Partial Identification and Confidence Intervals*

Jinyong Hahn
Department of Economics, UCLA

Geert Ridder
Department of Economics, USC

September 17, 2009

Abstract

We consider statistical inference on a single component of a parameter vector that satisfies a finite number of moment inequalities. The null hypothesis for this single component is given a dual characterization as a composite hypothesis regarding point identified parameters. We also are careful in the specification of the alternative hypothesis that also has a dual characterization as a composite hypothesis regarding point identified parameters. This setup substantially simplifies the conceptual basis of the inference problem. For an interval identified parameter we obtain closed form expressions for the confidence interval obtained by inverting the test statistic of the composite null against the composite alternative.

JEL code: C01, C12

Keywords: Partial identification, moment inequalities, composite hypotheses, confidence interval

*Financial support for this research was generously provided through NSF SES 0819612 and 0819638. Addresses: Jinyong Hahn, Department of Economics, Bunche Hall, UCLA, Los Angeles, CA 90095, hahn@econ.ucla.edu; Geert Ridder, Department of Economics, Kaprilian Hall, USC, Los Angeles, CA 90089, ridder@usc.edu. We thank participants in the Harvard/MIT econometrics workshop for helpful comments.

1 Introduction

There is a rapidly growing literature on inference on partially identified parameters. Manski (2003) surveys this literature. Early on Horowitz and Manski (2000) developed confidence regions that asymptotically cover the identified set with a fixed probability. Chernozhukov, Hong, and Tamer (2003) extended this to vector-valued parameters defined through minimization problems. Imbens and Manski (2004) developed confidence regions that cover the population parameter (as opposed to the set that contains it) with fixed probability. These pioneering ideas were subsequently followed by numerous technical innovations. A partial list includes Andrews, Berry, and Jia (2004), Moon and Schorfheide (2004), Pakes, Porter, Ho, and Ishii (2004), Romano, and Shaikh (2005), Andrews, and Soares (2007), Bugni (2007), Canay (2007), Fan and Park (2007), Galichon and Henry (2008), Beresteanu and Molinari (2008), Rosen (2008), and Chiburis (2009).

In this paper, we address issues that seem to have been ignored in the literature. The first issue is that we consider *one component* (or a subset) of the parameter vector that satisfies a finite number of moment inequalities. Unless we are mistaken, all the papers in the literature address the question of making inferences on the *entire parameter vector*. In many empirical applications, we focus on just one component of the parameter vector, and the remaining components can be treated as nuisance parameters. For example, applied researchers are typically interested in inference on just one slope coefficient in a linear regression model. One may argue that a confidence interval for one component of the parameter vector can be obtained by projecting the confidence region for the full parameter vector, but such a procedure in general produces a rather wide interval, especially when the dimension of the parameter vector is large.

Our approach differs from the recent literature in two other ways. First, we propose to consider an equivalent characterization of the null hypothesis on the parameter of interest in terms of *point identified* parameters. This "dual" characterization enables us to understand the inference in partially identified models in the standard framework of a multi-dimensional composite null hypothesis, which avoids some of the conceptual and technical problems in the recent literature. Second, we explicitly specify the alternative hypothesis that in general is a proper subset of the parameter space. As with the null we consider an equivalent characterization in terms of point identified parameters. By exploiting the structure of the alternative hypothesis, we can potentially increase the power of the test and reduce the length of the confidence interval. These two elements are further discussed in Section 2.

Our main contribution is the insight that inference for partially identified parameters can be conducted in terms of point identified parameters. No new conceptual framework is needed to understand the fundamentals of statistical inference with partially identified parameters, because we are dealing with a test of a composite null against a composite alternative hypothesis and the theory for such tests is well-known (see e.g. Lehmann (1986)). Although our insight is trivial ex post, we do not believe it was ex ante obvious. A byproduct of our setup is that the asymptotic analysis is deceptively simple, in particular much simpler than in most of the literature that develops special tools for this problem. Tests of a composite null against a composite alternative hypothesis are computationally challenging, depending on the nature of specific application. On the other hand, many tools proposed in the recent literature involve resampling methods such as subsampling. These techniques are computationally intensive as well, and therefore, we hope that the profession is willing to consider our conceptually simple but computationally challenging approach.

As discussed later, our setup is in many ways parallel to the weak instrument literature. Given the nonstandard nature of the weak instrument problem, such similarity should not be surprising. Because our objective is inference on a component of a parameter vector, our approach can be compared to Kleibergen and Mavroeidis (2008) in its scope. Our paper is not as ambitious as Andrews, Moreira and Stock (2006) in that we do not address optimality.

2 Dual Characterization and Alternative Hypothesis: An Example

We consider a simple toy model involving moment inequalities to illustrate our proposal. As was discussed in the introduction, our proposal consists of two elements. First, the null hypothesis is "marginalized" so that it is given a "dual" characterization in terms of *point identified* parameters. Second, the alternative hypothesis is explicitly spelled out. The alternative hypothesis maintains that the specification is correct, i.e., there exists some value of the parameter vector that satisfies the moment inequalities. We then marginalize the alternative hypothesis such that it is also given a dual characterization in terms of point identified parameters. The rationale behind these two steps is discussed later in Sections 2.3 and 2.4.

Our toy model has two parameters that satisfy the moment inequalities

$$E\left[X_1-\theta_1\right]\leq 0,\quad E\left[X_2-\theta_2\right]\leq 0,\quad \text{and}\quad E\left[X_3-(\theta_1+\theta_2)\right]\geq 0. \tag{1}$$

We are interested in testing $\theta_1=0$.

2.1 Characterization of the null and alternative hypothesis in terms of point identified parameters

We begin by discussing the first element of our proposal, i.e., the marginalization of the null hypothesis. The marginalization step starts with a reformulation of the test as a test of a composite null against a composite alternative hypothesis. In our toy example, the null hypothesis $\theta_1 = 0$ is equivalent to the statement "The moment inequality (1) is satisfied with $\theta_1 = 0$." Therefore, the null is satisfied if and only if there exists some θ_2 such that $\mu_1 \leq 0$, $\mu_2 \leq \theta_2$, and $\mu_3 \geq \theta_2$, where $\mu_j = E[X_j]$. The test of $\theta_1 = 0$ can then be interpreted as a specification test, i.e., a test of a null hypothesis that restricts the model parameters θ_1, θ_2 to a subset of the set of parameters that satisfy (1).

Having translated the null into an equivalent specification test format, we next provide a dual characterization in terms of point identified parameters. Recall that our specification test was about the existence of some θ_2 such that $\mu_1 \leq 0$, $\mu_2 \leq \theta_2$, and $\mu_3 \geq \theta_2$. It is trivial to recognize that such a θ_2 exists if and only if $\mu_2 \leq \mu_3$. In other words, our null hypothesis can be written as

$$H_0 : \mu_1 \leq 0, \quad \mu_2 \leq \mu_3. \tag{2}$$

Note that this dual characterization only involves the point identified parameters μ_1, μ_2, and μ_3.[1]

We now discuss the second element of our proposal, i.e., the explicit characterization of the alternative hypothesis. The alternative hypothesis that we consider is that the model is correctly specified, i.e., there exists a parameter vector (θ_1, θ_2) such that the moment inequalities (1) are satisfied. It is easy to see that there exists such a vector (θ_1, θ_2) if and only if $\mu_1 + \mu_2 \leq \mu_3$. Therefore, our alternative hypothesis can be written

$$H_1 : \mu_1 + \mu_2 \leq \mu_3 \text{ but } H_0 \text{ is not satisfied.} \tag{3}$$

Note that the alternative hypothesis is expressed only in terms of point identified parameters as well.

Let ω and Ω denote subsets of the parameter space of the point-identified parameter vector (μ_1, μ_2, μ_3), which without further restrictions is $\mathbb{R}^3$, such that ω is the collection of the parameter values that satisfy the null and Ω is the collection of the parameter values that satisfy the null or the alternative. We can then write $H_0 : (\mu_1, \mu_2, \mu_3) \in \omega$ and $H_1 : (\mu_1, \mu_2, \mu_3) \in \Omega \backslash \omega$

[1] In more complicated settings, Guggenberger, Hahn, and Kim (2008) provide a mathematical algorithm that converts the specification statement into a dual characterization.

with

$$\omega = \{(\mu_1, \mu_2, \mu_3) : \mu_1 \le 0, \mu_2 \le \mu_3\}$$
$$\Omega = \{(\mu_1, \mu_2, \mu_3) : \mu_1 + \mu_2 \le \mu_3\}$$

Standard tests, including the likelihood ratio (LR) test or Bayesian posterior odds, can be employed to distinguish the two hypotheses. We do not take a particular stand on which test should be adopted in practice. We anticipate that the LR test would probably be used more often due to its familiarity to the profession. On the other hand, we note that the LR test may not be admissible and optimality would require some version of a Bayesian procedure.

2.2 LR test

If the LR procedure is to be adopted, then the test statistic S takes the form

$$S = T_1^2 - T_2^2$$

where

$$T_1^2 = \inf_{t\in\omega} \left[n\left(\widehat{\mu} - t\right)' \widehat{\Sigma}^{-1} \left(\widehat{\mu} - t\right) \right], \quad \text{and} \quad T_2^2 = \inf_{t\in\Omega} \left[n\left(\widehat{\mu} - t\right)' \widehat{\Sigma}^{-1} \left(\widehat{\mu} - t\right) \right]$$

denote the squared distances of $\widehat{\mu}$ to ω and Ω, respectively, where $\widehat{\mu} = (\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\mu}_3)'$ denotes a $\sqrt{n}$-consistent estimator of $\mu = (\mu_1, \mu_2, \mu_3)'$ and the distances are defined using a consistent estimator $\widehat{\Sigma}$ of the asymptotic variance matrix Σ of $\widehat{\mu}$. The critical value s for a given size α is obtained by setting

$$\sup_{t^*\in\omega} \Pr\left[\inf_{t\in\omega} \left(Z + t^* - t\right)' \Sigma^{-1} \left(Z + t^* - t\right) - \inf_{t\in\Omega} \left(Z + t^* - t\right)' \Sigma^{-1} \left(Z + t^* - t\right) \ge s \right] = \alpha$$

where $Z \sim N(0, \Sigma)$. The maximization problem can be computationally intensive, and some approximation seems unavoidable in practice. In Sections 3.1 and 3.2 we will see that the maximization problem is relatively simple in the case of an interval identified parameter, because the maximum can only occur in three points. The asymptotic justification of this procedure, that will be discussed later, is deceptively simple.

2.3 Why a dual characterization of the null?

Why do we reformulate the null as a null hypothesis associated with a specification test, i.e., the hypothesis that the parameter vector (θ_1, θ_2) is in a (strict) subset of the parameter space of the model? It is because we would like to avoid forming a confidence interval on the component θ_1

by projecting the confidence region for (θ_1, θ_2). This is in principle possible, but it is expected to produce a wide confidence interval. This problem is well-known. For example, in the linear regression model $y_i = x_{i1}\beta_1 + \cdots + x_{iK}\beta_K + \varepsilon_i$ with normal and homoscedastic error ε_i with known variance, the confidence ellipsoid for or $(\beta_1, \ldots, \beta_K)$ is based on the χ^2_K distribution, and projection of this ellipsoid on the β_1-axis does not produce the usual confidence interval $\widehat{\beta}_1 \pm 1.96 \cdot \mathrm{se}(\widehat{\beta}_1)$. For instance if the covariates are orthogonal, then the projected confidence interval is $\widehat{\beta}_1 \pm \sqrt{\chi^2(K)_{.95}} \cdot \mathrm{se}(\widehat{\beta}_1)$ which is in general much wider than the usual one. On the other hand, we obtain the usual confidence interval if we invert the specification test of the hypothesis that both $\beta_1 = \beta_{10}$ and that there exists a parameter vector $(\beta_2, \ldots, \beta_K)$ such that

$$E\left[\begin{bmatrix} x_{i1} \\ \vdots \\ x_{iK} \end{bmatrix} (y_i - (x_{i1}\beta_{10} + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K))\right] = 0.$$

In this case the nuisance parameters $(\beta_2, \ldots, \beta_K)$ are point identified and it is easily seen that if we substitute the OLS estimators for these nuisance parameters, then the corresponding sample moment condition is proportional to $\widehat{\beta}_1 - \beta_{10}$, so that we indeed obtain the usual confidence interval if we invert the test statistic. The reformulation of the null $\theta_1 = 0$ as the specification test null hypothesis "There exists some θ_2 such that $\mu_1 \leq 0$, $\mu_2 \leq \theta_2$, and $\mu_3 \geq \theta_2$" plays an analogous role in the moment inequality setting.

2.4 Why an explicit characterization of the alternative?

Why do we provide an explicit characterization of the alternative hypothesis? It is again related to the power of the test and, because the confidence interval is obtained by inverting the test, the length of the confidence interval. This is probably best understood by an analogy to the weak instrument literature. In the early days of the weak IV literature, it was suggested that the confidence region for the structural parameters could be obtained by inverting the Anderson-Rubin statistic (Staiger and Stock (1997)). It was quickly recognized that the Anderson-Rubin statistic also has power against a violation of the overidentifying restrictions. This decreases the power of the test that the structural parameters have some pre-specified value. Moreira (2003) and Kleibergen (2005) suggested test statistics that did not suffer from this loss of power. This improvement was based on a careful consideration of the alternative hypothesis. To be specific, Moreira and Kleibergen considered the null hypothesis that the parameter vector is equal to some pre-specified value, *and* the alternative that the data satisfy overidentifying restrictions. By explicitly characterizing the alternative hypothesis, we expect to make analogous progress in the partial identification setting. The intuition is that if we do not explicitly characterize the alternative, then the power function of the test statistic of the hypothesis that the structural parameter has some pre-specified value is flatter than if we do specify the alternative properly. Hence the test has lower power and will upon inversion give a wider confidence interval.

In the IV setting some alternative hypothesis has in fact been implicit even in the conventional, i.e. without weak-instrument complications, case. Consider the linear regression model with one endogenous covariate $y_i = x_i\beta + \varepsilon_i$ with ε_i independent of the instruments $z_{i1}, \ldots, z_{iK}$. Inversion of the usual test of $H_0 : \beta = \beta_0$ gives a confidence interval of the form $\widehat{\beta} \pm 1.96 \cdot \mathrm{se}(\widehat{\beta}_1)$ with 1.96 the square root of the upper 95% percentile of the χ^2_1 distribution. The test of the moment condition

$$E\left[\begin{pmatrix} z_{i1} \\ \vdots \\ z_{iK} \end{pmatrix} (y_i - x_i\beta_0)\right] = 0$$

has asymptotically a χ^2_K distribution under the null, but in practice, we use the χ^2_1 distribution because we subtract the test of the overidentifying restrictions from this test statistic. The use of the χ^2_1 (as opposed to χ^2_K) distribution can be easily understood if we realize that we test H_0 :"$\beta = \beta_0$ and the overidentifying restrictions are satisfied for some β " against H_1 :"$\beta \neq \beta_0$ and the overidentifying restrictions are satisfied for some β". Maintaining the hypothesis that the overidentifying restrictions are satisfied yields a steeper power function and hence a more powerful test of the hypothesis that the structural parameters have some pre-specified value and for this reason this test is preferred in current econometric practice.

3 Application to an interval identified parameter

Although our approach can be applied in general moment inequality problems we will (re)consider a special case, the interval identified parameter problem that was studied by Imbens and Manski (2004). We show that the confidence interval obtained by inverting the LR test is similar (but not identical) to the one proposed by Imbens and Manski, except that our confidence interval overcomes the uniformity issue quite naturally. The uniformity problem in Imbens and Manski (2004) arises when the left and right end points of the estimated and/or identified sets are close to each other. Because of marginalization, our test is in terms of the point identified parameters, i.e., the left and right end points. By controlling the "size" of the test, where the size is defined as in Lehmann (1986), our approach deals with the uniformity problem directly. In this particular problem we obtain a closed-form expression for the confidence interval.

We consider the case that the parameter of interest θ satisfies the moment inequalities

$$E[X] \le \theta \le E[Y]$$

and we are interested in testing $H_0 : \theta = \theta_0$. Without loss of generality, we can set $\theta_0 = 0$. Initially we assume that we have a single observation X, Y from a normal population with

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right).$$

Our strategy is first to derive the LR test and the corresponding confidence interval for this normal population. The final step is to show that in large samples we obtain the same test and confidence intervals even if the population is not normal. By considering the finite sample inference problem and the asymptotic analysis separately, we can better understand the fundamental issues of the hypothesis test of a composite null against a composite alternative.

3.1 Finite Sample Problem: $\rho = 0, \sigma_1^2 = \sigma_2^2 = 1$

We first consider the finite sample problem if $\rho = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$. In that case the expressions are simpler so that the conceptual issues are easier to understand.

3.1.1 Test Statistic and Rejection Probability

The point identified parameters are $\mu_X = E[X]$ and $\mu_Y = E[Y]$. Implementation of our test requires that we characterize ω and Ω, where ω is the collection of (μ_X, μ_Y) satisfying the null and Ω is the collection of (μ_X, μ_Y) satisfying the null or the alternative. Obviously the null is satisfied if and only if $\mu_X \le 0 \le \mu_Y$, and therefore,

$$\omega = \{(\mu_X, \mu_Y) : \mu_X \le 0 \le \mu_Y\}.$$

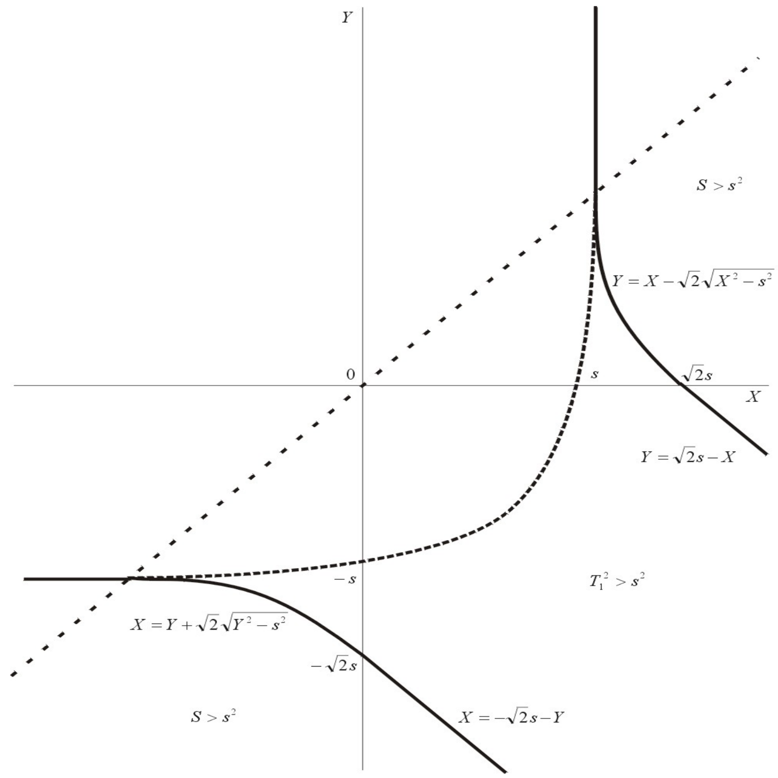
The model is correctly specified if and only if there exists some θ such that $\mu_X \le \theta \le \mu_Y$, which is equivalent to $\mu_X \le \mu_Y$. In other words,

$$\Omega = \{(\mu_X, \mu_Y) : \mu_X \le \mu_Y\}.$$

In order to test $H_0 : (\mu_X, \mu_Y) \in \omega$ against $H_1 : (\mu_X, \mu_Y) \in \Omega \backslash \omega$, we consider the LR statistic

$$S = T_1^2 - T_2^2 \tag{4}$$

Figure 1: Rejection regions $S > s^2$ and $T_1^2 > s^2$



where T_1 denotes the (Euclidean) distance from (X, Y) to ω, and T_2 denotes the distance (X, Y) to Ω. In other words,

$$T_1^2 = \begin{cases} 0 & \text{if } X \leq 0 \leq Y \\ X^2 & \text{if } X > 0 \text{ and } Y > 0 \\ X^2 + Y^2 & \text{if } X > 0 \text{ and } Y < 0 \\ Y^2 & \text{if } X \leq 0 \text{ and } Y < 0 \end{cases}$$

and

$$T_2^2 = \begin{cases} 0 & \text{if } X \leq Y \\ \frac{(X-Y)^2}{2} & \text{if } X > Y \end{cases}$$

The LR test rejects H_0 in favor of H_1 if S exceeds a critical value s^2. The rejection region is drawn in Figure 1. Note the difference between the rejection regions $S > s^2$ and $T_1^2 > s^2$ where the latter test also rejects if the model is misspecified.

Because the null is a composite hypothesis, the critical value s^2 is such that the supremum of the probability $\Pr(S > s^2)$ of the Type I error over ω is equal to a prespecified value, i.e., the size (Lehmann (1986)). The probability $\Pr\left(S > s^2\right)$ depends on (μ_X, μ_Y) and is equal to

$$\Pr(S > s^2) = \Phi(-s-\mu_X)\Phi(-s-\mu_Y) + \int_{-s}^{0} \Phi(-x-\sqrt{2}\sqrt{x^2+s^2}-\mu_Y)\phi(x-\mu_X)\mathrm{d}x + \quad (5)$$

$$\int_{0}^{\infty} \Phi(-\sqrt{2}s-x-\mu_Y)\phi(x-\mu_X)\mathrm{d}x + \Phi(-s+\mu_X)\Phi(-s+\mu_Y) +$$

$$\int_{0}^{s} \Phi(y-\sqrt{2}\sqrt{y^2+s^2}+\mu_X)\phi(y-\mu_Y)\mathrm{d}x + \int_{-\infty}^{0} \Phi(-\sqrt{2}s+y+\mu_X)\phi(y-\mu_Y)\mathrm{d}x$$

Note that the rejection region is symmetric with respect to $Y = -X$ and therefore if $\mu_Y = -\mu_X$ the first three terms are equal to the corresponding second three terms.

3.1.2 Least Favorable Parameter Values and Critical Value

The rejection probability as a function of (μ_X, μ_Y) can be computed either by simulation or by computing the integral in (5) numerically.[2] We plot these rejection probabilities both as a function of (μ_X, μ_Y) on $\mu_X < 0 < \mu_Y$ and as a function of $\mu_Y > 0$ for $\mu_X = 0$. In Appendix A we show that the probability of rejection cannot be maximal in a point of the interior of ω for any s^2, so that the maximal rejection probability is always attained on this section (or on the identical section with $\mu_Y = 0$ and $\mu_X < 0$). Therefore the latter graph shows what the value of the maximal rejection probability is and where it is attained.

The most unfavorable (μ_X, μ_Y) switches depending on the critical value s^2. For values $s^2 \leq 1.1284$, the most unfavorable combination is at $(0, 0)$ but for $s^2 > 1.1284$, the most unfavorable pair is at $\mu_X = -\infty$ or $\mu_Y = \infty$ and the other mean equal to 0. Because the type I error is equal to .1441 if $s^2 = 1.1284$ the latter is the relevant maximum at conventional significance levels. The maximal rejection probability is .05 if $s^2 = 1.645^2 = 2.7060$.

The intuition behind this critical value derives from the limit of the test statistic if $\mu_X = 0$ and $\mu_Y \to \infty$. Write $(X, Y) = (\mu_X + U, \mu_Y + V)$, where $(U, V)' \sim N(0, I_2)$. Combining the representation

$$S = U^2 1(U > 0) + (\mu_Y + V)^2 1(\mu_Y + V < 0) - \frac{(U - V - \mu_Y)^2}{2} 1(U - V - \mu_Y > 0)$$

with the fact that

$$\Pr\left((\mu_Y + V)^2 1(\mu_Y + V < 0) \neq 0\right) \leq \Pr(\mu_Y + V < 0) = \Pr(V < -\mu_Y) \to 0$$

$$\Pr\left(\frac{(U - V - \mu_Y)^2}{2} 1(U - V - \mu_Y > 0) \neq 0\right) \leq \Pr(U - V - \mu_Y > 0) = \Pr(U - V > \mu_Y) \to 0$$

[2] The simulation results (available from the authors) agree with the results of the numerical integration which is reassuring since in more complicated cases simulation may be the only viable method to compute the rejection probabilities.

we conclude that S converges in probability to $U^2 1(U > 0) \sim \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. This confirms that 1.645^2 is indeed the critical value for 5% size. Note that this approximation is relevant for all sizes up to 14.41%. By symmetry the same result obtains if $\mu_Y = 0$ and $\mu_X \to -\infty$.

3.1.3 Confidence Interval

With the conclusion that 1.645^2 is the critical value for 5% test, we turn to the problem of constructing a confidence interval by inverting the LR test. The confidence interval will contain values of θ_0 that are not rejected by the LR test of $H_0 : \mu_X \leq \theta_0 \leq \mu_Y$. Because the null can be equivalently written as $H_0 : \mu_X - \theta_0 \leq 0 \leq \mu_Y - \theta_0$, the test can be expressed in terms of $X - \theta_0$ and $Y - \theta_0$. Replacing (X, Y) by $(X - \theta_0, Y - \theta_0)$ in (4) we find the confidence interval (see Appendix B for details):

1. If $X < Y$, then
$$\theta_0 \in (X - 1.645, Y + 1.645) \tag{6}$$

2. If $X > Y$, and $X - Y \leq 1.645\sqrt{2}$, then
$$\theta_0 \in \left(X - \sqrt{1.645^2 + \frac{(X-Y)^2}{2}}, Y + \sqrt{1.645^2 + \frac{(X-Y)^2}{2}} \right) \tag{7}$$

3. If $X > Y$, and $X - Y > 1.645\sqrt{2}$, then
$$\theta_0 \in \left(\frac{X+Y}{2} - \frac{1.645}{\sqrt{2}}, \frac{X+Y}{2} + \frac{1.645}{\sqrt{2}} \right) \tag{8}$$

Our confidence interval is given by the above three expressions. Examination of (6) reveals that our confidence interval is identical to Imbens and Manski's (2004) if $X < Y$, although it differs when $X > Y$.

3.2 Finite Sample Problem: General Case

Imbens and Manski obtained the confidence interval for an interval identified parameter for the case that the bounds $(X, Y)'$ have an unrestricted variance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{9}$$

The expressions for the confidence interval are more complicated in this case as is the calculation of the least favorable alternative and the critical value. However the approach is the same as in

Section 3.1 and for that reason we give the expressions in Appendix C and only discuss some features of the results in Appendix C. The first difference with the results for the independent standard normal case is that the test statistic, least favorable alternative, critical value and confidence interval depend on the values of ρ, σ_1, σ_2. We distinguish between four cases:

(i) Positive correlation and variances of same order: $\rho \geq 0$ and $\rho \leq \frac{\sigma_1}{\sigma_2} < \frac{1}{\rho}$

(ii) Positive correlation and variance of X larger: $\rho \geq 0$ and $\frac{\sigma_1}{\sigma_2} \geq \frac{1}{\rho}$

(iii) Positive correlation and variance of Y larger: $\rho \geq 0$ and $\frac{\sigma_1}{\sigma_2} < \rho$

(iv) Negative correlation: $\rho < 0$

The rejection regions for cases (ii) and (iii) are not convex. As in Section 3.1 we obtain expressions for the rejection probabilities as a function of μ_X, μ_Y and these expressions involve a one dimensional integral that must be computed numerically. By numerical calculation we find that the rejection probability is maximal if $\mu_X = -\infty, \mu_Y = 0$, if $\mu_X = 0, \mu_Y = \infty$ or if $\mu_X = 0, \mu_Y = 0$. In the first case

$$S \xrightarrow{p} 1(V > 0)\frac{V^2}{\sigma_2^2}$$

with $U \sim N(0, \sigma_1^2)$ and in the second case

$$S \xrightarrow{p} 1(U > 0)\frac{U^2}{\sigma_1^2}$$

with $V \sim N(0, \sigma_2^2)$. This implies that for these least favorable alternatives the 5% critical value is 1.645^2. In the independent standard normal case the least favorable alternative is $\mu_X = 0, \mu_Y = 0$ only for values of s^2 that are so small that the type I error is much larger than the conventional values. In case (i) defined above the least favorable alternative is $\mu_X = 0, \mu_Y = 0$ for small s^2, but also if ρ is close to $\frac{\sigma_1}{\sigma_2}$ or to $\frac{\sigma_2}{\sigma_1}$. Because if $\rho = \frac{\sigma_1}{\sigma_2}$ or if $\rho = \frac{\sigma_2}{\sigma_1}$, the maximum is at $\mu_X = 0, \mu_Y = 0$, we find that in cases (ii) and (iii) the maximum is always at $\mu_X = 0, \mu_Y = 0$. In case (iv) the maximum is always at either $\mu_X = 0, \mu_Y = \infty$ or $\mu_X = -\infty, \mu_Y = 0$. If the least favorable alternative is $\mu_X = 0, \mu_Y = 0$ we find the critical value for a particular value of the type I error numerically. If $\sigma_1 = \sigma_2$ and $\rho \uparrow 1$, then if the type I error is 5% then $s^2 \uparrow 1.96^2$. Note that in this limit X and Y are the same with probability 1 and therefore θ is point identified with estimator $X \equiv Y$, so that 1.96^2 is relevant critical value. Hence if we are close to the point identified case the critical value approaches that for the point identified case.

If the type I error is .05 or equivalently the coverage probability of the confidence interval is .95, then for case (i) we compute first $\Pr(S > s^2)$ for $\mu_X = 0, \mu_Y = 0$. If this probability is less than .05 then the critical value $s^2 = 1.645^2$. If it is larger than .05 then the least favorable alternative is $\mu_X = 0, \mu_Y = 0$ and we compute the critical value numerically. In cases (ii) and (iii) we always use a critical value computed by numerical integration.

The expressions for the confidence interval are a direct generalization of the expressions for the independent standard normal case. They have a closed form expression and can be easily used in applications. The confidence interval is the same as that of Imbens and Manski (2004) if the estimated lower bound is smaller than the estimated upper bound. In all other cases our confidence interval differs from theirs. If the estimate of the lower bound is not much larger than the estimate of the upper bound the usual confidence interval is adjusted with the size of the adjustment dependent on $Y - X$. If the estimate of the lower bound is much larger than the estimate of the upper bound, the endpoints of the confidence interval are a convex combination of the lower and upper bounds with a constant subtracted and added, respectively.

4 Asymptotics

We have derived the confidence interval for the case that we have a single draw from a bivariate normal distribution. Now suppose that our data are a random sample $Z_i \equiv (X_i, Y_i)\,, i = 1, \ldots, n$ from a distribution with c.d.f. F. The means of X, Y are denoted by μ_X, μ_Y and their variance matrix by Σ. Let $\mu \equiv (\mu_X, \mu_Y)$. As before $\theta_0 = 0$ without loss of generality. Our LR test statistic is now $S = n\left(T_1^2 - T_2^2\right)$, where T_1 denotes the (Euclidean) distance from the sample means $\left(\overline{X}, \overline{Y}\right)$ to ω, and T_2 denotes the distance of $\left(\overline{X}, \overline{Y}\right)$ to Ω. The critical region is denoted by C and we reject the null hypothesis in favor of the alternative if and only if $\left(\sqrt{n}\overline{X}, \sqrt{n}\overline{Y}\right) \in C$. We first consider the case that X, Y are uncorrelated and have variances equal to 1. In that case the set C is convex as shown in Figure 1. We have to show that

$$\lim_{n\to\infty} \sup_{\mu\in\omega} \Pr\left(\left(\sqrt{n}\overline{X}, \sqrt{n}\overline{Y}\right) \in C\right) = \sup_{\mu\in\omega} \Pr\left(\mathcal{Z} \in C - \mu\right) \tag{10}$$

where $\mathcal{Z} \sim N\left(0, I_2\right)$ denotes the limiting standard normal random vector. First we observe that because $\mu \in \omega$ if and only if $\sqrt{n}\mu \in \omega$ we have

$$\sup_{\mu\in\omega} \Pr\left(\left(\sqrt{n}\overline{X}, \sqrt{n}\overline{Y}\right) \in C\right) = \sup_{\sqrt{n}\mu\in\omega} \Pr\left(\left(\sqrt{n}(\overline{X} - \mu_X), \sqrt{n}(\overline{Y} - \mu_Y)\right) \in C - \sqrt{n}\mu\right) =$$
$$\sup_{\mu\in\omega} \Pr\left(\left(\sqrt{n}(\overline{X} - \mu_X), \sqrt{n}(\overline{Y} - \mu_Y)\right) \in C - \mu\right)$$

In the uncorrelated and variance 1 case (Figure 1) the sets $C - \mu$ are convex for all $\mu \in \omega$. By the triangle inequality

$$\sup_{\mu\in\omega} \Pr\left(\mathcal{Z} \in C - \mu\right) - \inf_{\mu\in\omega} \left| \Pr\left(\left(\sqrt{n}(\overline{X} - \mu_X), \sqrt{n}(\overline{Y} - \mu_Y)\right) \in C - \mu\right) - \Pr\left(\mathcal{Z} \in C - \mu\right) \right| \leq$$

$$\sup_{\mu\in\omega} \Pr\left(\left(\sqrt{n}\overline{X}, \sqrt{n}\overline{Y}\right) \in C\right) \leq$$

$$\sup_{\mu\in\omega} \Pr\left(\mathcal{Z} \in C-\mu\right) + \sup_{\mu\in\omega} |\Pr\left(\left(\sqrt{n}(\overline{X}-\mu_X), \sqrt{n}(\overline{Y}-\mu_Y)\right) \in C-\mu\right) - \Pr\left(\mathcal{Z} \in C-\mu\right)|$$

The Berry-Esseen theorem in Götze (1991)[3] implies that because the sets $C-\mu$ are convex for all $\mu \in \omega$

$$\sup_{\mu\in\omega} |\Pr\left(\left(\sqrt{n}(\overline{X}-\mu_X), \sqrt{n}(\overline{Y}-\mu_Y)\right) \in C-\mu\right) - \Pr\left(\mathcal{Z} \in C-\mu\right)| \leq Cn^{-1/2}$$

so that (10) holds. The condition for this result is that

$$\sup_{\mu\in\omega} E(|Z|^3) < \infty$$

This is stronger than in Götze's original result, because we consider a set of population probability distributions indexed by μ.

In the case that X, Y are correlated and have arbitrary variances, we transform X, Y to the uncorrelated variables X^*, Y^* that have variance 1, i.e.

$$\begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \Sigma^{-1/2} \begin{pmatrix} X \\ Y \end{pmatrix} \qquad \begin{pmatrix} \mu_{X^*} \\ \mu_{Y^*} \end{pmatrix} = \Sigma^{-1/2} \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$$

Note that the transformed ω^* is such that $\mu^* = (\mu_{X^*}, \mu_{Y^*}) \in \omega^*$ if and only if $\sqrt{n}\mu^* \in \omega^*$. If X, Y have positive correlation and variances of same order, i.e. $\rho \geq 0$ and $\rho \leq \frac{\sigma_1}{\sigma_2} < \frac{1}{\rho}$ or if they are correlated negatively, then the transformed critical set is convex (see Appendix D), so that we can apply the same Berry-Esseen result from Götze (1991). In the two other cases the critical region is the union of a convex set C_1^* and a disjoint set C_2^* that has a complement that is convex. In those cases we compute the maximal rejection probability as

$$\sup_{\mu^*\in\omega^*} \Pr\left(\left(\sqrt{n}\overline{X^*}, \sqrt{n}\overline{Y^*}\right) \in C_1^*\right) + \sup_{\mu^*\in\omega^*} \Pr\left(\left(\sqrt{n}\overline{X^*}, \sqrt{n}\overline{Y^*}\right) \in C_2^*\right)$$

The first probability is approximated as before and for the second we have

$$\sup_{\mu^*\in\omega^*} \Pr\left(\mathcal{Z} \in C_2^*-\mu\right) - \inf_{\mu^*\in\omega^*} |\Pr\left(\left(\sqrt{n}(\overline{X^*}-\mu_{X^*}), \sqrt{n}(\overline{Y^*}-\mu_{Y^*})\right) \in C_2^{*c}-\mu^*\right) - \Pr\left(\mathcal{Z} \in C_2^{*c}-\mu^*\right)| \leq$$

$$\sup_{\mu^*\in\omega^*} \Pr\left(\left(\sqrt{n}\overline{X^*}, \sqrt{n}\overline{Y^*}\right) \in C_2^*\right) \leq$$

$$\sup_{\mu^*\in\omega^*} \Pr\left(\mathcal{Z} \in C_2^*-\mu^*\right) + \sup_{\mu^*\in\omega^*} |\Pr\left(\left(\sqrt{n}(\overline{X^*}-\mu_{X^*}), \sqrt{n}(\overline{Y^*}-\mu_{Y^*})\right) \in C_2^{*c}-\mu\right) - \Pr\left(\mathcal{Z} \in C_2^{*c}-\mu^*\right)|$$

[3] See expression (1.5) below his Theorem 1.3.

Because $C_2^{*c} - \mu^*$ is convex for all $\mu^* \in \omega^*$ the result follows.

If Σ is estimated, the critical region shown in Figure 6 depends on estimated parameters. If we denote the critical region for given Σ by C and with estimates substituted by $\hat{C}$, then we need to show that

$$\lim_{n\to\infty} \sup_{\mu\in\omega} \left| \Pr\left(\left(\sqrt{n}\overline{X}, \sqrt{n}\overline{Y}\right) \in \hat{C} \right) - \Pr\left(\left(\sqrt{n}\overline{X}, \sqrt{n}\overline{Y}\right) \in C \right) \right| = 0$$

We have if we denote $\ell_n = \left(\sqrt{n}(\overline{X} - \mu_X), \sqrt{n}(\overline{Y} - \mu_Y)\right)$ and $\ell_n^* = \left(\sqrt{n}(\overline{X^*} - \mu_{X^*}), \sqrt{n}(\overline{Y^*} - \mu_{Y^*})\right)$

$$\begin{aligned}
&\sup_{\mu\in\omega} \left| \Pr\left(\left(\sqrt{n}\overline{X}, \sqrt{n}\overline{Y}\right) \in \hat{C} \right) - \Pr\left(\left(\sqrt{n}\overline{X}, \sqrt{n}\overline{Y}\right) \in C \right) \right| = \\
&\sup_{\mu^*\in\omega^*} \left| \Pr\left(\ell_n^* \in \Sigma^{-1/2}\hat{C} - \mu^* \right) - \Pr\left(\ell_n^* \in \Sigma^{-1/2}C - \mu^* \right) \right| \leq \\
&\sup_{\mu^*\in\omega^*} \left| \Pr\left(\ell_n^* \in \Sigma^{-1/2}\hat{C} - \mu^* \right) - \Pr\left(\mathcal{Z} \in \Sigma^{-1/2}\hat{C} - \mu^* \right) \right| + \\
&\sup_{\mu^*\in\omega^*} \left| \Pr\left(\ell_n^* \in \Sigma^{-1/2}C - \mu^* \right) - \Pr\left(\mathcal{Z} \in \Sigma^{-1/2}C - \mu^* \right) \right| + \\
&\sup_{\mu^*\in\omega^*} \left| \Pr\left(\mathcal{Z} \in \Sigma^{-1/2}\hat{C} - \mu^* \right) - \Pr\left(\mathcal{Z} \in \Sigma^{-1/2}C - \mu^* \right) \right|
\end{aligned}$$

Because as is shown in Appendix D both $\Sigma^{-1/2}C - \mu^*$ and $\Sigma^{-1/2}\hat{C} - \mu^*$ are convex sets for all $\mu^* \in \omega^*$ Götze (1991) implies that the first two terms in the upper bound converge to 0.[4] Because $\Pr\left(\mathcal{Z} \in \Sigma^{-1/2}\hat{C} - \mu^* \right)$ is a continuous and bounded function of $\hat{\Sigma}$ and μ^*, the final term converges to 0 as well

5 Summary

In this paper, we propose a new method of inference for one component of a parameter vector characterized by moment inequalities. We translate the null hypothesis into a specification statement, which is in turn translated into a dual characterization involving only point identified parameters. We also state the alternative hypothesis explicitly, which is likewise given a dual characterization. Our innovation is to recognize the equivalent characterization involving only point identified parameters, which obviates many conceptual hurdles in the existing literature. We illustrate our procedure using the example in Imbens and Manski (2004), in which it is revealed that our procedure has a built-in protection against possible violations of uniformity.

[4]Because Götze's result is uniform the fact that the sets depend on estimated parameters does not matter.

References

[1] Andrews, D.W.K., S. Berry, & P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," mimeo, Yale University.

[2] Andrews, D.W.K. & G. Soares (2007): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," mimeo, Yale University.

[3] Andrews, D.W.K., M.J. Moreira, & J.H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," Econometrica 74, pp. 715–752.

[4] Beresteanu, A. & F. Molinari (2008): "Asymptotic properties for a class of partially identified models," Econometrica 76, 763–814.

[5] Bhattacharya, N.R. (1970): "Rates of Weak Convergence for the Multidimensional Central Limit Theorem," Theory of Probability and Its Applications 15, 68–86.

[6] Bugni, F. (2007): "Bootstrap Inference in Partially Identified Models," mimeo, Northwestern University.

[7] Canay, I.A. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," mimeo, University of Wisconsin.

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

[9] Chiburis, R.C. (2009): "Approximately Most Powerful Tests for Moment Inequalities," mimeo, Princeton University.

[10] Fan, Y. & S. Park (2007): "Confidence Sets for Some Partially Identified Parameters," mimeo, Vanderbilt University.

[11] Galichon, A. & M. Henry (2008): "A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters," Journal of Econometrics, forthcoming.

[12] Götze, F. (1991): "On the Rate of Convergence in the Multivariate CLT," Annals of Probability 19, 725–739.

[13] Guggenberger, P., J. Hahn, & K. Kim (2008): "Specification testing under moment inequalities," Economics Letters 99, 375–378.

[14] Horowitz, J., & C. Manski (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," Journal of the American Statistical Association 95, 77–84.

[15] Imbens, G. & C.F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," Econometrica 72, 1845–1857.

[16] Kleibergen, F. (2005): "Testing Parameters in GMM without Assuming that They are Identified," Econometrica 73, 1103–1124.

[17] Kleibergen, F., & S. Mavroeidis (2008): "Inference on Subsets of Parameters in GMM without Assuming Identification," mimeo, Brown University.

[18] Lehmann, E.L. (1986): Testing Statistical Hypotheses. New York: Springer-Verlag.

[19] Manski, C. (2003): Partial Identification of Probability Distributions. New York: Springer-Verlag.

[20] Moon, H. & F. Schorfheide (2004): "Empirical Likelihood Estimation with Inequality Moment Conditions," mimeo, University of Southern California.

[21] Moreira, M.J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," Econometrica 71, 1027–1048.

[22] Pakes, A., J. Porter, K. Ho, & J. Ishii (2004): "Applications of Moment Inequalities," mimeo, Harvard University.

[23] Michael D. Perlman, M.D. (1969): "One-Sided Testing Problems in Multivariate Analysis," Annals of Mathematical Statistics 40, 549–567.

[24] Romano, J.P. & A.M. Shaikh (2005): "Inference for the Identified Set in Partially Identified Econometric Models," mimeo, University of Chicago.

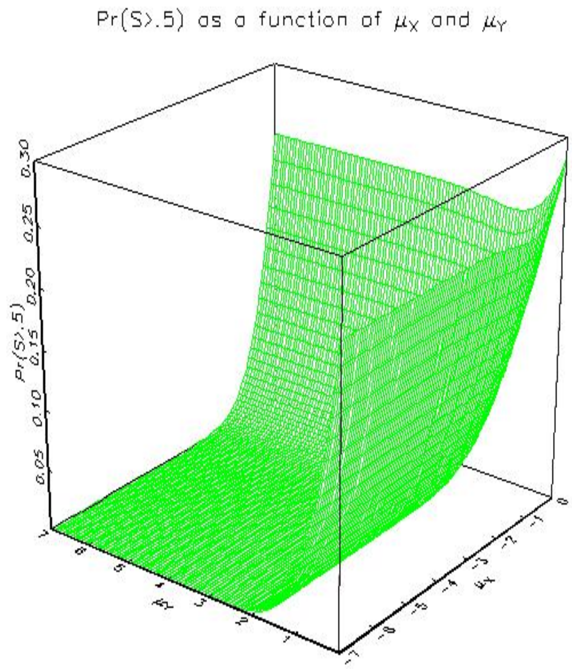
[25] Rosen, A.M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," Journal of Econometrics 146, 107–117.

[26] Staiger, D., & J.H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," Econometrica 65, 557–586.

[27] Takemura, A., & S. Kuriki (1997): "Weights of Chi-bar-square Distribution for Smooth or Piecewise Smooth Cone Alternatives," Annals of Statistics 25, 2368–2387.

[28] Warrack, G., & T. Robertson (1984): "A Likelihood Ratio Test Regarding Two Nested but Oblique Order-Restricted Hypotheses," Journal of the American Statistical Association 79, 881–886.

Figure 2: Rejection probability $\Pr(S > .5)$;$\sigma_1 = 1, \sigma_2 = 1, \rho = 0$



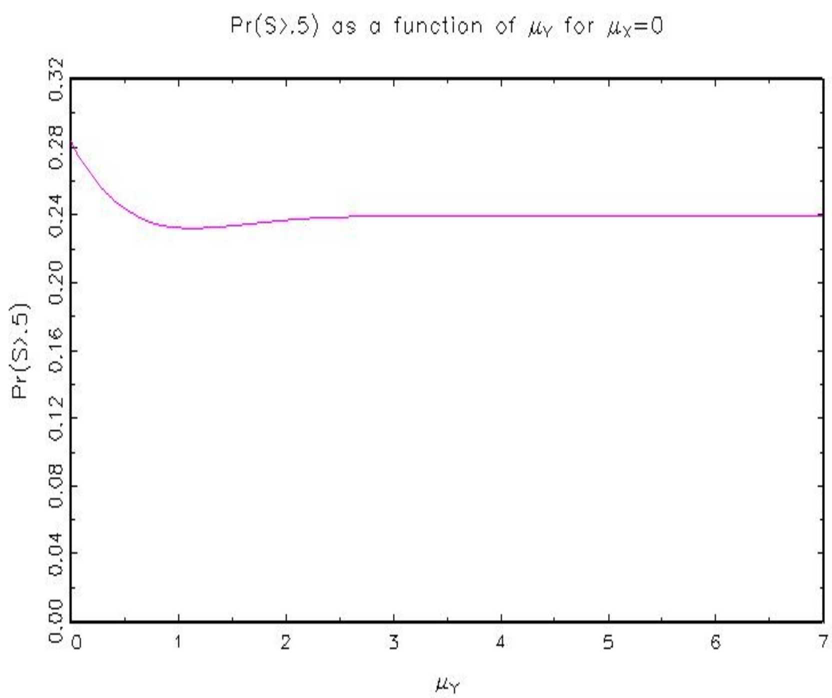
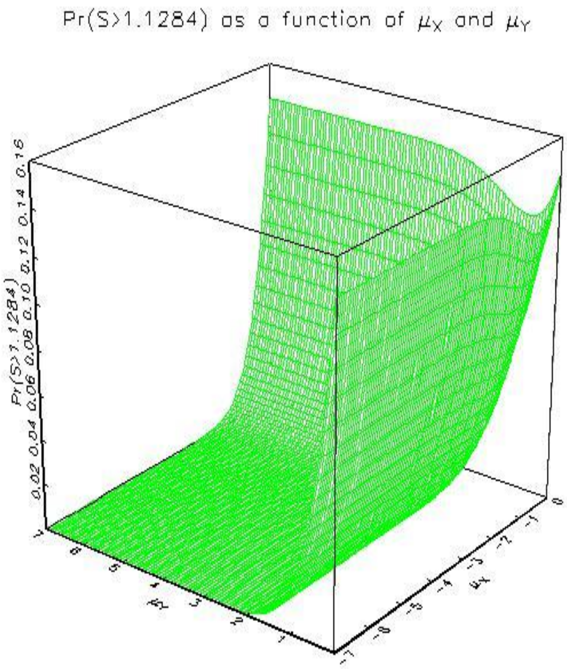


Figure 3: Rejection probability $\Pr(S > 1.1284)$;$\sigma_1 = 1, \sigma_2 = 1, \rho = 0$



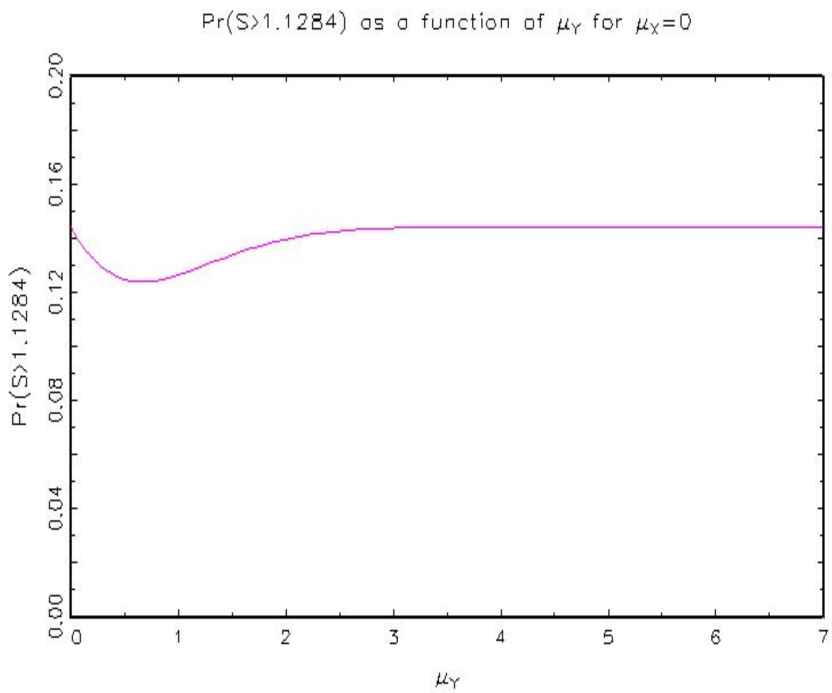
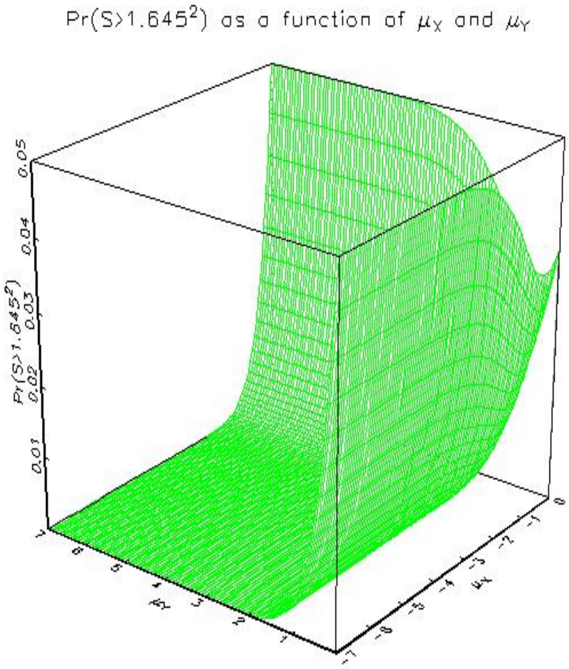


Figure 4: Rejection probability $\Pr(S > 2.7060)$; $\sigma_1 = 1, \sigma_2 = 1, \rho = 0$



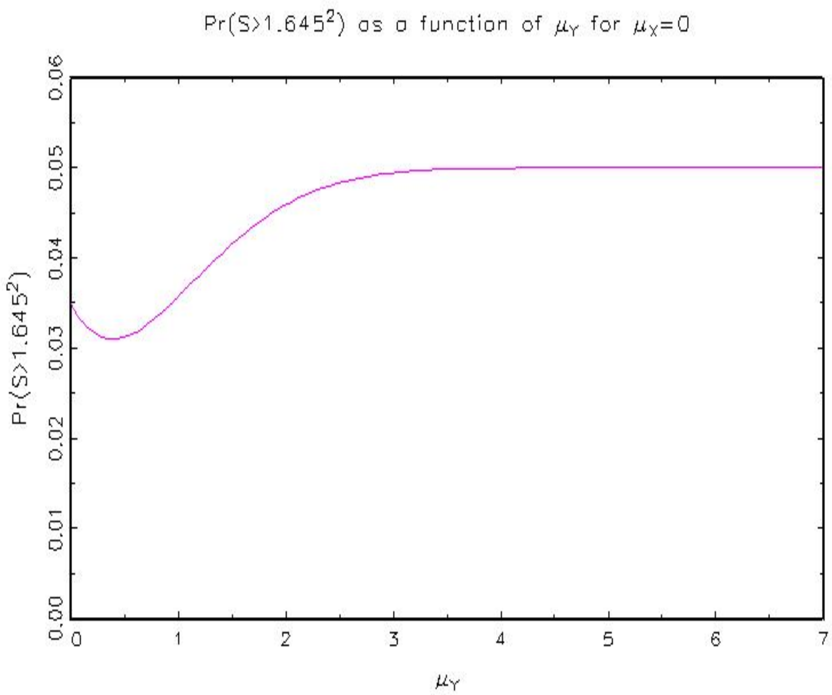
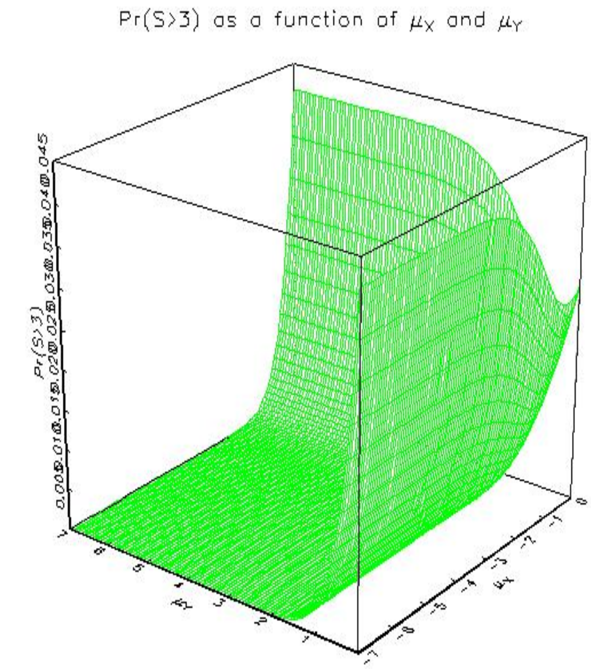


Figure 5: Rejection probability $\Pr(S > 3)$;$\sigma_1 = 1, \sigma_2 = 1, \rho = 0$



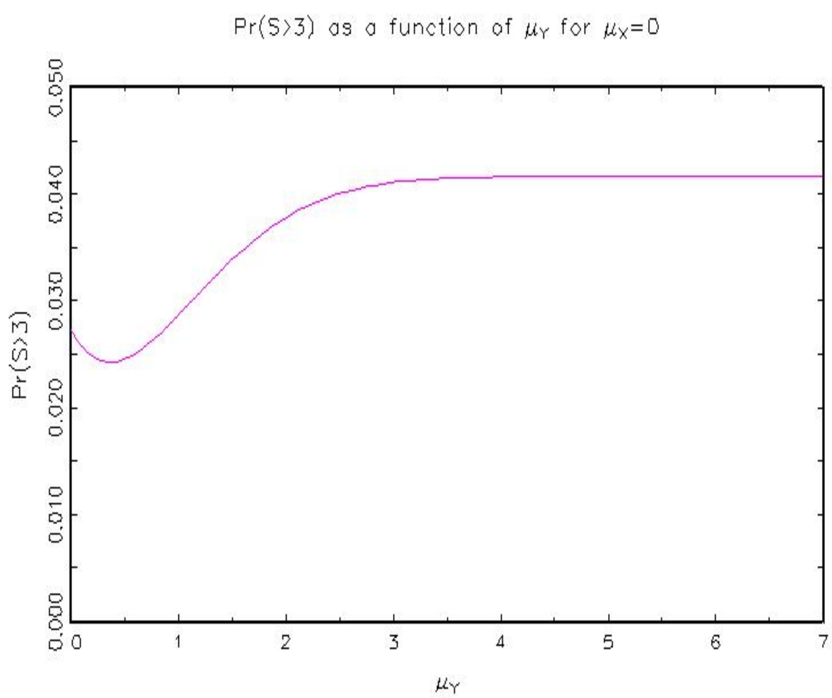


Figure 6: Critical region $S > s^2$ for case (i)

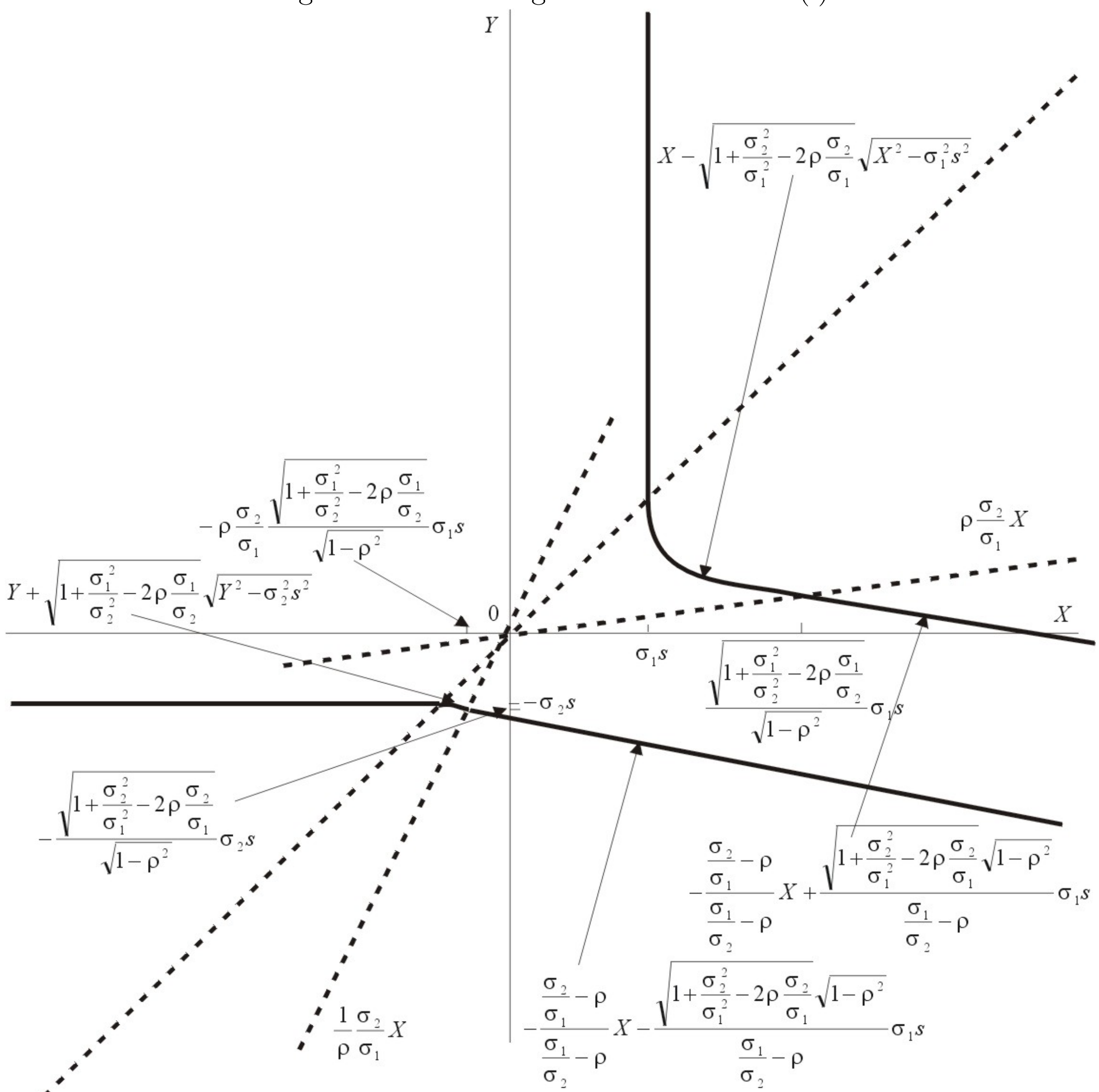
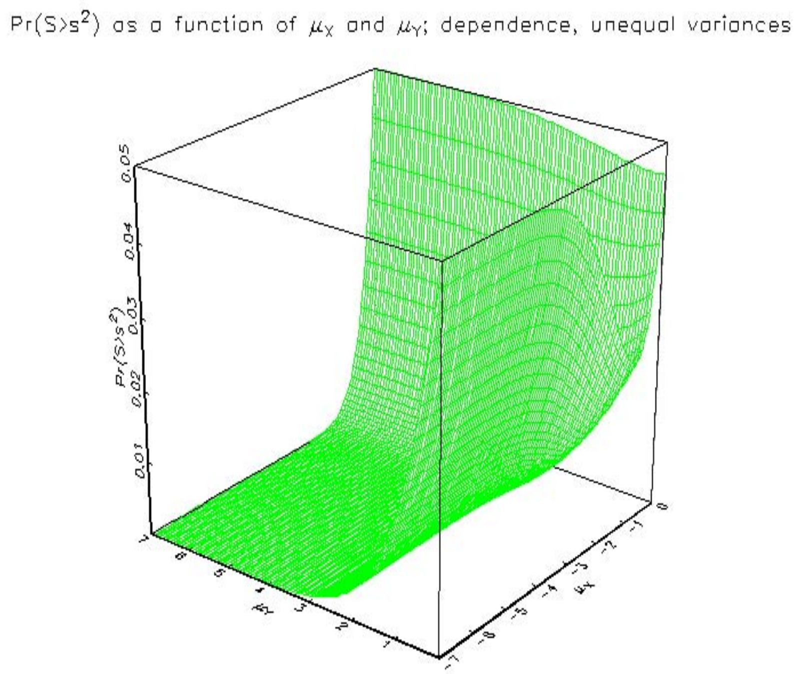


Figure 7: Rejection probability $\Pr(S > 2.7060)$; $\sigma_1 = 1, \sigma_2 = 2, \rho = .25$



Appendix

A Characterization of the Most Unfavorable (μ_X, μ_Y) for Independent Standard Normal X, Y

We show that $\Pr(S > s^2)$, viewed as a function of (μ_X, μ_Y), cannot achieve its maximum in the interior of the null set. Let

$$\begin{aligned}
u_1 &= \frac{z_2 - z_1}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \\
u_2 &= \frac{\sigma_2 - \rho\sigma_1}{\sqrt{1-\rho^2}\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \left(\frac{z_1}{\sigma_1} + \frac{\sigma_1 - \rho\sigma_2}{\sigma_2 - \rho\sigma_1} \frac{z_2}{\sigma_2} \right)
\end{aligned}$$

and note that the critical region can be characterized in terms of (u_1, u_2) in the form

$$\{u_1 \leq \psi(u_2), |u_2| \geq s\}$$

for some function $\psi(u_2)$, which implicitly depends on s. For example, when $\Sigma = I$,

$$\psi(u_2) = \begin{cases} |u_2| - \sqrt{2}s & \text{if} \quad |u_2| \geq \sqrt{2}s \\ -|u_2| + \sqrt{2}\sqrt{u_2^2 - s^2} & \text{if} \quad s \leq |u_2| \leq \sqrt{2}s \\ -\infty & \text{if} \quad |u_2| < s \end{cases}$$

Now, we evaluate the probability of rejection in terms of

$$\begin{aligned}
U_1 &= \frac{Y - X}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \\
U_2 &= \frac{\sigma_2 - \rho\sigma_1}{\sqrt{1-\rho^2}\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \left(\frac{X}{\sigma_1} + \frac{\sigma_1 - \rho\sigma_2}{\sigma_2 - \rho\sigma_1} \frac{Y}{\sigma_2} \right)
\end{aligned}$$

We note that

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \frac{\mu_Y - \mu_X}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \\ \frac{(\sigma_2 - \rho\sigma_1)\frac{\mu_X}{\sigma_1} + (\sigma_1 - \rho\sigma_2)\frac{\mu_Y}{\sigma_2}}{\sqrt{1-\rho^2}\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

We may therefore write that

$$\begin{aligned}
\Pr\left(S > s^2\right) &= \Pr\left[u_1 \leq \psi(u_2), |u_2| \geq s\right] \\
&= \int \Phi\left(\psi(u_2) - \frac{\mu_Y - \mu_X}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \right) \phi\left(u_2 - \frac{(\sigma_2 - \rho\sigma_1)\frac{\mu_X}{\sigma_1} + (\sigma_1 - \rho\sigma_2)\frac{\mu_Y}{\sigma_2}}{\sqrt{1-\rho^2}\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}} \right) du_2,
\end{aligned}$$

where Φ and ϕ are CDF and PDF of $N(0,1)$.

We now consider the maximization problem

$$\max_{\mu_X,\mu_Y} \Pr\left(S > s^2\right)$$

subject to

$$\frac{(\sigma_2 - \rho\sigma_1)\frac{\mu_X}{\sigma_1} + (\sigma_1 - \rho\sigma_2)\frac{\mu_Y}{\sigma_2}}{\sqrt{1-\rho^2}\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}} = c, \quad \mu_X \leq 0, \quad \mu_Y \geq 0.$$

For any given c, the maximization is equivalent to minimization of $\mu_Y - \mu_X$, and the solution always exists on the boundary of the null set ω. Therefore, $\Pr(S > s^2)$, viewed as a function of (μ_X, μ_Y), cannot achieve its maximum in the interior of the null set.

B Details of Confidence Interval

The confidence interval can be obtained by inverting the test statistic. Because the critical value is 1.645^2, we can see that the test does not reject for the following cases:

1. $X \leq \theta_0 \leq Y$
2. $(X-\theta_0)^2 \leq 1.645^2$ if $Y > X > \theta_0$
3. $(Y-\theta_0)^2 \leq 1.645^2$ if $X < Y \leq \theta_0$
4. $(X-\theta_0)^2 - \frac{(X-Y)^2}{2} \leq 1.645^2$ if $X > Y > \theta_0$
5. $(X-\theta_0)^2 + (Y-\theta_0)^2 - \frac{(X-Y)^2}{2} \leq 1.645^2$ if $X > \theta_0$ and $Y < \theta_0$
6. $(Y-\theta_0)^2 - \frac{(X-Y)^2}{2} \leq 1.645^2$ if $Y < X \leq \theta_0$

Combining Cases 1-3, we note that, if $X < Y$, the test does not reject for

$$\theta_0 \in (X - 1.645, Y + 1.645).$$

From Cases 4 and 6 , we can see that, if $X > Y$, and $X - Y \leq 1.645\sqrt{2}$, the test does not reject for

$$\theta_0 \in \left(X - \sqrt{1.645^2 + \frac{(X-Y)^2}{2}}, Y + \sqrt{1.645^2 + \frac{(X-Y)^2}{2}}\right).$$

From Case 6, we can see that, if $X > Y$, and $X - Y > 1.645\sqrt{2}$, the test does not reject for

$$\theta_0 \in \left(\frac{X+Y}{2} - \frac{1.645}{\sqrt{2}}, \frac{X+Y}{2} + \frac{1.645}{\sqrt{2}}\right).$$

C The General Case with an Unrestricted Variance Matrix

C.0.1 Test statistic and rejection probability

The test statistic depends on the values of ρ, σ_1, σ_2. Here we only consider the case of a positive correlation and variances of the same order[5] : $\rho \geq 0$ and $\rho \leq \frac{\sigma_1}{\sigma_2} < \frac{1}{\rho}$

$$
\begin{aligned}
S &= 0 && X \leq 0, Y \geq 0 \\
&= \tfrac{1}{\sigma_1^2} X^2 && X > 0, Y > X \\
&= \tfrac{1}{\sigma_1^2} X^2 - \frac{1}{\sigma_2^2\left(1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}\right)}(Y-X)^2 && X > 0, \rho\tfrac{\sigma_2}{\sigma_1} X < Y \leq X \\
&= \frac{\left(\left(\frac{\sigma_2}{\sigma_1}-\rho\right)X+\left(\frac{\sigma_1}{\sigma_2}-\rho\right)Y\right)^2}{(1-\rho^2)(\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2)} && X \geq 0, Y \leq \rho\tfrac{\sigma_2}{\sigma_1}X \text{ or } X < 0, Y < \tfrac{1}{\rho}\tfrac{\sigma_2}{\sigma_1}X \\
&= \tfrac{1}{\sigma_2^2} Y^2 - \frac{1}{\sigma_2^2\left(1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}\right)}(Y-X)^2 && X < 0, \tfrac{1}{\rho}\tfrac{\sigma_2}{\sigma_1} X \leq Y < X \\
&= \tfrac{1}{\sigma_2^2} Y^2 && X \leq Y < 0
\end{aligned}
$$

The rejection region $S > s^2$ is drawn in Figure 6. The rejection probability as a function of μ_X, μ_Y is

$$
\Pr(S > s^2) =
$$

$$
\int_{\sigma_1 s}^{\frac{\sqrt{1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}}}{\sqrt{1-\rho^2}}\sigma_1 s} \left[1-\Phi\left(\frac{x-\sqrt{1+\frac{\sigma_2^2}{\sigma_1^2}-2\rho\frac{\sigma_2}{\sigma_1}}\sqrt{x^2-\sigma_1^2 s^2}-\mu_Y-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_X)}{\sigma_2\sqrt{1-\rho^2}}\right)\right]\phi\left(\frac{x-\mu_X}{\sigma_1}\right)\mathrm{d}x+
$$

$$
\int_{\frac{\sqrt{1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}}}{\sqrt{1-\rho^2}}\sigma_1 s}^{\infty} \left[1-\Phi\left(\frac{-\frac{\frac{\sigma_2}{\sigma_1}-\rho}{\frac{\sigma_1}{\sigma_2}-\rho}x+\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}\sqrt{1-\rho^2}}{\frac{\sigma_1}{\sigma_2}-\rho}s-\mu_Y-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_X)}{\sigma_2\sqrt{1-\rho^2}}\right)\right]\phi\left(\frac{x-\mu_X}{\sigma_1}\right)\mathrm{d}x+
$$

$$
\int_{-\frac{\sqrt{1+\frac{\sigma_2^2}{\sigma_1^2}-2\rho\frac{\sigma_2}{\sigma_1}}}{\sqrt{1-\rho^2}}\sigma_2 s}^{-\sigma_2 s} \Phi\left(\frac{y+\sqrt{1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}}\sqrt{y^2-\sigma_2^2 s^2}-\mu_X-\rho\frac{\sigma_1}{\sigma_2}(y-\mu_Y)}{\sigma_1\sqrt{1-\rho^2}}\right)\phi\left(\frac{y-\mu_Y}{\sigma_2}\right)\mathrm{d}y+
$$

$$
\int_{-\infty}^{-\frac{\sqrt{1+\frac{\sigma_2^2}{\sigma_1^2}-2\rho\frac{\sigma_2}{\sigma_1}}}{\sqrt{1-\rho^2}}\sigma_2 s} \Phi\left(\frac{-\frac{\frac{\sigma_1}{\sigma_2}-\rho}{\frac{\sigma_2}{\sigma_1}-\rho}y-\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}\sqrt{1-\rho^2}}{\frac{\sigma_2}{\sigma_1}-\rho}s-\mu_X-\rho\frac{\sigma_1}{\sigma_2}(y-\mu_Y)}{\sigma_1\sqrt{1-\rho^2}}\right)\phi\left(\frac{y-\mu_Y}{\sigma_2}\right)\mathrm{d}y
$$

[5] The results for the other cases can be obtained from the authors.

C.0.2 Least favorable parameter values and critical value

The rejection probability is maximal if either $\mu_X = 0, \mu_Y = \infty$, $\mu_X = -\infty, \mu_Y = 0$ or $\mu_X = 0, \mu_Y = 0$. In the first two cases, because for $U = X - \mu_X, V = Y - \mu_Y$ and all c

$$\lim_{\mu_X \to -\infty} \Pr(X \geq cV) = \lim_{\mu_X \to -\infty} \Pr(U \geq cV - \mu_X) = 0$$

$$\lim_{\mu_Y \to \infty} \Pr(Y \leq cU) = \lim_{\mu_Y \to \infty} \Pr(V \leq cU - \mu_Y) = 0$$

we have that in all four cases S converges in probability to $1(U > 0)\frac{U^2}{\sigma_1^2}$ or to $1(V > 0)\frac{V^2}{\sigma_2^2}$. These two random variables have identical distributions so that the rejection probability is the same if either $\mu_X = 0, \mu_Y = \infty$ or $\mu_X = -\infty, \mu_Y = 0$. As in Section 3.1.2 the 5% critical value is 1.645^2 if the maximal rejection probability is at these values. In case (i) the maximum is at $\mu_X = 0, \mu_Y = 0$ if either s^2 is sufficiently small or ρ is close to either $\frac{\sigma_1}{\sigma_2}$ or to $\frac{\sigma_2}{\sigma_1}$. If $\sigma_1 \approx \sigma_2$ this occurs only if ρ is close to 1. Because if $\rho = \frac{\sigma_1}{\sigma_2}$ or if $\rho = \frac{\sigma_2}{\sigma_1}$, the maximum is at $\mu_X = 0, \mu_Y = 0$, we find that in cases (ii) and (iii) the maximum is always at $\mu_X = 0, \mu_Y = 0$. In case (iv) the maximum is always at either $\mu_X = 0, \mu_Y = \infty$ or $\mu_X = -\infty, \mu_Y = 0$. The figures 10-14 show the rejection probability for selected values of ρ, σ_1, σ_2 and $s^2 = 1.645^2$.

Therefore the computation of the 5% critical value if $\rho \geq 0$ and $\rho \leq \frac{\sigma_1}{\sigma_2} < \frac{1}{\rho}$ is as follows (the procedure is the same for all conventional significance levels). Compute the value of s^2 that satisfies

$$\Pr(S > s^2) =$$

$$\int_{\sigma_1 s}^{\frac{\sqrt{1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}}}{\sqrt{1-\rho^2}}\sigma_1 s} \left[1 - \Phi\left(\frac{x - \sqrt{1 + \frac{\sigma_2^2}{\sigma_1^2} - 2\rho\frac{\sigma_2}{\sigma_1}}\sqrt{x^2 - \sigma_1^2 s^2} - \rho\frac{\sigma_2}{\sigma_1}x}{\sigma_2\sqrt{1-\rho^2}}\right)\right] \phi\left(\frac{x}{\sigma_1}\right) \mathrm{d}x +$$

$$\int_{\frac{\sqrt{1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}}}{\sqrt{1-\rho^2}}\sigma_1 s}^{\infty} \left[1 - \Phi\left(\frac{-\frac{\frac{\sigma_2}{\sigma_1}-\rho}{\frac{\sigma_1}{\sigma_2}-\rho}x + \frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}\sqrt{1-\rho^2}}{\frac{\sigma_1}{\sigma_2}-\rho}s - \rho\frac{\sigma_2}{\sigma_1}x}{\sigma_2\sqrt{1-\rho^2}}\right)\right] \phi\left(\frac{x}{\sigma_1}\right) \mathrm{d}x +$$

$$\int_{-\frac{\sqrt{1+\frac{\sigma_2^2}{\sigma_1^2}-2\rho\frac{\sigma_2}{\sigma_1}}}{\sqrt{1-\rho^2}}\sigma_2 s}^{-\sigma_2 s} \Phi\left(\frac{y + \sqrt{1 + \frac{\sigma_1^2}{\sigma_2^2} - 2\rho\frac{\sigma_1}{\sigma_2}}\sqrt{y^2 - \sigma_2^2 s^2} - \rho\frac{\sigma_1}{\sigma_2}(y)}{\sigma_1\sqrt{1-\rho^2}}\right) \phi\left(\frac{y}{\sigma_2}\right) \mathrm{d}y +$$

$$\int_{-\infty}^{-\frac{\sqrt{1+\frac{\sigma_2^2}{\sigma_1^2}-2\rho\frac{\sigma_2}{\sigma_1}}}{\sqrt{1-\rho^2}}\sigma_2 s} \Phi\left(\frac{-\frac{\frac{\sigma_1}{\sigma_2}-\rho}{\frac{\sigma_2}{\sigma_1}-\rho}y - \frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}\sqrt{1-\rho^2}}{\frac{\sigma_2}{\sigma_1}-\rho}s - \mu_X - \rho\frac{\sigma_1}{\sigma_2}y}{\sigma_1\sqrt{1-\rho^2}}\right) \phi\left(\frac{y}{\sigma_2}\right) \mathrm{d}y = .05$$

The critical value is $\min\{s^2, 1.645^2\}$.

C.0.3 Confidence interval

If we take s^2 as in Section 3.2.2 the 95% confidence interval if $\rho \geq 0$, $\rho \leq \frac{\sigma_1}{\sigma_2} < \frac{1}{\rho}$ or $\rho < 0$ is given by:

If $\frac{\sigma_1}{\sigma_2} \geq \frac{\sigma_2}{\sigma_1}$

$$X - \sigma_1 s \leq \theta \leq Y + \sigma_2 s \qquad \text{if } \; Y - X \geq 0$$

$$X - \sigma_1 \sqrt{s^2 + \frac{(Y-X)^2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}} \leq \theta \leq Y + \sigma_2 \sqrt{s^2 + \frac{(Y-X)^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}}$$

$$\text{if } \; -s\frac{\frac{\sigma_2}{\sigma_1} - \rho}{\sqrt{1-\rho^2}}\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \leq Y - X < 0$$

$$\begin{aligned} X - \sigma_1 \sqrt{s^2 + \frac{(Y-X)^2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}} &\leq \theta \\ &\leq \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}X + \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}Y + s\sigma_1\sigma_2\frac{\sqrt{1-\rho^2}}{\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}} \end{aligned}$$

$$\text{if } \; -s\frac{\frac{\sigma_1}{\sigma_2} - \rho}{\sqrt{1-\rho^2}}\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \leq t < -s\frac{\frac{\sigma_2}{\sigma_1} - \rho}{\sqrt{1-\rho^2}}\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$$

$$\begin{aligned} \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}X + \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}Y - s\sigma_1\sigma_2\frac{\sqrt{1-\rho^2}}{\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}} &\leq \theta \\ &\leq \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}X + \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}Y + s\sigma_1\sigma_2\frac{\sqrt{1-\rho^2}}{\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}} \end{aligned}$$

$$\text{if } \; Y - X < -s\frac{\frac{\sigma_1}{\sigma_2} - \rho}{\sqrt{1-\rho^2}}\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$$

If $\frac{\sigma_1}{\sigma_2} < \frac{\sigma_2}{\sigma_1}$

$$X - \sigma_1 s \leq \theta \leq Y + \sigma_2 s \qquad \text{if } \; Y - X \geq 0$$

$$X - \sigma_1 \sqrt{s^2 + \frac{(Y-X)^2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}} \leq \theta \leq Y + \sigma_2 \sqrt{s^2 + \frac{(Y-X)^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}}$$

$$\text{if } \; -s\frac{\frac{\sigma_2}{\sigma_1} - \rho}{\sqrt{1-\rho^2}}\sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \leq Y - X < 0$$

$$\frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}X+\frac{\sigma_1^2-\rho\sigma_1\sigma_2}{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}Y-s\sigma_1\sigma_2\frac{\sqrt{1-\rho^2}}{\sqrt{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}}\le\theta$$
$$\le Y+\sigma_2\sqrt{s^2+\frac{(Y-X)^2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}$$
$$\text{if } -s\frac{\frac{\sigma_2}{\sigma_1}-\rho}{\sqrt{1-\rho^2}}\sqrt{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}\le t<-s\frac{\frac{\sigma_1}{\sigma_2}-\rho}{\sqrt{1-\rho^2}}\sqrt{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}$$

$$\frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}X+\frac{\sigma_1^2-\rho\sigma_1\sigma_2}{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}Y-s\sigma_1\sigma_2\frac{\sqrt{1-\rho^2}}{\sqrt{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}}\le\theta$$
$$\le\frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}X+\frac{\sigma_1^2-\rho\sigma_1\sigma_2}{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}Y+s\sigma_1\sigma_2\frac{\sqrt{1-\rho^2}}{\sqrt{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}}$$
$$\text{if } Y-X<-s\frac{\frac{\sigma_1}{\sigma_2}-\rho}{\sqrt{1-\rho^2}}\sqrt{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}$$

D Asymptotics

D.1 Critical Region for X^*, Y^* for known Σ

We only give the region for case (i): $\rho\ge 0$ and $\rho\le\frac{\sigma_1}{\sigma_2}<\frac{1}{\rho}$. The region is the union of two disjoint sets $\Sigma^{-1/2}C_1$ and $\Sigma^{-1/2}C_2$. The set $\Sigma^{-1/2}C_1$ is

$$Y^*>\frac{\frac{\sigma_1}{\sigma_2}-\rho}{\sqrt{1-\rho^2}}X^*-\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_2\sqrt{1-\rho^2}}\sqrt{X^{*2}-s^2}\qquad s<X^*\le\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_2\sqrt{1-\rho^2}}s$$
$$Y^*>-\frac{\sqrt{1-\rho^2}}{\frac{\sigma_1}{\sigma_2}-\rho}X^*+\frac{\sqrt{1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}}}{\frac{\sigma_1}{\sigma_2}-\rho}s\qquad X^*>\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_2\sqrt{1-\rho^2}}s$$

and the set $\Sigma^{-1/2}C_2$

$$X^*<\frac{\frac{\sigma_2}{\sigma_1}-\rho}{\sqrt{1-\rho^2}}Y^*+\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_1\sqrt{1-\rho^2}}\sqrt{Y^{*2}-s^2}\qquad -\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_1\sqrt{1-\rho^2}}s<Y^*\le -s$$
$$X^*<-\frac{\frac{\sigma_1}{\sigma_2}-\rho}{\sqrt{1-\rho^2}}Y^*-\frac{\sqrt{1+\frac{\sigma_1^2}{\sigma_2^2}-2\rho\frac{\sigma_1}{\sigma_2}}}{\sqrt{1-\rho^2}}s\qquad Y^*<-\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_1\sqrt{1-\rho^2}}s$$

Because for $s<X^*\le\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_2\sqrt{1-\rho^2}}s$ the function

$$g(X^*)=\frac{\frac{\sigma_1}{\sigma_2}-\rho}{\sqrt{1-\rho^2}}X^*-\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_2\sqrt{1-\rho^2}}\sqrt{X^{*2}-s^2}$$

has $g'(X^*) < 0$ and $g''(X^*) > 0$, the set $\Sigma^{-1/2}C_1$ is convex. Because for $-\frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_1\sqrt{1-\rho^2}}s < Y^* \leq -s$ the function

$$h(Y^*) = \frac{\frac{\sigma_2}{\sigma_1}-\rho}{\sqrt{1-\rho^2}}Y^* + \frac{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}{\sigma_1\sqrt{1-\rho^2}}\sqrt{Y^{*2}-s^2}$$

has $h'(Y^*) < 0$ and $h''(Y^*) < 0$, the set $\Sigma^{-1/2}C_2$ is also convex.

D.2 Critical Region for X^*, Y^* for estimated Σ

We only give the region for case (i): $\hat\rho \geq 0$ and $\hat\rho \leq \frac{\hat\sigma_1}{\hat\sigma_2} < \frac{1}{\hat\rho}$. The region is the union of two disjoint sets $\Sigma^{-1/2}\hat C_1$ and $\Sigma^{-1/2}\hat C_2$. The set $\Sigma^{-1/2}\hat C_1$ is

$$\begin{aligned}
Y^* &> \frac{\frac{\sigma_1}{\sigma_2}-\rho}{\sqrt{1-\rho^2}}X^* - \frac{\sigma_1}{\hat\sigma_1}\frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\sigma_2\sqrt{1-\rho^2}}\sqrt{X^{*2}-\frac{\hat\sigma_1^2}{\sigma_1^2}s^2} && \frac{\hat\sigma_1}{\sigma_1}s < X^* \leq \frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\hat\sigma_2\sqrt{1-\hat\rho^2}}\frac{\hat\sigma_1}{\sigma_1}s \\
Y^* &> -\frac{\left(\frac{\hat\sigma_2}{\hat\sigma_1}-\hat\rho\right)\sigma_1+\left(\frac{\hat\sigma_1}{\hat\sigma_2}-\hat\rho\right)\rho\sigma_2}{\left(\frac{\hat\sigma_1}{\hat\sigma_2}-\hat\rho\right)\sigma_2\sqrt{1-\rho^2}}X^* + \frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\sigma_2\sqrt{1-\rho^2}\left(\frac{\hat\sigma_1}{\hat\sigma_2}-\rho\right)}s && X^* > \frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\hat\sigma_2\sqrt{1-\hat\rho^2}}\frac{\hat\sigma_1}{\sigma_1}s
\end{aligned}$$

and the set $\Sigma^{-1/2}\hat C_2$

$$\begin{aligned}
X^* &< \frac{\frac{\sigma_2}{\sigma_1}-\rho}{\sqrt{1-\rho^2}}Y^* + \frac{\sigma_2}{\hat\sigma_2}\frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\sigma_1\sqrt{1-\rho^2}}\sqrt{Y^{*2}-\frac{\hat\sigma_2^2}{\sigma_2^2}s^2} && -\frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\hat\sigma_1\sqrt{1-\hat\rho^2}}\frac{\hat\sigma_2}{\sigma_2}s < Y^* \leq -\frac{\hat\sigma_2}{\sigma_2}s \\
X^* &< -\frac{\left(\frac{\hat\sigma_1}{\hat\sigma_2}-\hat\rho\right)\sigma_1+\left(\frac{\hat\sigma_2}{\hat\sigma_1}-\hat\rho\right)\rho\sigma_1}{\left(\frac{\hat\sigma_2}{\hat\sigma_1}-\hat\rho\right)\sigma_1\sqrt{1-\rho^2}}Y^* - \frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\sigma_1\sqrt{1-\rho^2}\left(\frac{\hat\sigma_2}{\hat\sigma_1}-\rho\right)}s && Y^* < -\frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\hat\sigma_1\sqrt{1-\hat\rho^2}}\frac{\hat\sigma_2}{\sigma_2}s
\end{aligned}$$

Because for $\frac{\hat\sigma_1}{\sigma_1}s < X^* \leq \frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\hat\sigma_2\sqrt{1-\hat\rho^2}}\frac{\hat\sigma_1}{\sigma_1}s$ the function

$$g(X^*) = \frac{\frac{\sigma_1}{\sigma_2}-\rho}{\sqrt{1-\rho^2}}X^* - \frac{\sigma_1}{\hat\sigma_1}\frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\sigma_2\sqrt{1-\rho^2}}\sqrt{X^{*2}-\frac{\hat\sigma_1^2}{\sigma_1^2}s^2}$$

has $g''(X^*) > 0$ (the function can actually be increasing) , the set $\Sigma^{-1/2}\hat C_1$ is convex, because it is bounded by a vertical line at $\frac{\hat\sigma_1}{\sigma_1}s$, a convex function for $\frac{\hat\sigma_1}{\sigma_1}s < X^* \leq \frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\hat\sigma_2\sqrt{1-\hat\rho^2}}\frac{\hat\sigma_1}{\sigma_1}s$ and a straight line for $X^* > \frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\hat\sigma_2\sqrt{1-\hat\rho^2}}\frac{\hat\sigma_1}{\sigma_1}s$. Because for $-\frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\hat\sigma_1\sqrt{1-\hat\rho^2}}\frac{\hat\sigma_2}{\sigma_2}s < Y^* \leq -\frac{\hat\sigma_2}{\sigma_2}s$

$$h(Y^*) = \frac{\frac{\sigma_2}{\sigma_1}-\rho}{\sqrt{1-\rho^2}}Y^* + \frac{\sigma_2}{\hat\sigma_2}\frac{\sqrt{\hat\sigma_1^2+\hat\sigma_2^2-2\hat\rho\hat\sigma_1\hat\sigma_2}}{\sigma_1\sqrt{1-\rho^2}}\sqrt{Y^{*2}-\frac{\hat\sigma_2^2}{\sigma_2^2}s^2}$$

has $h''(Y^*) < 0$ (the function can be increasing) , the set $\Sigma^{-1/2}\hat C_2$ is also convex using an analogous argument as for $\Sigma^{-1/2}\hat C_1$.