A permutation evaluation of the robustness of a high-dimensional test

By Nils Eckerdal

Department of Statistics
Uppsala University

Supervisor: Rauf Ahmad

2018
Abstract

The present thesis is a study of the robustness and performance of a test applicable in the high-dimensional context \((p > n)\) whose components are unbiased statistics (U-statistics). This test (the U-test) has been shown to perform well under a variety of circumstances and can be adapted to any general linear hypothesis. However, the robustness of the test is largely unexplored. Here, a simulation study is performed, focusing particularly on violations of the assumptions the test is based on. For extended evaluation, the performance of the U-test is compared to its permutation counterpart. The simulations show that the U-test is robust, performing poorly only when the permutation test does so as well. It is also discussed that the U-test does not inevitably rest on the assumptions originally imposed on it.

Key words: \((n,p)\)-asymptotics, permutation test, U-statistics.
Table of Contents

1 Introduction ........................................................................................................................................... 1

2 Methods .................................................................................................................................................. 2
  2.1 The U-test ......................................................................................................................................... 2
  2.2 The Permutation Test ...................................................................................................................... 5

3 Simulations ............................................................................................................................................ 7
  3.1 Simulation Under Ideal Conditions ................................................................................................. 7
  3.2 Robustness Evaluation .................................................................................................................... 9
    3.2.1 Violation of Assumption 1 ........................................................................................................... 9
    3.2.2 Violations of Assumptions 2 and 3 ............................................................................................ 10

4 Conclusions ......................................................................................................................................... 15

References ............................................................................................................................................. 16

Appendix A – Additional Results ........................................................................................................... 18

Appendix B – The Spiked Covariance Matrix ....................................................................................... 19
1 Introduction

Inference in the high-dimensional context, where the sample size is smaller than the number of parameters, is an important part of empirical research in many fields. A number of methods have been proposed, with applications in fields such as genetics (Barry et al., 2008; Efron and Tibshirani, 2007), medical imaging (Sjöstrand et al., 2008) and finance (Ng et al., 2015; Zou et al., 2015). While tests designed explicitly for the high-dimensional case date back to the 1950’s (Dempster, 1958), interest in the area has greatly increased since the 1990’s. Tests have been developed for varying hypotheses, with examples including two-sample mean tests of location (Dempster, 1958; Bai and Saranadasa, 1996; Chen and Qin, 2010; Ahmad, 2014) and tests of general linear hypotheses (Ahmad 2008; Zhou et al., 2017). Similarly, tests applicable in the high-dimensional case for special covariance structures e.g. sphericity have also been developed (Chen et al., 2010; Ahmad, 2017). The tests suggested in literature are based on a number of approaches valid under high-dimensionality (Hu and Bai, 2016).

Permutation tests are a class of nonparametric tests that can be defined for a wide variety of hypotheses, whose underlying concept is accredited to RA Fisher (see, e.g., Fisher, 1935). These tests are computationally intensive, to the point where they have only been made possible by modern computational development. They have particularly been used as a kind of benchmark for hypothesis testing. With the arrival of high-speed computer routines, however, permutation tests have now been made a feasible option of analysis. Requiring only few distributional assumptions, these tests have been suggested as a robust alternative to many parametric tests (see e.g. Good, 1994).

The aim of this thesis is to examine the robustness of a one-sample test of location (where \( H_0: \mu = 0 \)). Ahmad (2018) proposed an approach to mean vector testing applicable to the high-dimensional case and explored testing this hypothesis. The size and power of the test was inspected by the authors and it was found to enjoy high power in varying situations where its assumptions hold. In this thesis, this test is examined, and the main property of interest is the robustness of its size and power under violated assumptions. A simulation study is performed and in all explored cases, the test is compared with its permutation counterpart.

Among high-dimensional test Among high-dimensional tests of location in the literature, the one-sample test suggested by Ahmad (2018) is based on a relatively weak set of assumptions. It does not assume normality. Its assumptions on the covariance structure of the population also allow for many of the most commonly used structures, including the compound symmetric case.
which violates assumptions of tests such as those suggested by Bai and Saranadasa (1996) or Chen and Qin (2010). The fact that this test suggested by Ahmad (2018) is based on a relatively weak set of assumptions provides a reasonable ground to study the robustness of its performance.

The outline of the thesis is as follows. Section 2 provides the details of the compared tests, focusing on their assumptions. Section 3 is on the simulation strategy and the results. A discussion of the results concludes the thesis, in Section 4.

2 Methods

2.1 The U-test

In this thesis, the one-sample test of location proposed by Ahmad (2018) is referred to as the U-test, since it is based on components which are unbiased statistics (U-statistics). Its test statistic is referred to as $Z_U$.

The statistic was originally proposed by Ahmad (2008), where it was examined for the case of one-sample tests of location for high-dimensional repeated measures data under normality. The test is, however, adaptable to any general linear hypothesis. In Ahmad (2018), it is shown that the test holds under more general conditions than shown in Ahmad (2008). Primarily, it is shown that the originally stated assumption of multivariate normality is not needed. While simulations have provided support for the tests good performance in previous publications, these have not examined the robustness of the test under violations of the assumptions. This section gives the expression of the test statistic $Z_U$ as given in Ahmad (2018) and presents the assumptions of the test. Further, it discusses some conditions under which these assumptions do not hold. In this thesis, the U-test is evaluated only as it is derived under the null.

To arrive at an expression for $Z_U$, let $X_k = (X_{1k} \ldots X_{pk})^\prime$ be the $k^{th}$ $p$-dimensional observation vector, where $k = 1 \ldots n$ and $n$ is the size of the sample. Let also $\bar{X}$ be the observed vector of averages, $\bar{X} = (1/n)\sum_{k=1}^{n}X_k$. Further, let $A_k$ and $A_{kl}$ be the quadratic and bilinear forms, respectively; $A_k = X_k^\prime X_k$, $A_{kl} = X_k^\prime X_l$, where $k, l = 1 \ldots n, k \neq l$.

This lets us write out the components $Q, E_1, E_2$ and $E_3$. 
\[ Q = n \bar{X}' \bar{X}, \quad E_1 = \frac{1}{n} \sum_{k=1}^{n} A_k, \quad E_2 = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{l=1 \atop k \neq l}^{n} A_k A_l, \quad E_3 = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{l=1}^{n} A_{kl}^2. \]

These components define the statistic \( T \) and its variance estimator \( \hat{\text{Var}}(T) \):

\[ T = \frac{Q}{E_1}, \quad \text{Var}(T) = \frac{2E_2}{E_3}, \]

where \( E(T) \approx 1 \). Then,

\[ Z_U = \frac{T - E(T)}{\sqrt{\text{Var}(T)}} \xrightarrow{d} N(0,1), \quad \text{as } n,p \to \infty. \]  

(2)

This convergence holds under the following assumptions.

\[ E(X_{k,s}^4) \leq \gamma < \infty, \ s = 1 \ldots p, \text{ for some finite constant } \gamma \text{ independent of } p. \]  

(A1)

For \( p \to \infty \), let \( \frac{\text{tr}(\Sigma)}{p} = O(1) \).  

(A2)

For \( p \to \infty \), let \( \frac{\text{tr}(\Sigma^2)}{p^2} = O(\delta) \), where \( 0 < \delta \leq 1 \).  

(A3)

For proofs, see Ahmad (2018). Readers are referred to this publication for full details as well as a more comprehensive motivation of the test. In this section, we focus on the assumptions related above.

Under assumption (A1), the first four moments are finite. This is needed for \( Z_U \), a test statistic containing both the quadratic forms of \( T \) and its variance estimator. To understand assumptions (A2) and (A3), note that the convergence of \( Z_U \) depends on the asymptotic theory of U-statistics. This becomes relevant due to the following equations.

\[ T = 1 + \frac{E_0}{E_1/p}, \text{where} \]

(3)
\[ E_0 = \frac{1}{n} \sum_{k=1}^{n} \sum_{l=1, k \neq l}^{n} \frac{1}{p} A_{kl} = (n - 1) \left( \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{l=1, k \neq l}^{n} \frac{1}{p} A_{kl} \right) = (n - 1)U_n \]  

(4)

Note that \( \frac{E_1}{p} \) is an unbiased and consistent estimator of \( tr(\Sigma)/p \) under assumption (A2) which uniformly bounds \( tr(\Sigma)/p \) away from 0 and \( \infty \), as \( p \to \infty \).

Assumption (A3) is more directly connected to the \( U \)-statistic \( U_n \), and serves to make sure that it does not become degenerate nor that it diverges as \((n,p) \to \infty \). It can be shown that \( Var(U_n) = \frac{2}{n(n-1)} \cdot tr(\Sigma^2)/p^2 \), and that under (A3), \( Var(nU_n) \) is uniformly bounded, allowing the non-degenerate limit of \( Z_U \) given in equation (2).

Assumption (A2) and (A3) both relates to the eigenvalues of \( \Sigma \) (and \( \Sigma^2 \)) since the trace of a matrix equals the sum of its eigenvalues. Note that these will be violated if the sum of the eigenvalues diverges too quickly, such that \( tr(\Sigma)/p \to \infty \) or \( tr(\Sigma^2)/p^2 \to \infty \) as \( p \to \infty \). These assumptions can also be violated if the opposite is true, yielding convergence to zero with increasing dimension. For assumption (A3), although not for assumption (A2), an example that leads to violation through convergence to zero is \( \Sigma = I \).

Spiked covariance matrices are also of interest regarding the latter two assumptions. Such covariance matrices have one or a few eigenvalues much larger than the other positive eigenvalues. In some cases, these eigenvalues grow with increasing dimension, and in extreme cases the quantities controlled in (A2) and (A3) may diverge. Highly spiked covariance matrices have been the subject of several studies to examine the behavior of their population and sample eigenvalues, see e.g. Johnstone (2001) and Ahn et al. (2007). In particular, Ahn et al. (2007) examined a situation in which the largest eigenvalue grows exponentially with \( p \), which may cause violations of assumption (A2) or (A3). Simulations under conditions similar to this are performed, with results presented in Section 3.2.2. A case of spiked covariance matrices is the compound symmetric (CS) structure, defined as \( \Sigma_{CS} = \sigma^2 \rho J_{p \times p} + (\sigma^2 - \sigma^2 \rho)I_{p \times p} \), where \( J \) is a matrix of ones. That is, all diagonal elements are equal to \( \sigma^2 \) and all off-diagonal elements are equal to \( \sigma^2 \rho \). This is a simple and commonly discussed spiked covariance structure, one for which the assumptions of the test hold.
2.2 The Permutation Test

In this thesis, a permutation test is used as a benchmark to which the U-test is compared. Many tests have permutation counterparts, including the U-test. The present section provides a brief introduction to permutation tests.

When computing a permutation test, observations are permuted in all possible ways according to some method. For the present test, it is to either change the sign of each observation or not, finding all possible permutations doing this – numbering $2^n$ in total. For example, with the two univariate observations \{1,2\}, the full set of sign permutations is \{1,2\}, \{-1,2\}, \{1,-2\}, and \{-1,-2\}. In the case of this one-sample multivariate test of location, the signs of all elements of each observation is changed simultaneously.

All permutation tests are based on the assumption of exchangeability. In general, it holds if the probability of any given observation is the same under permutation. That is, if a given observed observation is denoted as $X_k$ and a given permuted observation as $\Pi(X_k)$, exchangeability holds if and only if $P(X_k) = P(\Pi(X_k))$, for all possible permutations. Under this assumption, all permutations of the data are equally likely to be drawn from the population. Therefore, if the test statistic for all possible permutations is calculated, the set of values of the test statistic thus found are all equally probable to observe. This set of values of the test statistic is referred to as the permutation distribution. In the above example, the permutation distribution contains only four values, one for each permutation of signs. However, for an example where the sample size is 10, the permutation distribution consists of 1024 elements. For such a sample size, under exchangeability and the null, the probability that the observed data would yield the largest value in the permutation distribution is $1/1024$, and this constitutes the p-value of this example. However, if the true mean is much larger, the probability of observing a large (extreme) value of the test statistic based on the observed data is large, thus this test is exact under the null but its power increases with increasing deviations from the null.

Note that for this one-sample test of location, in which the mean under the null without loss of generality is zero, exchangeability means that $P(X_k) = P(-X_k)$, for all observations ($k = 1 \ldots n$). It follows that exchangeability is in this case equivalent to multivariate symmetry.

The permutation test counterpart to the U-test is the one where $Z_U$ is computed for every permutation to find the permutation distribution, and the p-value is the share of values in the
permutation distribution that are as large as $Z_U$ for the observed data. Two issues are worth mentioning for the computation of this permutation test.

The first is the possibility of simplifying $Z_U$ without affecting the validity of the test. In fact, to perform the permutation test corresponding to the U-test, one need only calculate a test statistic that is order equivalent to $Z_U$ under permutation of signs. Here, we refer to the simplest such test statistic as $Z_{perm}$. This order equivalence ensures that the p-value of the test will be the same regardless of whether $Z_{perm}$ or $Z_U$ is used. Recall $Z_U$ and its components, given in equations (1) – (2). Notice that the quadratic form $A_k = X_k'X_k$ is unchanged by sign changes in the observation $X_k$. Further, regarding $A_{kl}$, notice that neither sign changes in $X_k$ nor in $X_l$ affects the absolute value of $A_{kl}$. This statistic appears only in $E_3$, in which it is squared, thus only the absolute value of $A_{kl}$ affects $E_3$, meaning that it is invariant to sign changes. Knowing this about $A_k$ and $A_{kl}$, we notice that the only factor in $Z_U$ that is not invariant under sign changes is $\bar{X}'\bar{X}$ in $Q$. Every other part of $Z_U$ either scales or shifts this quantity in a way that is constant across sign changes. Looking closer at this part of $Q$, we notice that $\bar{X}'\bar{X} = \sum_{j=1}^{p} \bar{X}_j^2 = \Sigma_{j=1}^{p} \left( \frac{1}{n} \sum_{i=1}^{n} X_{ij} \right)^2 = \frac{1}{n^2} \Sigma_{j=1}^{p} \left( \Sigma_{i=1}^{n} X_{ij} \right)^2$. Thus, the sum of squares of means, $\bar{X}'\bar{X}$, is proportional to the sum of squares of sums, $\sum_{j=1}^{p} \left( \Sigma_{i=1}^{n} X_{ij} \right)^2$, meaning that these are order-equivalent – to each other and to $Z_U$. This lets us define $Z_{perm}$ as

$$Z_{perm} = \sum_{j=1}^{p} \left( \sum_{i=1}^{n} X_{ij} \right)^2.$$  

(5)

This is a much simpler statistic to compute than $Z_U$, allowing for large increases in computation speed.

The second computational note of interest arises from the fact that $Z_{perm}$ is a sum of squares of the sums of the variables. Note that, for a given permutation of signs in the data matrix $X$, inverting all signs of the matrix also inverts the sign of each observed variable-wise sum. Thus, since the absolute value does not change for any of these elements, the sum of squares of sums is also unchanged. This yields the fact that for every sign permutation of the data matrix $X$, there is one other that yields the same value of $Z_{perm}$. In a sense, not all sign permutations are unique. Omitting the calculation of these non-unique permutations will not change the resulting p-value of the test, and this has been done in all simulations in this thesis. This changes the total number of permutations from $2^n$ to $2^{n-1}$.
While $2^{n-1}$ is only half as big as $2^n$, it quickly grows to a size impossible to handle. Calculating $Z_{perm}$ for $2^9 = 512$ unique permutations is trivial, but 100 observations yield $2^{99} \approx 6 \cdot 10^{29}$ unique permutations. However, not all permutations need to be used in order to provide an approximately exact test. In the simulations presented in Section 3, $Z_{perm}$ have been calculated for no more then 10 000 unique permutations. For sample sizes where a larger number of unique permutations are available, 10 000 randomly selected unique permutations are used.

3 Simulations

3.1 Simulation Under Ideal Conditions

The size and power of the tests being compared are first examined under ideal conditions, under which the assumptions of both tests hold. In these simulations, data are generated from a multivariate normal distribution, where the covariance matrix follow a CS structure with $\sigma^2 = 1$ and $\rho = 0.3$. For all simulations in this thesis, the size simulations (under the null) are made with 5000 repetitions, and power simulations (under the alternative) are made with 1000 repetitions. A greater number of size simulations are run to increase the accuracy of these results, seeing as relatively small changes in size can affect the usefulness of the test to a large degree. Time constraints on these computationally heavy simulations prevented this greater accuracy to be used in the power simulations as well. All simulations are run in R version 3.4.3 or later.

Throughout this thesis, the significance level ($\alpha$) of all tests is set to 5%.

The results for the test size are given in Table 1. The permutation test seems to be conservative more often than liberal, especially for the moderately large sample size of 50. The U-test does not seem to be consistently conservative nor liberal but does rather shows rejection rates close to the nominal significance level in general.

To examine the power of the tests, the mean vector $\mu$ is set by letting $\mu_j = \delta \cdot j/p$, for all elements, $j = 1, \ldots, p$. Therefore, letting $\delta$ equal zero corresponds to the null, which lets the sizes of the tests be measured. The results for power are shown in Figure 1. In each of the nine graphs included in Figure 1, seven values of $\delta$ are evaluated: 0, 0.2, 0.4, 0.6, 0.8, 1 and 1.2. These values of $\delta$ are consistently evaluated for all power simulations in this thesis.

In most cases, the power of the two tests is very similar. In the small sample case where $n = 10$, the test will only reliably reject relatively large deviations from the null, but the moderately small ($n = 20$) and moderately large ($n = 50$) sample sizes show favorable power curves. Based on the properties of the U-test as explored in Ahmad (2008) and Ahmad (2018), the
power of the test should increase with increasing dimension. This might also be expected from this simulation design, since the distance between $\mu$ and $0_{p \times 1}$ (the mean vector and the null) increases with dimension, as measured by e.g. the Euclidian distance. This is not immediately visible from these results, however.

Table 1: Size. Rejection rates of the U-test and the permutation test under the null, $\alpha = 0.05$. Multivariate normal data, CS covariance matrix with $\sigma^2 = 1$ and $\rho = 0.3$.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>n = 10</th>
<th>n = 20</th>
<th>n = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 20$</td>
<td>$Z_U$</td>
<td>0.042</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>$Z_{\text{perm}}$</td>
<td>0.046</td>
<td>0.046</td>
</tr>
<tr>
<td>$p = 100$</td>
<td>$Z_U$</td>
<td>0.056</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>$Z_{\text{perm}}$</td>
<td>0.054</td>
<td>0.051</td>
</tr>
<tr>
<td>$p = 500$</td>
<td>$Z_U$</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>$Z_{\text{perm}}$</td>
<td>0.048</td>
<td>0.049</td>
</tr>
</tbody>
</table>

Figure 1: Power. Rejection rates (y axes) over varying values of $\delta$ (x axes) for the U-test (solid line) and the permutation test (dashed line), $\alpha = 0.05$. Multivariate normal data, CS covariance matrix with $\sigma^2 = 1$ and $\rho = 0.3$. 
3.2 Robustness Evaluation

3.2.1 Violation of Assumption 1

In Table 2 and Figure 2, the results of simulations from the multivariate Cauchy (MVC) distribution are given. This is a cursory examination of the performance of the U-test in cases where not all of the first four moments exists, which is required to fulfil assumption (A1). However, the Cauchy distribution must be seen as an extreme violation of this assumption since all moments of the distribution are undefined. Further, the heavy-tailed nature of this distribution makes hypothesis testing difficult in general. This can be seen in the results, since both the U-test and the permutation test perform poorly, even though the assumption of the permutation test is fulfilled.

The data generated from the multivariate Cauchy distribution are generated with the intention of being as similar to the data generated in the ideal case as possible, in terms of location and dispersion. The data are generated with the function rmvc, from the LaplacesDemon package. Here, the MVC data are generated by first generating multivariate normal data, then each observation vector is divided by the square root of a $\chi^2_1$ distributed observation. These $\chi^2_1$-distributed observations are independently generated. Controlling the “mean” and “covariance matrix” of the MVC distribution is done, respectively, by adding the mean vector to all observations in a last step of data generation and by setting the covariance matrix of the multivariate normal data to the desired matrix. This provides some control over the location and dispersion of the resulting distribution, even though the true mean vector and covariance matrix are undefined and not directly controllable. In these simulations, these parameters are set in the same way as in the ideal settings, with the “mean” a function of $\delta$ and the covariance matrix being a CS correlation matrix with $\sigma^2 = 1$ and $\rho = 0.3$.

Regarding the size of the test, the permutation test performs similarly to ideal conditions, with rejection rates close to $\alpha$ for $n = 10$ as well as $n = 20$ and mostly conservative results for $n = 50$. The U-test is highly conservative in all cases. However, both tests show very poor results in terms of power, a result that barely improves with increasing sample size or dimension. For comparison, the power performance of the U-test test under ideal conditions is also shown in Figure 2.
Table 2: Size. Rejection rates of the U-test and the permutation test under the null. α = 0.05. Multivariate Cauchy distributed data.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>n = 10</th>
<th>n = 20</th>
<th>n = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>p = 20</td>
<td>$Z_U$</td>
<td>0.009</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.047</td>
<td>0.047</td>
</tr>
<tr>
<td>p = 100</td>
<td>$Z_U$</td>
<td>0.014</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.045</td>
<td>0.039</td>
</tr>
<tr>
<td>p = 500</td>
<td>$Z_U$</td>
<td>0.017</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.049</td>
<td>0.063</td>
</tr>
</tbody>
</table>

Figure 2: Power. Rejection rates (y axes) over varying values of $\delta$ (x axes) for the U-test (solid line) and the permutation test (dashed line), $\alpha = 0.05$. Multivariate Cauchy distributed data. For comparison: The U-test under ideal conditions (dotted line).

3.2.2 Violations of Assumptions 2 and 3

As indicated in Section 2.1, assumption (A3) is violated if $\Sigma = I$, since $tr(I^2)/p^2 = p/p^2$ will converge to 0 as $p \to \infty$. In one set of robustness simulations, data are generated from the multivariate normal distribution where $\Sigma = I$. To examine whether the U-test is robust to cases where this is nearly true, multivariate normal data are also generated where the covariance matrix is a compound symmetry correlation matrix where $\sigma^2 = 1$ and $\rho = 0.1$. As mentioned
in Section 2.1., this violation of assumption (A3) is not a violation of assumption (A2). In Table 3 and Figure 3 these conditions are compared with the results under the ideal conditions presented in Section 3.1, in which \( \rho = 0.3 \).

Examining the result of these simulations, the U-test is conservative for cases where the sample size is ten. For larger sample sizes and dimensions, the test is decreasingly conservative, and its rejection rate is very close to the nominal significance level of 5 % in the case where \( n = 50 \) and \( p = 500 \). The permutation test is neither consistently conservative nor liberal for the small or moderately small sample sizes (\( n = 10 \) and \( n = 20 \)) but is conservative for the largest sample size under consideration. We also notice that the type I error rate is most often lower in the case where \( \rho = 0.1 \) as compared to \( \rho = 0.3 \), and lower still where \( \rho = 0 \). However, the lower rejection rate under the null that seems to be a result of this violation to the assumption does not come with a decrease in power; rather, the power increases substantially as \( \rho \) is decreasing from 0.3 to 0.1 and from 0.1 to 0. This improvement in performance, with generally lower type I error rates and substantially higher power, is not an expected result to this violation of assumption (A3) for the U-test.

**Table 3: Size. Rejection rates of the U-test and the permutation test under the null, \( \alpha=0.05 \). Multivariate normal data, CS covariance matrix with varying \( \rho \).**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>( \rho = 0 )</th>
<th>( \rho = 0.1 )</th>
<th>( \rho = 0.3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 10, p = 20 )</td>
<td>( Z_{U} ) 0.034</td>
<td>0.043</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>( Z_{perm} ) 0.049</td>
<td>0.055</td>
<td>0.046</td>
</tr>
<tr>
<td>( n = 20, p = 100 )</td>
<td>( Z_{U} ) 0.040</td>
<td>0.046</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>( Z_{perm} ) 0.057</td>
<td>0.050</td>
<td>0.051</td>
</tr>
<tr>
<td>( n = 50, p = 500 )</td>
<td>( Z_{U} ) 0.051</td>
<td>0.052</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>( Z_{perm} ) 0.036</td>
<td>0.037</td>
<td>0.036</td>
</tr>
</tbody>
</table>
Figure 3: Power. Rejection rates (y axes) over varying values of \( \delta \) (x axes) for the U-test (solid line) and the permutation test (dashed line). \( \alpha = 0.05 \). Multivariate normal data, CS covariance matrix with varying \( p \).

The robustness of the U-test is also being explored under a condition with a highly spiked covariance matrix. Unlike the case where \( \Sigma = I \), in which the quantity being controlled in assumption (A3) converges to 0, this case is defined such that both quantities controlled in the assumptions (A2) and (A3) diverges with increasing dimension. The definition of this highly spiked covariance matrix follows closely from a condition discussed in Ahn et al. (2007), in which a highly spiked covariance matrix under high-dimensionality was considered. The authors examined the properties of a diagonal covariance matrix where all diagonal elements are 1, with the first of these elements being the only exception. This first diagonal element was instead defined as \( p^e \), where \( e \) was some constant greater than one. This is considered an extreme case of spiked covariance matrices.

For the simulations being presented here, a similar covariance matrix is considered. To provide a closer comparison to the ideal case discussed in Section 3.1, the off-diagonal entries are not 0. Instead, they are chosen such that the corresponding correlation matrix is compound
symmetric with the correlation coefficient $\rho = 0.3$. This means that every off-diagonal entry in the first row and column of the covariance matrix is defined as $p^{e/2} \cdot 0.3$, and all other off-diagonal entries are 0.3. The diagonal entries are all defined as in Ahn et al. (2007). In cases where $e > 1$, both $tr(\Sigma)/p$ and $tr(\Sigma^2)/p^2$ diverges as $p \to \infty$ for this spiked covariance matrix, the details of which are given in Appendix B. Three cases are compared in Table 4 and Figure 4. In one, $e = 0.5$. This does not mark a violation of the assumptions of the U-test but serves to compare the cases where (A2) and (A3) are violated through a highly spiked covariance matrix with the case of a spiked covariance matrix that still satisfies the assumptions. Further, simulations are made with $e = 1.1$ and $e = 1.5$, which constitutes slight and moderate violations of (A2) and (A3). Note, though, that all three cases must be considered spiked covariance matrices. While only the second and third are highly spiked in the sense that they violate the assumptions of the U-test, the first would also be considered severe in many analysis situations.

For additional ease of comparison, Table 4 also contains results from the ideal conditions, which corresponds to the case where $e = 0$. Further, in Figure 4, the power of the U-test under ideal conditions is also provided.

The rejection rate under the null seem to be lower in the ideal case as compared to the case where $e = 0.5$, but the differences are neither very large nor consistent. In terms of power, the tests perform even more similarly, with barely noticeable differences between $e = 0$ and $e = 0.5$. In the cases with slight and moderate violations of the two assumptions, the performance seems to follow the same patterns in terms of size as it does under ideal conditions; the permutation test is not consistently liberal or conservative for the smaller sample sizes but conservative for $n = 50$. The U-test rejects relatively close to the nominal significance level for the cases where $e = 1.1$ and $e = 1.5$. However, the power of both tests suffers under the slight violation. Under the moderate violation, the power of the tests never increases far beyond the nominal significance level in conditions of simulations presented here – the U-test never reaches a rejection rate above 12.9%. This poor performance comes as no surprise, since the variation of the first variable is bound to dominate the test statistic even for relatively small dimensions in cases where no mean is larger than $\delta$, which is at most set to 1.2. After all, the variances of the first variables are approximately 89, 1 000 and 11 180 in the conditions presented in Figure 4 where $e = 1.5$. Rather, it might be seen as more surprising that the tests perform relatively well even in the case of the slight violation (where $e = 1.1$), despite the
variances of approximately 27, 158 and 931 that the first variables have in this case. In this case, the power of the U-test is still acceptable in the case of the moderately large sample size (n = 50) for larger deviations from the null.

**Table 4**: Rejection rates of the U-test and the permutation test under the null, $\alpha = 0.05$. Multivariate normal data, spiked covariance matrix with varying $e$.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>$e = 0$</th>
<th>$e = 0.5$</th>
<th>$e = 1.1$</th>
<th>$e = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 10$, $p = 20$</td>
<td>$Z_{U}$</td>
<td>0.042</td>
<td>0.048</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.046</td>
<td>0.052</td>
<td>0.045</td>
</tr>
<tr>
<td>$n = 20$, $p = 100$</td>
<td>$Z_{U}$</td>
<td>0.049</td>
<td>0.057</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.051</td>
<td>0.063</td>
<td>0.046</td>
</tr>
<tr>
<td>$n = 50$, $p = 500$</td>
<td>$Z_{U}$</td>
<td>0.059</td>
<td>0.054</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.036</td>
<td>0.042</td>
<td>0.031</td>
</tr>
</tbody>
</table>

**Figure 4**: Power. Rejection rates (y axes) over varying values of $\delta$ (x axes) for the U-test (solid line) and the permutation test (dashed line). $\alpha = 0.05$. Multivariate normal data, spiked covariance matrix with varying $e$. For comparison: The U-test under ideal conditions (dotted line).
4 Conclusions

The robustness of the U-test is examined under one condition violating assumption (A1), where data are generated from a multivariate Cauchy distribution. Further, two conditions violating assumption (A3) are examined, such that the quantity controlled therein \((tr(\Sigma^2)/p^2)\) converges to zero and diverges as \(p \to \infty\), respectively. Under (A3), this quantity is instead uniformly bounded away from zero and infinity. The divergence condition is also a violation of assumption (A2), since \(tr(\Sigma)/p\) also diverges with increasing dimension under this condition.

The performance of the U-test under these robustness simulations is compared to the permutation counterpart of the test, as well as its performance under ideal conditions.

In general, the U-test performs poorly only in the cases under which hypothesis testing is difficult in general, and all cases of poor performance of the U-test are also cases in which the permutation test exhibits lacking performance. These results are found in spite of the fact that the assumption of the permutation test holds in all examined cases. The conditions being examined under which the tests perform poorly are for data generated from the multivariate Cauchy distribution, which is characterized by very heavy tails, and for data generated with a highly spiked covariance structure. However, in the case of a more moderately spiked covariance matrix which does not violate the assumptions of the U-test, the performance of the test is still good. All in all, it seems like the U-test is quite robust to the violations examined herein, since the alternate test has not outperformed the U-test substantially in any simulation performed.

The last examined condition, in which \(\Sigma = I\) and \(tr(\Sigma^2)/p^2\) converges to zero with increasing dimension, was also expected to yield decreasing performance from the U-test, since assumption (A3) is violated. However, the performance is instead found to increase dramatically, with rejection rates under the null remaining near to the nominal significance level but with a considerable increase in power compared to the ideal case. One possible reason is that the conditions denoted ideal in these simulations, under which all assumptions hold, are still defined with a spiked covariance matrix of compound symmetric structure. The fact that assumption (A3) of the U-test is violated might be of less practical importance than the fact that the identity covariance structure is not a spiked covariance matrix. It is possible that conditions that violate assumption (A3) but under which assumption (A2) is still intact has little effect on performance in general, but the results in this thesis are not sufficiently general or exhaustive to provide conclusive proof of this point.
References


Appendix A – Additional Results

Here, some additional simulation results are given.

In Table 5 and Figure 5, results are presented for multivariate normal data, where the covariance matrix is compound symmetric with $\sigma^2 = 1$ and $\rho = 0.3$. Note that these results are very similar to those presented in Section 3.1.

Table 5: Size. Rejection rates of the U-test and the permutation test under the null, $\alpha = 0.05$. Multivariate normal data, CS covariance matrix with $\sigma^2 = 1$ and $\rho = 0.5$.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>n = 10</th>
<th>n = 20</th>
<th>n = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 20$</td>
<td>$Z_U$</td>
<td>0.051</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>$p = 100$</td>
<td>$Z_U$</td>
<td>0.052</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.051</td>
<td>0.054</td>
</tr>
<tr>
<td>$p = 500$</td>
<td>$Z_U$</td>
<td>0.051</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>$Z_{perm}$</td>
<td>0.050</td>
<td>0.051</td>
</tr>
</tbody>
</table>

Figure 5: Power. Rejection rates (y axes) over varying values of $\delta$ (x axes) for the U-test (solid line) and the permutation test (dashed line). $\alpha = 0.05$. Multivariate normal data, CS covariance matrix with $\sigma^2 = 1$ and $\rho = 0.5$. 
Appendix B – The Spiked Covariance Matrix

In this appendix, it is shown that $\text{tr}(\Sigma)/p$ and $\text{tr}(\Sigma^2)/p^2$, which are positive and finite under assumption (A2) and (A3), respectively, diverges in cases where $e > 1$ for the spiked covariance matrix defined in Section 3.2.2. It is also shown that these quantities do not diverge in cases where $e \leq 1$. This spiked covariance matrix is defined as

$$
\Sigma = \begin{pmatrix}
p^e & 0.3p^{e/2} & \cdots & 0.3p^{e/2} \\
0.3p^{e/2} & 1 & \cdots & 0.3 \\
\vdots & \vdots & \ddots & \vdots \\
0.3p^{e/2} & 0.3 & \cdots & 1
\end{pmatrix}_{p \times p}.
$$

It follows that

$$
\frac{\text{tr}(\Sigma)}{p} = p^{e-1} + \frac{(p-1)}{p}.
$$

First, note that the term $(p - 1)/p$ converges to 1 regardless of $e$. In the case where $e < 1$, $p^{e-1}$ converges to 0 as $p \to \infty$ and there is no term that diverges with increasing dimension. If $e = 1$, $p^{e-1} = 1$, and there is again no diverging term. If, instead, $e > 1$, $p^{e-1} \to \infty$ as $p \to \infty$. Thus, it is shown that assumption (A2) is violated by the spiked covariance matrix if and only if $e > 1$.

From the definition of the spiked covariance matrix, the diagonal entries of $\Sigma^2$ can also be found. Let $d_1 \ldots d_p$ denote these diagonal entries. Then,

$$
d_1 = p^{2e} + (p - 1)0.09p^e \\
d_2 = 0.09p^e + (p - 2)0.09 + 1 \\
\vdots \\
d_p = 0.09p^e + (p - 2)0.09 + 1.
$$

It follows that $\text{tr}(\Sigma^2)$ is found to be

$$
\text{tr}(\Sigma^2) = d_1 + (p - 1)d_2 = \\
= p^{2e} + (p - 1)0.09p^e + (p - 1)0.09p^e + (p - 1)(p - 2)0.09 + p - 1 = \\
= p^{2e} + 0.18p^{e+1} - 0.18p^e + 0.09p^2 - 0.27p + p - 1.27.
$$

Thus, $\text{tr}(\Sigma^2)/p^2$ can be expressed as

$$
\frac{\text{tr}(\Sigma^2)}{p^2} = p^{2e-2} + 0.18p^{e-1} - 0.18p^{e-2} + 0.09 - 0.27p^{-1} + p^{-1} - 1.27p^{-2}.
$$

Clearly, as $p \to \infty$, the last three terms all approach 0, and the constant term (0.09) neither diverges or converges to zero. Further, in cases where $e < 1$, the first three terms also approach 0, and $\text{tr}(\Sigma^2)/p^2$ will not diverge. If $e = 1$, these three terms will either be constants or converging to 0, also yielding non-divergence of $\text{tr}(\Sigma^2)/p^2$. If, instead, $e > 1$, the first two terms diverge as $p \to \infty$, and the speed of this divergence will depend on $e$. As stated in Section 3.2.2, it is thus shown that assumption (A3) is also violated by the spiked covariance matrix if and only if $e > 1$. 

19