eta_2 = -4, -1, 1, 4, \infty, \infty$. Power curves of the test based on the Short sample moments are in blue, power curves of the test based on the Long sample moments are in green, and power curves of the test based on the Adjusted sample moments are in red.

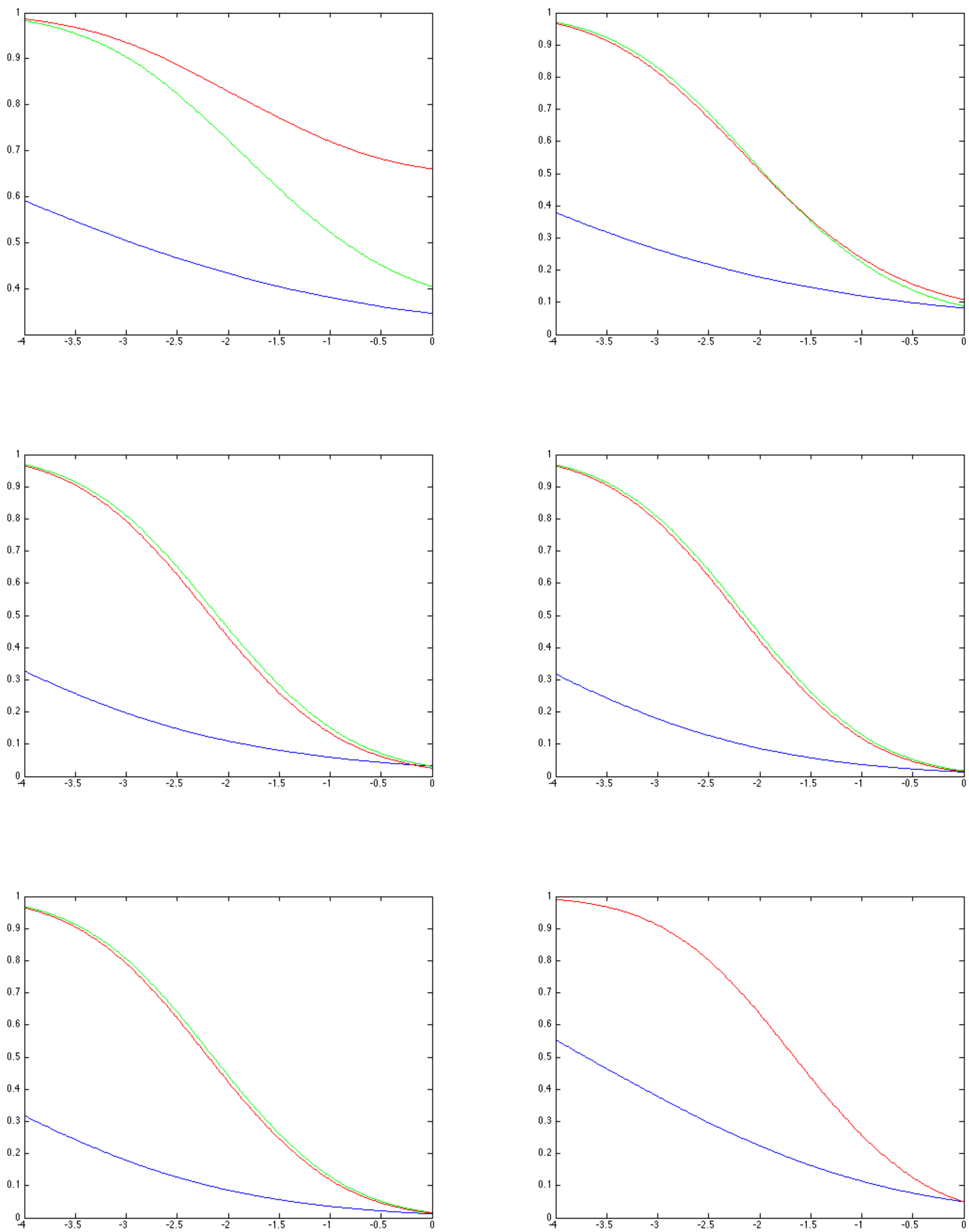

Figure 2: Power curves of the three tests with $\rho = 0.8, \lambda = 0.2, h_{1,1} - \eta_1 = -4, -1, 1, 4, \infty, \infty$. Power curves of the test based on the Short sample moments are in blue, power curves of the test based on the Long sample moments are in green, and power curves of the test based on the Adjusted sample moments are in red.

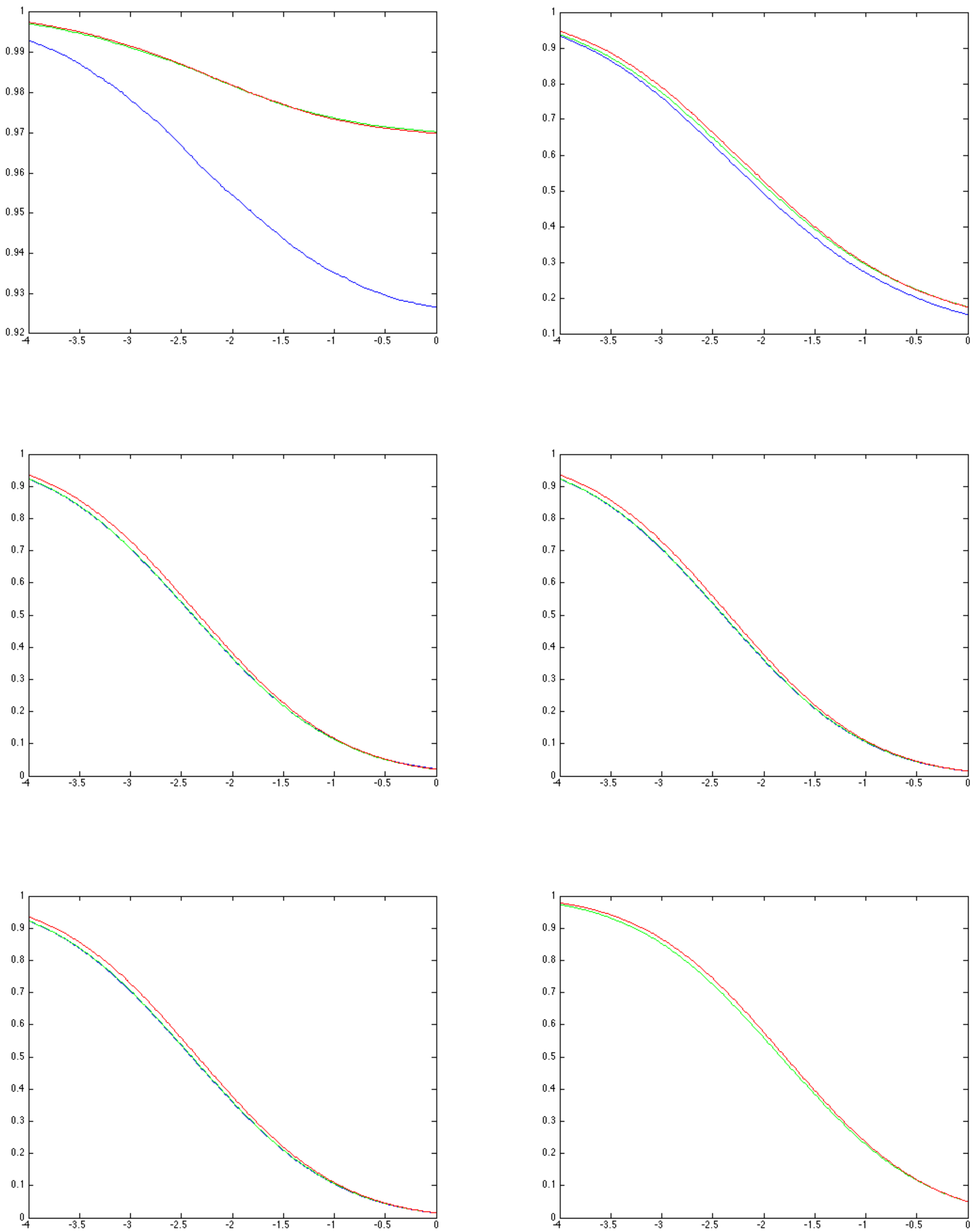

Figure 3: Power curves of the three tests with $\rho = 0.5, \lambda = 0.8, h_{1,2} - \eta_2 = -4, -1, 1, 4, \infty, \infty$. Power curves of the test based on the Short sample moments are in blue, power curves of the test based on the Long sample moments are in green, and power curves of the test based on the Adjusted sample moments are in red.

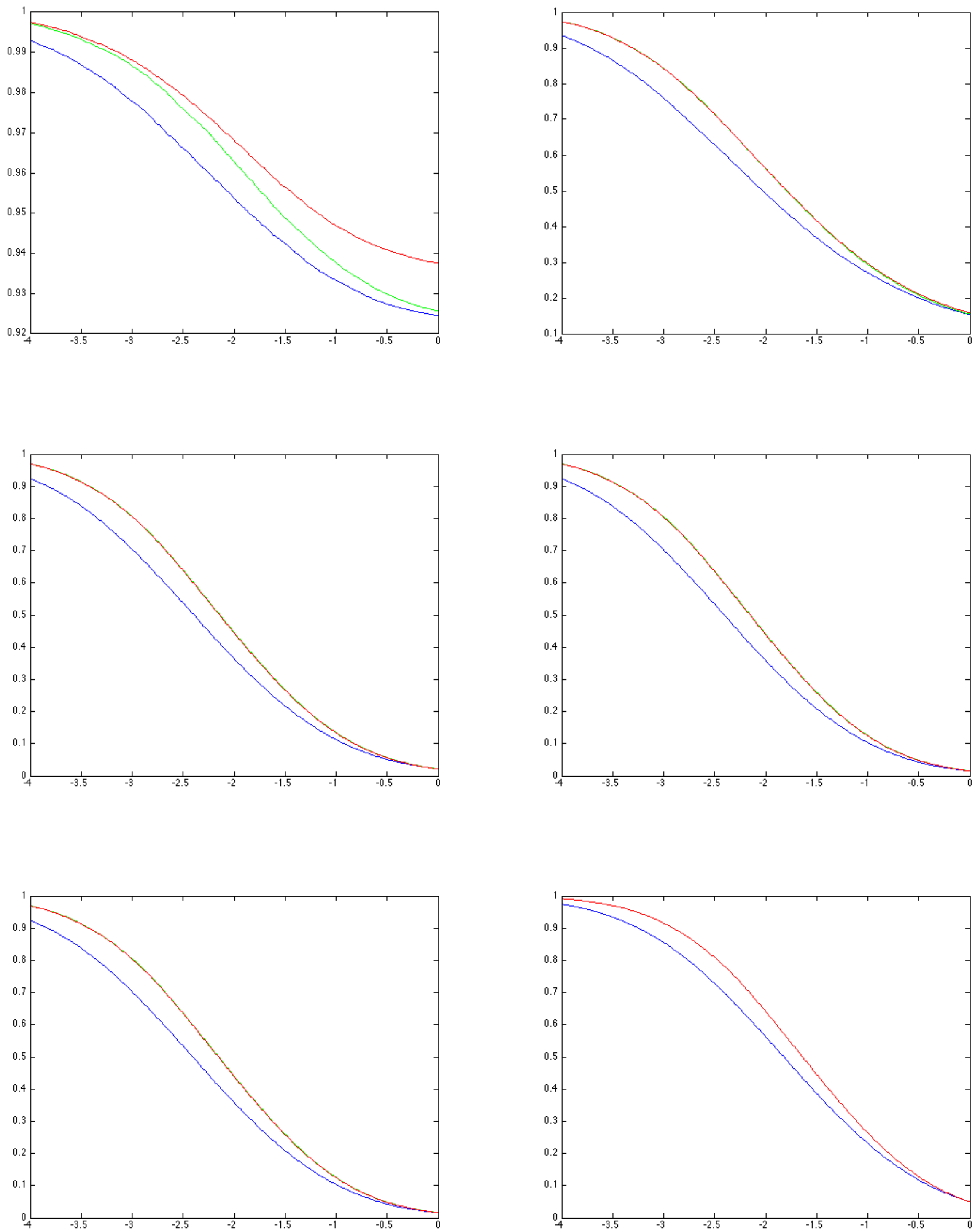

Figure 4: Power curves of the three tests with $\rho = 0.5, \lambda = 0.8, h_{1,1} - \eta_1 = -4, -1, 1, 4, \infty, \infty$. Power curves of the test based on the Short sample moments are in blue, power curves of the test based on the Long sample moments are in green, and power curves of the test based on the Adjusted sample moments are in red.

Appendix 1 Assumptions

1.1 Test Statistics Assumptions

This subsection contains the assumptions we impose on the test function S, and they are maintained from Andrews and Soares (2010).

Assumption T1:

(a) $S((m_{11}, m_{12}, m_{21}, m_{22}), \Sigma)$ is nonincreasing in m_{11} and m_{21} for all $m_{11} \in R^{p_1}$, $m_{12} \in R^{v_1}$, $m_{21} \in R^{p_2}$, $m_{22} \in R^{v_2}$ and variance matrix $\Sigma \in R^{k \times k}$.

(b) $S(m, \Sigma) = S(Dm, D\Sigma D)$ for all $m \in R^k$, $\Sigma \in R^{k \times k}$, and positive definite $D \in R^{k \times k}$.

(c) $S(m, \Omega) \geq 0$ for all $m \in R^k$ and $\Omega \in \Psi$.

(d) $S(m, \Omega)$ is continuous at all $m \in R^p_{[+\infty]} \times R^v$ and $\Omega \in \Psi$.

Assumption T2: For all $h_{11} \in R^{p_1}_{[+,\infty]}$, all $h_{12} \in R^{p_2}_{[+,\infty]}$, all $\Omega \in \Psi$ and $Z \sim N(0_k, \Omega)$, the degree of freedom of $S(Z + (h_{11}, 0_{v_1}, h_{12}, 0_{v_2}), \Omega)$ at $x \in R$ is

(a) continuous for $x > 0$,

(b) strictly increasing for $x > 0$ unless $v_1 = v_2 = 0$ and $h_{11} = \infty^{p_1}, h_{12} = \infty^{p_2}$,

(c) less than or equal to $1/2$ at $x = 0$ whenever $v \geq 1$ or $h_{11} = 0_{p_1}, h_{12} = 0_{p_2}$.

Assumption T3: $S(m, \Omega) > 0$ if and only if $m_j < 0$ for some $j = 1, ..., p_1, k_1 + 1, ..., k_1 + p_2$ or $m_j \neq 0$ for some $j = p_1 + 1, ..., k_1, k_1 + p_2 + 1, ..., k$, where $m = (m_1, ..., m_k)'$ and $\Omega \in \Psi$.

Assumption T4:

(a) The degree of freedom of $S(Z, \Omega)$ is continuous at its $(1 - \alpha)$ quantile, $c(\Omega, 1 - \alpha)$, for all $\Omega \in \Psi$, where $Z \sim N(0_k, \Omega)$ and $\alpha \in (0, 1/2)$.

(b) $c(\Omega, 1 - \alpha)$ is continuous in Ω uniformly for $\Omega \in \Psi$.

Assumption T5:

(a) For all $g \in R^{p_1}_{[+\infty]} \times R^{v_1} \times R^{p_2}_{[+\infty]} \times R^{v_2}$, all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the degree of freedom of $S(Z + g, \Omega)$ at x is continuous for $x > 0$ and strictly increasing for $x > 0$ unless $v = 0$ and $g = \infty^{p_1 + p_2}$.

(b) $P(S(Z + (m_1, 0_{v_1}, m_2, 0_{v_2}), \Omega) \leq x) < P(S(Z + (m_1^*, 0_{v_1}, m_2^*, 0_{v_2}), \Omega) \leq x)$ for all $x > 0$ for all $m_1, m_1^* \in R^{p_1}_{+,\infty}$, $m_2, m_2^* \in R^{p_2}_{+,\infty}$ with $(m_1, m_2) < (m_1^*, m_2^*)$.

Assumption T6: For some $\chi > 0$, $S(am, \Omega) = a^{\chi} S(m, \Omega)$ for all scalars $a > 0$, $m \in R^k$, and $\Omega \in \Psi$.

For $(\theta, F) \in \mathcal{F}$, define $h_{1,j} = \infty$ if $\mathrm{E}_F\, m_j(X_i, Y_i, \theta) > 0$ and $h_{1,j} = 0$ if $\mathrm{E}_F\, m_j(X_i, Y_i, \theta) = 0$ for $j = 1, ..., p_1$, and define $h_{1,j} = \infty$ if $\mathrm{E}_F\, m_j(Y_i, \theta) > 0$ and $h_{1,j} = 0$ if $\mathrm{E}_F\, m_j(Y_i, \theta) = 0$ for $j = k_1 + 1, ..., k_1 + p_2$. Let $h_{11}(\theta, F) = (h_{1,1}(\theta, F), ..., h_{1,p_1}(\theta, F))'$, $h_{12}(\theta, F) = (h_{1,k_1+1}(\theta, F), ..., h_{1,k_1+p_2}(\theta, F))'$, and $\Omega(\theta, F) = \mathrm{Corr}_F((m_1^{\star}(X_i, Y_i, \theta)', m_2^{\star}(Y_i, \theta_0)')')$.

Assumption T7: For some $(\theta, F) \in \mathcal{F}$, the degree of freedom of $S(Z + (h_{11}(\theta, F), 0_{v_1}, h_{12}(\theta, F), 0_{v_2}), \Omega(\theta, F))$ is continuous at its $1-\alpha$ quantile, where $Z \sim N(0_k, \Omega(\theta, F))$.

1.2 GMS Assumptions

This subsection contains the assumptions on the function φ and the constants $\kappa_n : n \geq 1$ that define a GMS procedure, and they are maintained from Andrews and Soares (2010).

Assumption GMS1:

(a) $\varphi_j(\xi, \Omega)$ is continuous at all $(\xi, \Omega) \in (R^{p_1}_{[+\infty]} \times R^{v_1}_{[\pm\infty]} \times R^{p_2}_{[+\infty]} \times R^{v_2}_{[\pm\infty]}) \times \Psi$ with $\xi_j = 0$, where $\xi = (\xi_1, ..., \xi_k)'$, for $j = 1, ..., p_1, k_1 + 1, ..., k_1 + p_2$.

(b) $\varphi_j(\xi, \Omega) = 0$ for all $(\xi, \Omega) \in (R^{p_1}_{[+\infty]} \times R^{v_1}_{[\pm\infty]} \times R^{p_2}_{[+\infty]} \times R^{v_2}_{[\pm\infty]}) \times \Psi$ with $\xi_j = 0$, where $\xi = (\xi_1, ..., \xi_k)'$, for $j = 1, ..., p_1, k_1 + 1, ..., k_1 + p_2$.

(c) $\varphi_j(\xi, \Omega) = 0$ for all $j = p_1 + 1, ..., k_1, k_1 + p_2 + 1, ..., k_2$ for all $(\xi, \Omega) \in (R^{p_1}_{[+\infty]} \times R^{v_1}_{[\pm\infty]} \times R^{p_2}_{[+\infty]} \times R^{v_2}_{[\pm\infty]}) \times \Psi$.

Assumption GMS2: $\kappa_n \to \infty$.

Assumption GMS3:$\varphi_j(\xi, \Omega) \to \infty$ as $(\xi, \Omega) \to (\xi_*, \Omega_*)$ for all $(\xi_*, \Omega_*) \in (R^{p_1}_{[+\infty]} \times R^{v_1}_{[\pm\infty]} \times R^{p_2}_{[+\infty]} \times R^{v_2}_{[\pm\infty]}) \times cl(\Psi)$ with $\xi_{*,j} = \infty$, where $\xi_* = (\xi_{*,1}, ..., \xi_{*,k})'$, for $j = 1, ..., k$.

Assumption GMS4: $\kappa_n^{-1} \sqrt{n} \to \infty$.

Assumption GMS6: $\varphi_j(\xi, \Omega) \geq 0$ for all $(\xi, \Omega) \in (R^{p_1}_{[+\infty]} \times R^{v_1}_{[\pm\infty]} \times R^{p_2}_{[+\infty]} \times R^{v_2}_{[\pm\infty]}) \times \Psi$ for $j = p_1 + 1, ..., k_1, k_1 + p_2 + 1, ..., k_2$.

Assumption GMS7: $\varphi_j(\xi, \Omega) \geq \min\{\xi_j, 0\}$ for all $(\xi, \Omega) \in (R^{p_1}_{[+\infty]} \times R^{v_1}_{[\pm\infty]} \times R^{p_2}_{[+\infty]} \times R^{v_2}_{[\pm\infty]}) \times \Psi$ for $j = p_1 + 1, ..., k_1, k_1 + p_2 + 1, ..., k_2$.

References

ACEMOGLU, D., U. AKCIGIT, N. BLOOM, AND W. KERR (2013): "Innovation, Reallocation and Growth," Working Paper.

ANDREWS, D. W. K. AND P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," forthcoming at *Econometrica.*

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 G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119-158.

BANERJEE, A., A. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013): "The Diffusion of Microfinance," *Science*, vol. 341 no. 6144 1236498.

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

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

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

DUPAS, P., A. KEATS, AND J. ROBINSON (2015): "The Effect of Savings Accounts on Interpersonal Financial Relationships: Evidence from a Field Experiment in Rural Kenya," NBER Working Paper.

LYNCH, A. W. AND J. WACHTER (2013): "Using Samples of Unequal Length in Generalized Method of Moments Estimation," *Journal of Financial and Quantitative Analysis*, 48 (1), 277 - 307.

POLITIS, D. N. and J. P. ROMANO (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annual of Statistics*, 22, 2031-2050.

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.

ROMANO, J. P. AND A. M. SHAIKH (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169-212.

ROMANO, J. P., A. M. SHAIKH, AND M. WOLF(2014): "A Practical Two-Step Methods for Testing Moment Inequalities," forthcoming in *Econometrica.*

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

SILVAPULLE, M. J., AND P. K. SEN (2005): *Constrained Statistical Inference*. Wiley, New York.

VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.

WOLLMAN, T. (2014): "Trucks without bailouts: Equilibrium product characteristics for commercial vehicles," Working Paper.