Approaches for Testing the Joint Hypotheses for Multivariate Normal Distribution. Applications in Panel Data Models

Inauguraldissertation
zur Erlangung des akademischen Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Volkswirtschaftlichen Fakultät
an der Ludwig-Maximilians-Universität München

vorgelegt von
Valerii Dashuk

2021

Referent: Prof. Dr. Joachim Winter
Korreferentin: Prof. Dr. Derya Uysal
Promotionsabschlussberatung: 14.07.2021

Acknowledgments

Writing this dissertation has become possible thanks to my colleagues, mentors, family, and friends. They all contributed to this research in numerous ways. I am grateful to all of them for their help.

First and foremost, I would like to thank my supervisor, Prof. Dr. Joachim Winter. He was always ready to help me in my initiatives and provided me a lot of academic freedom. I am very grateful to him for all the support and supervision he provided. Secondly, I would like to thank my first-year supervisor, Prof. Dr. Flaig, who provided me a lot of support in my research and teaching. I am also grateful to my second supervisor, Prof. Dr. Derya Uysal. She provided me with very important guidance in large and small obstacles and questions in my research. I would also like to thank Prof. Dr. Uwe Jensen, who agreed to be my oral examiner. I am very grateful to him for supervising me during my master's studies. Further, the result I obtained during his supervision has become a basis for the research, developed in this dissertation.

Next, I would like to thank the Munich Graduate School of Economics and particularly the Research Training Group GRK 1928 for funding, great working conditions, that make it possible to center on my research. I would like to express my gratitude to the administration and particularly to Prof. Dr. Carsten Eckel, Prof. Dr. Florian Englmaier, Prof. Dr. Uwe Sunde, and Prof. Dr. Joachim Winter, who were responsible for many official issues and for organizing different meetings, workshops, seminars. And special thanks for the possibility to take part both as an organizer and as a participant in Munich Young Economists' Meeting 2019, which took place in Kochel. I would also like to thank Julia Zimmermann for her great help with all administrative questions during my first years in Munich. Furthermore, I appreciate Caroline Benkert helping with the administrative part of my dissertation. Additionally, I am grateful to all my colleagues who supported me during my doctoral studies, especially Lukas and Felix.

Besides, I would like to thank my teacher, Serhii Matviychuk, who motivated me to progress in studying economics, math, and physics. Thanks to him I took part in more than 50 science competitions and became the best student economist in Ukraine. It motivated me to continue my education and growth further in the Christian-Albrecht University of Kiel and in the Ludwig Maximilian University of Munich.

Last but not least, I would like to thank my mother, my uncle, and my grandparents for their incredible everyday support. Unfortunately, my grandfather did not live just a little to this moment. But I can still feel his care and I believe he would be proud of me.

Valerii Dashuk
Munich, March 2021

Contents

List of Figures

List of Tables

Preface

One cannot imagine modern economics without collecting and handling data. For example, microeconomics works with data collected from individuals, households, companies. Based on the data collected, one tries to find the best-fitted model among the already existing ones or develop a new model, that may describe certain aspects of economics on the micro-level more thoroughly. The theoretical basis of microeconomics includes diverse models, that aim to maximize the profit of a company (see Varian (1992), p.23), the total utility of the household (see Varian (1992), p.94), or to find the optimal price that will equalize demand and supply on the certain market (for competitive markets see Varian (1992), p.215, for monopoly see Varian (1992), p.233).

The same methodology that connects data and theoretical model works for macroeconomics. But it operates more with aggregate data on the state or country level. This could be data on Gross Domestic Product, on Gross National Product, on inflation rate or unemployment rate. But all the parts and branches of economics at some point of the analysis refer to econometrics to handle the data observed. It provoked the origin of microeconometrics, macroeconometrics, financial econometrics, spatial econometrics. Each of the branches of econometrics demands approaches for the specific assumptions and therefore treatment of the data. For example, the possibility to work with the heteroscedasticity of the individuals and firms is the most vivid requirement for the methods that are applied in microeconometrics (see Cameron, Trivedi (2005), p.5).

Simultaneously, increasing amounts of data collected together with the computerization in all life aspects provoked the progress in econometric science. New methods for the treatment of large multidimensional statistics arose and thus, provided better evidence, more effective estimations, and more exact forecasts. Given the growing demand from a range of different branches of economics for the methods that may efficiently analyze data and a supply from the statistics and econometrics, they stimulate their mutual progress. And my dissertation contributes to this development, outlined below.

To begin with, testing statistical hypotheses is one of the ways to verify the estimation results, the choice of the model, or the significance of the independent variables included in the model. Starting from the most popular Student's t-test and Fisher F-test, a huge variety of testing techniques are available nowadays for different specifications (see Lehmann and Romano (2005) describing many of the existing tests).

In this dissertation I aim to development of the joint tests for the parameters of the multivariate normal distribution. Firstly, an exact test for mean and variance of the normal distribution, constructed by Mood (1950), p.227, described and investigated the issue for the univariate case. Univariate approaches were also presented by

Douglas (1993), Jensen (1995), Arnold & Shavelle (1998), etc. In my master's thesis, I have also derived a joint test, called CCR, for the parameters of univariate normal distribution (see Appendix 1.III). All of these tests were designed to verify whether two univariate normal distributions (usually one of them is a sample distribution) are close enough to say that they may be treated as equal. Extending the possibility to test not only univariate but also multivariate normal distribution and its parameters is the essential question that has appeared. My dissertation contributes to this subject in various ways. In particular, Part 1 and Part 3 introduce new approaches to test the parameters of the bivariate normal distribution with possible extension to the multivariate case, and Part 2 suggest an implementation of the test to the panel data models.

Following paragraphs provide a summary of all three Parts. Part 1 presents a CCR approach for the bivariate normal distributions. Part 2 is the continuation of Part 1 for the further application of the CCR method in panel data models. Part 1 together with Part 2 are self-contained and may be read independently. Part 3 presents the other testing technique for parameters of the bivariate normal distribution, based on the Mood test (Mood (1950), p.227). Part 3 is also self-contained and may be read independently. Separate bibliographies, appendices, and programming codes reproducing the algorithms are presented after each Part of the dissertation.

The first Part advances the CCR approach for using with the bivariate samples. It constructs a measure, that allows checking whether the sample of bivariate random vectors may originate from a normal distribution with some predefined parameter values. Such a sample may be obtained as a composition of the data from two variables. Testing this bivariate sample may describe the connections between two variables (this connection is expressed in correlation coefficient), as well as the individual characteristics of the variables (means and variances). A large piece of the research in this Part is dedicated to 16 different cases of the CCR test measure. On one hand, it establishes a smooth and unimodal measure. On the other hand, the analysis of the cases gives a possibility to apply the CCR approach for testing the distribution parameters both individually and jointly. Furthermore, in Part 1 I propose a procedure for using the CCR test also for multivariate samples. Then I analyze the properties of the derived bivariate CCR method concerning the shape of the confidence set, the ratio of the rejected samples, robustness to samples from non-normal distribution and samples with outliers. At the end of Part 1, I present an example of how the CCR procedure may be implemented in Seemingly Unrelated Regressions (SUR) models.

Second part of my dissertation contributes to the methods that test for serial correlation in panel data models. A variety of such tests was developed in recent decades. For example, Wooldridge-Drukker test (Wooldridge (2002) and Drukker (2003)), tests based on the Lagrange Multiplier (Baltagi & Li (1995), Born & Breitung (2016)), etc. The serial correlation test, developed in this dissertation, is based on the CCR approach. I describe the testing scheme and compare it to the known testing techniques. An important result that was obtained is the robustness of the test to panels with heteroscedasticity. Unlike most of the other tests, the CCR-based test proved

itself as a method that may work with heteroscedastic panels, which is one of the attributes of microeconomic data (Cameron, Trivedi (2005), p.5).
Third part of my dissertation advances another path for jointly testing the parameters of the bivariate normal distribution. This path is based on the idea of Mood (1950), who proposed a joint test for mean and variance of normal distribution by merging two individual tests. I appealed to his work and combined two multivariate tests, Hotelling t^2 for the means and Wishart for the covariance matrix. As a result, a joint mean-variance test for the bivariate normal distribution, named the bivariate Mood test, was obtained. Additionally, the volume of the confidence set produced by the bivariate Mood approach was minimized for different values of significance level and sample size. Furthermore, Part 3 compared the bivariate CCR and Mood test for the real significance levels, computation speed, shape of the confidence set, and robustness to samples with outliers and from other distributions (i.e. not from the normal one). The result of this comparison showed the bivariate Mood approach as a good alternative to the CCR method, with a not optimal form of the confidence set, but with a significant advantage in computation speed.

References

Arnold, B., Shavelle, R., 1998. Joint confidence sets for the mean and variance of a Normal distribution. *The American Statistician*; 52; 133-140.

Baltagi, B., Qi, L., 1995. Testing AR(1) against MA(1) disturbances in an error component model. *Journal of Econometrics*; 68(1); 133-151.

Born, B., Breitung, J., 2016. Testing for Serial Correlation in Fixed-Effects Panel Data Models. *Econometric Reviews*; 35; 1290-1316.

Cameron, C., Trivedi, P., 2005. Microeconometrics Methods and Applications. Cambridge University Press; New York; 2005.

Douglas, J., 1993. Confidence regions for parameter pairs. *The American Statistician*; 47; 43-45.

Drukker, D., 2003. Testing for Serial Correlation in Linear Panel-data Models. *The Stata Journal*; 3; 168-177.

Jensen, U., 1995. A review of the derivation and calculation of Rao distances with an application to portfolio theory. In Maddala, G., Phillips, P., Srinivasan, T. (eds.), Advances in econometrics and quantitative economics. Blackwell; 2008; 413-462.

Lehmann, E., Romano, J., 2005. Testing Statistical Hypotheses. Springer Science; New York; 2005.

Mood, A., 1950. Introduction to the Theory of Statistics. McGraw-Hill; New York; 1950.

Varian, H., 1992. Microeconomic analysis. W.W. Norton & Company, Inc.; New York; 1992.

Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel Data, first edition. The MIT Press; Cambridge, Massachusetts; 2002.

Part 1 Confidence Regions for Bivariate Normal Distributions. Approach Based on Cumulative Distribution Function

1.1 Introduction

To study the behavior of a variable and understand how it depends on the other variables, one usually has to construct a model that reflects real-world connections. It is a necessity for the cases to start from the simpler ones (e.g. how distance depends on velocity or how earnings depend on education level) up to more complex models with nonlinear functional forms and big panel data.

Parameter estimation of the model is the next step that one has to do, either before or after the model selection. During the estimation process one's task is not only to find values that fit the model, but also to construct confidence regions and check the significance of the estimated values. Both of these tasks may be performed using different tests. For example t-test and Wald test can check the significance of one or group of the estimators, while F-test may even help one to select the model.

In this Part I will present an extended test called the Cumulative Distribution Function Confidence Region (CCR). It may be used in testing and constructing joint confidence sets for serial correlation, inter-temporal and cross-equation relations. This makes it an essential instrument for the estimation of models in time-series analysis, panel econometrics, portfolio analysis, etc. The CCR test works as an alternative to portmanteau tests. For example, the Box-Pierce test for autocorrelation, introduced by Box & Pierce (1970) and its improvement introduced by Ljung & Box (1978). Furthermore, CCR test in this Part is a possible substitute for the LM based Breush-Godfrey test, introduced by Godfrey (1978) and a range of tests, described by Born & Breitung (2016).

The CCR test uses a technique that is based on the difference between the probability densities of two distributions. Firstly, this approach was introduced in my master thesis, but only for a univariate case. Find a brief description of a univariate CCR approach in Appendix 1.III. In this Part I will extend the CCR method to the bivariate case and demonstrate how it may be used on any finite dimensionality of the sample. A proposed extension of the test is non-parametric and therefore its generality makes it a good alternative to a multitude of existing tests. The CCR approach also gives new testing opportunities, demonstrated in this Part.

To introduce hypothesis that CCR technique tests, suppose an independent and identically distributed (iid) sample of m-dimensional random vectors $\{\mathbf{X}_i\}_{i=1}^n = \mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ from multivariate normal distribution with unknown parameters $N(\cdot, \cdot)$. Additionally, note that mean is defined as $\vec{\mu} \in \mathbb{R}^m$ and covariance matrix is defined as $\Sigma \in \mathbb{R}^{m \times m}$. Based on the given sample $\{\mathbf{X}_i\}_{i=1}^n$, the CCR test checks whether this sample with estimated mean $\vec{\mu}_1$ and covariance matrix Σ_1 may follow normal

distribution with the proposed parameters $(\vec{\mu}_2, \Sigma_2)$. Therefore, the Null hypothesis is formulated as (1.1):

$$H_0 : (\vec{\mu}_1, \Sigma_1) = (\vec{\mu}_2, \Sigma_2) \tag{1.1}$$

Given the Null hypothesis, defined in statement (1.1), one can also set up a confidence region $\mathcal{R}(\mathbf{X})$ for a pair $(\vec{\mu}, \Sigma)$, as a set of all points, that are not rejected by H_0 with a given confidence level $(1 - \gamma)$, where γ is significance level. Formally this means (1.2):

$$P((\vec{\mu}, \Sigma)) \in \mathcal{R}(\mathbf{X})) = 1 - \gamma \tag{1.2}$$

To be able to perform the CCR test in the form, introduced in (1.1) and (1.2), three important cases need to be separated. Firstly, the univariate case with $m = 1$, that is described in Appendix 1.III. For a given sample of iid random numbers from Normal distribution, it jointly tests whether sample parameters (mean and variance) may be equal to some specified values. Implementation of the univariate CCR test resulted in a confidence set in a form of an ellipse, which coincided with majority of the other univariate tests. In fact, the CCR test was developed as an alternative for previously existing univariate tests. For instance, the exact test derived by Mood (1950), p.227 as a conjunction of t-test and χ^2-test, and an assortment of tests based on Lagrange Multiplier (LM), introduced by Arnold & Shavelle (1998). Or, a test that uses non-Euclidean properties of variance, introduced by Jensen (1995). Additionally Frey, Marrero, Norton (2009) and Zhang (2017) derived confidence sets with a minimum possible area.

A comparison of all these tests, with respect to their efficiency, computation speed, confidence set area (using already mentioned tests designed to minimize this area), robust properties and distribution misspecification showed CCR as one of the best options and gave a motivation to expand CCR test for $m > 1$.

In bivariate case with $m = 2$, which is the main aim of Part 1, the CCR approach (1.1) enables us to jointly test two means, two variances and a covariance between dimensions of the bivariate random sample. These 5 mentioned parameters are tested with the bivariate CCR method and produce a confidence set in a form of 5-dimensional ellipse. It is a natural extension of univariate confidence set with the same properties. Such a massive increase in number of variables entangles the procedure of the bivariate CCR method. Nevertheless, the multi-variability of the CCR method brings many benefits. Firstly, with $m > 1$ there are no alternative ways to test (1.1), except of bivariate Mood approach that will be described in Part 3. Secondly, by restricting some parameters to be equal to a constant, one can separately test means, variances or correlation coefficient of the sample. After developing the CCR method for bivariate case in this Part, I also test it for efficiency, the difference between real and theoretical significance level γ, robustness and stability when working with non-normal samples and samples with outliers.

The third case that I separate comes from the distinct advantage of the CCR method. There is a possibility to extend it to m-dimensional case with any $m > 2$, which was not possible in the alternative approaches, such as Frey, Marrero, Norton (2009) or Arnold & Shavelle (1998) (they worked only with $m = 1$). In a situation when $m > 2$, the CCR approach allows us to run the test (1.1) by arranging all possible pairs of variables (totally $\binom{m}{2} = \frac{m(m-1)}{2}$ pairs) and testing them similarly as in the bivariate case. Therefore, development of the CCR technique to multivariate case with $m > 2$ increases the complexity of the calculations with the speed of $O(m^2)$. This coincides with the speed of increase of the number of parameters tested in (1.1) and makes the CCR test reasonable to use in multivariate cases.

Such a pairwise performing of the CCR test helps to implement the approach and interpret the results. Besides, in this Part I will exhibit how to apply the CCR technique in Seemingly Unrelated Regressions (SUR) model, introduced by Zellner (1962). In SUR models, connections between error terms of different equations is the essential property on which estimation is based. Pairwise comparison of these error terms helps to determine the group of independent equations and another group of dependent ones. Hence, this comparative procedure serves also as a decision-making tool. It divides all the equations into 2 groups, that may include more subgroups with the same properties. One of the group may be empty as well. For the first group one uses an estimation method that assumes independence between equations to get better results, e.g. ordinary least squares (OLS). For the second group one should apply a technique that allows dependencies between equations, e.g. feasible generalized least squares (FGLS), to achieve better estimation results. Depending on the correlations between error terms, the CCR approach applied in the SUR model at the end presented a reasonable improvement in estimation results.

The derived CCR method is a general and extensive instrument, which provides an alternative to already existing tests in standard situations. For instance, the Rao distance based test, derived by Jensen (1995) or even basic t-tests. Simultaneously it gives an opportunity to perform tests, construct confidence sets and make decisions on nonstandard multidimensional problems and on large amounts of data.

1.2 Model Construction

1.2.1 Comparison of Univariate and Bivariate Cases

Before the introduction of the model in this subsection I show why there is a need for an individual technique that constructs a bivariate confidence region and why combining two or more univariate confidence regions is inadvisable.

Suppose there is a sample $\{\mathbf{X}_i\}_{i=1}^n$ of two-dimensional iid random vectors. It is possible to define and construct a separate confidence set for each of the two subsamples $\{\mathbf{X}_{x,i}\}_{i=1}^n$ and $\{\mathbf{X}_{y,i}\}_{i=1}^n$ (here and further indexes x and y stand for the first and second dimension of the vector). Firstly, merging two confidence sets does not always give a possible confidence set for both dimensions, as it does not take the correlation between x and y into account. Secondly, a confidence set for only one of the dimensions is restricted because other variables are not taken into account. Thus,

it provides the same confidence set for any value of the variable, that is not included. Thirdly, merging separate confidence sets together does not take into account the fractions of significance level for each of the sets. To illustrate my points, I plotted confidence sets for univariate and bivariate cases on Figure 1.1.

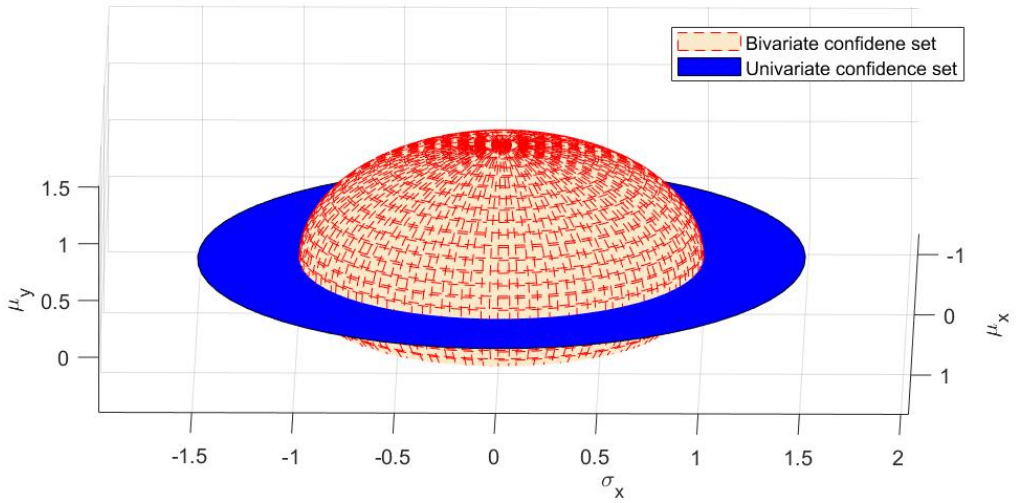

Figure 1.1: Univariate and Bivariate CCR Confidence Sets in 3-D

On Figure 1.1 I plotted univariate and bivariate confidence regions for some arbitrary sample, defined in $\mathbb{R}^2$. The univariate confidence set is for mean and variance of dimension x, and bivariate is projected on the same parameters plus mean of dimension y. It can be clearly seen that for different values of μ_y, the bivariate method gives different confidence sets for mean and variance of x as a cross-section of the volumetric figure. Univariate confidence set does not take dimension y into account, producing the same confidence set for each value of μ_y. It does not have any information about relations between variables, and if there is a strong positive or negative correlation, the univariate technique does not react to this fact. Moreover, the univariate confidence set distributes the non-rejection points only in (μ_x, σ_x) dimensions. And does not take into account that some points outside of this 2-D plane are more likely not to be rejected than the points that are already in the univariate confidence set.

Therefore, development of a more complex bivariate technique is an expedient task that may capture additional factors that influence the confidence set and p-values of the test thus producing more effective procedure.

1.2.2 Derivation of the Test Measure W

In this subsection I describe a procedure of constructing measure W, that is used for performing the CCR test. Additionally, I examine this measure for different properties and convergence.

I start with an iid sample from normal distribution $\{\mathbf{X}_i\}_{i=1}^n \in \mathbb{R}^2$. This sample may be described with its sample mean μ_1 and covariance matrix Σ_1. The aim of the CCR technique is to check whether the supposed sample may be considered as a sample from normal distribution with given parameters $N(\mu_2, \Sigma_2)$ and with a predefined

significance level γ. Hence, I can construct two probability density functions (PDFs) of the bivariate normal distributions, $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$, for their further comparison (1.3):

$$f_i(x,y) = \frac{1}{2\pi\sigma_{x,i}\sigma_{y,i}\sqrt{1-\rho_i^2}} e^{-\frac{1}{2(1-\rho_i^2)}\left(\frac{(x_i-\mu_{x,i})^2}{\sigma_{x,i}^2}+\frac{(y_i-\mu_{y,i})^2}{\sigma_{y,i}^2}-\frac{2\rho_i(x_i-\mu_{x,i})(y_i-\mu_{y,i})}{\sigma_{x,i}\sigma_{y,i}}\right)} \tag{1.3}$$
$$i = 1, 2$$

Plotting two PDFs from equation (1.3) in Figure 1.2 illustrates how they may intersect, defining two regions where $f_1(x, y)$ has a higher likelihood and similarly where $f_2(x, y)$ has a higher likelihood. The larger likelihood difference at each point ergo the smaller probability for the sample to be drawn from the proposed distribution $N(\mu_2, \Sigma_2)$.

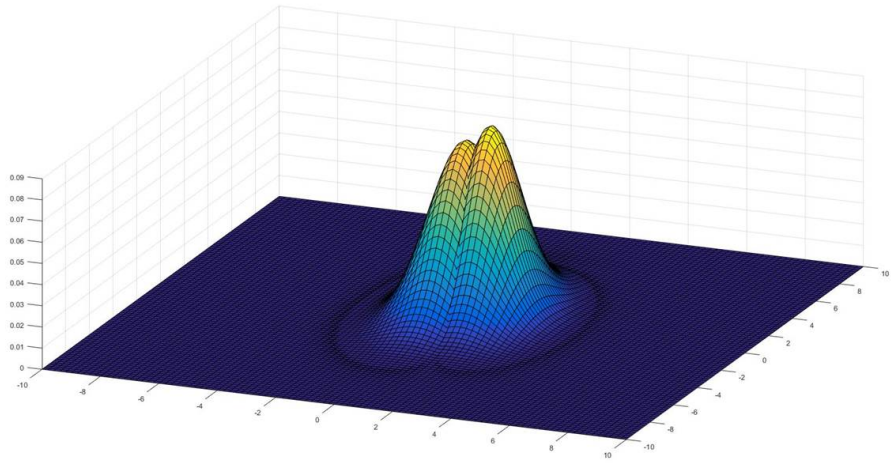

Figure 1.2: Intersection of Two PDFs of Bivariate Normal Distribution

Figure 1.2 gives the idea of how a measure W between two distributions may be constructed. Divide the plain with a large number of vertical and horizontal lines. Take some small rectangular region $(\triangle\ x_i, \triangle\ y_j)$ from the domain of PDFs and calculate the difference between probabilities of getting into this rectangle. For a small enough region one can calculate this probability as the likelihood at one of the points (x_i, y_j) of the rectangle multiplied by the area of the rectangle $\triangle\ x_i * \triangle\ y_j$. In the next step, avoid negative values by defining W as the sum of absolute differences of probabilities to get into each of the formed rectangles of the domain ($(x, y) \in \mathbb{R}^2$ for normal distribution). Equation (1.4) designates measure W and transforms it from a function of probabilities to a function of PDFs using the definition of the integral sums:

$$
\begin{gathered}
W = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |P_{1ij} - P_{2ij}| = \\
\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |f_1(x_i, y_j) \triangle x_i \triangle y_j - f_2(x_i, y_j) \triangle x_i \triangle y_j| \overset{\triangle x_i, \triangle y_j > 0 \ \forall i,j}{=} \\
\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |(f_1(x_i) - f_2(x_i))| \triangle x_i \triangle y_j \overset{\triangle x_i, \triangle y_j \to 0}{\longrightarrow} \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f_1(x,y) - f_2(x,y)| dxdy
\end{gathered}
\tag{1.4}
$$

For the next step remember, that the whole domain may be divided into 2 subsets, depending on which of the PDFs is higher. Denote one of the regions by $\mathbb{A}$, then the second one is $\mathbb{R}^2 \backslash \mathbb{A}$. Note that for symmetry reasons I do not define which of the PDFs should have higher likelihood in the set $\mathbb{A}$, but use absolute values instead. Points, where $f_1(x,y) = f_2(x,y)$, i.e. intersection of PDFs, either define a 2-D curve, therefore have zero area, or cover the whole plane and give a degenerate case, when two PDFs coincide. Using the region $\mathbb{A}$ and normalization property of PDF, simpler formula for measuring W is obtained in equation (1.5):

$$
\begin{gathered}
W = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f_1(x,y) - f_2(x,y)| dxdy = \\
\left| \iint_{\mathbb{A}} (f_1(x,y) - f_2(x,y)) dxdy \right| + \left| \iint_{\mathbb{R}^2 \backslash \mathbb{A}} (f_2(x,y) - f_1(x,y)) dxdy \right| = \\
\left| \iint_{\mathbb{A}} (f_1(x,y) - f_2(x,y)) dxdy \right| + \left| \iint_{\mathbb{A}} ((1 - f_2(x,y)) - (1 - f_1(x,y))) dxdy \right| = \\
2 \left| \iint_{\mathbb{A}} (f_1(x,y) - f_2(x,y)) dxdy \right|
\end{gathered}
\tag{1.5}
$$

From equation (1.5) I conclude that because of the normalization property, it does not matter which of the PDFs is higher on the set $\mathbb{A}$ and one may integrate either trough $\mathbb{A}$ or $\mathbb{R}^2 \backslash \mathbb{A}$ sets.

1.2.3 Properties of the Measure W

To be a measure for testing a hypothesis (1.1), W should observe certain properties. Firstly, it describes distance in a non-Euclidean space. Following statement (1.6), increasing the covariance matrix decreases the 'distance' between two distributions, assuming that the means stay unchanged. Because the variance-covariance matrix describes the level of dispersion of the variable around the mean, increasing of the dispersion level brings more likelihood to the points further from the mean and vice versa. In degenerate case influence of means on W is removed and zero variance produces the upper bound of measure W, while infinite variance provides $W = 0$.

$$\begin{gathered}\forall \mu_1, \mu_2, \quad \forall ||\tilde{\Sigma}|| > ||\Sigma|| : \\ W(N(\mu_1, \tilde{\Sigma}), N(\mu_2, \tilde{\Sigma})) < W(N(\mu_1, \Sigma), N(\mu_2, \Sigma))\end{gathered} \tag{1.6}$$

Secondly, W is a unimodal function of its parameters μ_1, Σ_1, μ_2 and Σ_2 with a minimum $W = 0$ achieved at $(\mu_1, \Sigma_1) = (\mu_2, \Sigma_2)$. This statement is clear for a general case, because W becomes zero only when two PDFs overlap. However, restricting some of the variables does not change the unimodality of W. This fact will be presented in latter sections of Part 1.

Thirdly, measure W is symmetrical in means. Since normal distribution is also symmetrical, changing the places of means of two distributions does not change their intersection curve, therefore keeps measure W the same.

The next point is the boundedness of measure W. To show the possible bounds from below and above, use the normalization property of PDF (1.7):

$$\iint_{\mathbb{A}} f_1(x, y) dx dy \in [0, 1] \tag{1.7}$$

In equation (1.8) I use a term from (1.7) to derive the bounds of expression inside the modulus in (1.5):

$$\iint_{\mathbb{A}} (f_1(x, y) - f_2(x, y)) dx dy = \left(\iint_{\mathbb{A}} f_1(x, y) dx dy - \iint_{\mathbb{A}} f_2(x, y) dx dy \right) \in [-1, 1] \tag{1.8}$$

Therefore, W is bounded with lower and upper bounds that are defined in (1.9). In practice both of the bounds are unachievable because, as mentioned in the first property of W, only degenerate cases with infinite and zero variances brings to the lower and upper bounds respectively.

$$W = 2 \left| \iint_{\mathbb{A}} (f_1(x, y) - f_2(x, y)) dx dy \right| \in [0, 2] \tag{1.9}$$

Note that bounds defined for W in (1.9) also work as the bounds for critical values (critical values will be derived in the latter sections of Part 1).

As a final point of this subsection I discuss the convergence properties of measure W for sample size n and significance level γ. Following the Central Limit Theorem increased number of observations n decrease the variance of sample mean μ_1. Meaning that by increasing n sample parameters will get closer to its theoretical values. Thus, measure W, as well as critical values, monotonically decrease with $n \uparrow$. As a result, according to the monotone convergence theorem bounded and monotonically decreasing measure W converges to zero for $n \to \infty$. The same conclusion follows for critical values, that converge to zero by increasing sample size n.

A monotone decrease of measure W and critical values with the increase of significance level γ follows directly from a sequence of nested confidence regions defined by measure W. As significance level increase provokes the decrease of confidence region, each next confidence region lies fully inside the old confidence region. As a result, confidence regions are nested inside each other and with the significance level increase they converge to the only common point. Therefore measure W and critical values converge to zero with the increase of significance level γ.
To conclude this subsection, convergence and unimodality properties described here allow us to use measure W as an instrument for testing joint mean-variance hypothesis (1.1) and construct appropriate confidence regions.

1.2.4 Intersection of PDFs

The region of integration $\mathbb{A}$ for measuring W was previously defined as the set of all points where one of the PDFs has a higher likelihood. Hence, the bounds of this region may be obtained by equating two PDFs of bivariate normal distribution, given in equation (1.10):

$$f_1(x, y) = f_2(x, y) \tag{1.10}$$

Plugging in PDFs from (1.3) and rearranging similar terms provides a second order curve in equation (1.11):

$$\begin{gathered}
x^2\left(\frac{1}{2(1-\rho_1^2)\sigma_{x1}^2}-\frac{1}{2(1-\rho_2^2)\sigma_{x2}^2}\right)+y^2\left(\frac{1}{2(1-\rho_1^2)\sigma_{y1}^2}-\frac{1}{2(1-\rho_2^2)\sigma_{y2}^2}\right)+ \\
2xy\left(\frac{\rho_2}{2(1-\rho_2^2)\sigma_{x2}\sigma_{y2}}-\frac{\rho_1}{2(1-\rho_2^2)\sigma_{x1}\sigma_{y1}}\right)+ \\
2x\left(\frac{\mu_{x2}}{2(1-\rho_2^2)\sigma_{x2}^2}-\frac{\rho_2\mu_{y2}}{2(1-\rho_2^2)\sigma_{x2}\sigma_{y2}}-\frac{\mu_{x1}}{2(1-\rho_1^2)\sigma_{x1}^2}+\frac{\rho_1\mu_{y1}}{2(1-\rho_1^2)\sigma_{x1}\sigma_{y1}}\right)+ \\
2y\left(\frac{\mu_{y2}}{2(1-\rho_2^2)\sigma_{y2}^2}-\frac{\rho_2\mu_{x2}}{2(1-\rho_2^2)\sigma_{x2}\sigma_{y2}}-\frac{\mu_{y1}}{2(1-\rho_1^2)\sigma_{y1}^2}+\frac{\rho_1\mu_{x1}}{2(1-\rho_1^2)\sigma_{x1}\sigma_{y1}}\right)+ \\
\left(\frac{\mu_{x1}^2}{2(1-\rho_1^2)\sigma_{x1}^2}+\frac{\mu_{y1}^2}{2(1-\rho_1^2)\sigma_{y1}^2}-\frac{\rho_1\mu_{x1}\mu_{y1}}{(1-\rho_1^2)\sigma_{x1}\sigma_{y1}}-\frac{\mu_{x2}^2}{2(1-\rho_2^2)\sigma_{x2}^2}-\right. \\
\left.\frac{\mu_{y2}^2}{2(1-\rho_2^2)\sigma_{y2}^2}+\frac{\rho_2\mu_{x2}\mu_{y2}}{(1-\rho_2^2)\sigma_{x2}\sigma_{y2}}\right)=\log\left(\frac{\sigma_{x2}\sigma_{y2}\sqrt{1-\rho_2^2}}{\sigma_{x1}\sigma_{y1}\sqrt{1-\rho_1^2}}\right)
\end{gathered} \tag{1.11}$$

Denote the second order curve from equation (1.11) as Ω for further use. However, analysis and implementation of such a massive function is complicated. Thus, in the next subsection I propose a procedure to simplify Ω by getting rid of correlations.

1.2.5 Transformation Procedure

Earlier I derived the intersection Ω of two PDFs from (1.3) and demonstrated that this is a second-order curve. To simplify and decrease the number of different possible

curve shapes as a result of intersection, I use linear transformation to switch to another coordinate system. This transformation may affect distances and volumes, but keep invariant the ratios between them. Additionally, I impose restrictions on the Cartesian coordinates after the transformation:

- correlations ρ_1 and ρ_2 between pairs of random variables x_1 and y_1/ x_2 and y_2 respectively should be zeros
- normal distribution $N(\mu_2, \Sigma_2)$ should be standardized, i.e. $\mu_{x2} = \mu_{y2} = 0$ and $\sigma_{x2} = \sigma_{y2}$

In other words, I construct such a coordinate system that comparable distribution $N(\mu_2, \Sigma_2)$ becomes standard normal for any hypothesis (1.1). Therefore, the measure between two distributions W converts into $\widetilde{W}$ - distance between the distribution transformed from $N(\mu_1, \Sigma_1)$ and standard normal distribution. As far as linear transformations will not affect ratios between volumes and distances, measure $\widetilde{W}$ will be just a scaled measure W with the same properties.
In the next step I present the transformation process of a new coordinate system. Transformation coefficients are obtained by solving the next nonlinear system of equations (1.12) for v_{11}, v_{12}, v_{21} and v_{22}:

$$\begin{cases} \sum_{i=1}^{2}\sum_{j=1}^{2} v_{1i}v_{2j}\Sigma_1(i,j) = 0; \\ \sum_{i=1}^{2}\sum_{j=1}^{2} v_{1i}v_{2j}\Sigma_2(i,j) = 0; \\ \sum_{i=1}^{2}\sum_{j=1}^{2} v_{1i}v_{1j}\Sigma_2(i,j) = 1; \\ \sum_{i=1}^{2}\sum_{j=1}^{2} v_{2i}v_{2j}\Sigma_2(i,j) = 1 \end{cases} \tag{1.12}$$

Where $\Sigma_1(i,j)$ and $\Sigma_2(i,j)$ are just the elements of covariance matrices. This type of referring is used to simplify the notation of system (1.12). For example, $\Sigma_1(1,1) = \sigma_{x1}^2$, $\Sigma_2(1,2) = \rho_2\sigma_{x2}\sigma_{y2}$.
The structure of system (1.12) is next. The first and second equations of the system verify that correlations of transformed distributions will be zeros. The third and fourth part are optional and help to avoid underidentification of the system by equating variances of the comparable distribution $N(\mu_2, \Sigma_2)$ to 1.
Solution of system (1.12) may be obtained numerically using nonlinear solvers (e.g. $fsolve$ in MatLab). One can also use exact solution of this system:

- if $\sigma_{y1}^2\rho_2\sigma_{x2}\sigma_{y2} - \rho_1\sigma_{x1}\sigma_{y1}\sigma_{y2}^2 \neq 0$
  - define

$$\begin{aligned} D &= (\sigma_{y1}^2\sigma_{x2}^2 - \sigma_{x1}^2\sigma_{y2}^2)^2 - 4(\rho_2\sigma_{y1}^2\sigma_{x2}\sigma_{y2} - \rho_1\sigma_{x1}\sigma_{y1}\sigma_{y2}^2)* \\ &\quad (\rho_1\sigma_{x1}\sigma_{y1}\sigma_{x2}^2\rho_2\sigma_{x1}^2\sigma_{x2}\sigma_{y2}) \\ E &= \frac{\sigma_{x1}^2\sigma_{y2}^2 - \sigma_{y1}^2\sigma_{x2}^2 + \sqrt{D}}{2(\rho_2\sigma_{y1}^2\sigma_{x2}\sigma_{y2} - \rho_1\sigma_{x1}\sigma_{y1}\sigma_{y2}^2)} \\ F &= \frac{2(\rho_2\sigma_{y1}^2\sigma_{x2}\sigma_{y2} - \rho_1\sigma_{x1}\sigma_{y1}\sigma_{y2}^2)}{\sigma_{x1}^2\sigma_{y2}^2 - \sigma_{y1}^2\sigma_{x2}^2 - \sqrt{D}} \end{aligned} \tag{1.13}$$

– calculate the solution

$$\begin{cases} v_{11} = \frac{1}{\sqrt{\sigma_{x2}^2+2E\rho_2\sigma_{x2}\sigma_{y2}+E^2\sigma_{y2}^2}}; \\ v_{12} = \frac{E}{\sqrt{\sigma_{x2}^2+2E\rho_2\sigma_{x2}\sigma_{y2}+E^2\sigma_{y2}^2}}; \\ v_{21} = \frac{1}{\sqrt{\sigma_{x2}^2+2F\rho_2\sigma_{x2}\sigma_{y2}+F^2\sigma_{y2}^2}}; \\ v_{22} = \frac{F}{\sqrt{\sigma_{x2}^2+2F\rho_2\sigma_{x2}\sigma_{y2}+F^2\sigma_{y2}^2}}; \end{cases} \tag{1.14}$$

- if $\sigma_{y1}^2\rho_2\sigma_{x2}\sigma_{y2} - \rho_1\sigma_{x1}\sigma_{y1}\sigma_{y2}^2 = 0$

  – define

$$G = -\frac{\sigma_{x1}^2 + \rho_1\sigma_{x1}\sigma_{y1}}{\sigma_{y1}^2 + \rho_1\sigma_{x1}\sigma_{y1}} \tag{1.15}$$

  – calculate the solution

$$\begin{cases} v_{11} = \frac{1}{\sqrt{\sigma_{x2}^2+2\rho_2\sigma_{x2}\sigma_{y2}+\sigma_{y2}^2}} \\ v_{12} = \frac{1}{\sqrt{\sigma_{x2}^2+2\rho_2\sigma_{x2}\sigma_{y2}+\sigma_{y2}^2}} \\ v_{21} = \frac{1}{\sqrt{\sigma_{x2}^2+2G\rho_2\sigma_{x2}\sigma_{y2}+G^2\sigma_{y2}^2}} \\ v_{22} = \frac{G}{\sqrt{\sigma_{x2}^2+2G\rho_2\sigma_{x2}\sigma_{y2}+G^2\sigma_{y2}^2}} \end{cases} \tag{1.16}$$

Computed coefficients v_{ij} are the elements of linear transformation matrix V, which defines new random variables in (1.17)-(1.18):

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = V\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - V\begin{pmatrix} \mu_{x2} \\ \mu_{y2} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\tilde{x}} \\ \mu_{\tilde{y}} \end{pmatrix}, \begin{pmatrix} \sigma_{\tilde{x}}^2 & 0 \\ 0 & \sigma_{\tilde{y}}^2 \end{pmatrix}\right) \tag{1.17}$$

$$V\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} - V\begin{pmatrix} \mu_{x2} \\ \mu_{y2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \tag{1.18}$$

After the process of transformation, hypothesis (1.1) is reduced to (1.19), i.e. one needs to compute the measure $\widetilde{W}$ for comparison of converted sample $\{\widetilde{\mathbf{X}}_i\}_{i=1}^n$ with bivariate standard normal distribution, depending on 4 parameters: $\mu_{\tilde{x}}$, $\mu_{\tilde{y}}$, $\sigma_{\tilde{x}}^2$, $\sigma_{\tilde{y}}^2$.

$$\widetilde{H}_0 : (\overrightarrow{\widetilde{\mu}}, \widetilde{\Sigma}) = (\overrightarrow{0}, I) \tag{1.19}$$

In the next stage intersection curve $\widetilde{\Omega}$ of two PDFs from hypothesis (1.19) is derived in new coordinates. One plugs in zero correlations ρ_1, ρ_2 and unite standard deviations of the comparable distribution σ_{x2} and σ_{y2} into the general formula of intersection curve Ω (1.11). New intersection $\widetilde{\Omega}$ is defined in equation (1.20). It is a simplified (comparing to Ω) second order curve with respect to $(\tilde{x}, \tilde{y})$:

$$\frac{(\tilde{x}-\mu_{\tilde{x}})^2}{2\sigma_{\tilde{x}}^2}-\frac{\tilde{x}^2}{2}+\frac{(\tilde{y}-\mu_{\tilde{y}})^2}{2\sigma_{\tilde{y}}^2}-\frac{\tilde{y}^2}{2}=\log\left(\frac{1}{\sigma_{\tilde{x}}\sigma_{\tilde{y}}}\right) \tag{1.20}$$

1.2.6 Calculation of Measure $\widetilde{W}$

To compute after-transformation measure $\widetilde{W}$, all possible shapes of interception curve $\widetilde{\Omega}$ are to be considered in Figure 1.3. The cases depend on the values of parameters $\mu_{\tilde{x}}$, $\mu_{\tilde{y}}$, $\sigma_{\tilde{x}}$ and $\sigma_{\tilde{y}}$ after the transformation. Degenerate case stands for the situation of the same distributions with zero 'distance' W between them.

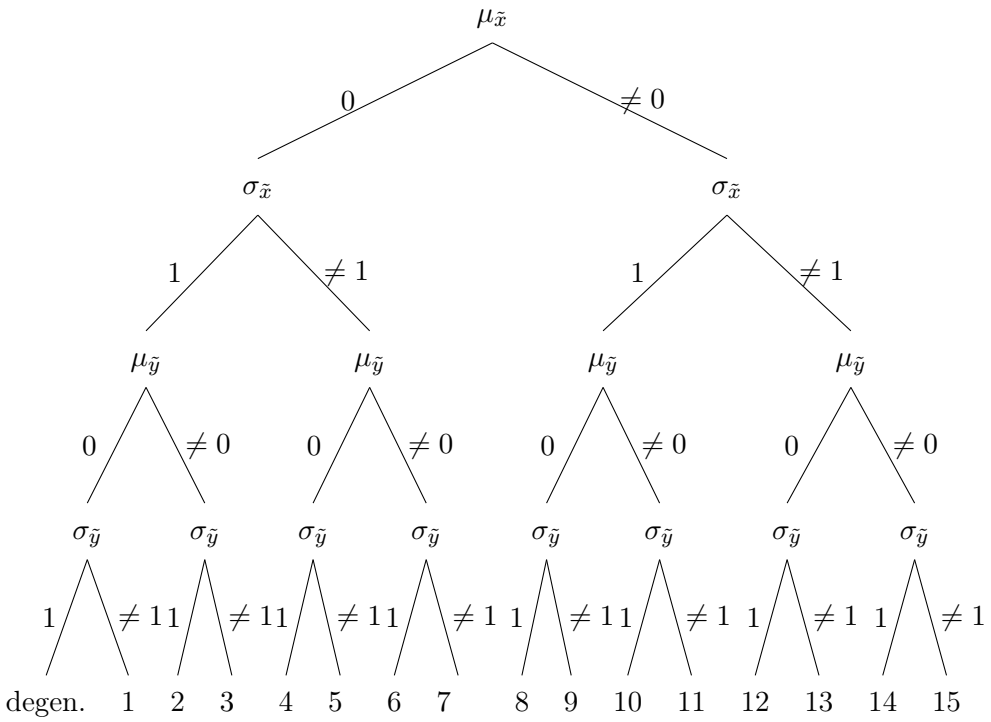

Figure 1.3: Extensive Tree for Cases of Intersection of PDFs that Define Measure $\widetilde{W}$ for the CCR Approach

Next I go through all the 15 pure cases and provide form of curve $\widetilde{\Omega}$ that forms an integration region for $\widetilde{W}$. The measure obtained from each of the cases is also described and simplified where it is possible.

(1) Case 1 arises when only $\sigma_{\tilde{y}}$ deviates from pattern (standard normal distribution). Curve $\widetilde{\Omega}$ is given by equation (1.21):

$$\tilde{y}_{1,2}=\pm\frac{\sigma_{\tilde{y}}}{\sqrt{|1-\sigma_{\tilde{y}}^2|}}\log(\sigma_{\tilde{y}}^2) \tag{1.21}$$

In this case the curve simplifies to two horizontal lines and produces a measure (1.22):

$$\widetilde{W} = 2|F_{\tilde{y}}(y_1) - F_{\tilde{y}}(y_2) - \Phi(y_1) + \Phi(y_2)| \tag{1.22}$$

Where $F_{\tilde{y}}(\cdot)$ represents marginal cumulative distribution function (CDF) of $\tilde{y}$ and $\Phi(\cdot)$ represents CDF of standard normal distribution.

(2) Case 2 occurs when only $\mu_{\tilde{y}}$ differs from the mean of standard normal distribution. Then produced intersection curve $\widetilde{\Omega}$ is a horizontal line given by (1.23):

$$y^* = \frac{\mu_{\tilde{y}}}{2} \tag{1.23}$$

And derived measure in (1.24):

$$\widetilde{W} = 2|F_{\tilde{y}}(y^*) - \Phi(y^*)| \tag{1.24}$$

(3) The third case presents the situation where both mean and variance of second variable $\tilde{y}$ deviate from standard normal. Then intersection $\widetilde{\Omega}$ is described by two horizontal lines (1.25):

$$y_{1,2} = \frac{\mu_{\tilde{y}}}{1-\sigma_{\tilde{y}}^2} \pm \sigma_{\tilde{y}} \sqrt{\left(\frac{\mu_{\tilde{y}}}{1-\sigma_{\tilde{y}}^2}\right)^2 - \frac{\log(\sigma_{\tilde{y}}^2)}{1-\sigma_{\tilde{y}}^2}} \tag{1.25}$$

Thus, measure is given as (1.26):

$$\widetilde{W} = 2|F_{\tilde{y}}(y_1) - F_{\tilde{y}}(y_2) - \Phi(y_1) + \Phi(y_2)| \tag{1.26}$$

(4) The fourth case illustrates the situation when only variance of $\tilde{x}$ differs from 1. Due to symmetry, this case is similar to the first one with $\widetilde{\Omega}$ simplified to 2 vertical lines in equation (1.27):

$$\tilde{x}_{1,2} = \pm \frac{\sigma_{\tilde{x}}}{\sqrt{|1-\sigma_{\tilde{x}}^2|}} \log(\sigma_{\tilde{x}}^2) \tag{1.27}$$

Measure $\widetilde{W}$ is also the same, just with marginal CDF of $\tilde{x}$ instead of $\tilde{y}$ (1.28):

$$\widetilde{W} = 2|F_{\tilde{x}}(x_1) - F_{\tilde{x}}(x_2) - \Phi(x_1) + \Phi(x_2)| \tag{1.28}$$

(5) The fifth case is produced by both variances of $\tilde{x}$ and $\tilde{y}$ been different from 1. Then PDFs' intercept $\widetilde{\Omega}$ is an ellipse (1.29):

$$\frac{\tilde{x}^2}{a^2} + \frac{\tilde{y}^2}{b^2} = 1, \text{ with } \begin{cases} a^2 = \frac{\sigma_{\tilde{x}}^2 \log(\sigma_{\tilde{x}}^2 \sigma_{\tilde{y}}^2)}{\sigma_{\tilde{x}}^2 - 1} \\ b^2 = \frac{\sigma_{\tilde{y}}^2 \log(\sigma_{\tilde{x}}^2 \sigma_{\tilde{y}}^2)}{\sigma_{\tilde{y}}^2 - 1} \end{cases} \tag{1.29}$$

And measure $\widetilde{W}$ is a double integral (1.9) over the area $\mathbb{A}$ bounded by this ellipse.

(6) Case number 6 describes the situation when the mean of one variable $\tilde{y}$ and variance of the other variable $\tilde{x}$ deviate from standard normal. This gives an intersect $\widetilde{\Omega}$ that takes the form of a parabola (1.30):

$$\tilde{y} = a\tilde{x}^2 + b, \text{ with } \begin{cases} a = \frac{1-\sigma_{\tilde{x}}^2}{2\sigma_{\tilde{x}}^2 \mu_{\tilde{y}}} \\ b = \frac{1}{2}\mu_{\tilde{y}} + \frac{\log(\sigma_{\tilde{x}})}{\mu_{\tilde{y}}} \end{cases} \tag{1.30}$$

Further, measure $\widetilde{W}$ is defined as a double integral (1.9) over the region $\mathbb{A}$ bounded by a parabola. As stated previously, it does not matter whether integrate over the area inside or outside the parabola.

(7) The seventh case sets up a situation when only the mean of variable $\tilde{x}$ coincides with 0, which is the mean of standard normal distribution. Generated intersection $\widetilde{\Omega}$ is showed in equation (1.31):

$$\frac{\tilde{x}^2}{c/a} + \frac{\left(y - \frac{\mu_{\tilde{y}}}{1-\sigma_{\tilde{y}}^2}\right)^2}{c/b} = 1, \text{ with } \begin{cases} a = (1 - \sigma_{\tilde{x}}^2)/\sigma_{\tilde{x}}^2 \\ b = (1 - \sigma_{\tilde{y}}^2)^2/\sigma_{\tilde{y}}^2 \\ c = \mu_{\tilde{y}}^2/(1 - \sigma_{\tilde{y}}^2) - \log(\sigma_{\tilde{x}}^2 \sigma_{\tilde{y}}^2) \end{cases} \tag{1.31}$$

The shape of $\widetilde{\Omega}$, defined in equation (1.31), is an ellipse, if both variances are simultaneously larger or smaller than 1. Otherwise, it takes a form of hyperbola. This allows us to calculate measure $\widetilde{W}$ using formula (1.9).

(8) In the eighth case the intersection curve behaves similarly to case (2), producing a vertical line (1.32):

$$x^* = \frac{\mu_{\tilde{x}}}{2} \tag{1.32}$$

And measure $\widetilde{W}$ is also obtained as the absolute difference of two CDFs (1.33):

$$\widetilde{W} = 2|F_{\tilde{x}}(x^*) - \Phi(x^*)| \tag{1.33}$$

(9) This case describes the behavior of intersection when mean of variable $\tilde{x}$ differs from 0 and variance of $\tilde{y}$ differs from 1. Hence, the form of $\widetilde{\Omega}$ is a parabola, defined in equation (1.34). The parabola is similar to the case (6), but rotated around $\tilde{x} = \tilde{y}$ line:

$$\tilde{x} = a\tilde{y}^2 + b, \text{ with } \begin{cases} a = \frac{1-\sigma_{\tilde{y}}^2}{2\sigma_{\tilde{y}}^2 \mu_{\tilde{x}}} \\ b = \frac{1}{2}\mu_{\tilde{x}} + \frac{\log(\sigma_{\tilde{y}})}{\mu_{\tilde{x}}} \end{cases} \tag{1.34}$$

Measure $\widetilde{W}$ is calculated by integration over the area $\mathbb{A}$ bounded by this parabola (1.9).

(10) The tenth case covers the situation with standard variances, but deviating means for both of the variables and gives the line intersect, presented in equation (1.35):

$$\tilde{y} = a\tilde{x} + b, \text{ with } \begin{cases} a = -\frac{\mu_{\tilde{x}}}{\mu_{\tilde{y}}} \\ b = \frac{\mu_{\tilde{x}}^2}{2\mu_{\tilde{y}}} + \frac{1}{2}\mu_{\tilde{y}} \end{cases} \tag{1.35}$$

Integrating in the equation (1.9) over one of the half-planes bounded by (1.35) results in the desired measure $\widetilde{W}$.

(11) The eleventh case sets up a situation, when only the variance of $\tilde{x}$ equals 1, while all the other parameters deviate from the standard normal distribution. This gives an intersection $\widetilde{\Omega}$ in form of a parabola along x-axis (1.36):

$$\tilde{x} = k(\tilde{y} - a)^2 + b, \text{ with } \begin{cases} k = (1 - \sigma_{\tilde{y}}^2)/(2\sigma_{\tilde{y}}^2 \mu_{\tilde{x}}) \\ a = \mu_{\tilde{y}}/(1 - \sigma_{\tilde{y}}^2) \\ b = \mu_{\tilde{y}}^2/(2\mu_{\tilde{x}}(\sigma_{\tilde{y}}^2 - 1)) + \frac{1}{2}\mu_{\tilde{x}} + \log(\sigma_{\tilde{y}})/\mu_{\tilde{x}} \end{cases} \tag{1.36}$$

Measure $\widetilde{W}$ is obtained by double integration, as defined in equation (1.9). Region of integration $\mathbb{A}$ is bounded by given parabola (1.36).

(12) In case 12, restrictions are similar to those already described in case (3), but now variable $\tilde{y}$ has a standard mean of 0 and a variance of 1. The second variable $\tilde{x}$ differs in both mean and variance from the comparable one in hypothesis (1.19). This provides an intersection $\widetilde{\Omega}$ of the PDFs along two vertical lines (1.37):

$$\tilde{x}_{1,2} = \frac{\mu_{\tilde{x}}}{1-\sigma_{\tilde{x}}^2} \pm \sigma_{\tilde{x}} \sqrt{\left(\frac{\mu_{\tilde{x}}}{1-\sigma_{\tilde{x}}^2}\right)^2 - \frac{\log(\sigma_{\tilde{x}}^2)}{1-\sigma_{\tilde{x}}^2}} \tag{1.37}$$

And $\widetilde{W}$ is simplified to equation (1.38), using CDFs of the normal distribution:

$$\widetilde{W} = 2|F_{\tilde{x}}(x_1) - F_{\tilde{x}}(x_2) - \Phi(x_1) + \Phi(x_2)| \tag{1.38}$$

(13) This case describes the situation when only the mean of variable $\tilde{y}$ coincides with the right-hand side of the transformed hypothesis (1.19), making the picture similar to the already described case (7) but 90 degrees rotated clockwise. Obtained intersection $\widetilde{\Omega}$ is presented in equation (1.39):

$$\frac{\left(x - \frac{\mu_{\tilde{x}}}{1-\sigma_{\tilde{x}}^2}\right)^2}{c/a} + \frac{\tilde{y}^2}{c/b} = 1, \text{ with } \begin{cases} a = (1-\sigma_{\tilde{x}}^2)/\sigma_{\tilde{x}}^2 \\ b = (1-\sigma_{\tilde{y}}^2)^2/\sigma_{\tilde{y}}^2 \\ c = \mu_{\tilde{x}}^2/(1-\sigma_{\tilde{x}}^2) - \log(\sigma_{\tilde{x}}^2\sigma_{\tilde{y}}^2) \end{cases} \tag{1.39}$$

The intersect may take elliptical form, if both $\sigma_{\tilde{x}}$ and $\sigma_{\tilde{y}}$ are simultaneously larger or smaller than 1, otherwise it is a hyperbola. And the measure $\widetilde{W}$ is calculated over the bounds, defined in (1.39).

(14) The fourteenth case inherits the properties of case (11), while only the variance of variable $\tilde{y}$ coincides with standard normal distribution. Therefore, intersection curve $\widetilde{\Omega}$ is the same parabola, but rotated counterclockwise to be along the y-axis (1.40):

$$\tilde{y} = k(\tilde{x}-a)^2 + b, \text{ with } \begin{cases} k = (1-\sigma_{\tilde{x}}^2)/(2\sigma_{\tilde{x}}^2\mu_{\tilde{y}}) \\ a = \mu_{\tilde{x}}/(1-\sigma_{\tilde{x}}^2) \\ b = \mu_{\tilde{x}}^2/(2\mu_{\tilde{y}}(\sigma_{\tilde{x}}^2-1)) + \frac{1}{2}\mu_{\tilde{y}} + \log(\sigma_{\tilde{x}})/\mu_{\tilde{y}} \end{cases} \tag{1.40}$$

And measure $\widetilde{W}$ is a double integral (1.9) over the area $\mathbb{A}$ bounded by the parabola (1.40).

(15) The last case describes the situation when all the parameters differ from the right-hand side of hypothesis (1.19). This gives the intersection curve $\widetilde{\Omega}$ presented in equation (1.20), with nothing to be reduced. To clearly see the form of the curve, I rewrite it as equation (1.41):

$$\frac{(\tilde{x}-a)^2}{b^2}+\frac{(\tilde{y}-c)^2}{d^2}=1,\ \text{with}\ \begin{cases} a=\frac{\mu_{\tilde{x}}}{1-\sigma_{\tilde{x}}^2} \\ b^2=\frac{\sigma_{\tilde{x}}^2}{1-\sigma_{\tilde{x}}^2}\left(\frac{\mu_{\tilde{x}}^2}{1-\sigma_{\tilde{x}}^2}+\frac{\mu_{\tilde{y}}^2}{1-\sigma_{\tilde{y}}^2}-\log(\sigma_{\tilde{x}}^2\sigma_{\tilde{y}}^2)\right) \\ c=\frac{\mu_{\tilde{y}}}{1-\sigma_{\tilde{y}}^2} \\ d^2=\frac{\sigma_{\tilde{y}}^2}{1-\sigma_{\tilde{y}}^2}\left(\frac{\mu_{\tilde{x}}^2}{1-\sigma_{\tilde{x}}^2}+\frac{\mu_{\tilde{y}}^2}{1-\sigma_{\tilde{y}}^2}-\log(\sigma_{\tilde{x}}^2\sigma_{\tilde{y}}^2)\right) \end{cases} \tag{1.41}$$

Analysis of the curve $\widetilde{\Omega}$, defined in (1.41), showed that when both variances of $\tilde{x}$ and $\tilde{y}$ are smaller or larger than 1 at the same time, then the intersection is an ellipse. Otherwise it takes the shape of a hyperbola. Measure $\widetilde{W}$ as previously is obtained by integrating over the region $\mathbb{A}$, bounded by $\widetilde{\Omega}$ (1.9).

Described algorithm of obtaining $\widetilde{W}$ may be used for the realization in software, like MatLab, Python, etc. To optimize and clarify the computations, one should additionally consider the area around the means. 3-σ rule may not give a sufficient precision thus, I used 5-σ rule, which can guarantee a precision up to the 7^{th} sign. Then one should integrate only through the common area bounded by $\widetilde{\Omega}$ and 5-σ region, and the remaining area may be negotiated. If the curve $\widetilde{\Omega}$ does not intersect with the 5-σ region, then the compared distributions are too far away from each other and hypothesis (1.19) may be rejected with p-value of less than 0.00001%.
Furthermore, all 15 cases, presented above not only simplify the calculations of test measure $\widetilde{W}$, but also are useful in the simpler hypotheses, when one wants to compare only some of the parameters. For example, case (8) compares only the means of variable $\tilde{x}$ and may be used as an alternative to the t-test. Equivalently, case (1) compares the variances of $\tilde{y}$. Thus, one may apply it instead of the χ^2-test.
Moreover, studies of $\widetilde{W}$ showed that differentiation into the cases does not bring any brakes to it, keeping $\widetilde{W}$ a smooth function. This will be demonstrated in the next subsections.

1.2.7 Critical Values

After the analysis of the possible cases that measure $\widetilde{W}$ can follow, critical values should be calculated. They will be used for decision making on hypothesis (1.19), as well as construction of confidence sets. To find the critical values, I generated a range of samples from normal distribution and calculated their measure $\widetilde{W}$ in comparison to the distribution from which these samples were generated. 10 million repetitions were computed for each of the sample sizes n from 10 to 50000, producing altogether 120 million repetitions. Then the calculated measures were sorted in descending order so that each i^{th} quantile defines critical value for i% significance level. With a step of 0.5% this procedure was performed to cover all the possible significance levels γ. Table 1.1 presents an extract from the empirically obtained critical values for different n and γ. The complete table comprises 2388 knots, for which the critical value was computed. See Appendix 1.I for the full table.

Table 1.1: Empirical Critical Values for the CCR Approach for Different Sample Sizes n and Significance Levels γ

n \ γ	0.2	0.1	0.05	0.025	0.01	0.005
10	0.6619	0.7640	0.8536	0.9349	1.0315	1.0980
25	0.3942	0.4494	0.4983	0.5428	0.5973	0.6357
50	0.2739	0.3106	0.3426	0.3721	0.4077	0.4330
100	0.1922	0.2176	0.2396	0.2593	0.2833	0.3002
250	0.1209	0.1366	0.1502	0.1624	0.1772	0.1873
500	0.0854	0.0965	0.1061	0.1146	0.1249	0.1321
1000	0.0603	0.0681	0.0748	0.0808	0.0880	0.0930
2000	0.0426	0.0480	0.0525	0.0566	0.0623	0.0650
5000	0.0270	0.0306	0.0335	0.0363	0.0398	0.0419
10000	0.0191	0.0215	0.0236	0.0256	0.0278	0.0291
20000	0.0135	0.0151	0.0167	0.0181	0.0200	0.0210
50000	0.0085	0.0096	0.0106	0.0114	0.0124	0.0130

The effect of sample size and significance level on the measure $\widetilde{W}$, mentioned earlier in the properties of $\widetilde{W}$, is clearly established in Table 1.1. Firstly, increasing n decreases dispersion of the sample around the mean thus, also decreases critical values. Secondly, decrease of the significance level γ has a positive effect on the confidence level $1-\gamma$, hence allowing more points to get into the confidence set and increasing the critical value.

After the table of critical values was obtained, another question arises; what if one demands a critical value that is not listed in the table? Exploring data may bring to the case when the sample size is in between of two points, or a specific significance level is requested. One may also need to obtain an exact p-value. Therefore, to use CCR on the real data, one needs a continuous function that provides the critical value for each possible point. As mentioned in properties of measure W, critical values have an inverse relationship with sample size n, as well as with significance level γ. Other dependencies may not be excluded at this point. Before the construction of functional form, I plotted critical values for different significance levels γ and sample sizes n. Investigation of critical values in Figure 1.4 provided the points for fixed sample size $n = 10$, that showed a clear non-linear trend. The same shape of an exponential function is observed for all n. Thus, for each fixed sample size it makes sense to assume a form of a priori function as $\log(z) = f(\log(\gamma))$, with a critical value denoted by z.

Next is the dependence of z on sample size n. Exploring critical values as a function of sample size is presented in Figure 1.5. Different significance levels are marked on the graphs. The shapes of figures in the upper row also look similar to an exponential function. Therefore, in the lower row I plotted the same points, using logarithmic scale for both critical values and sample sizes. The obtained result shows a clear linear dependence after the logarithm was embedded. Thus, a priori function form for fixed significance level is $\log(z) = f(\log(n))$.

Combining both of the prior functions into $\log(z) = f(\log(n), \log(\gamma))$ gives a good starting point in the estimation of the functional form, that fits the critical value

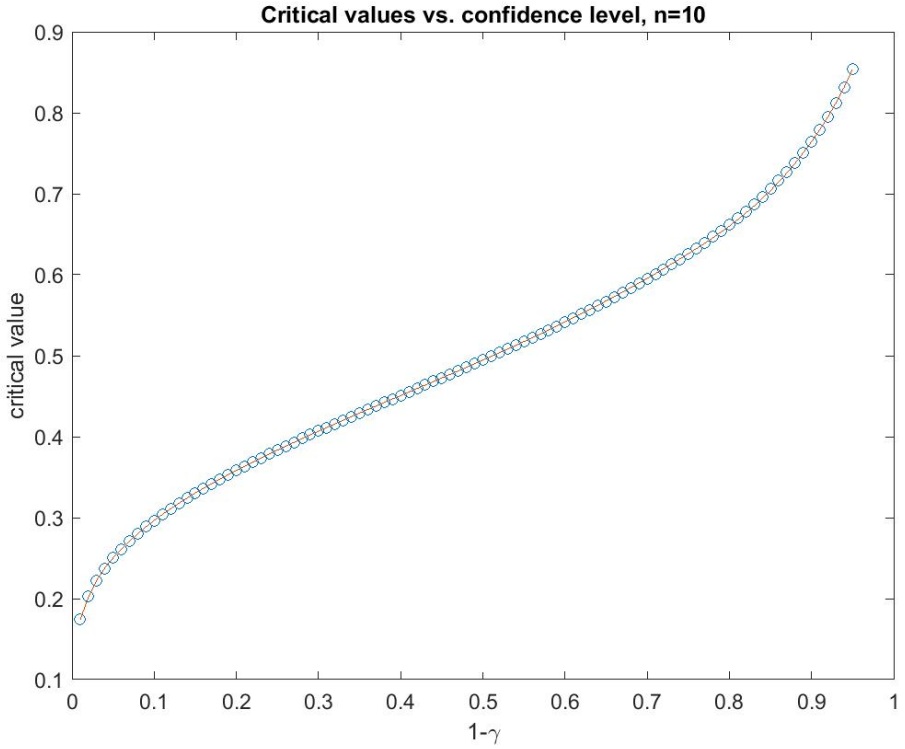

Figure 1.4: Critical Value Points Estimated from the CCR Approach in Comparison with a Function of Significance γ for Fixed $n = 10$

points the best. I explored different linear, logarithmic and exponential functions of n and γ and ended up with 14 of them. All their combinations, totally $2^{14} = 16384$, were monitored and estimated with OLS using 2388 knots from the extended table of critical values (see Appendix 1.I). The obtained estimation results were compared for the goodness of fit, using residual sum of squares (RSS), coefficients of determination R^2 and adjusted R^2. I have extracted the best combinations for each number of estimators from 1 to 14 and collected them in Table 1.2.

Table 1.2: Estimated Parameters for Functional Forms that are Potentially a Part of the Function of Critical Values for the CCR Approach

N	Const.	$\log(\gamma)$	$\log(n)$	$\gamma\log(\gamma)$	γ	n	$n\log(\gamma)$	$\gamma\log(n)$	$n\log(n)$	γ^2	$\log^2(\gamma)$	n^2	$\log^2(n)$	e^{γ}	$e^{-\gamma}$	RSS	R^2
1	0.43		-0.51													170.58	0.83
2	1.63		-0.51												-1.90	7.56	0.9927
3	0.49	0.23	-0.51							0.51						1.36	0.9987
4	-0.26	0.26	-0.51		-0.98									0.83		1.07	0.9994
5	-17.13		-0.51							-10.17	-0.02			9.41	7.63	0.60	0.9994
6	-17.01		-0.57							-10.17	-0.02		0.01	9.41	7.63	0.38	0.9996
7	-83.42	0.24	-0.57		33.39					-38.63			0.01	25.08	58.99	0.31	0.9997
8	-83.45	0.24	-0.56		33.46			-0.01		-38.63			0.01	25.08	58.99	0.25	0.9998
9	-183.96		-0.56	-4.37	90.62			-0.01		-74.32	-0.01		0.01	43.14	140.44	0.21	0.9998
10	-183.90		-0.60	-4.37	90.62	0.00		-0.01		-74.32	-0.01		0.01	43.14	140.44	0.19	0.9998
11	-296.82	-1.63	-0.60	-14.64	159.03	0.00		-0.01		-107.49	-0.16		0.01	59.27	232.44	0.17	0.9998
12	**-296.84**	**-1.63**	**-0.60**	**-14.64**	**159.07**	**0.00**	**0.00**	**-0.02**		**-107.49**	**-0.16**		**0.01**	**59.27**	**232.44**	**0.16**	**0.9999**
13	-296.77	-1.63	-0.65	-14.64	159.07	0.00	0.00	-0.02	0.00	-107.49	-0.16		0.02	59.27	232.44	0.15	0.9999
14	-296.68	-1.63	-0.71	-14.64	159.07	-0.01	0.00	-0.02	0.00	-107.49	-0.16	0.00	0.04	59.27	232.44	0.15	0.9999

Table 1.2 demonstrates combinations of functions that were used. Other composite functions that were also considered in the estimation are not displayed as they showed worse performance comparing to the displayed models with the best fit.
The first thing to mention in Table 1.2 is significance of all parameters at the 1% level.

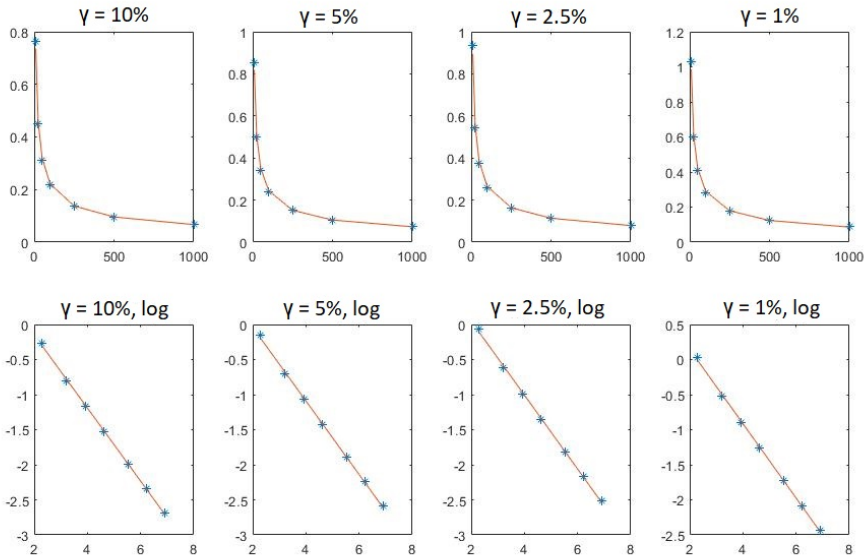

Figure 1.5: Critical Values Estimated from the CCR Approach in Comparison with a Function of Sample Size n in the Upper Graphs and Log of Sample Size $\log(n)$ in the Below Graphs for Different Values of γ

Even small estimated values, for example for n and $n\log(n)$ (columns 7-8 of Table 1.2) are still highly significant. These coefficients are small, but not to be neglected, because of different scales of the variables. For example, n goes up to 50000, while γ is between 0 and 1. Since the best combination with respect to RSS and R^2 is chosen in each case, combinations with worse R^2 are just skipped.
Secondly, even with two regressors one can already achieve $R^2 = 99\%$ but residuals are still too large, according to RSS (row 2 of Table 1.2). Therefore, it is logical to include more regressors that take into account not only the variance from the sample size and significance level, but also variance that comes from their interaction, as $\gamma\log(n)$ and $n\log(\gamma)$ (columns 8-9 of Table 1.2).
Thirdly, if the regressor is included in the best fitted model, it does not mean that it will still be included in more complex model with larger number of parameters. For example, regressor γ (row 6 of Table 1.2) is included in the best model with 4 regressors, but not with 5 and neither with 6. Combinations with γ^2 and $\log^2(\gamma)$ produce better goodness of fit for 5 and 6 regressors, as together they may take care of the larger part of the variance than only γ. Starting from the model with 7 regressors, γ is back in the best fitted model.
Finally, including all regressors in case 14 gives the best R^2 and smallest RSS. However, does it make the model 14 the best choice? I used the F-test for comparison of restricted and full model and analyzed the results to choose the overall best model. The essential assumption for this test is nested structure, meaning that one model (restricted) contains only the subset of terms of the other model (full). Unless some of the pairwise comparisons do not work, as the models refuse nested structure, they can still be compared. The procedure starts from the bottom full model with all the regressors included. It can obviously be compared with all the previous models, as they contain only the same variables as in the full model. Then the hypothesis for

each pair of models can be formulated as H_0: restricted model is preferable to the full model. Therefore, non-rejection of H_0 means that the full model does not bring any significant improvement to estimation performance. And if H_0 is rejected by the F-test for each of the 13 restricted models, then the full model is the best choice for estimation of functional form. Otherwise, I make one step up and consider second to last model as full and so on, till I find the model which will be preferred to all restricted models. The results of analysis with the F-test is presented in Table 1.3.

Table 1.3: F-Test Statistics for Comparison of Potential Critical Value Functions for the CCR Approach

N \ N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	30006	86450	73249	98476	124583	128464	133185	137580	140869	138092	135615	129835	121173
2	0	0	6332	4204	5375	6568	6565	6663	6777	6856	6651	6477	6155	5706
3	0	0	0	374.6	882.3	1197	1193	1212	1234	1248	1205	1170	1106	1018
4	0	0	0	0	1095	1267	1155	1120	1108	1098	1043	999.7	935.0	852.4
5	0	0	0	0	0	805.1	662.7	631.2	621.4	614.7	578.6	551.7	511.9	461.9
6	0	0	0	0	0	0	329.5	344.7	354.8	359.2	337.8	322.7	297.8	265.5
7	0	0	0	0	0	0	0	290.9	297.0	298.4	274.8	259.8	236.5	207.3
8	0	0	0	0	0	0	0	0	250.7	249.8	222.8	208.5	186.6	160.0
9	0	0	0	0	0	0	0	0	0	210.9	177.0	164.7	144.6	120.2
10	0	0	0	0	0	0	0	0	0	0	124.2	123.0	106.4	84.75
11	0	0	0	0	0	0	0	0	0	0	0	111.8	89.53	65.76
12	0	0	0	0	0	0	0	0	0	0	0	0	**62.28**	39.60
13	0	0	0	0	0	0	0	0	0	0	0	0	0	**16.24**
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Starting the analysis from the very right column of Table 1.3, I compared model 14 with all the others. The value of statistic 16.24, highlighted in the table, is inside of the confidence interval for the F-test and H_0 may not rejected with 1% significance in comparison to model 13. Hence, the second column from the right, which compares model 13 with all the others, is considered for further analysis with the F-test. Note that n^2, which is the only term, not included in the model 13, does not enter the other model either, except of already rejected model 14. Therefore assumption about nested structure of the models is not violated. H_0 is this case was also non-rejected with a 1% significance in comparison with model 12 (see highlighted F-statistic 62.28 in the second column from the right in Table 1.3). In the third step, I compared model 12 with 11 restricted models left. Assumption about nested structure of the models is not violated in this situation. The terms n^2 and $n\log(n)$ that were excluded from the model 12, were also excluded from all the models 1-11. In this case null hypothesis is rejected at a 1% significance level for each of the pairwise comparisons with the models 1-11. Meaning that model 12 is significantly better in estimation performance than the other 11 restricted models. Thus, making the model 12 preferable for the restoring of functional form of critical value, with $RSS = 0.16$ and $R^2 = 99.99\%$ on more than 2000 observations. And the functional form estimated by model 12 is displayed in equation (1.42):

$$z = e^{\beta_0+\beta_4\gamma+\beta_5 n+\beta_9\gamma^2+\beta_{13}e^{\gamma}+\beta_{14}e^{-\gamma}} * \gamma^{\beta_1+\beta_3\gamma+\beta_6 n+\beta_{10}\log(\gamma)} * n^{\beta_2+\beta_7\gamma+\beta_{12}\log(n)} \tag{1.42}$$

Betas in equation (1.42) are denoted in the order they appear in the highlighted row 12 in Table 1.2. The intersection is denoted as β_0, while β_8 and β_{11} are skipped as they show the effect of the functions $n\log(n)$ (column 10) and n^2 (column 13) respectively, excluded from the model 12.
Consequently, both measure $\widetilde{W}$ and critical values are obtained. An investigation of their properties and a comparison with theoretical results is the next step. Figure 1.6 presents the interaction of measure $\widetilde{W}$ with 5% and 10% significance level, marked by darker and lighter dashed lines respectively. Each graph illustrates how the value of $\widetilde{W}$ changes with the alteration of one parameter ceteris paribus. Set of points, where measure $\widetilde{W}$ is below the dashed line defines the corresponding confidence set (95% or 90%).

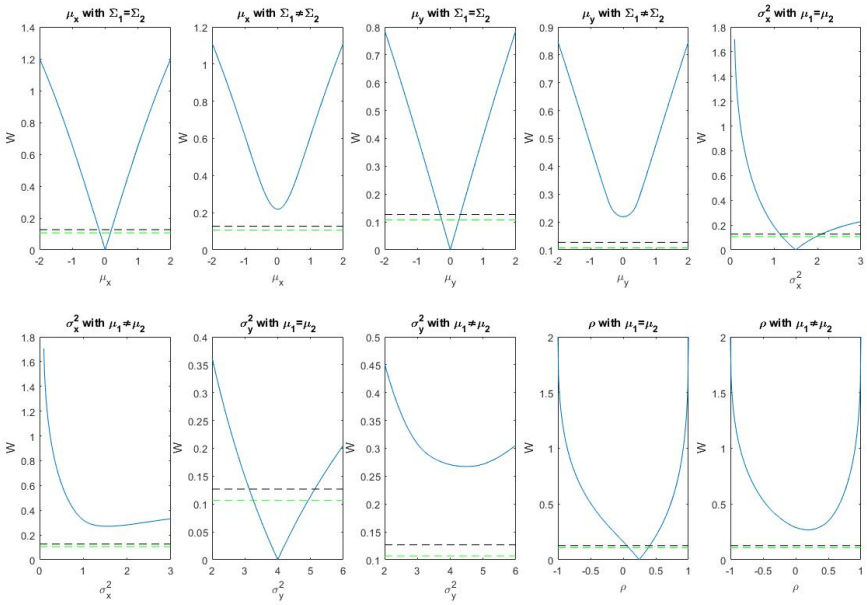

Figure 1.6: Behavior of Measure $\widetilde{W}$ (Solid Lines) from the CCR Approach Compared with the Critical Values for a 5% (Upper Dashed Lines) and a 10% (Lower Dashed Lines) Significance Level with Different Restricted Parameters

In Figure 1.6 graphs come in pairs which describe how one parameter reacts to the different values of the other parameters in terms of measure $\widetilde{W}$. Graph 1 shows how the dynamics of measure $\widetilde{W}$ depends on the mean of the first variable of random vector, holding variance matrices of compared distributions the same. Minimum of $\widetilde{W}$ is achieved at zero mean, indicating that the comparable distribution also has a zero mean. Thereafter, the confidence set is constructed around this point. The second graph reveals the same situation, but with distinct variance matrices. As a result, being far from the zero mean, measure $\widetilde{W}$ is almost the same as in graph 1. When the mean gets closer to zero, the influence of difference in variances has a clear effect and strongly increases measure $\widetilde{W}$. The minimum is still at zero mean. But $\widetilde{W}$ does not reach critical value, neither for a 10% significance level, nor for a 5%. Therefore, the confidence set is empty and a null hypothesis will be always rejected. Graphs 3 and 4 demonstrate a similar case to the one in graphs 1-2, but for the mean of the second variable of the vector. Displayed relations are the same, forming an

unreachable confidence set, when variance matrices differ too much, and rejecting a null hypothesis for any value of the mean. Another necessary aspect to mention regarding graphs 1-4 is the symmetry with respect to the mean. Symmetry still holds independently of the values of other parameters, contained in variance matrix.
Graphs 5-6, together with graphs 7-8, describe behavior of variances of the first and second variable respectively. Graphs 5 and 7 display measure $\widetilde{W}$ depending on variance and holding both means equal to the ones in the comparable distribution. While graphs 6 and 8 on contrary include means that alter from comparable distribution. Note that in graphs 5-6 the comparable variance from hypothesis (1.1) is $\sigma_2^2 = 1.5$, while in graphs 7-8 this variance is $\sigma_2^2 = 4$. Which may be also seen in the minimum points of the graphs. Next, all 4 graphs are skewed, displaying the property of 'distance' $\widetilde{W}$ between two distributions with the increasing variance (formula (1.6) describes this property). In other words, the right-hand side of all the graphs is lower than the left-hand side. This occurs due to the fact that the increase of variance has a negative effect on measure $\widetilde{W}$. However, the value of $\widetilde{W}$ still grows, but much more slowly.
The last two graphs demonstrate how $\widetilde{W}$ depends on the correlation coefficient ρ between two variables of one vector. Graph 9 is constructed assuming equal left and right-hand side means of the null hypothesis (1.1), while graph 10 assumes that the means are different. The minimum is achieved in the both graphs at the same point of equal correlations $\rho_1 = \rho_2 = 0.25$, that are compared. However, in the case of different means, measure $\widetilde{W}$ does not manage to get as low as critical value. Therefore, the means differ a lot and hypothesis will always be rejected, even if correlations are the same.

1.2.8 Extension to Multivariate Case

Development of the CCR approach to samples of random vectors with more than 2 dimensions is highly profitable. It gives the possibility of applying CCR on large amounts of data and simultaneously comparing any number of dimensions. There are at least two ways to extend the CCR technique with respect to the dimensionality. Firstly, computing measure $\widetilde{W}$ as an absolute difference between multivariate PDFs on the domain from $\mathbb{R}^m$. However, the main complication of this approach is the intersection of PDFs. The second order curve from equation (1.11) will get even more complex in m-dimensional space with a growing number of different interaction terms.
Another way to extend the CCR method is a pairwise comparison. One can easily pick out two scalars from each vector of the sample $\{\mathbf{X}_i\}_{i=1}^n$, find their means and covariance matrix, making it possible to test this subsample. Then, all $\frac{m(m-1)}{2}$ pairs of scalars from m-dimensional sample are compared. Hypothesis (1.1) for the full sample is rejected, if at least for one of the pairs of scalar null hypothesis is rejected. Otherwise hypothesis for the full sample may not be rejected. This extension of CCR technique will be implemented further in the SUR example.

1.3 Qualitative and Quantitative Analysis of Developed CCR Approach

In this section I describe the main qualities of the CCR method, which will also be confirmed with computations. This includes: the form of confidence set, preciseness of critical values and the significance level, reaction on non-standard normally distributed samples and samples from other distributions, as well as effectiveness of approach with respect to computation time.

1.3.1 Shape of the Confidence Region

The confidence region, produced by the CCR technique, includes the values of parameters, where measure W is smaller than the critical value for some predefined significance level γ. To visualize this set, I used Monte-Carlo simulations and generated 10 million uniformly distributed points in a 5-dimensional space $(\mu_x, \sigma_x^2, \mu_y, \sigma_y^2, \rho)$. Then I omitted points that were rejected by the CCR test in comparison with bivariate standard normal distribution for $n = 100$ and $\gamma = 0.1$. This simulation resulted in a figure, which I have projected on all possible 2-D planes. The center circles show the parameters for which confidence set is constructed. Here I describe the most interesting ones and all the others may be found in Appendix 1.II.

Figure 1.7 demonstrates the confidence set, projected on a (μ_x, σ_x^2) plane. It is symmetrical in means and its shape is close to an ellipse. However, the larger upper half makes it asymmetrical in variance. This comes from the non-Euclidean property of variance. Increasing it produces more space for alteration of mean and therefore also increases critical values from above.

Next, Figure 1.8 shows the cross section of confidence region by (μ_x, σ_y^2) plane. On one hand, the figure remains symmetrical with respect to mean and has a larger top half than a bottom half. On the other hand, its form is narrowed in the upper part, in comparison to Figure 1.7, where the wider upper part stems from the effect of increased variance. The smaller top part of Figure 1.8, compared to Figure 1.7, is due to the fact that the effect of increasing the mean of x is to a lesser extent compensated by increasing variance of y, than it was compensated by the variance of x.

Furthermore, Figure 1.9 describes the joint behavior of variance of y together with the correlation between x and y. A similar effect of compensated variance makes Figure 1.9 larger from above, as the increase of variance provides more space for the correlation to differ from zero.

Other cross-sections, that can be found in Appendix 1.II, show the symmetry with respect to means and correlation, and keeping the effect of variance to enlarge the upper part as well.

1.3.2 Significance Level

In this subsection real and theoretical significance levels are compared. This investigation is possible since all the samples are generated from the known distribution and the share of rejected samples defines the real significance level. In such a way

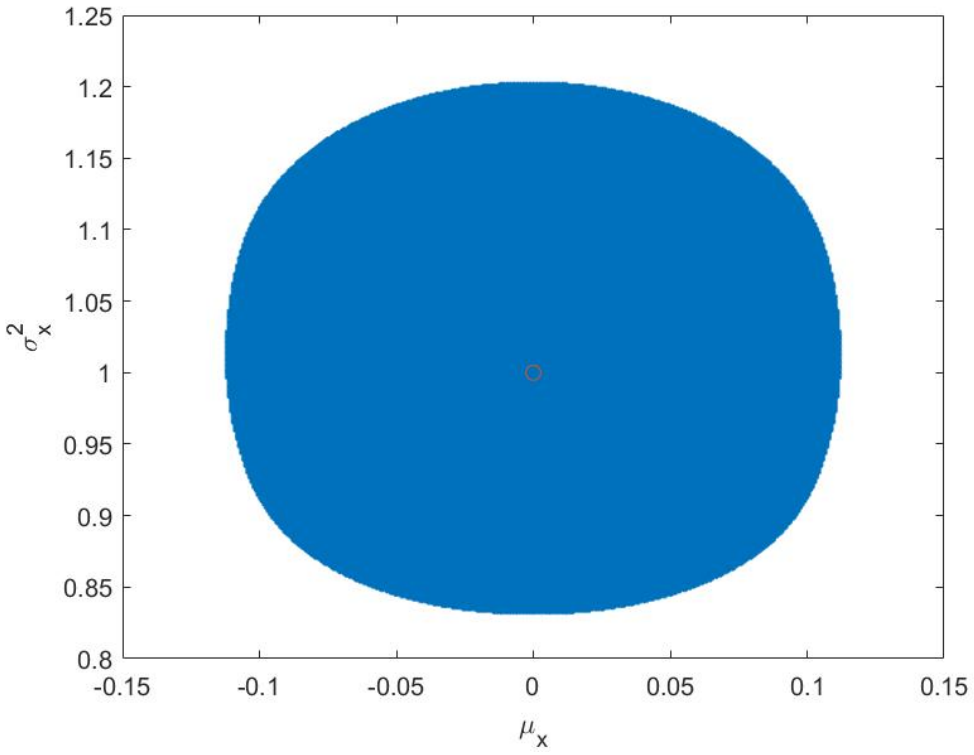

Figure 1.7: Projection of the CCR Confidence Set on (μ_x, σ_x^2)-Plane ($n = 100$, $\gamma = 0.1$)

empirically obtained critical values for specific significance level γ and sample size n are examined to produce the same significance (as defined in Table 1.1) as well as in extended tables of critical values in Appendix 1.I. As far as I also presented critical values as a function of n and γ in equation (1.42), non-tabular values of n and γ are used for the investigation as well. This control is important to determine whether any bias related to critical values takes place in the CCR method.

Tables 1.4, 1.5 and 1.6 present the obtained real significance levels. For each of the sample sizes n, 100.000 replications of the CCR test were performed.

Table 1.4 shows the results for small n's up to 250. Table 1.5 shows the results for larger n's up to 1000. And Table 1.6 displays results for the largest sample sizes up to 50.000.

A closer look at Table 1.4 demonstrates that for $n = 10$ (column 2) the maximum deviation of real significance level from theoretical is slightly more than 1% for $\gamma = 0.5$. And the accuracy slightly decreases with a larger sample size. For example, for $n = 100$ (column 5) maximum deviation of real significance is 0.81% at $\gamma = 0.3$. Columns 6 and 7 of Table 1.4 show real significances for non-tabular sample sizes of 150 and 200. Even though the real significance levels of these two columns were calculated with the critical value function from equation (1.42), they still present reasonably precise results. The maximum deviation for $n = 150$ is around 2%, when $\gamma = 0.5$, and for $n = 200$ the maximum deviation is 0.78% at $\gamma = 0.3$, which is even better than for $n = 100$.

Table 1.5 has only 2 columns with tabular sample sizes: $n = 500$ and $n = 1000$. The other 5 columns use estimated functional form of critical value. Nevertheless, the best

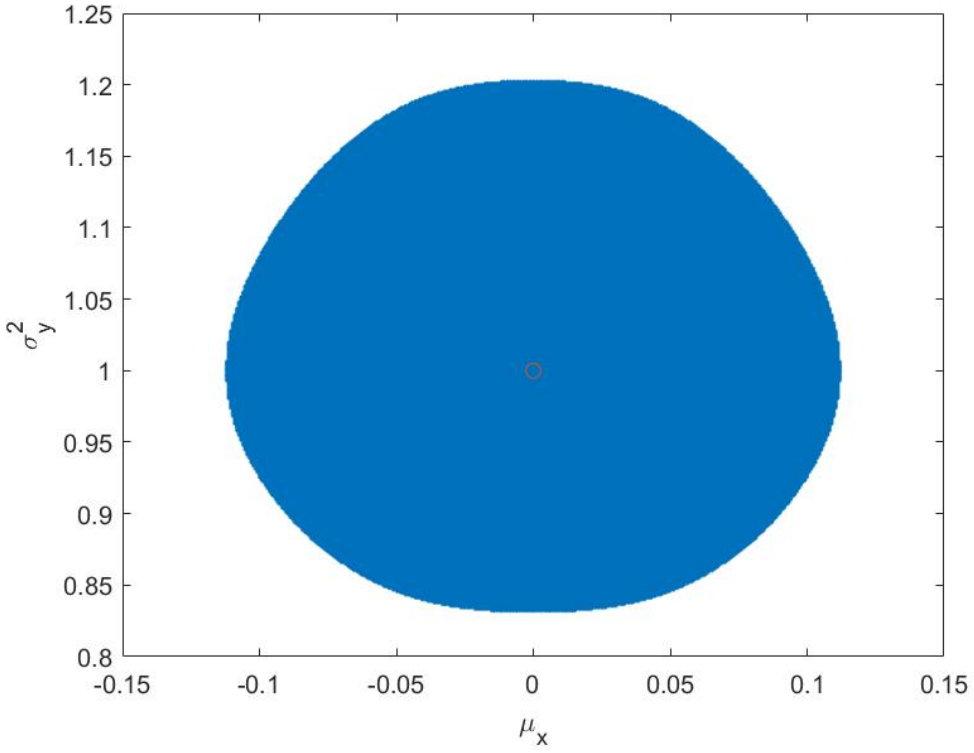

Figure 1.8: Projection of the CCR Confidence Set on (μ_x, σ_y^2)-Plane ($n = 100$, $\gamma = 0.1$)

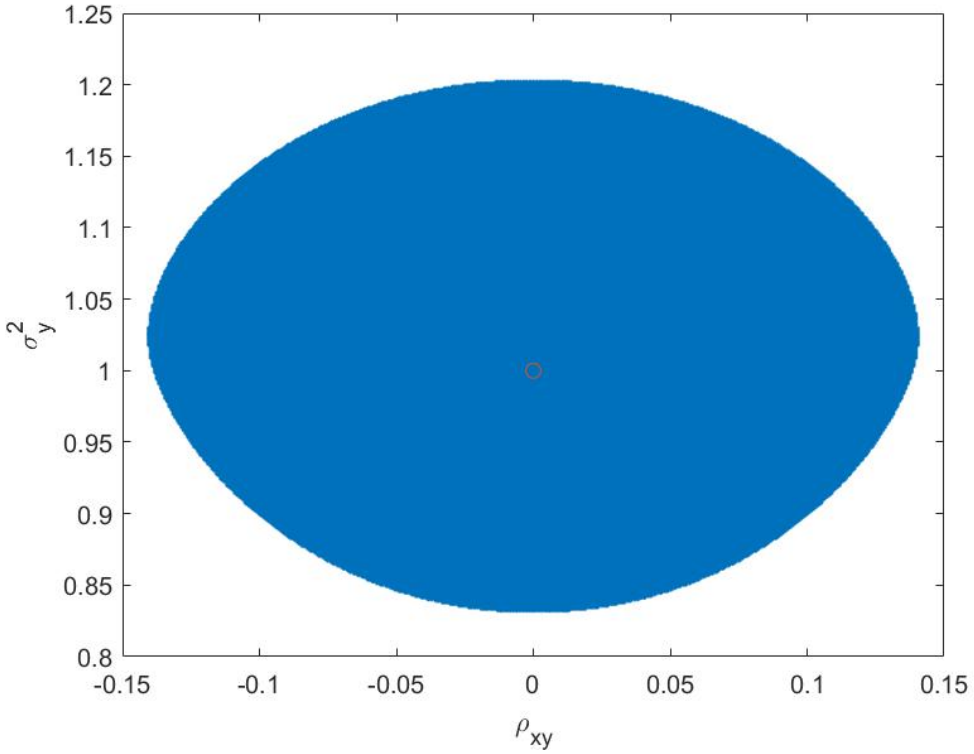

Figure 1.9: Projection of the CCR Confidence Set on (ρ, σ_y^2)-Plane ($n = 100$, $\gamma = 0.1$)

fitted real significance levels are presented for $n = 400$ (column 4) with maximum deviation from the theoretical significance being equal to 0.44% at $\gamma = 0.1$. At the

Table 1.4: Theoretical (Left Column) and Real Significance Levels of the CCR Approach for Small Samples from $n = 10$ to $n = 250$

γ \ n	10	25	50	100	150	200	250
0.99	0.9915	0.9910	0.9903	0.9918	0.9883	0.9905	0.9909
0.95	0.9505	0.9522	0.9543	0.9521	0.9458	0.9495	0.9500
0.9	0.8991	0.9034	0.9019	0.9005	0.8913	0.8997	0.8982
0.75	0.7516	0.7535	0.7502	0.7486	0.7408	0.7487	0.7454
0.5	0.5114	0.5018	0.4990	0.5012	0.4796	0.4991	0.4995
0.3	0.3106	0.2981	0.3028	0.3081	0.2834	0.2922	0.2999
0.2	0.2050	0.1987	0.2001	0.2030	0.1887	0.1944	0.1998
0.1	0.0979	0.1033	0.0982	0.0996	0.0919	0.0981	0.1002
0.05	0.0526	0.0537	0.0505	0.0512	0.0449	0.0490	0.0515
0.025	0.0268	0.0279	0.0260	0.0258	0.0204	0.0233	0.0244
0.01	0.0091	0.0105	0.0116	0.0107	0.0089	0.0102	0.0099

Table 1.5: Theoretical (Left Column) and Real Significance Levels of the CCR Approach for Samples from $n = 300$ to $n = 1000$

γ \ n	300	350	400	450	500	750	1000
0.99	0.9909	0.9899	0.9901	0.9905	0.9905	0.9901	0.9885
0.95	0.9488	0.9488	0.9484	0.9539	0.9499	0.9515	0.9507
0.9	0.8970	0.9046	0.8989	0.9000	0.8974	0.9072	0.9001
0.75	0.7495	0.7567	0.7467	0.7509	0.7558	0.7646	0.7551
0.5	0.4955	0.5114	0.4972	0.5096	0.4995	0.5221	0.5003
0.3	0.2993	0.3080	0.3023	0.3096	0.2920	0.3195	0.2978
0.2	0.2024	0.2074	0.2018	0.2073	0.1961	0.2249	0.1986
0.1	0.0996	0.1023	0.1044	0.1027	0.0963	0.1203	0.0996
0.05	0.0515	0.0495	0.0501	0.0531	0.0457	0.0614	0.0504
0.025	0.0279	0.0242	0.0259	0.0272	0.0249	0.0329	0.0257
0.01	0.0110	0.0096	0.0107	0.0111	0.0100	0.0129	0.0104

same time, the furthest real significance value from the theoretical one is for case $n = 750$ (column 2 from the right) that also uses non-tabular critical values. At $\gamma = 0.2$ its deviation is 2.49% from the theoretical significance level.

Table 1.6 with its large sample sizes also demonstrates a very good outcome. Only one case, in the whole table, differs from the theoretical significance level by more than 1% for $n = 10.000$ and $\gamma = 0.5$.

Finally, an analysis of significance levels made in this subsection verifies two issues in the CCR method. The first is the preciseness of the critical values table, derived earlier. And the second issue is the correctness of functional form, that may be used for not tabulated values of n and γ.

1.3.3 Robustness

In this section I investigate how precisely samples from different distributions may be tested and how well they may produce confidence regions. Firstly, I take normal distribution with non-standard parameters and generate a sample from it. Then I calculate measure W, that compares this sample with the distribution from which it

Table 1.6: Theoretical (Left Column) and Real Significance Levels of the CCR Approach for Large Samples from $n = 2000$ to $n = 50000$

γ \ n	2000	5000	10000	20000	50000
0.99	0.9893	0.9902	0.9889	0.9885	0.9913
0.95	0.9507	0.9497	0.9501	0.9474	0.9521
0.9	0.8957	0.8938	0.9017	0.9027	0.9011
0.75	0.7463	0.7445	0.7458	0.7558	0.7478
0.5	0.5030	0.4968	0.4883	0.5100	0.4991
0.3	0.2961	0.2915	0.2937	0.3065	0.3003
0.2	0.1972	0.1962	0.1961	0.2028	0.1991
0.1	0.0991	0.0953	0.0993	0.1085	0.1001
0.05	0.0529	0.0480	0.0480	0.0516	0.0476
0.025	0.0265	0.0231	0.0233	0.0248	0.0257
0.01	0.0099	0.0087	0.0091	0.0079	0.0093

was taken. After performing 100.000 replications I can obtain a critical value for this exact distribution. Table 1.7 presents the results of this procedure performed for 7 normal distributions with different parameters, i.e. alternating correlation, variances and means.

The first row of Table 1.7 gives significance levels for which critical values were calculated. The second row exhibits critical values of standard normal distribution, obtained earlier. And they are used as a template for comparison with the other columns. The next 7 columns display all possible permutations of standard and non-standard means, variances and covariance. To illustrate, column 3 gives critical values for samples generated from normal distribution with $\mu_x = 7$, $\mu_y = 3$, $\sigma_x^2 = 0.8$, $\sigma_y^2 = 1.2$ and $\rho = -0.3/(\sqrt{1.2 * 0.8}) = -0.306$. This results in particularly precise critical values, with the largest difference from template of 0.0006 for $\gamma = 0.01$. Other distributions considered in Table 1.7 also provide precise critical values, with an average deviation from the template of less than 0.0002.

The results demonstrated in Table 1.7 not only check the stability of the CCR test, with respect to samples from non-standard normal distributions, they also check the smoothness of measure $\widetilde{W}$ since values of parameters of normal distribution cover most of the 15 cases for which the measure was calculated in subsection 1.2.6.

The following aspect of the analysis covers the samples from non-normal distributions. Table 1.8 portrays these results, with each column based on 100.000 replications. In Table 1.8 I cover the most demanded values of γ (column 1). And in comparison to the theoretical significance levels, I compute the real ones from the samples from the t-distribution, the Poisson distribution and the samples with outliers.

Columns 2-5 of Table 1.8 show the ratios of rejected samples from the t-distribution with growing degrees of freedom from 3 to 100. The essential property of t-distribution is the convergence to normal one with increasing degrees of freedom. And this property of t-distribution can be traced in the table: fractions of rejected samples converge to the real significance levels with a growing number of degrees of freedom. When number of degrees of freedom is 3 (column 2) and tails of t-distribution are extremely thick then there is no possibility to perform the CCR test properly. For example, at

Table 1.7: CCR Critical Values Obtained from Non-Standard Normal Distributions in Comparison with the CCR Critical Values from Standard Normal Distribution (Second Column)

γ	$N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right)$	$N\left(\begin{pmatrix}7\\3\end{pmatrix},\begin{pmatrix}0.8&-0.3\\-0.3&1.2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}0.8&-0.3\\-0.3&1.2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}0.8&0\\0&1.2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&-0.3\\-0.3&1\end{pmatrix}\right)$	$N\left(\begin{pmatrix}7\\3\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right)$	$N\left(\begin{pmatrix}7\\3\end{pmatrix},\begin{pmatrix}0.8&0\\0&1.2\end{pmatrix}\right)$	$N\left(\begin{pmatrix}7\\3\end{pmatrix},\begin{pmatrix}1&-0.3\\-0.3&1\end{pmatrix}\right)$
0.99	**0.0525**	0.0525	0.0525	0.0526	0.0525	0.0523	0.0528	0.0521
0.975	**0.0641**	0.0641	0.0642	0.0641	0.0639	0.0642	0.0640	0.0638
0.95	**0.0753**	0.0752	0.0753	0.0751	0.0751	0.0754	0.0752	0.0754
0.9	**0.0893**	0.0893	0.0892	0.0891	0.0892	0.0891	0.0893	0.0892
0.85	**0.0993**	0.0994	0.0997	0.0992	0.0992	0.0989	0.0993	0.0994
0.8	**0.1077**	0.1078	0.1079	0.1076	0.1077	0.1075	0.1079	0.1079
0.75	**0.1152**	0.1154	0.1153	0.1150	0.1152	0.1151	0.1153	0.1152
0.7	**0.1221**	0.1223	0.1222	0.1219	0.1221	0.1219	0.1222	0.1220
0.5	**0.1475**	0.1475	0.1475	0.1476	0.1476	0.1470	0.1475	0.1474
0.25	**0.1830**	0.1827	0.1832	0.1832	0.1830	0.1828	0.1830	0.1827
0.1	**0.2176**	0.2178	0.2180	0.2178	0.2175	0.2174	0.2175	0.2173
0.05	**0.2396**	0.2397	0.2397	0.2399	0.2391	0.2394	0.2392	0.2393
0.025	**0.2593**	0.2591	0.2597	0.2597	0.2597	0.2590	0.2592	0.2592
0.01	**0.2833**	0.2827	0.2838	0.2831	0.2834	0.2838	0.2839	0.2840

Table 1.8: CCR Real Significance Levels for Samples from t-Distribution, Poisson Distribution and Samples with Outliers

γ	t_3	t_{10}	t_{50}	t_{100}	$Pois(3)$	$Pois(10)$	5% out	10% out	20% out
0.2	0.8891	0.3119	0.2212	0.2116	0.2346	0.2173	0.2009	0.2585	0.731
0.1000	0.8329	0.1871	0.1151	0.1089	0.1269	0.1101	0.0953	0.1427	0.5919
0.0500	0.7786	0.1064	0.0568	0.0509	0.0733	0.0552	0.0461	0.0764	0.4572
0.0250	0.7260	0.0611	0.0298	0.0244	0.0409	0.0278	0.0205	0.041	0.3444
0.0100	0.6607	0.0312	0.0120	0.0109	0.0182	0.0098	0.0128	0.0174	0.2315

a 10% predefined significance level, it rejects more than 83% of the samples. Even at $\gamma = 0.01$, more than 66% of samples are rejected. The opposite situation may be seen for 100 degrees of freedom (column 5). At a 10% significance level, only 10.89% of samples from t_{100} are rejected. And at $\gamma = 0.01$, only 1.09% of the samples are rejected.

Columns 6 and 7 of Table 1.8 describe the behavior of the CCR approach on samples from the Poisson distribution. Although, Poisson distribution, as a discrete one, cannot directly converge to the normal distribution, with increasing value of its parameter λ, its probability mass function gets closer to the bell-shaped form of the normal distribution. Hence, at $\lambda = 3$ the samples generated from Poisson distribution are rejected with a bit higher rate than they should. For example, at $\gamma = 0.05$ more than 7% of the samples are rejected. An increase of λ to 10 shows a reasonable growth in preciseness of the CCR method. For a 5% significance level, slightly more that 5.5% of the samples are rejected.

A study of the CCR approach on samples not from the normal distribution, demonstrated that the method is sufficiently robust to the change of distribution. However, for the distributions that are far from the normal bell-shaped form, CCR is not to be used.

The last 3 columns of Table 1.8 present the results of using the CCR approach on samples with outliers (5, 10 and 20% of outliers respectively). Outliers are added to the sample as random numbers to avoid additional correlation in the sample. Based on the last 3 columns, it is clearly seen that 20% of the outliers bring too much noise to the sample and much more samples are rejected. For example, at $\gamma = 0.2$ more than 73% of the samples are rejected (the last column from the right). With 10% of outliers, results of the CCR method get closer to real significance levels, but still errors are present. To illustrate, at $\gamma = 0.2$ almost 26% of samples are rejected. A much better performance of the CCR approach is demonstrated on samples with 5% of outliers. At a 20% significance level it rejects only 0.09% more samples, than it was predefined by γ. Thus, even on the samples with a reasonable proportion of outliers (around 5%), CCR can stay robust and correctly test the joint hypothesis (1.1).

1.3.4 Computation Time

The time, that the CCR method needs to calculate measure W and perform the test is not a critical parameter for small samples. But when it comes to large amounts of data, the computation speed of the algorithm should be also taken into account, as it may play a crucial role. Therefore, in this subsection I check how the sample size

affects computation time and whether it could be important in the implementation of the CCR method.
Table 1.9 presents computation time[1] for 100 replications of the CCR approach for different sample sizes n.

Table 1.9: Computation Time for 100 Replications of the CCR Method for Different Sample Sizes n

n	10	25	50	100	200	250	500
time (sec)	20.74	21.07	27.53	27.94	26.95	24.91	49.74
n	750	1000	2000	5000	10000	20000	50000
time (sec)	53.98	42.38	44.34	52.73	42.21	43.00	35.78

According to Table 1.9 one can distinguish two groups. The first aggregates n up to 250 with the average time to perform 1 CCR test of $2-3$ seconds. The second group, for n starting from 500, their computation time is doubled thus being around $4-5$ seconds for 1 repetition of the CCR test. But a further increase of the sample size from 500 to 50.000 does not lead to any significant change in computation speed. Actually, in this example performing CCR tests on samples of size 50.000 took 35.78 seconds. This is sufficiently less than the time spent on performing the same task on samples of size 5000, which took 52.73 seconds.
Table 1.9 additionally investigates how non-tabular values of sample size influence the computation speed. Samples of size 200 and 750 are out of critical values Table 1.1. For these two entries in Table 1.9, there is no significant difference in computation times compared to the others with tabular sample sizes. Although, for $n = 750$ computation time is the largest in the entire table, it took only 1.25 sec more to calculate the CCR test on this samples than on the samples of size 5000.

1.4 Example

The CCR method presented in this Part may be used as an alternative for testing a lot of issues in the fields of econometrics and economics. In this section I demonstrate how it is applied to the choice of estimator in SUR models. Usually two main estimation techniques are considered in SUR models: system OLS (SOLS) and feasible generalized least squares (FGLS). Depending on the available correlation across equations, one or another estimator will deliver better estimates: SOLS in case of uncorrelated error terms between the equations, otherwise FGLS ia a more efficient estimator. In the example I show how CCR is applied in order to choose the correct estimator among the two.
Suppose there are two models with 5 equations each. Every equation includes 3 unique regressors. Therefore, 15 parameters β in each model are to be estimated. In the first step for both of the models, I randomly define parameters β and two covariance matrices Ω that describe cross-equation relations of error terms. Next 1000 observations for all 10 equations are generated, considering covariance matrices Ω.

[1]Computer specifications: CPU Intel Core i5-6200U with 2.3 Ghz and 8 GB RAM type DDR4

Table 1.10: Predefined Theoretical Covariance Matrices Ω for Use in CCR Approach for 2 SUR Model Examples: without (Left Matrix) and with (Right Matrix) Significant Correlations Between Equations

(a) SUR model 1

1	0.05	-0.1	0	0
0.05	1	-0.2	0	0
-0.1	-0.2	1	0.05	0
0	0	0.05	1	0.05
0	0	0	0.05	1

(b) SUR model 2

1	0.8	-0.1	0.45	0
0.8	1	-0.1	0	-0.2
-0.1	-0.1	1	0.5	0
0.45	0	0.5	1	0.15
0	-0.2	0	0.15	1

Tables 1.10a and 1.10b show the covariance matrices for both of the models. In the first SUR I choose Ω so that only small correlations across equations are present. For the second SUR I added considerably large correlations between equations 1 and 2 (0.8), 1 and 3 (0.45) and 3 and 4 (0.5).

Table 1.11: Comparison of p-Values from Testing the Correlations of Estimated Covariance Matrices $\hat{\Omega}$ with the CCR Method for 2 SUR Models

(a) SUR model 1

0	0.998	0.211	0.998	0.999
0	0	0.18	0.894	0.999
0	0	0	0.999	0.999
0	0	0	0	0.948
0	0	0	0	0

* p<.1; ** p<.05; *** p<.001

(b) SUR model 2

0	0.001***	0.999	0.001***	0.997
0	0	0.999	0.991	0.358
0	0	0	0.001***	0.999
0	0	0	0	0.695
0	0	0	0	0

* p<.1; ** p<.05; *** p<.001

Tables 1.11a and 1.11b present the p-values of pairwise testing of error terms of the equations with the CCR technique. I tested whether each pair of error terms may be considered as uncorrelated, i.e. their covariance matrix is diagonal. In Table 1.11a all p-values are larger than $\gamma = 0.05$, therefore the null hypothesis saying that there is no correlation between error terms cannot be rejected. All the correlations in Ω of the first SUR are insignificant. In the second SUR the same null hypothesis for error terms following uncorrelated normal distribution should be rejected for 3 pairs: equations (1, 2), equations (1, 3) and equations (3, 4). Their p-values in Table 1.11b are smaller than the significance level $\gamma = 0.05$.

Moreover, the estimation of both SUR models is done with SOLS and FGLS. The results are demonstrated in Tables 1.12a and 1.12b. As far as data was generated using known theoretical parameters β, they may be easily compared with estimated $\hat{\beta}_{SOLS}$ and $\hat{\beta}_{FGLS}$.

Table 1.12a for the SUR model, with insignificant cross-equation correlations, shows that FGLS produced better estimates only in 3 out of 15 cases (rows 1, 3 and 15). Additionally, I compare the estimators by the sum of absolute differences from the real β. For SOLS this sum is 0.442, while for FGLS it is 0.554 thus, making the FGLS estimation 25% worse than the SOLS.

Table 1.12a, with 3 cross-equation correlations significantly different from zero, gives the opposite results: FGLS is closer to real β in 11 out of 15 cases. The exceptions

Table 1.12: Comparison of SOLS ($\hat{\beta}_{SOLS}$) and FGLS ($\hat{\beta}_{FGLS}$) Estimated Parameters for 2 SUR Models with the Predefined Values β

(a) SUR model 1

#	β	$\hat{\beta}_{SOLS}$	$\hat{\beta}_{FGLS}$
1	0	-0.0653	-0.0543
2	4	4.0953	4.1033
3	5	5.1311	5.1022
4	3	3.2711	3.3127
5	5	4.7820	4.7049
6	2	1.8485	1.8400
7	1	0.8536	0.8438
8	5	5.1020	5.1525
9	4	3.9187	3.8904
10	1	0.8272	0.8130
11	3	3.2828	3.3161
12	-5	-5.0450	-5.0509
13	-1	-1.1581	-1.1622
14	-3	-2.7493	-2.7392
15	0	0.1706	0.1688

(b) SUR model 2

#	β	$\hat{\beta}_{SOLS}$	$\hat{\beta}_{FGLS}$
1	-1	-0.7058	-0.9487
2	-4	-4.0822	-3.9423
3	-2	-2.5043	-2.1540
4	-4	-3.9208	-3.9697
5	-2	-2.2211	-1.9609
6	2	2.1723	2.0216
7	-2	-2.0587	-2.1211
8	-2	-1.8336	-1.8537
9	-3	-2.9064	-2.7718
10	-5	-5.0055	-5.0657
11	4	3.9157	4.0577
12	-5	-4.8989	-4.9224
13	-4	-3.9290	-4.0052
14	3	2.9451	3.0983
15	-5	-4.9931	-4.9915

are rows 9, 10, 14 and 15, for which SOLS gives a better estimation. And the total sum of absolute differences from β for SOLS is 0.498 while for FGLS it is 0.144. This result in FGLS estimation being 71% better than SOLS.

In the real applications comparison, with unknown parameters β, this comparison is impossible. But use of the CCR approach may help in the choice of a better estimator. Results of this section demonstrate that small, negligible cross-equation correlations of error terms imply the use of SOLS estimator. In the case of significant correlations, the FGLS estimator is preferable, even if only some of the correlations between error terms show their significance.

1.5 Conclusion

In this Part extension of the CCR approach to bivariate case was introduced. Bivariate extension of the CCR approach is much more broad in its application than the univariate version, presented in Appendix 1.III, since it covers not only the parameters of a variable itself, but also relations to the other variables. Additionally, it gives the opportunity to test samples of any dimensionality.

This technique is based on the absolute difference between two normal distributions and is designed to jointly test mean vector and variance matrix. I presented how measure W for the CCR approach is obtained, based on the different possible intersections of PDFs of normal distribution. Furthermore, properties of measure W were discussed and checked in the examples, including unimodality, smoothness and convergence. The effect of decreasing 'distance' between distributions with increasing variance was discussed and demonstrated on the examples as well.

Next, the transformation method for measure W was introduced, that converts W to the space with zero correlation. It simplified the form of the intersection curve and

therefore computation of measure W. The transformation also showed how the CCR approach may be applied to simpler hypotheses as an alternative to the t-test or the χ^2-test.

Besides, 2388 critical values for measure W were received for a wide number of sample sizes from 10 to 50.000 and significance levels from 0.005 to 0.995. To cover the all the possible values of n and γ between given intervals, I analyzed more than 16.000 different functional forms and choose the best one to fit. Chosen equation (1.42) for critical values depending on n and γ provides R^2 of 99.99%.

In the next section I studied the CCR approach for the shape of the confidence region it produces, compared real and theoretical significance levels, investigated CCR robustness properties and checked the efficiency with respect to its computation time. The CCR method demonstrated good performance on the samples from non-standard normal distribution. Therefore, proved the preciseness of transformation and computed critical values. Application of the CCR approach on samples not from the normal distribution demonstrated that it can be implemented, if the distribution is close in shape to normal distribution, even if it is a discrete one. Additionally, the CCR approach was tested on samples with outliers and demonstrated good performance on the samples with 5% of outliers. Therefore, discussed properties of the CCR approach make it possible to use in wide array of scientific fields for better analysis and estimation of data.

In final section I showed how the CCR technique may be applied to SUR model estimation, depending on the cross-equation relations. In both examples presented here, CCR helped to make the correct choice of estimator. This improved the estimations on average by 48%.

Further development vector of the CCR technique may be found in implementation to problems in panel data econometrics (e.g. serial correlation testing) or portfolio theory (mean-variance analysis), in which one could benefit from generic CCR approach to problem solving.

References

Arnold, B., Shavelle, R., 1998. Joint confidence sets for the mean and variance of a Normal distribution. *The American Statistician*; 52; 133-140.

Born, B., Breitung, J., 2016. Testing for Serial Correlation in Fixed-Effects Panel Data Models. *Econometric Reviews*; 35; 1290-1316.

Box, G., Pierce, D., 1970. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*; 65(332); 1509–1526.

Frey, J., Marrero, O., Norton, D., 2009. Minimum-area confidence sets for a Normal distribution. *Journal of Statistical Planning and Inference*; 139; 1023-1032.

Godfrey, L., 1978. Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*; 46; 1293–1301.

Jensen, U., 1995. A review of the derivation and calculation of Rao distances with an application to portfolio theory. In Maddala, G., Phillips, P., Srinivasan, T. (eds.), Advances in econometrics and quantitative economics. Blackwell; 2008; 413-462.

Ljung, G., Box, G., 1978. On a measure of lack of fit in time series models. *Biometrika*; 65(2); 297–303.

Mood, A., 1950. Introduction to the Theory of Statistics. McGraw-Hill; New York; 1950.

Zhang, J., 2017. Minimum volume confidence sets for parameters of normal distributions. *AStA Advances in Statistical Analysis*; 101; 309–320.

Zellner, A., 1962. An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association*; 57(298); 348–368.

Appendix 1.I: Tables

Table 1.13: Empirical Critical Values for the CCR Method for Different n and γ from 86.5% to 99.5%

n \ γ	0.995	0.990	0.985	0.980	0.975	0.970	0.965	0.960	0.955
10	0.1499	0.1741	0.1903	0.2027	0.2133	0.2223	0.2304	0.2377	0.2443
25	0.0919	0.1065	0.1161	0.1238	0.1303	0.1358	0.1407	0.1451	0.1491
50	0.0640	0.0744	0.0811	0.0865	0.0909	0.0948	0.0982	0.1013	0.1042
100	0.0451	0.0525	0.0572	0.0610	0.0641	0.0668	0.0692	0.0714	0.0734
250	0.0285	0.0330	0.0360	0.0384	0.0404	0.0421	0.0437	0.0450	0.0463
500	0.0201	0.0233	0.0254	0.0271	0.0285	0.0298	0.0308	0.0318	0.0327
1000	0.0142	0.0164	0.0180	0.0192	0.0202	0.0210	0.0218	0.0225	0.0231
2000	0.0105	0.0119	0.0127	0.0134	0.0141	0.0147	0.0153	0.0159	0.0163
5000	0.0063	0.0072	0.0080	0.0087	0.0091	0.0094	0.0098	0.0101	0.0103
10000	0.0046	0.0053	0.0057	0.0061	0.0064	0.0067	0.0069	0.0072	0.0073
20000	0.0032	0.0038	0.0041	0.0043	0.0045	0.0047	0.0049	0.0050	0.0052
50000	0.0020	0.0023	0.0025	0.0027	0.0028	0.0030	0.0031	0.0032	0.0033
n \ γ	0.950	0.945	0.940	0.935	0.930	0.925	0.920	0.915	0.910
10	0.2504	0.2561	0.2614	0.2665	0.2713	0.2760	0.2804	0.2847	0.2889
25	0.1528	0.1564	0.1597	0.1628	0.1657	0.1686	0.1713	0.1740	0.1765
50	0.1069	0.1093	0.1117	0.1139	0.1160	0.1180	0.1199	0.1218	0.1235
100	0.0753	0.0770	0.0786	0.0802	0.0816	0.0830	0.0843	0.0856	0.0869
250	0.0474	0.0486	0.0496	0.0505	0.0514	0.0523	0.0532	0.0540	0.0547
500	0.0335	0.0343	0.0350	0.0357	0.0364	0.0370	0.0376	0.0382	0.0387
1000	0.0237	0.0242	0.0248	0.0252	0.0257	0.0261	0.0265	0.0270	0.0273
2000	0.0167	0.0172	0.0175	0.0179	0.0182	0.0186	0.0189	0.0192	0.0195
5000	0.0106	0.0109	0.0111	0.0113	0.0116	0.0118	0.0120	0.0122	0.0124
10000	0.0075	0.0077	0.0078	0.0080	0.0081	0.0083	0.0085	0.0086	0.0087
20000	0.0054	0.0055	0.0056	0.0057	0.0058	0.0059	0.0060	0.0060	0.0061
50000	0.0033	0.0034	0.0035	0.0036	0.0036	0.0037	0.0038	0.0038	0.0039
n \ γ	0.905	0.900	0.895	0.890	0.885	0.880	0.875	0.870	0.865
10	0.2928	0.2966	0.3005	0.3041	0.3076	0.3111	0.3146	0.3179	0.3212
25	0.1789	0.1813	0.1836	0.1858	0.1880	0.1901	0.1921	0.1942	0.1961
50	0.1252	0.1269	0.1285	0.1301	0.1316	0.1331	0.1345	0.1359	0.1373
100	0.0881	0.0893	0.0904	0.0915	0.0926	0.0936	0.0946	0.0955	0.0965
250	0.0555	0.0562	0.0569	0.0576	0.0583	0.0590	0.0596	0.0602	0.0609
500	0.0393	0.0398	0.0403	0.0408	0.0413	0.0417	0.0422	0.0426	0.0430
1000	0.0277	0.0281	0.0284	0.0288	0.0291	0.0294	0.0298	0.0301	0.0304
2000	0.0197	0.0200	0.0202	0.0204	0.0207	0.0209	0.0212	0.0214	0.0216
5000	0.0126	0.0128	0.0129	0.0131	0.0132	0.0134	0.0135	0.0137	0.0138
10000	0.0088	0.0089	0.0090	0.0091	0.0092	0.0094	0.0094	0.0095	0.0096
20000	0.0062	0.0063	0.0063	0.0064	0.0065	0.0065	0.0066	0.0067	0.0067
50000	0.0039	0.0040	0.0040	0.0041	0.0041	0.0042	0.0042	0.0043	0.0043

Table 1.14: Empirical Critical Values for the CCR Method for Different n and γ from 73% to 86%

n \ γ	0.860	0.855	0.850	0.845	0.840	0.835	0.830	0.825	0.820
10	0.3243	0.3273	0.3304	0.3334	0.3363	0.3392	0.3420	0.3449	0.3476
25	0.1981	0.1999	0.2017	0.2036	0.2054	0.2072	0.2090	0.2107	0.2124
50	0.1387	0.1400	0.1413	0.1426	0.1438	0.1451	0.1463	0.1475	0.1487
100	0.0975	0.0984	0.0993	0.1002	0.1011	0.1020	0.1028	0.1037	0.1045
250	0.0615	0.0620	0.0626	0.0632	0.0638	0.0643	0.0649	0.0654	0.0659
500	0.0435	0.0439	0.0443	0.0447	0.0451	0.0455	0.0459	0.0462	0.0466
1000	0.0307	0.0310	0.0313	0.0316	0.0318	0.0321	0.0324	0.0326	0.0329
2000	0.0217	0.0220	0.0222	0.0224	0.0225	0.0227	0.0229	0.0231	0.0233
5000	0.0140	0.0141	0.0142	0.0143	0.0145	0.0146	0.0147	0.0149	0.0150
10000	0.0097	0.0098	0.0099	0.0100	0.0101	0.0102	0.0103	0.0104	0.0104
20000	0.0068	0.0069	0.0070	0.0070	0.0071	0.0071	0.0072	0.0073	0.0073
50000	0.0043	0.0044	0.0044	0.0045	0.0045	0.0045	0.0046	0.0046	0.0047
n \ γ	0.815	0.810	0.805	0.800	0.795	0.790	0.785	0.780	0.775
10	0.3504	0.3530	0.3557	0.3584	0.3610	0.3636	0.3661	0.3686	0.3712
25	0.2140	0.2157	0.2173	0.2189	0.2204	0.2220	0.2236	0.2251	0.2267
50	0.1499	0.1510	0.1521	0.1532	0.1544	0.1555	0.1565	0.1576	0.1587
100	0.1054	0.1062	0.1069	0.1077	0.1085	0.1093	0.1101	0.1108	0.1115
250	0.0664	0.0669	0.0674	0.0679	0.0684	0.0689	0.0694	0.0699	0.0704
500	0.0470	0.0473	0.0477	0.0480	0.0484	0.0487	0.0490	0.0494	0.0497
1000	0.0331	0.0334	0.0336	0.0339	0.0341	0.0344	0.0346	0.0349	0.0351
2000	0.0235	0.0237	0.0238	0.0240	0.0242	0.0244	0.0245	0.0248	0.0249
5000	0.0151	0.0152	0.0153	0.0154	0.0155	0.0155	0.0156	0.0157	0.0158
10000	0.0105	0.0106	0.0107	0.0108	0.0108	0.0109	0.0110	0.0111	0.0111
20000	0.0074	0.0074	0.0075	0.0076	0.0076	0.0077	0.0077	0.0078	0.0078
50000	0.0047	0.0047	0.0048	0.0048	0.0048	0.0049	0.0049	0.0049	0.0050
n \ γ	0.770	0.765	0.760	0.755	0.750	0.745	0.740	0.735	0.730
10	0.3737	0.3761	0.3786	0.3811	0.3835	0.3860	0.3883	0.3907	0.3931
25	0.2282	0.2297	0.2311	0.2326	0.2340	0.2355	0.2369	0.2383	0.2397
50	0.1597	0.1608	0.1618	0.1629	0.1639	0.1649	0.1659	0.1669	0.1678
100	0.1123	0.1130	0.1137	0.1145	0.1152	0.1159	0.1166	0.1173	0.1180
250	0.0708	0.0713	0.0718	0.0722	0.0726	0.0731	0.0735	0.0740	0.0744
500	0.0500	0.0504	0.0507	0.0510	0.0513	0.0517	0.0520	0.0523	0.0526
1000	0.0353	0.0356	0.0358	0.0360	0.0363	0.0365	0.0367	0.0369	0.0371
2000	0.0251	0.0253	0.0254	0.0255	0.0257	0.0259	0.0260	0.0261	0.0263
5000	0.0159	0.0160	0.0161	0.0163	0.0163	0.0164	0.0165	0.0166	0.0167
10000	0.0112	0.0113	0.0113	0.0114	0.0115	0.0116	0.0116	0.0117	0.0118
20000	0.0079	0.0079	0.0080	0.0080	0.0081	0.0081	0.0082	0.0082	0.0082
50000	0.0050	0.0050	0.0051	0.0051	0.0051	0.0052	0.0052	0.0052	0.0053

Table 1.15: Empirical Critical Values for the CCR Method for Different n and γ from 59.5% to 72.5%

n \ γ	0.725	0.720	0.715	0.710	0.705	0.700	0.695	0.690	0.685
10	0.3954	0.3978	0.4001	0.4024	0.4047	0.4069	0.4092	0.4113	0.4136
25	0.2411	0.2425	0.2439	0.2453	0.2467	0.2480	0.2494	0.2508	0.2521
50	0.1688	0.1698	0.1708	0.1717	0.1727	0.1736	0.1746	0.1755	0.1765
100	0.1187	0.1194	0.1200	0.1207	0.1214	0.1221	0.1228	0.1234	0.1241
250	0.0748	0.0753	0.0757	0.0761	0.0766	0.0770	0.0774	0.0778	0.0782
500	0.0529	0.0532	0.0535	0.0538	0.0541	0.0544	0.0547	0.0550	0.0553
1000	0.0374	0.0376	0.0378	0.0380	0.0382	0.0384	0.0386	0.0388	0.0390
2000	0.0264	0.0266	0.0268	0.0269	0.0271	0.0272	0.0274	0.0275	0.0277
5000	0.0168	0.0169	0.0170	0.0171	0.0172	0.0173	0.0174	0.0175	0.0176
10000	0.0118	0.0119	0.0119	0.0120	0.0121	0.0122	0.0122	0.0123	0.0124
20000	0.0083	0.0083	0.0084	0.0084	0.0085	0.0085	0.0086	0.0086	0.0087
50000	0.0053	0.0053	0.0053	0.0054	0.0054	0.0054	0.0055	0.0055	0.0055
n \ γ	0.680	0.675	0.670	0.665	0.660	0.655	0.650	0.645	0.640
10	0.4159	0.4181	0.4203	0.4225	0.4248	0.4270	0.4292	0.4314	0.4336
25	0.2535	0.2548	0.2562	0.2575	0.2588	0.2600	0.2613	0.2627	0.2640
50	0.1774	0.1783	0.1793	0.1802	0.1811	0.1820	0.1829	0.1838	0.1847
100	0.1248	0.1254	0.1260	0.1267	0.1273	0.1280	0.1286	0.1292	0.1299
250	0.0786	0.0790	0.0794	0.0799	0.0803	0.0807	0.0811	0.0815	0.0819
500	0.0556	0.0559	0.0561	0.0564	0.0567	0.0570	0.0573	0.0576	0.0579
1000	0.0392	0.0394	0.0396	0.0399	0.0401	0.0403	0.0405	0.0407	0.0409
2000	0.0278	0.0280	0.0281	0.0282	0.0284	0.0285	0.0286	0.0287	0.0289
5000	0.0177	0.0177	0.0178	0.0179	0.0180	0.0181	0.0182	0.0183	0.0184
10000	0.0125	0.0125	0.0126	0.0127	0.0128	0.0128	0.0129	0.0129	0.0130
20000	0.0087	0.0088	0.0088	0.0088	0.0089	0.0089	0.0090	0.0090	0.0090
50000	0.0055	0.0056	0.0056	0.0056	0.0057	0.0057	0.0057	0.0057	0.0058
n \ γ	0.635	0.630	0.625	0.620	0.615	0.610	0.605	0.600	0.595
10	0.4357	0.4379	0.4401	0.4423	0.4445	0.4467	0.4489	0.4511	0.4532
25	0.2653	0.2666	0.2679	0.2692	0.2704	0.2717	0.2730	0.2743	0.2755
50	0.1857	0.1866	0.1875	0.1884	0.1893	0.1902	0.1911	0.1920	0.1928
100	0.1305	0.1312	0.1318	0.1324	0.1331	0.1337	0.1343	0.1350	0.1356
250	0.0823	0.0827	0.0831	0.0835	0.0839	0.0842	0.0846	0.0851	0.0854
500	0.0581	0.0584	0.0587	0.0590	0.0593	0.0596	0.0598	0.0601	0.0604
1000	0.0411	0.0413	0.0415	0.0416	0.0418	0.0420	0.0422	0.0424	0.0426
2000	0.0291	0.0292	0.0293	0.0295	0.0296	0.0297	0.0299	0.0300	0.0301
5000	0.0184	0.0185	0.0186	0.0187	0.0188	0.0189	0.0190	0.0190	0.0191
10000	0.0131	0.0131	0.0132	0.0133	0.0134	0.0134	0.0135	0.0136	0.0136
20000	0.0091	0.0091	0.0092	0.0092	0.0093	0.0093	0.0094	0.0094	0.0095
50000	0.0058	0.0058	0.0058	0.0059	0.0059	0.0059	0.0060	0.0060	0.0060

Table 1.16: Empirical Critical Values for the CCR Method for Different n and γ from 46% to 59%

n \ γ	0.590	0.585	0.580	0.575	0.570	0.565	0.560	0.555	0.550
10	0.4554	0.4575	0.4597	0.4619	0.4641	0.4662	0.4684	0.4706	0.4728
25	0.2768	0.2781	0.2794	0.2807	0.2820	0.2833	0.2846	0.2859	0.2871
50	0.1937	0.1946	0.1955	0.1964	0.1973	0.1982	0.1990	0.1999	0.2008
100	0.1362	0.1369	0.1375	0.1381	0.1387	0.1393	0.1400	0.1406	0.1412
250	0.0858	0.0862	0.0866	0.0870	0.0874	0.0878	0.0882	0.0886	0.0890
500	0.0607	0.0609	0.0612	0.0615	0.0618	0.0620	0.0623	0.0626	0.0629
1000	0.0428	0.0430	0.0432	0.0434	0.0436	0.0438	0.0440	0.0442	0.0444
2000	0.0303	0.0305	0.0306	0.0307	0.0308	0.0309	0.0311	0.0312	0.0313
5000	0.0193	0.0194	0.0194	0.0195	0.0196	0.0197	0.0198	0.0199	0.0199
10000	0.0137	0.0138	0.0138	0.0139	0.0139	0.0140	0.0140	0.0141	0.0142
20000	0.0095	0.0096	0.0096	0.0096	0.0097	0.0097	0.0098	0.0098	0.0099
50000	0.0060	0.0061	0.0061	0.0061	0.0061	0.0062	0.0062	0.0062	0.0063
n \ γ	0.545	0.540	0.535	0.530	0.525	0.520	0.515	0.510	0.505
10	0.4749	0.4771	0.4793	0.4814	0.4836	0.4857	0.4879	0.4901	0.4923
25	0.2884	0.2897	0.2910	0.2923	0.2935	0.2948	0.2961	0.2974	0.2987
50	0.2017	0.2026	0.2035	0.2044	0.2053	0.2062	0.2071	0.2079	0.2088
100	0.1418	0.1424	0.1431	0.1437	0.1443	0.1449	0.1456	0.1462	0.1469
250	0.0894	0.0898	0.0902	0.0905	0.0909	0.0913	0.0917	0.0921	0.0925
500	0.0631	0.0634	0.0637	0.0640	0.0643	0.0645	0.0648	0.0651	0.0654
1000	0.0446	0.0448	0.0450	0.0452	0.0454	0.0456	0.0458	0.0460	0.0462
2000	0.0315	0.0316	0.0317	0.0319	0.0320	0.0322	0.0323	0.0324	0.0326
5000	0.0201	0.0201	0.0202	0.0203	0.0204	0.0205	0.0205	0.0206	0.0207
10000	0.0142	0.0143	0.0143	0.0144	0.0145	0.0145	0.0146	0.0146	0.0147
20000	0.0099	0.0099	0.0100	0.0100	0.0101	0.0101	0.0101	0.0102	0.0102
50000	0.0063	0.0063	0.0063	0.0064	0.0064	0.0064	0.0064	0.0065	0.0065
n \ γ	0.500	0.495	0.490	0.485	0.480	0.475	0.470	0.465	0.460
10	0.4945	0.4968	0.4991	0.5013	0.5036	0.5058	0.5081	0.5104	0.5127
25	0.3000	0.3013	0.3026	0.3039	0.3052	0.3066	0.3078	0.3092	0.3105
50	0.2097	0.2106	0.2115	0.2124	0.2133	0.2142	0.2152	0.2161	0.2170
100	0.1475	0.1481	0.1488	0.1494	0.1500	0.1507	0.1513	0.1519	0.1526
250	0.0929	0.0933	0.0937	0.0941	0.0945	0.0949	0.0952	0.0956	0.0961
500	0.0656	0.0659	0.0662	0.0665	0.0668	0.0670	0.0673	0.0676	0.0679
1000	0.0464	0.0466	0.0468	0.0470	0.0472	0.0474	0.0476	0.0478	0.0480
2000	0.0327	0.0329	0.0330	0.0331	0.0332	0.0334	0.0335	0.0336	0.0337
5000	0.0208	0.0209	0.0210	0.0211	0.0212	0.0213	0.0213	0.0214	0.0215
10000	0.0148	0.0148	0.0149	0.0149	0.0150	0.0151	0.0152	0.0152	0.0153
20000	0.0103	0.0103	0.0104	0.0104	0.0105	0.0105	0.0105	0.0106	0.0106
50000	0.0065	0.0066	0.0066	0.0066	0.0067	0.0067	0.0067	0.0067	0.0068

Table 1.17: Empirical Critical Values for the CCR Method for Different n and γ from 32.5% to 45.5%

n \ γ	0.455	0.450	0.445	0.440	0.435	0.430	0.425	0.420	0.415
10	0.5150	0.5173	0.5196	0.5220	0.5243	0.5267	0.5291	0.5315	0.5339
25	0.3118	0.3132	0.3145	0.3159	0.3173	0.3186	0.3200	0.3214	0.3228
50	0.2179	0.2188	0.2198	0.2207	0.2216	0.2226	0.2235	0.2245	0.2254
100	0.1532	0.1539	0.1545	0.1552	0.1558	0.1565	0.1571	0.1578	0.1584
250	0.0965	0.0969	0.0973	0.0977	0.0981	0.0985	0.0989	0.0993	0.0998
500	0.0682	0.0685	0.0688	0.0690	0.0693	0.0696	0.0699	0.0702	0.0705
1000	0.0482	0.0484	0.0486	0.0488	0.0490	0.0492	0.0494	0.0496	0.0498
2000	0.0339	0.0340	0.0342	0.0344	0.0346	0.0347	0.0348	0.0350	0.0351
5000	0.0217	0.0217	0.0218	0.0219	0.0220	0.0221	0.0222	0.0223	0.0224
10000	0.0153	0.0154	0.0154	0.0155	0.0156	0.0156	0.0157	0.0158	0.0158
20000	0.0107	0.0107	0.0108	0.0108	0.0109	0.0109	0.0110	0.0110	0.0111
50000	0.0068	0.0068	0.0069	0.0069	0.0069	0.0070	0.0070	0.0070	0.0070
n \ γ	0.410	0.405	0.400	0.395	0.390	0.385	0.380	0.375	0.370
10	0.5363	0.5388	0.5413	0.5438	0.5462	0.5487	0.5513	0.5538	0.5563
25	0.3242	0.3256	0.3270	0.3283	0.3297	0.3311	0.3325	0.3339	0.3354
50	0.2264	0.2273	0.2283	0.2292	0.2302	0.2312	0.2322	0.2332	0.2342
100	0.1591	0.1598	0.1604	0.1611	0.1618	0.1625	0.1632	0.1639	0.1646
250	0.1002	0.1006	0.1010	0.1014	0.1019	0.1023	0.1027	0.1031	0.1036
500	0.0708	0.0711	0.0714	0.0717	0.0720	0.0723	0.0726	0.0729	0.0732
1000	0.0500	0.0502	0.0504	0.0506	0.0508	0.0510	0.0513	0.0515	0.0517
2000	0.0353	0.0354	0.0356	0.0357	0.0359	0.0360	0.0362	0.0363	0.0365
5000	0.0225	0.0226	0.0226	0.0228	0.0228	0.0230	0.0231	0.0232	0.0233
10000	0.0159	0.0160	0.0160	0.0161	0.0162	0.0162	0.0163	0.0164	0.0164
20000	0.0111	0.0112	0.0112	0.0113	0.0113	0.0113	0.0114	0.0115	0.0115
50000	0.0071	0.0071	0.0071	0.0071	0.0072	0.0072	0.0072	0.0073	0.0073
n \ γ	0.365	0.360	0.355	0.350	0.345	0.340	0.335	0.330	0.325
10	0.5589	0.5615	0.5641	0.5667	0.5694	0.5721	0.5748	0.5775	0.5803
25	0.3368	0.3383	0.3398	0.3413	0.3428	0.3443	0.3458	0.3474	0.3490
50	0.2352	0.2362	0.2372	0.2382	0.2392	0.2403	0.2413	0.2424	0.2434
100	0.1653	0.1660	0.1667	0.1674	0.1681	0.1688	0.1696	0.1703	0.1710
250	0.1040	0.1044	0.1049	0.1053	0.1058	0.1062	0.1067	0.1071	0.1076
500	0.0735	0.0738	0.0741	0.0744	0.0748	0.0751	0.0754	0.0757	0.0761
1000	0.0519	0.0521	0.0523	0.0526	0.0528	0.0530	0.0532	0.0535	0.0537
2000	0.0367	0.0368	0.0370	0.0372	0.0374	0.0376	0.0377	0.0378	0.0380
5000	0.0234	0.0234	0.0235	0.0236	0.0237	0.0238	0.0240	0.0241	0.0241
10000	0.0165	0.0165	0.0166	0.0167	0.0167	0.0168	0.0169	0.0169	0.0170
20000	0.0115	0.0116	0.0117	0.0117	0.0118	0.0118	0.0118	0.0119	0.0119
50000	0.0073	0.0073	0.0074	0.0074	0.0074	0.0075	0.0075	0.0075	0.0076

Table 1.18: Empirical Critical Values for the CCR Method for Different n and γ from 19% to 32%

n \ γ	0.320	0.315	0.310	0.305	0.300	0.295	0.290	0.285	0.280
10	0.5830	0.5858	0.5887	0.5916	0.5945	0.5975	0.6005	0.6034	0.6065
25	0.3506	0.3522	0.3538	0.3554	0.3570	0.3586	0.3603	0.3620	0.3637
50	0.2445	0.2455	0.2466	0.2477	0.2488	0.2499	0.2510	0.2522	0.2534
100	0.1718	0.1725	0.1733	0.1741	0.1748	0.1756	0.1764	0.1772	0.1780
250	0.1081	0.1085	0.1090	0.1095	0.1100	0.1105	0.1110	0.1114	0.1119
500	0.0764	0.0767	0.0771	0.0774	0.0777	0.0781	0.0784	0.0788	0.0791
1000	0.0539	0.0542	0.0544	0.0546	0.0549	0.0551	0.0554	0.0556	0.0559
2000	0.0382	0.0383	0.0385	0.0387	0.0388	0.0390	0.0391	0.0393	0.0395
5000	0.0243	0.0244	0.0245	0.0246	0.0247	0.0248	0.0249	0.0250	0.0251
10000	0.0171	0.0171	0.0172	0.0173	0.0174	0.0174	0.0175	0.0176	0.0177
20000	0.0120	0.0121	0.0121	0.0122	0.0122	0.0123	0.0123	0.0124	0.0125
50000	0.0076	0.0076	0.0077	0.0077	0.0077	0.0078	0.0078	0.0078	0.0079
n \ γ	0.275	0.270	0.265	0.260	0.255	0.250	0.245	0.240	0.235
10	0.6095	0.6127	0.6159	0.6192	0.6224	0.6258	0.6292	0.6325	0.6360
25	0.3654	0.3672	0.3690	0.3708	0.3726	0.3743	0.3762	0.3780	0.3799
50	0.2545	0.2557	0.2569	0.2581	0.2593	0.2605	0.2617	0.2630	0.2644
100	0.1788	0.1796	0.1805	0.1813	0.1821	0.1830	0.1838	0.1847	0.1856
250	0.1125	0.1130	0.1135	0.1140	0.1145	0.1151	0.1156	0.1162	0.1167
500	0.0795	0.0798	0.0802	0.0806	0.0809	0.0813	0.0817	0.0821	0.0825
1000	0.0561	0.0564	0.0566	0.0569	0.0571	0.0574	0.0577	0.0579	0.0582
2000	0.0397	0.0399	0.0401	0.0402	0.0405	0.0407	0.0408	0.0410	0.0412
5000	0.0252	0.0253	0.0255	0.0256	0.0257	0.0259	0.0260	0.0261	0.0262
10000	0.0177	0.0178	0.0179	0.0180	0.0181	0.0181	0.0182	0.0183	0.0184
20000	0.0125	0.0126	0.0126	0.0127	0.0128	0.0128	0.0129	0.0130	0.0130
50000	0.0079	0.0080	0.0080	0.0080	0.0081	0.0081	0.0081	0.0082	0.0082
n \ γ	0.230	0.225	0.220	0.215	0.210	0.205	0.200	0.195	0.190
10	0.6395	0.6432	0.6468	0.6504	0.6541	0.6579	0.6619	0.6658	0.6699
25	0.3818	0.3838	0.3858	0.3879	0.3899	0.3920	0.3942	0.3964	0.3986
50	0.2657	0.2670	0.2683	0.2697	0.2711	0.2725	0.2739	0.2754	0.2768
100	0.1865	0.1874	0.1883	0.1893	0.1902	0.1912	0.1922	0.1932	0.1943
250	0.1173	0.1179	0.1185	0.1191	0.1196	0.1203	0.1209	0.1215	0.1221
500	0.0829	0.0833	0.0837	0.0841	0.0845	0.0850	0.0854	0.0858	0.0863
1000	0.0585	0.0588	0.0591	0.0594	0.0597	0.0600	0.0603	0.0606	0.0609
2000	0.0414	0.0416	0.0418	0.0420	0.0422	0.0424	0.0426	0.0428	0.0430
5000	0.0263	0.0264	0.0264	0.0266	0.0267	0.0268	0.0270	0.0271	0.0272
10000	0.0185	0.0186	0.0187	0.0188	0.0189	0.0190	0.0191	0.0192	0.0193
20000	0.0131	0.0131	0.0132	0.0133	0.0133	0.0134	0.0135	0.0135	0.0136
50000	0.0083	0.0083	0.0083	0.0084	0.0084	0.0085	0.0085	0.0086	0.0086

Table 1.19: Empirical Critical Values for the CCR Method for Different n and γ from 5.5% to 18.5%

n \ γ	0.185	0.180	0.175	0.170	0.165	0.160	0.155	0.150	0.145
10	0.6739	0.6782	0.6825	0.6869	0.6915	0.6961	0.7009	0.7058	0.7107
25	0.4008	0.4031	0.4054	0.4078	0.4103	0.4129	0.4154	0.4181	0.4207
50	0.2783	0.2799	0.2815	0.2831	0.2847	0.2863	0.2881	0.2898	0.2917
100	0.1953	0.1964	0.1975	0.1986	0.1998	0.2010	0.2022	0.2034	0.2046
250	0.1228	0.1235	0.1241	0.1248	0.1255	0.1263	0.1270	0.1277	0.1285
500	0.0868	0.0872	0.0877	0.0882	0.0887	0.0892	0.0897	0.0903	0.0908
1000	0.0612	0.0616	0.0619	0.0622	0.0626	0.0629	0.0633	0.0637	0.0641
2000	0.0433	0.0435	0.0437	0.0440	0.0442	0.0445	0.0447	0.0449	0.0452
5000	0.0273	0.0275	0.0277	0.0278	0.0280	0.0282	0.0284	0.0285	0.0288
10000	0.0194	0.0195	0.0196	0.0197	0.0198	0.0199	0.0201	0.0202	0.0203
20000	0.0137	0.0138	0.0138	0.0139	0.0140	0.0140	0.0141	0.0142	0.0143
50000	0.0086	0.0087	0.0087	0.0088	0.0088	0.0089	0.0089	0.0090	0.0091
n \ γ	0.140	0.135	0.130	0.125	0.120	0.115	0.110	0.105	0.100
10	0.7159	0.7213	0.7268	0.7324	0.7382	0.7444	0.7507	0.7572	0.7640
25	0.4235	0.4264	0.4293	0.4324	0.4356	0.4389	0.4423	0.4457	0.4494
50	0.2936	0.2955	0.2974	0.2995	0.3016	0.3037	0.3060	0.3082	0.3106
100	0.2059	0.2072	0.2085	0.2099	0.2114	0.2129	0.2144	0.2160	0.2176
250	0.1293	0.1301	0.1309	0.1318	0.1327	0.1336	0.1345	0.1355	0.1366
500	0.0914	0.0920	0.0925	0.0931	0.0938	0.0944	0.0951	0.0958	0.0965
1000	0.0645	0.0649	0.0653	0.0657	0.0662	0.0666	0.0671	0.0676	0.0681
2000	0.0455	0.0457	0.0460	0.0462	0.0466	0.0469	0.0473	0.0476	0.0480
5000	0.0289	0.0291	0.0293	0.0295	0.0297	0.0299	0.0302	0.0303	0.0306
10000	0.0204	0.0205	0.0206	0.0208	0.0209	0.0210	0.0212	0.0213	0.0215
20000	0.0143	0.0145	0.0145	0.0146	0.0147	0.0148	0.0149	0.0150	0.0151
50000	0.0091	0.0092	0.0092	0.0093	0.0094	0.0094	0.0095	0.0095	0.0096
n \ γ	0.095	0.090	0.085	0.080	0.075	0.070	0.065	0.060	0.055
10	0.7710	0.7784	0.7862	0.7941	0.8026	0.8115	0.8209	0.8311	0.8417
25	0.4532	0.4572	0.4614	0.4657	0.4703	0.4752	0.4805	0.4860	0.4920
50	0.3132	0.3158	0.3185	0.3214	0.3244	0.3276	0.3309	0.3346	0.3384
100	0.2193	0.2211	0.2230	0.2249	0.2270	0.2293	0.2316	0.2340	0.2367
250	0.1376	0.1387	0.1399	0.1411	0.1424	0.1438	0.1452	0.1467	0.1484
500	0.0973	0.0980	0.0989	0.0997	0.1006	0.1016	0.1026	0.1036	0.1048
1000	0.0686	0.0692	0.0697	0.0704	0.0710	0.0717	0.0724	0.0731	0.0739
2000	0.0484	0.0488	0.0492	0.0496	0.0500	0.0503	0.0508	0.0512	0.0518
5000	0.0308	0.0311	0.0313	0.0316	0.0318	0.0322	0.0325	0.0328	0.0332
10000	0.0216	0.0218	0.0220	0.0222	0.0224	0.0226	0.0229	0.0231	0.0234
20000	0.0152	0.0153	0.0155	0.0156	0.0158	0.0160	0.0161	0.0163	0.0165
50000	0.0097	0.0098	0.0098	0.0099	0.0100	0.0101	0.0102	0.0104	0.0105

Table 1.20: Empirical Critical Values for the CCR Method for Different n and γ from 0.5% to 5%

n \ γ	0.050	0.045	0.040	0.035	0.030
10	0.8536	0.8665	0.8805	0.8968	0.9145
25	0.4983	0.5053	0.5131	0.5217	0.5318
50	0.3426	0.3473	0.3525	0.3581	0.3645
100	0.2396	0.2427	0.2461	0.2499	0.2543
250	0.1502	0.1521	0.1542	0.1565	0.1592
500	0.1061	0.1074	0.1088	0.1105	0.1123
1000	0.0748	0.0757	0.0768	0.0779	0.0793
2000	0.0525	0.0532	0.0540	0.0548	0.0556
5000	0.0335	0.0340	0.0345	0.0350	0.0356
10000	0.0236	0.0240	0.0243	0.0247	0.0252
20000	0.0167	0.0169	0.0172	0.0175	0.0178
50000	0.0106	0.0107	0.0108	0.0110	0.0112
n \ γ	0.025	0.020	0.015	0.010	0.005
10	0.9349	0.9592	0.9899	1.0315	1.0980
25	0.5428	0.5565	0.5737	0.5973	0.6357
50	0.3721	0.3812	0.3926	0.4077	0.4330
100	0.2593	0.2654	0.2731	0.2833	0.3002
250	0.1624	0.1660	0.1706	0.1772	0.1873
500	0.1146	0.1172	0.1204	0.1249	0.1321
1000	0.0808	0.0826	0.0849	0.0880	0.0930
2000	0.0566	0.0581	0.0597	0.0623	0.0650
5000	0.0363	0.0372	0.0382	0.0398	0.0419
10000	0.0256	0.0262	0.0269	0.0278	0.0291
20000	0.0181	0.0186	0.0193	0.0200	0.0210
50000	0.0114	0.0116	0.0120	0.0124	0.0130

Appendix 1.II: Figures

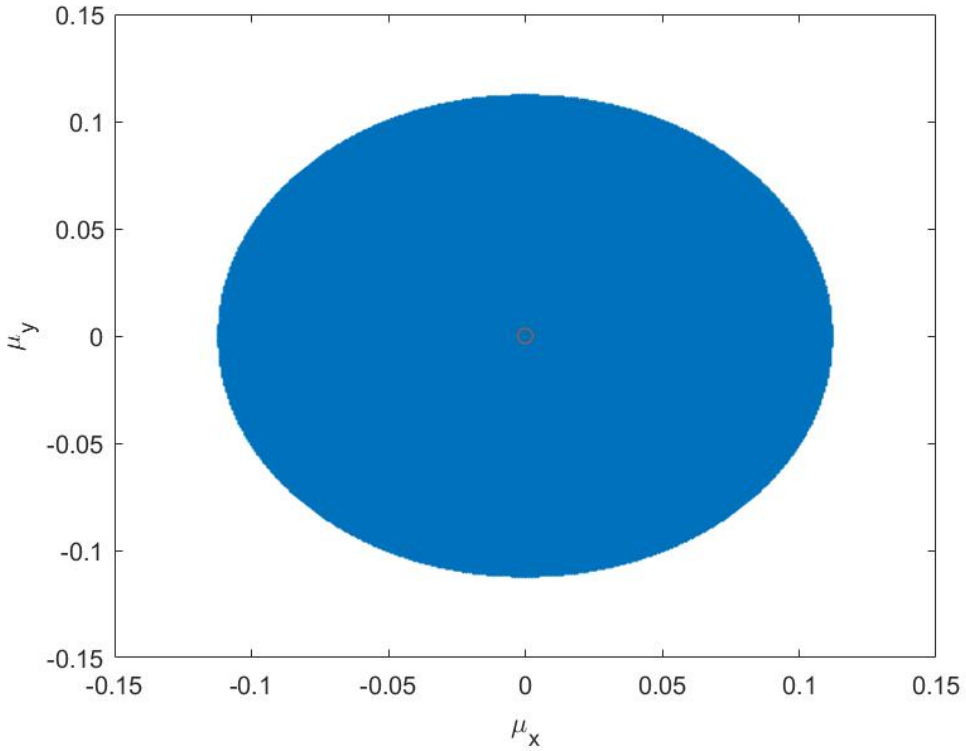

Figure 1.10: Projection of the CCR Confidence Set on (μ_x, μ_y)-Plane ($n = 100$, $\gamma = 0.1$)

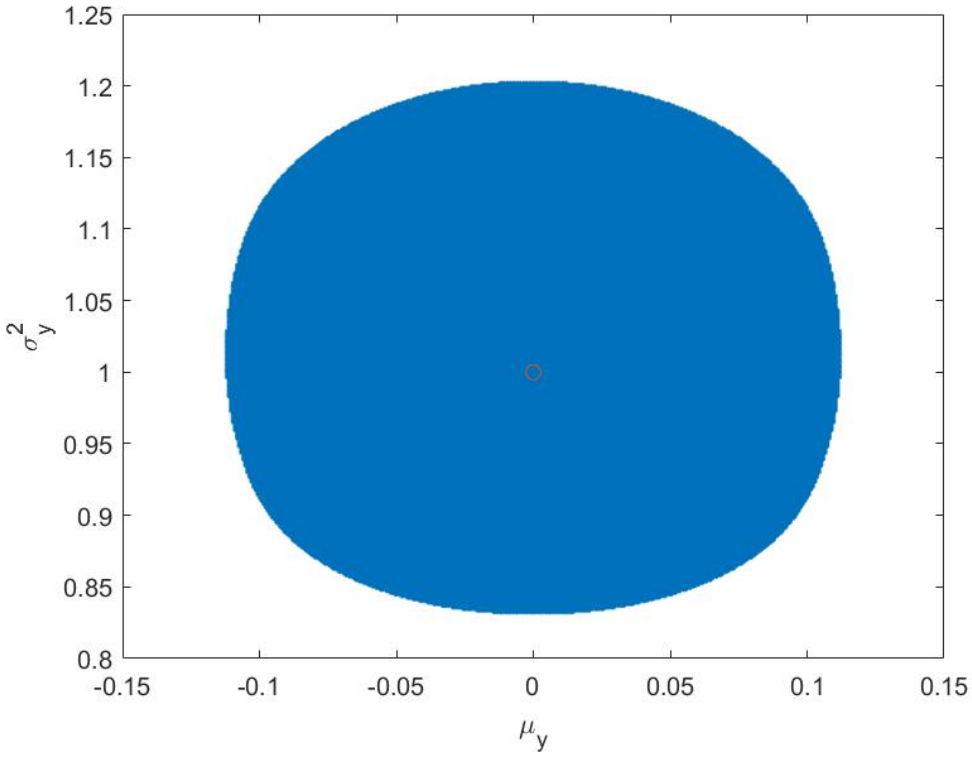

Figure 1.11: Projection of the CCR Confidence Set on (μ_y, σ_y^2)-Plane ($n = 100$, $\gamma = 0.1$)

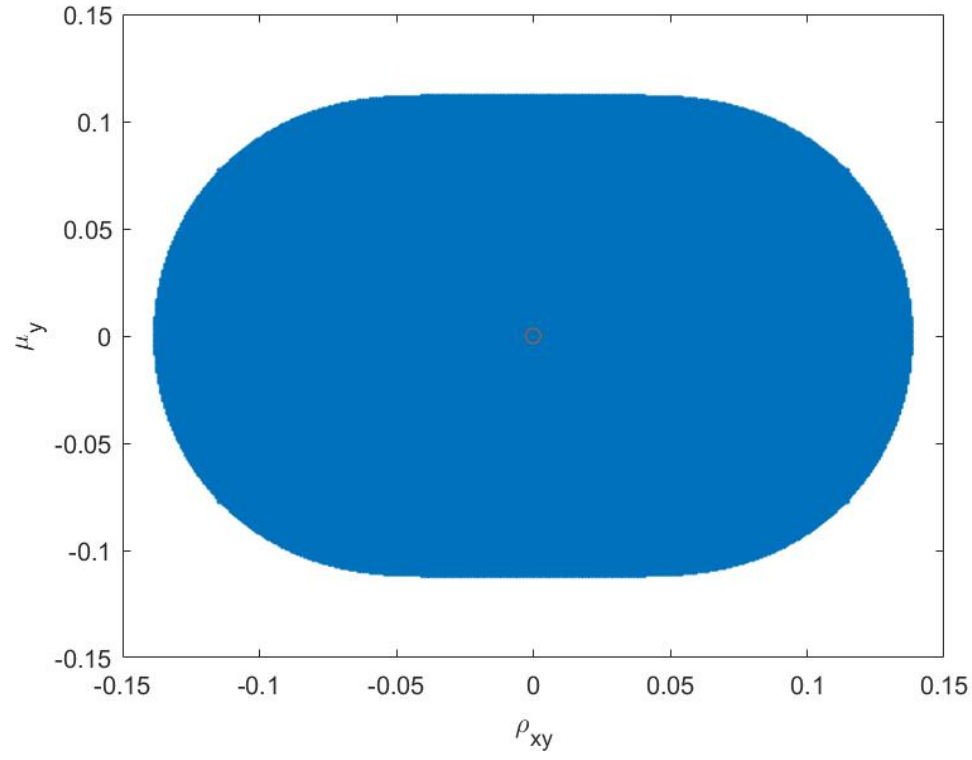

Figure 1.12: Projection of the CCR Confidence Set on (ρ, μ_y)-Plane ($n = 100$, $\gamma = 0.1$)

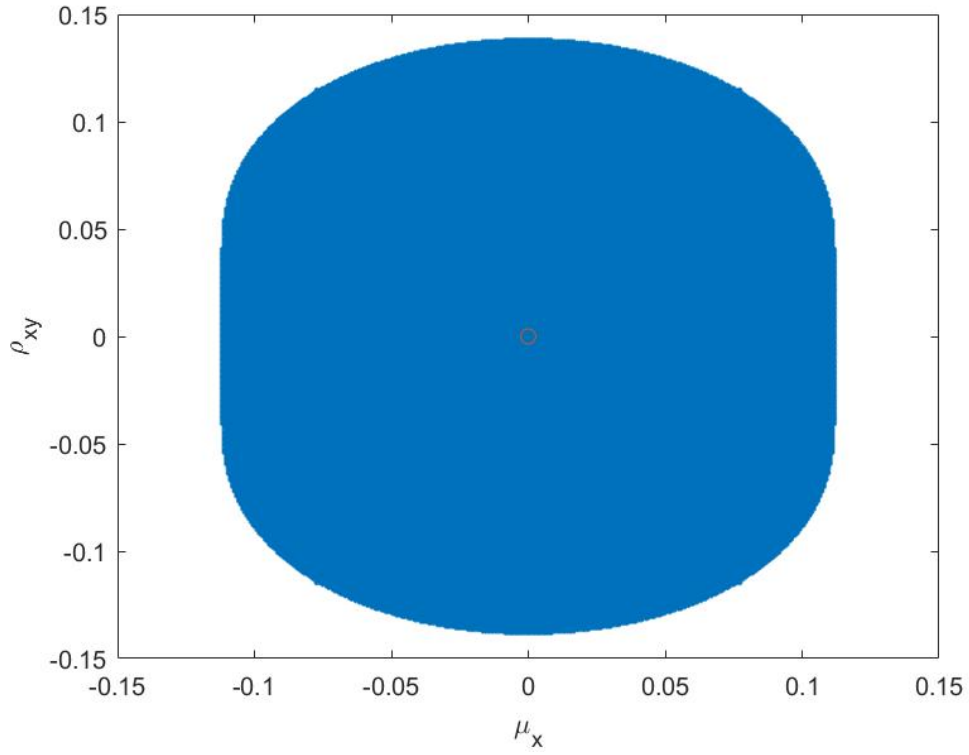

Figure 1.13: Projection of the CCR Confidence Set on (μ_x, ρ)-Plane ($n = 100$, $\gamma = 0.1$)

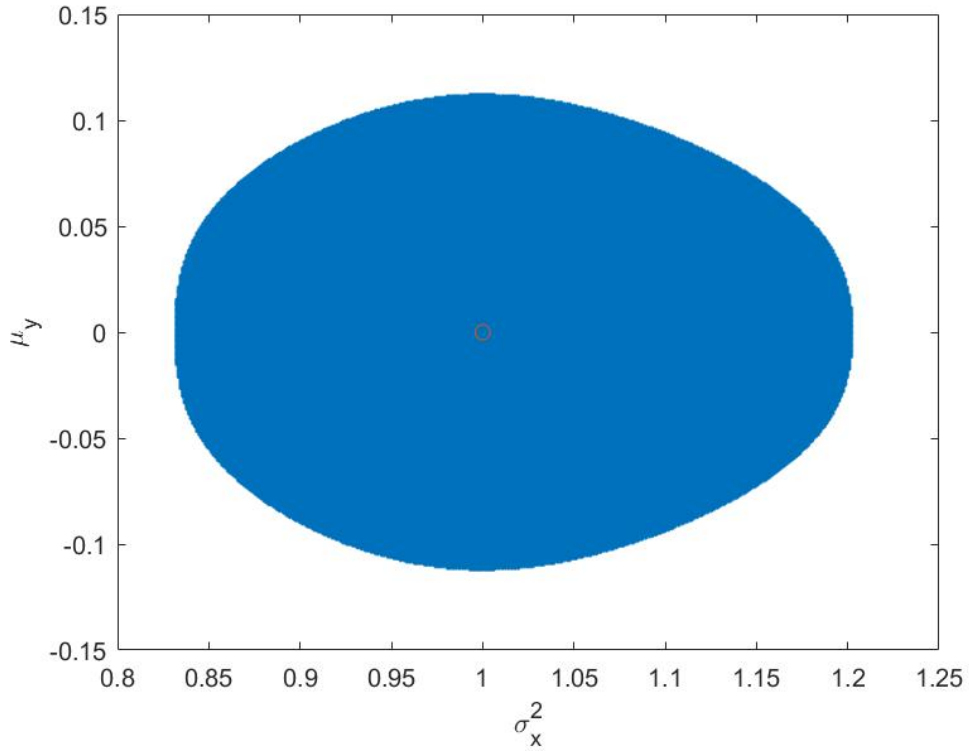

Figure 1.14: Projection of the CCR Confidence Set on (σ_x^2, μ_y)-Plane ($n = 100$, $\gamma = 0.1$)

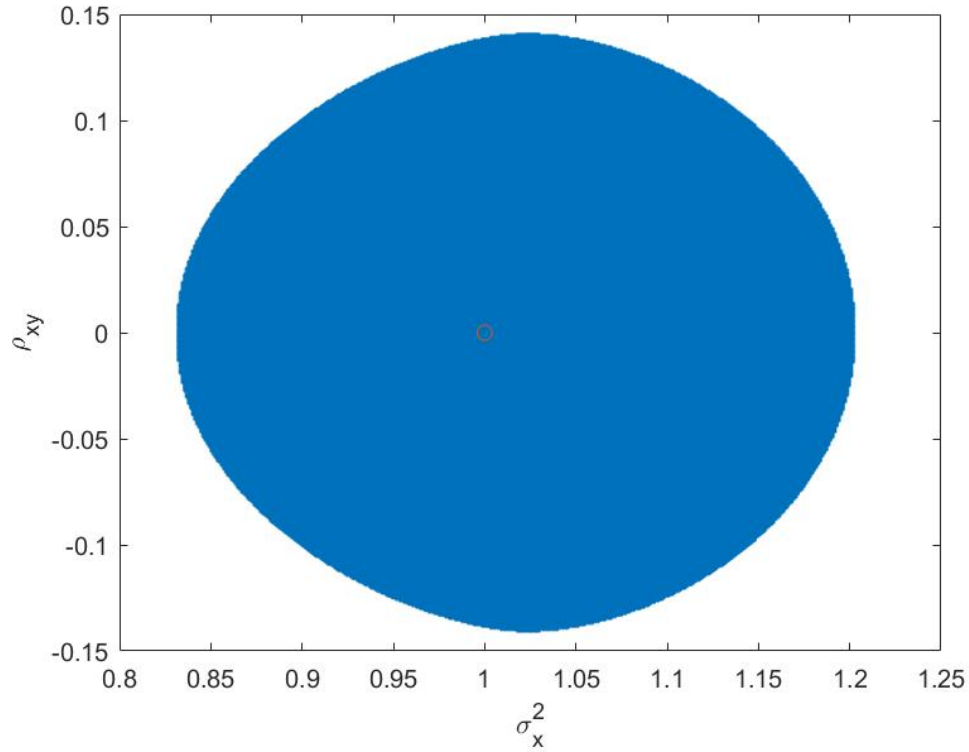

Figure 1.15: Projection of the CCR Confidence Set on (σ_x^2, ρ)-Plane ($n = 100$, $\gamma = 0.1$)

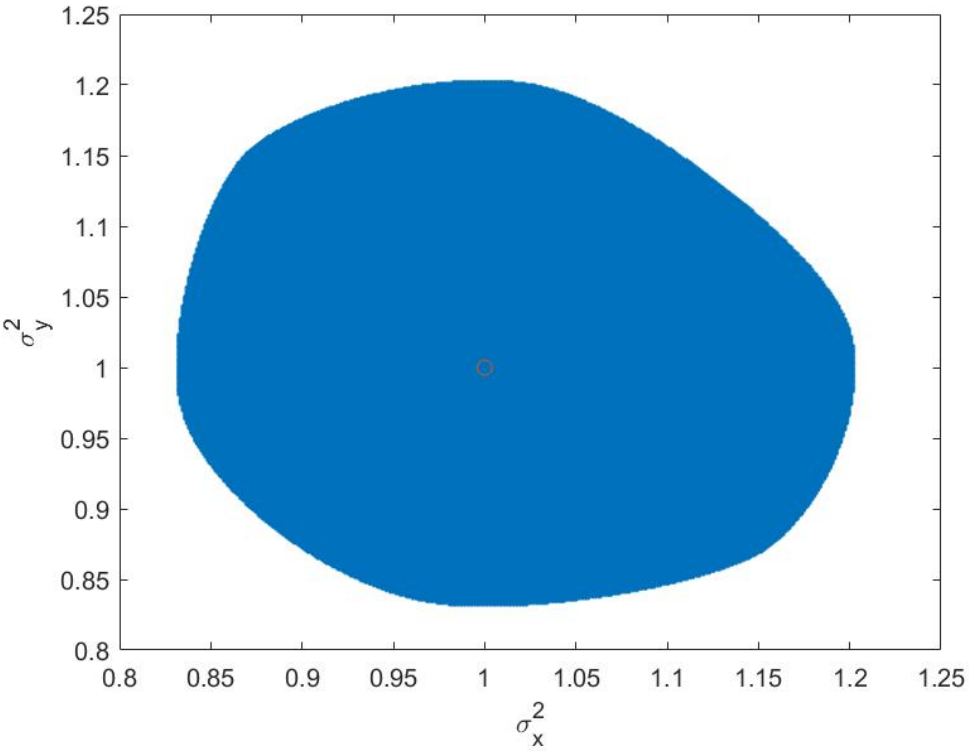

Figure 1.16: Projection of the CCR Confidence Set on (σ_x^2, σ_y^2)-Plane ($n = 100$, $\gamma = 0.1$)

Appendix 1.III: CCR Approach for Univariate Normal Distribution

Primary Theory for CCR Approach

I start with the introduction of a theoretical background that will be used to derive this technique. Assume two Normal distributions, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. Their PDFs look like (1.43):

$$f_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}, \ i = 1, 2 \tag{1.43}$$

If one plots both of the density functions on one coordinate plane, one gets a figure similar to one displayed in Figure 1.17:

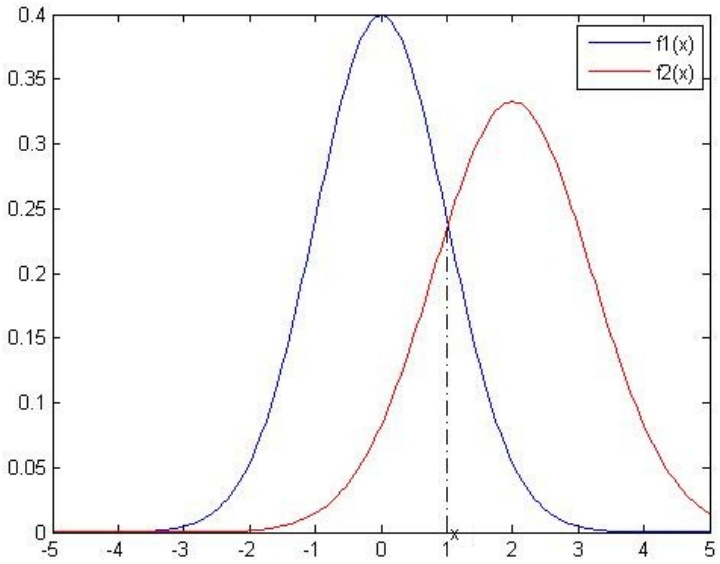

Figure 1.17: Density Functions of $N(0, 1)$ and $N(2, 1.44)$

To determine the difference between two distributions, one can compare the probabilities at each point where the densities are defined (domain). For this I will divide the whole domain of the PDFs into many small intervals $\triangle x_i$. Measure $\breve{W}$ is defined as a sum of the absolute differences between the two probabilities in each of the intervals of the domain. This is possible because of the fact that for Normal distribution domain of the PDF is invariant for μ and σ^2 and is a whole set of real numbers $f_{1,2} : \mathbb{R} \to \mathbb{R}^+ \backslash \{0\}$. It means that the domains are always the same for both density functions. Thus, one can write (1.44)-(1.45):

$$\breve{W} = \sum_{i=-\infty}^{\infty} |P_{1i} - P_{2i}| = \sum_{i=-\infty}^{\infty} |f_1(x_i) \triangle x_i - f_2(x_i) \triangle x_i| \overset{\triangle x_i \geq 0 \ \forall i}{=} \tag{1.44}$$

$$\sum_{i=-\infty}^{\infty} |(f_1(x_i) - f_2(x_i))| \triangle x_i \overset{\triangle x_i \to 0}{\longrightarrow} \int_{-\infty}^{\infty} |f_1(x) - f_2(x)| dx \tag{1.45}$$

Therefore, the aim is the difference in the areas of the PDFs. First, one can estimate

the bounds for the obtained integral (1.45). The lower bound is naturally 0, because the integrand is an absolute value and therefore non-negative. This is the case of two Normal distributions with $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, when the difference inside the modulus is 0. The upper bound is based on the normalization property of the PDF $\int_{-\infty}^{\infty} f(x) = 1$ and the subadditivity property of the absolute value $|a + b| \leq |a| + |b|$. Thus, according to (1.46)-(1.47), $\breve{W}$ cannot exceed 2. However, 2 is in this case unreachable since negative numbers and 0 are out of the range of the PDFs of Normal distribution:

$$\breve{W} = \int_{-\infty}^{\infty} |f_1(x) - f_2(x)|dx < \int_{-\infty}^{\infty} (|f_1(x)| + |-f_2(x)|)dx \overset{f_i(x)\geq 0 \ \forall i}{=} \tag{1.46}$$

$$\int_{-\infty}^{\infty} f_1(x) + \int_{-\infty}^{\infty} f_2(x) = 2 \tag{1.47}$$

Thus, $\breve{W} \in [0, 2)$. Further, one has to eliminate the absolute values from (1.45). This is not a trivial problem and it will be dealt with in the next subsection.

Computing the Function $\breve{W}$

As was claimed, to be able to calculate $\breve{W}$, in the next step one has to get rid of the modulus in (1.45). To do that, one has to find the interception points of the PDFs, where the expression under the absolute value changes its sign. These points can be found from the equation $f_1(x) = f_2(x)$, assuming $\sigma_1 \geq \sigma_2$ for convenience. One can always rename the distributions and their parameters to achieve this inequality. I distinguish two cases that define the number of interceptions. For $\sigma_1 > \sigma_2$ two solutions are possible (1.48):

$$x^{1,2} = \frac{\sigma_1^2\mu_2 - \sigma_2^2\mu_1}{\sigma_1^2 - \sigma_2^2} \pm \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2 - \sigma_2^2}} \sqrt{\frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 - \sigma_2^2} + 2\log\frac{\sigma_1}{\sigma_2}} \tag{1.48}$$

For clarification I denote the roots of the equation so that $x^1 < x^2$. If $\sigma_1 = \sigma_2$, then one proceeds to the second case where only one interception, and therefore only one solution is possible (1.49):

$$\breve{x} = \frac{\mu_1 + \mu_2}{2} \tag{1.49}$$

I do not consider a degenerate situation $(\mu_1, \sigma_1) = (\mu_2, \sigma_2)$ since the distributions are exactly the same and then $\breve{W}$, defined in (1.45), turns into 0. Hence the two cases shown above claim that there are exactly two points of interception for different variances and only one for the same variances.

Before the reestablishment of the derivation of $\breve{W}$, one has to mention that it is not clear in which of the intervals $(-\infty, x^1]$, $(x^1, x^2]$ or (x^2, ∞) for 2 interceptions and $(-\infty, \breve{x}]$ or $(\breve{x}, \infty)$ for 1 interception the integrand is smaller or larger than 0. It it only for sure that it changes its sign. Hence, one has to calculate an absolute value

of $\breve{W}$. I continue the calculation of the area difference, started in (1.45) for the case $\sigma_1 > \sigma_2$, in the following way (1.50)-(1.53):

$$|\breve{W}| = \Big| \int_{-\infty}^{\infty} |f_1(x) - f_2(x)|dx \Big| = \Big| \int_{-\infty}^{x^1} (f_1(x) - f_2(x))dx + \tag{1.50}$$

$$\int_{x^1}^{x^2} (f_2(x) - f_1(x))dx + \int_{x^2}^{\infty} (f_1(x) - f_2(x))dx \Big| = \Big| F_1(x^1) - F_2(x^1) + \tag{1.51}$$

$$(F_2(x^2) - F_2(x^1)) - (F_1(x^2) - F_1(x^1)) + (1 - F_1(x^2)) - \tag{1.52}$$

$$(1 - F_2(x^2)) \Big| = 2 \big| F_1(x^1) - F_1(x^2) - F_2(x^1) + F_2(x^2) \big| \tag{1.53}$$

Where $F_i(x) = F(x; \mu_i \sigma_i^2)$ is a CDF of the Normal distribution $N(\mu_i, \sigma_i^2)$ for $i = 1, 2$. For the case $\sigma_1 = \sigma_2$, according to equation (1.45), one obtains (1.54)-(1.56):

$$|\breve{W}| = \Big| \int_{-\infty}^{\infty} |f_1(x) - f_2(x)|dx \Big| = \Big| \int_{-\infty}^{\breve{x}} (f_1(x) - f_2(x))dx + \tag{1.54}$$

$$\int_{\breve{x}}^{\infty} (f_2(x) - f_1(x))dx \Big| = \Big| F_1(\breve{x}) - F_2(\breve{x}) + (1 - F_2(\breve{x})) - \tag{1.55}$$

$$(1 - F_1(\breve{x})) \Big| = 2 \big| F_1(\breve{x}) - F_2(\breve{x}) \big| \tag{1.56}$$

After completing the calculation of the area differences, one has to compare them to some value, say α, so that if $|\breve{W}|$ is smaller than this value then one can accept that the distributions are the same for some confidence level $1 - \gamma$.

Estimation of the Critical Value α

Since the distribution of $|\breve{W}|$ is unknown, specific α for the given γ is also unknown and it is a natural step to try to forecast the behavior of α using theory. The critical value α should decrease for a larger sample size n, because the sample variance is inversely proportional to the sample size, meaning that larger n induces S^2 to decrease and therefore decreases the dispersion of the points around the mean. At the same time, α should also decrease for growing significance level γ, as far as a larger γ means that one leaves more points out of the confidence set. Now one can estimate the behavior of alpha. To ensure these theoretical guesses I continue the calculations for both cases (1.53) and (1.56) in equations (1.57) and (1.58) respectively:

$$-\frac{\alpha}{2} < F_1(x^1) - F_1(x^2) - F_2(x^1) + F_2(x^2) < \frac{\alpha}{2}, \quad \sigma_1 > \sigma_2 \tag{1.57}$$

$$-\frac{\alpha}{2} < F_1(\breve{x}) - F_2(\breve{x}) < \frac{\alpha}{2}, \quad \sigma_1 = \sigma_2 \tag{1.58}$$

As it is not possible to define the critical value α analytically, I simulated samples with sizes ranging from 5 to 10000 and calculated α's for different significance levels γ. All this was done for number of replications, rp, from 10^4 for large n to 10^5 for

small sample sizes. The results are introduced in Table 1.21. They confirm that the estimated limits for $\breve{W}$ were correct because all the critical values α got in the interval $(2, 0]$. The critical values in Table 1.21 are descending for growing sample size n and significance level γ, as it was predicted above by the theory.
Because Table 1.21 includes only the discrete points for critical value α, it would be beneficial to have the results for any n and γ since it would raise calculation precision. To do this, I used the bilinear interpolation method. Suppose that one wants to get a critical value for n, γ such that they are located between known points: $n1 < n < n2$ and $\gamma_1 < \gamma < \gamma_2$. Then one can use the formula given in statement (1.59):

$$\alpha(n,\gamma) = \frac{1}{(n_2-n_1)(\gamma_2-\gamma_1)} \times (n_2-n, n-n_1)\times \begin{pmatrix} \alpha(n_1,\gamma_1) & \alpha(n_1,\gamma_2) \\ \alpha(n_2,\gamma_1) & \alpha(n_2,\gamma_2) \end{pmatrix} \times \begin{pmatrix} \gamma_2-\gamma \\ \gamma-\gamma_1 \end{pmatrix} \tag{1.59}$$

As a result, the critical values α are defined for any n up to 10000 and for any significance level γ from 1% to 99%. Therefore, in the next step I can define an appropriate CCR confidence set.

Realization of the Univariate CCR Confidence Set

Since the point of interest is the confidence set for a random sample, one redefines $(\mu, \sigma^2) = (\mu_1, \sigma_1^2)$ and $(\overline{x}, S^2) = (\mu_2, \sigma_2^2)$ in (1.48), (1.49) and gains a fully determined $100(1-\gamma)\%$ CCR confidence region in the next form (1.60), (1.61):

$$\mathfrak{C}_{CCR} = \Big\{(\mu,\sigma^2) : \big|F(x^1;\mu,\sigma^2) - F(x^2;\mu,\sigma^2) - F(x^1;m,S^2) + F(x^2;m,S^2)\big| < \alpha(n,\gamma)\Big\}, \quad \sigma^2 \neq S^2 \tag{1.60}$$

$$\mathfrak{C}_{CCR} = \Big\{(\mu,\sigma^2) : \big|F(\breve{x};\mu,\sigma^2) - F(\breve{x};m,S^2)\big| < \alpha(n,\gamma)\Big\}, \quad \sigma^2 = S^2 \tag{1.61}$$

Because of using the absolute value of $\breve{W}$, shown in (1.56), one does not need the condition $\sigma_1 > \sigma_2$ anymore, it could be changed to $\sigma_1 \neq \sigma_2$ without any loss. This was done in (1.60). After the description of the CCR technique, it is reasonable to demonstrate an example of the confidence region obtained with the help of this approach. I simulated a random sample with 100 elements from $N(0, 1)$ and constructed a 90% CCR confidence set. The obtained result is displayed in Figure 1.18.
The introduced set looks like an ellipse, however this can not be specified by the equation. This 'pseudo-ellipse' has a center, marked by the circle in Figure 1.18. It is shifted down, thus demonstrates the non-Euclidean property of variance. This leads to the conclusion: CCR allows for more deviation in the mean for variances that are larger than the sample variance.

Table 1.21: Critical Values α for Calculation of the Univariate CCR Confidence Regions of (μ, σ^2)

n \ γ	0.01	0.025	0.05	0.10	0.15	0.20	0.25	0.30	0.50	0.75	0.90	0.95	0.975	0.99
5	1.2813	1.1406	1.0156	0.8672	0.7715	0.7041	0.6426	0.5898	0.4336	0.2734	0.1650	0.1152	0.0820	0.0508
10	0.8125	0.7266	0.6406	0.5508	0.4941	0.4487	0.4136	0.3831	0.2871	0.1816	0.1104	0.0771	0.0527	0.0352
15	0.6406	0.5625	0.5020	0.4316	0.3887	0.3555	0.3281	0.3047	0.2290	0.1465	0.0884	0.0615	0.0430	0.0273
20	0.5469	0.4766	0.4258	0.3672	0.3315	0.3027	0.2813	0.2617	0.1958	0.1257	0.0762	0.0526	0.0371	0.0234
25	0.4766	0.4219	0.3770	0.3262	0.2930	0.2695	0.2490	0.2319	0.1743	0.1118	0.0679	0.0469	0.0332	0.0215
30	0.4375	0.3828	0.3398	0.2954	0.2666	0.2440	0.2266	0.2109	0.1582	0.1016	0.0620	0.0430	0.0303	0.0195
40	0.3672	0.3281	0.2910	0.2534	0.2288	0.2100	0.1948	0.1807	0.1366	0.0879	0.0532	0.0366	0.0254	0.0166
50	0.3281	0.2891	0.2588	0.2246	0.2041	0.1870	0.1729	0.1612	0.1211	0.0781	0.0474	0.0322	0.0234	0.0146
60	0.2969	0.2617	0.2354	0.2051	0.1855	0.1699	0.1577	0.1469	0.1112	0.0715	0.0430	0.0303	0.0210	0.0137
80	0.2539	0.2266	0.2021	0.1768	0.1598	0.1470	0.1362	0.1267	0.0959	0.0616	0.0371	0.0259	0.0186	0.0117
100	0.2266	0.2031	0.1804	0.1572	0.1426	0.1313	0.1221	0.1133	0.0857	0.0552	0.0332	0.0229	0.0164	0.0107
250	0.1406	0.1270	0.1138	0.0991	0.0898	0.0825	0.0767	0.0713	0.0540	0.0348	0.0210	0.0146	0.0103	0.0068
500	0.0996	0.0889	0.0798	0.0698	0.0632	0.0581	0.0542	0.0504	0.0382	0.0246	0.0146	0.0103	0.0073	0.0044
750	0.0811	0.0726	0.0655	0.0571	0.0518	0.0476	0.0439	0.0410	0.0310	0.0200	0.0121	0.0083	0.0061	0.0039
1000	0.0703	0.0631	0.0566	0.0498	0.0450	0.0410	0.0382	0.0353	0.0268	0.0171	0.0105	0.0073	0.0051	0.0029
10000	0.0215	0.0199	0.0176	0.0157	0.0142	0.0129	0.0122	0.0112	0.0085	0.0054	0.0034	0.0023	0.0017	0.0010

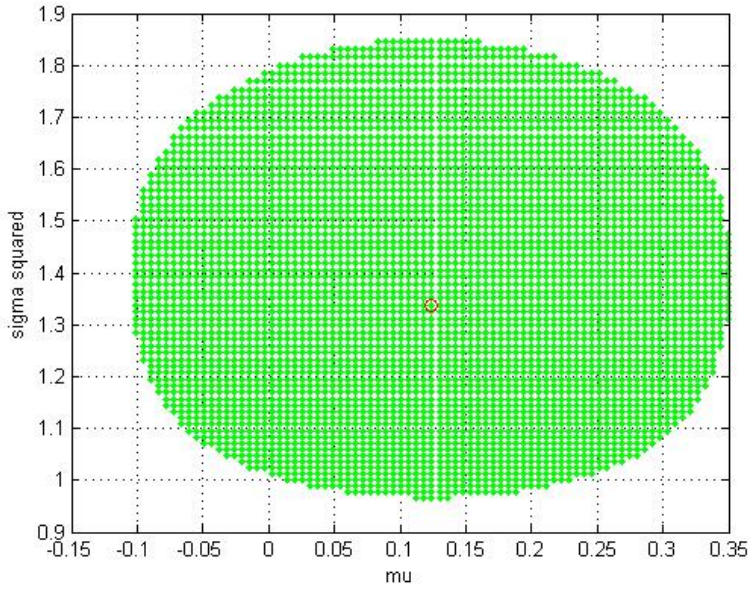

Figure 1.18: Univariate CCR Confidence Region for $N(0,1)$ with the Sample Size $n = 100$ and Significance Level $\gamma = 0.1$

Appendix 1.IV: MatLab Code

Main subroutine for CCR test:

```
function significance  = CCR(mu1, C1, mu2, C2, n)

% function significance.m calculates singificance level alpha for
    testing
% joint hypothesis (mu1,C1) = (mu2,C2) with the sample size n, using
    CCR
% technique

% define accuracy eps

eps = 0.0001;

sign_given = 0.1; % define significance level

% calculate measure W

W = statistic2(mu1,C1,mu2,C2);

% find the exact significance with bisection method

alpha_low = 0.001;
alpha_high = 0.999;
alpha = (alpha_high + alpha_low)/2;
crit_val = regress2(n,alpha);
crit_val_high = regress2(n,alpha_high);
crit_val_low = regress2(n,alpha_low);

if W< crit_val_low

% uncomment to see the testing results in command window
   disp('Hypothesis about the equality of mean and covariance <strong>
       may be accepted</strong>');
   disp(['with the significance p > ', num2str(0.99), '%']);

    significance = 1-alpha_low;
elseif W> crit_val_high

% uncomment to see the testing results in command window
   disp('Hypothesis about the equality of mean and covariance <strong>
       may not be accepted</strong>');
   disp(['significance p <', num2str(0.001), '%']);

    significance = 1-alpha_high;
else
    while abs(alpha_high - alpha_low) > eps
        if W> crit_val
            alpha_low = alpha;
            alpha = (alpha_high + alpha_low)/2;
            crit_val = regress2(n,alpha);
        elseif W< crit_val
            alpha_high = alpha;
```

```
48              alpha = (alpha_high + alpha_low)/2;
49              crit_val = regress2(n,alpha);
50          end
51      end
52
53      significance = 1-alpha;
54
55  % uncomment to see the testing results in command window
56      if significance>sign_given
57          disp('Hypothesis about the equality of mean and covariance <
                strong>may be accepted</strong>');
58          disp(['significance p = ', num2str( significance*100), '%']);
59      else
60          disp('Hypothesis about the equality of mean and covariance <
                strong>may not be accepted</strong>');
61          disp(['significance p = ', num2str( significance*100), '%']);
62      end
63
64  end
65
66  end
```

Subroutine for calculation of the measure $\widetilde{W}$:

```
1   function  W = statistic2(mu1,C1,mu2,C2)
2   % calcultes the difference in areas of two normally distributed random
3   % vectors from R2:
4   % (x1;y1) ~ N([mux1;muy1],[sx1^2, rho1*sx1*sy1; rho1*sx1*sy1, sy1^2])
5   % (x2;y2) ~ N([mux2;muy2],[sx2^2, rho2*sx2*sy2; rho2*sx2*sy2, sy2^2])
6
7
8   %% Step 1 - find transformation coefficients to get rid of correlation
9   % C1 = [sx1^2, rho1*sx1*sy1; rho1*sx1*sy1, sy1^2];
10  % C2 = [sx2^2, rho2*sx2*sy2; rho2*sx2*sy2, sy2^2];
11  % mu1 = [mux1; muy1];
12  % mu2 = [mux2; muy2];
13  % v = Trans2(mu1,C1,mu2,C2);
14
15  % Step 2 - calculate new parameters of the Normal Distributions to
        compare
16  % (no correlation between variables)
17
18  % Check for the correlations first, whether we need a transformation or
        not
19
20  eps = 0.000001;% to compare real numbers
21
22  if (abs(C1(1,2))<eps) && (abs(C2(1,2))<eps)
23
24      mux1 = mu1(1);
25      muy1 = mu1(2);
26      mux2 = mu2(1);
27      muy2 = mu2(2);
28      sx1 = sqrt(C1(1,1));
29      sy1 = sqrt(C1(2,2));
```

```
30      sx2 = sqrt(C2(1,1));
31      sy2 = sqrt(C2(2,2));
32      % disp("no transformation");
33  else
34
35      [nu1,O1,~,~] = Distr_new(mu1,C1,mu2,C2);
36
37      mux1 = nu1(1);
38      muy1 = nu1(2);
39      mux2 = 0;
40      muy2 = 0;
41      sx1 = sqrt(O1(1,1));
42      sy1 = sqrt(O1(2,2));
43      sx2 = 1;
44      sy2 = 1;
45      % disp("transformation");
46  end
47
48
49  m = zeros(16,1);% check run of all methods
50
51  %tic
52  if abs(mux1-mux2)<eps
53      mux = mux1;
54      clear mux1 mux2
55      if abs(sx1-sx2)<eps
56          sx = sx1;
57          clear sx1 sx2
58          if abs(muy1-muy2)<eps
59              muy = muy1;
60              clear muy1 muy2
61              if abs(sy1-sy2)<eps %degenerate case
62                  % disp('degenerate case');
63                  m(16) = 1;
64                  W = 0;
65              else %case 1.1.1
66                  % disp('case 1.1.1');
67                  m(1) = 1;
68                  y1 = muy - sy1*sy2*sqrt(log(sy2^2/(sy1^2))/(sy2^2-sy1
                        ^2));
69                  y2 = muy + sy1*sy2*sqrt(log(sy2^2/(sy1^2))/(sy2^2-sy1
                        ^2));
70                  W = 2*abs(F(y1,muy,sy1) - F(y2,muy,sy1) - F(y1,muy,sy2)
                         + F(y2,muy,sy2));
71              end
72          else
73              if abs(sy1-sy2)<eps %case 1.1.3
74                  % disp('case 1.1.3');
75                  m(3) = 1;
76                  sy = sy1;
77                  clear sy1 sy2
78                  y_star = (muy1 + muy2)/2;
79                  W = 2*abs(F(y_star,muy1,sy) - F(y_star,muy2,sy));
80              else %case 1.2.2
```

```
81                  m(6) = 1;
82                  % disp('case 1.2.2');
83                  y1 = (muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2) - sy1*sy2
                        *...
84                      sqrt(((muy2 - muy1)/(sy2^2-sy1^2))^2 + log(sy2^2/(
                            sy1^2))/(sy2^2 - sy1^2));
85                  y2 = (muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2) + sy1*sy2
                        *...
86                      sqrt(((muy2 - muy1)/(sy2^2-sy1^2))^2 + log(sy2^2/(
                            sy1^2))/(sy2^2 - sy1^2));
87                  W = 2*abs(F(y1,muy1,sy1) - F(y2,muy1,sy1) - F(y1,muy2,
                        sy2) + F(y2,muy2,sy2));
88              end
89          end
90      else
91          if abs(muy1-muy2)<eps
92              muy = muy1;
93              clear muy1 muy2
94              if abs(sy1-sy2)<eps %case 1.1.2
95                  % disp('case 1.1.2');
96                  m(2) = 1;
97                  sy = sy1;
98                  clear sy1 sy2
99                  x1 = mux - sx1*sx2*sqrt(log(sx2^2/(sx1^2))/(sx2^2-sx1
                        ^2));
100                 x2 = mux + sx1*sx2*sqrt(log(sx2^2/(sx1^2))/(sx2^2-sx1
                        ^2));
101                 W = 2*abs(F(x1,mux,sx1) - F(x2,mux,sx1) - F(x1,mux,sx2)
                         + F(x2,mux,sx2));
102             else %case 1.2.1
103                 m(5) = 1;
104                 % disp('case 1.2.1');
105                 W = 2*abs(W1(mux,muy,sx1,sx2,sy1,sy2) - W2(mux,muy,sx1,
                        sx2,sy1,sy2));
106             end
107         else
108             if abs(sy1-sy2)<eps %case 1.2.3
109                 % disp('case 1.2.3');
110                 m(7) = 1;
111                 sy = sy1;
112                 clear sy1 sy2
113                 W = 2*abs(W3(mux,muy1,muy2,sx1,sx2,sy) - W4(mux,muy1,
                        muy2,sx1,sx2,sy));
114             else %case 1.3.1
115                 % disp('case 1.3.1');
116                 m(11) = 1;
117                 W = 2*abs(W9(mux,muy1,muy2,sx1,sx2,sy1,sy2) - W10(mux,
                        muy1,muy2,sx1,sx2,sy1,sy2));
118             end
119         end
120     end
121 else
122     if abs(sx1-sx2)<eps
123         sx = sx1;
```

```
124         clear sx1 sx2
125         if abs(muy1-muy2)<eps
126             muy = muy1;
127             clear muy1 muy2
128             if abs(sy1-sy2)<eps %case 1.1.4
129                 % disp('case 1.1.4');
130                 m(4) = 1;
131                 sy = sy1;
132                 clear sy1 sy2
133                 x_star = (mux1 + mux2)/2;
134                 W = 2*abs(F(x_star,mux1,sx) - F(x_star,mux2,sx));
135             else %case 1.2.4
136                 % disp('case 1.2.4');
137                 m(8) = 1;
138                 W = 2*abs(W5(mux1,mux2,muy,sx,sy1,sy2) - W6(mux1,mux2,
                        muy,sx,sy1,sy2));
139             end
140         else
141             if abs(sy1-sy2)<eps %case 1.2.6
142                 % disp('case 1.2.6');
143                 m(10) = 1;
144                 sy = sy1;
145                 clear sy1 sy2
146                 W = 2*abs(W7(mux1,mux2,muy1,muy2,sx,sy) - W8(mux1,mux2,
                        muy1,muy2,sx,sy));
147             else %case 1.3.3
148                 % disp('case 1.3.3');
149                 m(13) = 1;
150                 W = 2*abs(W13(mux1,mux2,muy1,muy2,sx,sy1,sy2) - W14(
                        mux1,mux2,muy1,muy2,sx,sy1,sy2));
151             end
152         end
153     else
154         if abs(muy1-muy2)<eps
155             muy = muy1;
156             clear muy1 muy2
157             if abs(sy1-sy2)<eps %case 1.2.5
158                 % disp('case 1.2.5');
159                 m(9) = 1;
160                 sy = sy1;
161                 clear sy1 sy2
162                 x1 = (mux1*sx2^2 - mux2*sx1^2)/(sx2^2 - sx1^2) - sx1*
                        sx2*...
163                     sqrt(((mux2 - mux1)/(sx2^2-sx1^2))^2 + log(sx2^2/(
                            sx1^2))/(sx2^2 - sx1^2));
164                 x2 = (mux1*sx2^2 - mux2*sx1^2)/(sx2^2 - sx1^2) + sx1*
                        sx2*...
165                     sqrt(((mux2 - mux1)/(sx2^2-sx1^2))^2 + log(sx2^2/(
                            sx1^2))/(sx2^2 - sx1^2));
166                 W = 2*abs(F(x1,mux1,sx1) - F(x2,mux1,sx1) - F(x1,mux2,
                        sx2) + F(x2,mux2,sx2));
167             else %case 1.3.2
168                 % disp('case 1.3.2');
169                 m(12) = 1;
```

```
                    W = 2*abs(W11(mux1,mux2,muy,sx1,sx2,sy1,sy2) - W12(mux1
                        ,mux2,muy,sx1,sx2,sy1,sy2));
                end
            else
                if abs(sy1-sy2)<eps %case 1.3.4
                    % disp('case 1.3.4');
                    m(14) = 1;
                    sy = sy1;
                    clear sy1 sy2
                    W = 2*abs(W15(mux1,mux2,muy1,muy2,sx1,sx2,sy) - W16(
                        mux1,mux2,muy1,muy2,sx1,sx2,sy));
                else %case 1.4
                    % disp('case 1.4');
                    m(15) = 1;
                    W = 2*abs(W17(mux1,mux2,muy1,muy2,sx1,sx2,sy1,sy2) -
                        W18(mux1,mux2,muy1,muy2,sx1,sx2,sy1,sy2));
                end
            end
        end
    end
%toc
end
```

Subroutine for calculation of critical values, using derived functional form (1.42):

```
function z = regress2(n,alpha)

% Note - alpha is confidence level
% using table of critical values Crit2 this function uses a functional
    form
% of (n,alpa) from constructed regression (R^2=0.9992) to find the
    critical
% value for exact given alpha (confidence level) and sample size n

load('results\Regression_critical_values.mat', 'beta');

z = exp(beta(1) + beta(2)*log(alpha) + beta(3)*log(n) + beta(4)*alpha*
    log(alpha) +...
    beta(5)*alpha + beta(6)*n + beta(7)*n*log(alpha) + beta(8)*alpha*
        log(n) +...
    beta(9)*alpha^2 + beta(10)*(log(alpha))^2 + beta(11)*(log(n))^2
        +...
    beta(12)*exp(alpha) + beta(13)*exp(-alpha));

end
```

Subroutine for calculation of the transformation parameters from system (1.12):

```
function [v,d_mu] = Trans2(~, C1, mu2, C2)

% Calculates the transformation coefficients v and constant d_mu to
    switch
% from the case N(mu1, C1) and N(mu2, C2) to the case of non-corretaled
% Random Vectors with second variable N(0,I)
% Input data as two vectors (2*1) of means mu1, mu2 and two covariance
% matrices (2*2) C1, C2
```

```
% check the uncorr by v*C1*v' and v*C2*v'
%tic

sx1 = sqrt(C1(1,1));
sx2 = sqrt(C2(1,1));
sy1 = sqrt(C1(2,2));
sy2 = sqrt(C2(2,2));
rho1 = C1(1,2)/(sx1*sy1);
rho2 = C2(1,2)/(sx2*sy2);

eps = 0.00001;

if abs(sy1^2*rho2*sx2*sy2 - rho1*sx1*sy1*sy2^2)>eps

    % Discriminant
    D = (sy1.^2.*sx2.^2 - sx1.^2.*sy2.^2).^2 - ...
        4*(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2).*...
        .*(rho1.*sx1.*sy1.*sx2.^2 - rho2.*sx1.^2.*sx2.*sy2);

    % Auxiliary variables (case 1)

    phi1 = 1;
    phi2 = (rho1.*sx1.*sy1.*sx2.^2 - rho2.*sx1.^2.*sx2.*sy2).*...
        ./(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2);
    phi3 = (sx1.^2.*sy2.^2 - sy1.^2.*sx2.^2 - sqrt(D)).*...
        ./(2*(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2));
    phi4 = (sx1.^2.*sy2.^2 - sy1.^2.*sx2.^2 + sqrt(D)).*...
        ./(2*(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2));


    %% Auxiliary variables (case 2)
    %
    % phi1 = 1;
    % phi2 = (rho1.*sx1.*sy1.*sx2.^2 - rho2.*sx1.^2.*sx2.*sy2).*...
    %     ./(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2^2);
    % phi3 = (sx1.^2.*sy2.^2 - sy1.^2.*sx2.^2 + sqrt(D)).*...
    %     ./(2*(rho2.*sy1.^2.*sx2.*sy2-rho1.*sx1.*sy1.*sy2.^2));
    % phi4 = (sx1.^2.*sy2.^2 - sy1.^2.*sx2.^2 - sqrt(D)).*...
    %     ./(2*(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2));

    % Transformation coefficients (case 1)

    b = phi4;
    a = ones(size(b,1),size(b,2));
    c = ones(size(b,1),size(b,2));
    d = phi3./phi1;

    %% Transformation coefficients (case 2)
    %
    % a = phi1./phi2;
    % b = phi1./phi3;
    % c = phi2;
    % d = phi2.*phi3./phi1;

```

```
elseif abs(sy1^2*sx2^2 - sx1^2*sy2^2)<eps

    a = 1;
    b = 1;
    c = 1;
    d =  -(sx1.^2 + rho1.*sx1.*sy1)./(sy1.^2 + rho1.*sx1.*sy1);

else

    sx1 = sx1+eps;
    a = 1;
    b = 1;
    c = 1;
    d =  -(sx1.^2 + rho1.*sx1.*sy1)./(sy1.^2 + rho1.*sx1.*sy1);

end

v = [a,b;c,d];

% additionally transform the solution so that new covariance matrix C2
    will
% be identity matrix and second vector standard normally distributed

C2_1 = v*C2*v';
k1 = sqrt(1/C2_1(1,1));
k2 = sqrt(1/C2_1(2,2));
r = [k1, 0; 0, k2];
v = r*v;

% calculate the constant term to make means of the second Random
    Variable
% zeros

d_mu = v*mu2;

%toc
end
```

Subroutines for different intersections, that define measure $\widetilde{W}$, according to Figure 1.3:

```
function out = W1(mux,muy,sx1,sx2,sy1,sy2)
% calculates first part of the integral for the case 1.2.1

%a2 = 2*sx1^2*sx2^2*log(sx2*sy2/(sx1*sy1))/(sx2^2-sx1^2);

b2 = 2*sy1^2*sy2^2*log(sx2*sy2/(sx1*sy1))/(sy2^2-sy1^2);

out = integral2(@(x,y) exp(-(((x - mux).^2)/(2*sx1^2) + ((y - muy).^2)
    /(2*sy1^2))),...
    mux - 4*sx1,...
    mux + 4*sx1,...
    @(x) max(muy - sqrt(b2 - ((sy1^2*sy2^2*(sx2^2-sx1^2))...
        /(sx1^2*sx2^2*(sy2^2-sy1^2)))*(x - mux).^2).*((b2 - ((sy1^2*sy2
            ^2*(sx2^2-sx1^2))...
```

```
13          /(sx1^2*sx2^2*(sy2^2-sy1^2)))*(x - mux).^2)>0), muy - 4*sy1)
                ,...
14      @(x) min(muy + sqrt(b2 - ((sy1^2*sy2^2*(sx2^2-sx1^2))/...
15          (sx1^2*sx2^2*(sy2^2-sy1^2)))*(x - mux).^2).*((b2 - ((sy1^2*sy2
                ^2*(sx2^2-sx1^2))...
16          /(sx1^2*sx2^2*(sy2^2-sy1^2)))*(x - mux).^2)>0), muy + 4*sy1)
                ,...
17      'Method', 'iterated');
18
19  out = abs(out)/(2*pi*sx1*sy1);
20  %disp(out)
21
22  end
```

```
1   function out = W2(mux,muy,sx1,sx2,sy1,sy2)
2   % calculates second part of the integral for the case 1.2.1
3
4   a2 = 2*sx1^2*sx2^2*log(sx2*sy2/(sx1*sy1))/(sx2^2-sx1^2);
5
6   b2 = 2*sy1^2*sy2^2*log(sx2*sy2/(sx1*sy1))/(sy2^2-sy1^2);
7
8   out = integral2(@(x,y) exp(-(((x - mux).^2)/(2*sx2^2) + ((y - muy).^2)
        /(2*sy2^2))),...
9       mux - 4*sx2,...
10      mux + 4*sx2,...
11      @(x) max(muy - sqrt(b2 - ((sy1^2*sy2^2*(sx2^2-sx1^2))...
12          /(sx1^2*sx2^2*(sy2^2-sy1^2)))*(x-mux).^2).*((b2 - ((sy1^2*sy2
                ^2*(sx2^2-sx1^2))...
13          /(sx1^2*sx2^2*(sy2^2-sy1^2)))*(x - mux).^2)>0), muy - 4*sy2)
                ,...
14      @(x) min(muy + sqrt(b2 - ((sy1^2*sy2^2*(sx2^2-sx1^2))...
15          /(sx1^2*sx2^2*(sy2^2-sy1^2)))*(x-mux).^2).*((b2 - ((sy1^2*sy2
                ^2*(sx2^2-sx1^2))...
16          /(sx1^2*sx2^2*(sy2^2-sy1^2)))*(x - mux).^2)>0), muy + 4*sy2)
                ,...
17      'Method', 'iterated');
18
19  out = abs(out)/(2*pi*sx2*sy2);
20  %disp(out)
21
22  end
```

```
1   function out = W3(mux,muy1,muy2,sx1,sx2,sy)
2   % calculates first part of the integral for the case 1.2.3
3
4   alpha = (sx2^2 - sx1^2)*sy^2/(2*sx1^2*sx2^2*(muy1 - muy2));
5
6   c = (muy1 + muy2)/2 + sy^2*log(sx1/sx2)/(muy1 - muy2);
7
8   out = integral2(@(x,y) exp(-(((x - mux).^2)/(2*sx1^2) + ((y - muy1).^2)
        /(2*sy^2))),...
9       mux - 4*sx1,...
10      mux + 4*sx1,...
11      @(x) max(alpha.*(x - mux).^2 + c, muy1 - 4*sy),...
12      muy1 + 4*sign(alpha)*sy,...
```

```
13      'Method', 'iterated');
14
15  out = abs(out)/(2*pi*sx1*sy); % we take absolute value of out, because
        with
16                                % muy2>muy1 -> alpha<0 -> tails of
                                      parabola go
17                                % down and we have inverse integration
                                      interval
18
19  %disp(out);
20
21  end
```

```
1   function out = W4(mux,muy1,muy2,sx1,sx2,sy)
2   % calculates second part of the integral for the case 1.2.3
3
4   alpha = (sx2^2 - sx1^2)*sy^2/(2*sx1^2*sx2^2*(muy1 - muy2));
5
6   c = (muy1 + muy2)/2 + sy^2*log(sx1/sx2)/(muy1 - muy2);
7
8   out = integral2(@(x,y) exp(-(((x - mux).^2)/(2*sx2^2) + ((y - muy2).^2)
        /(2*sy^2))),...
9       mux - 4*sx2,...
10      mux + 4*sx2,...
11      @(x) max(alpha.*(x - mux).^2 + c, muy2 - 4*sy),...
12      muy2 + 4*sign(alpha)*sy,...
13      'Method', 'iterated');
14
15  out = abs(out)/(2*pi*sx2*sy); % we take absolute value of out, because
        with
16                                % muy2>muy1 -> alpha<0 -> tails of
                                      parabola go
17                                % down and we have inverse integration
                                      interval
18
19  %disp(out);
20
21  end
```

```
1   function out = W5(mux1,mux2,muy,sx,sy1,sy2)
2   % calculates first part of the integral for the case 1.2.4
3
4   beta = (sy2^2 - sy1^2)*sx^2/(2*sy1^2*sy2^2*(mux1 - mux2));
5
6   d = (mux1 + mux2)/2 + sx^2*log(sy1/sy2)/(mux1 - mux2);
7
8   out = integral2(@(x,y) exp(-((x - mux1).^2)/(2*sx^2) - ((y - muy).^2)
        /(2*sy1^2)),...
9       max(d, mux1 - 4*sx),...
10      mux1 + 4*sign(beta)*sx,...
11      @(x) max(muy - sqrt((x-d)/beta), muy - 4*sy1),...
12      @(x) min(muy +sqrt((x-d)/beta), muy + 4*sy1),...
13      'Method', 'iterated');
14
```

```
15 out = abs(out)/(2*pi*sx*sy1); % we take absolute value of out, because
       with
16                                % mux2>mux1 -> beta<0 and we have inverse
17                                % integration interval
18
19 %disp(out);
20
21 end
```

```
1  function out = W6(mux1,mux2,muy,sx,sy1,sy2)
2  % calculates second part of the integral for the case 1.2.4
3
4  beta = (sy2^2 - sy1^2)*sx^2/(2*sy1^2*sy2^2*(mux1 - mux2));
5
6  d = (mux1 + mux2)/2 + sx^2*log(sy1/sy2)/(mux1 - mux2);
7
8  out = integral2(@(x,y) exp(-((x - mux2).^2)/(2*sx^2) - ((y - muy).^2)
       /(2*sy2^2)),...
9      max(d, mux2 - 4*sx),...
10     mux2 + 4*sign(beta)*sx,...
11     @(x) max(muy - sqrt((x-d)/beta), muy - 4*sy2),...
12     @(x) min(muy +sqrt((x-d)/beta), muy + 4*sy2),...
13     'Method', 'iterated');
14
15 out = abs(out)/(2*pi*sx*sy2); % we take absolute value of out, because
       with
16                                % mux2>mux1 -> beta<0, we have inverse
17                                % integration interval
18
19 %disp(out);
20
21 end
```

```
1  function out = W7(mux1,mux2,muy1,muy2,sx,sy)
2  % calculates first part of the integral for the case 1.2.6
3
4  k = -sy^2*(mux2 - mux1)/(sx^2*(muy2 - muy1));
5
6  b = (mux2^2 - mux1^2)*sy^2/(2*(muy2 - muy1)*sx^2) + (muy1 + muy2)/2;
7
8  out = integral2(@(x,y) exp(-(((x - mux1).^2)/(2*sx^2) + ((y - muy1).^2)
       /(2*sy^2))),...
9      mux1 - 4*sx,...
10     mux1 + 4*sx,...
11     muy1 - 4*sy,...
12     @(x) min(k*x + b, muy1 + 4*sy),...
13     'Method', 'iterated');
14
15 % out = integral2(@(x,y) exp(-(((x - mux1).^2)/(2*sx^2) + ((y - muy1)
       .^2)/...
16 %      (2*sy^2))), mux1-5*sx, mux1+5*sx, @(x) max(k*x+b,mux1-5*sx), mux1
       +5*sx);
17
18
19 out = out/(2*pi*sx*sy);
```

```
20
21 %disp(out);
22
23 end
```

```
1  function out = W8(mux1,mux2,muy1,muy2,sx,sy)
2  % calculates second part of the integral for the case 1.2.6
3
4  k = -sy^2*(mux2 - mux1)/(sx^2*(muy2 - muy1));
5
6  b = (mux2^2 - mux1^2)*sy^2/(2*(muy2 - muy1)*sx^2) + (muy1 + muy2)/2;
7
8  out = integral2(@(x,y) exp(-(((x - mux2).^2)/(2*sx^2) + ((y - muy2).^2)
       /(2*sy^2))),...
9      mux2 - 4*sx,...
10     mux2 + 4*sx,...
11     muy2 - 4*sy,...
12     @(x) min(k*x + b, muy2 + 4*sy),...
13     'Method', 'iterated');
14
15 % out = integral2(@(x,y) exp(-(((x - mux2).^2)/(2*sx^2) + ((y - muy2)
       .^2)/...
16 %      (2*sy^2))), mux2-5*sx, mux2+5*sx, @(x) max(k*x + b, muy2-5*sy),
       muy2+5*sy, 'Method', 'iterated');
17
18
19 out = out/(2*pi*sx*sy);
20
21 %disp(out);
22
23 end
```

```
1  function out = W9(mux,muy1,muy2,sx1,sx2,sy1,sy2)
2  % calculates first part of the integral for the case 1.3.1
3
4  a = (sx2^2 - sx1^2)/(sx1^2*sx2^2);
5
6  b = (sy2^2 - sy1^2)/(sy1^2*sy2^2);
7
8  c = ((muy2 - muy1)^2)/(sy2^2 - sy1^2) + log((sx2^2*sy2^2)/(sx1^2*sy1^2)
       );
9
10 out = integral2(@(x,y) exp(-(((x - mux).^2)/(2*sx1^2) + ((y - muy1).^2)
       /(2*sy1^2))),...
11     mux - 4*sx1,...
12     mux + 4*sx1,...
13     @(x) max((muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2) - ...
14         sqrt(c/b - a*((x-mux).^2)/b).*((c/b - a*((x-mux).^2)/b)>0),
               muy1 - 4*sy1),...
15     @(x) min((muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2) + ...
16         sqrt(c/b - a*((x-mux).^2)/b).*((c/b - a*((x-mux).^2)/b)>0),
               muy1 + 4*sy1),...
17     'Method', 'iterated');
18
19 out = out/(2*pi*sx1*sy1);
```

```
20
21 %disp(out);
22
23 end
```

```
1  function out = W10(mux,muy1,muy2,sx1,sx2,sy1,sy2)
2  % calculates second part of the integral for the case 1.3.1
3
4  a = (sx2^2 - sx1^2)/(sx1^2*sx2^2);
5
6  b = (sy2^2 - sy1^2)/(sy1^2*sy2^2);
7
8  c = ((muy2 - muy1)^2)/(sy2^2 - sy1^2) + log((sx2^2*sy2^2)/(sx1^2*sy1^2)
       );
9
10 out = integral2(@(x,y) exp(-(((x - mux).^2)/(2*sx2^2) + ((y - muy2).^2)
       /(2*sy2^2))),...
11     mux - 4*sx2,...
12     mux + 4*sx2,...
13     @(x) max((muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2) - ...
14         sqrt(c/b - a*((x-mux).^2)/b).*((c/b - a*((x-mux).^2)/b)>0),
               muy2 - 4*sy2),...
15     @(x) min((muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2) + ...
16         sqrt(c/b - a*((x-mux).^2)/b).*((c/b - a*((x-mux).^2)/b)>0),
               muy2 + 4*sy2),...
17     'Method', 'iterated');
18
19 out = out/(2*pi*sx2*sy2);
20
21 %disp(out);
22
23 end
```

```
1  function out = W11(mux1,mux2,muy,sx1,sx2,sy1,sy2)
2  % calculates first part of the integral for the case 1.3.2
3
4  a = (sx2^2 - sx1^2)/(sx1^2*sx2^2);
5
6  b = (sy2^2 - sy1^2)/(sy1^2*sy2^2);
7
8  c = ((mux2 - mux1)^2)/(sx2^2 - sx1^2) + log((sx2^2*sy2^2)/(sx1^2*sy1^2)
       );
9
10 out = integral2(@(x,y) exp(-(((x - mux1).^2)/(2*sx1^2) + ((y - muy).^2)
       /(2*sy1^2))),...
11     mux1 - 4*sx1,...
12     mux1 + 4*sx1,...
13     @(x) max(muy - sqrt(c/b - a*((x-(mux1*sx2^2-mux2*sx1^2)/(sx2^2-sx1
           ^2)).^2)/b).*...
14         ((c/b - a*((x-(mux1*sx2^2-mux2*sx1^2)/(sx2^2-sx1^2)).^2)/b)>0),
               muy - 4*sy1),...
15     @(x) min(muy + sqrt(c/b - a*((x-(mux1*sx2^2-mux2*sx1^2)/(sx2^2-sx1
           ^2)).^2)/b).*...
16         ((c/b - a*((x-(mux1*sx2^2-mux2*sx1^2)/(sx2^2-sx1^2)).^2)/b)>0),
               muy + 4*sy1),...
```

```
17      'Method', 'iterated');
18
19  out = out/(2*pi*sx1*sy1);
20
21  %disp(out);
22
23  end
```

```
1   function out = W12(mux1,mux2,muy,sx1,sx2,sy1,sy2)
2   % calculates second part of the integral for the case 1.3.2
3
4   a = (sx2^2 - sx1^2)/(sx1^2*sx2^2);
5
6   b = (sy2^2 - sy1^2)/(sy1^2*sy2^2);
7
8   c = ((mux2 - mux1)^2)/(sx2^2 - sx1^2) + log((sx2^2*sy2^2)/(sx1^2*sy1^2)
        );
9
10  out = integral2(@(x,y) exp(-(((x - mux2).^2)/(2*sx2^2) + ((y - muy).^2)
        /(2*sy2^2))),...
11      mux2 - 4*sx2,...
12      mux2 + 4*sx2,...
13      @(x) max(muy - sqrt(c/b - a*((x-(mux1*sx2^2-mux2*sx1^2)/(sx2^2-sx1
            ^2)).^2)/b).*...
14          ((c/b - a*((x-(mux1*sx2^2-mux2*sx1^2)/(sx2^2-sx1^2)).^2)/b)>0),
                muy - 4*sy2),...
15      @(x) min(muy + sqrt(c/b - a*((x-(mux1*sx2^2-mux2*sx1^2)/(sx2^2-sx1
            ^2)).^2)/b).*...
16          ((c/b - a*((x-(mux1*sx2^2-mux2*sx1^2)/(sx2^2-sx1^2)).^2)/b)>0),
                muy + 4*sy2),...
17      'Method', 'iterated');
18
19  out = out/(2*pi*sx2*sy2);
20
21  %disp(out);
22
23  end
```

```
1   function out = W13(mux1,mux2,muy1,muy2,sx,sy1,sy2)
2   % calculates first part of the integral for the case 1.3.3
3
4   k = (sy2^2 - sy1^2)*sx^2/(2*sy1^2*sy2^2*(mux1 - mux2));
5
6   a = (muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2);
7
8   b = sx^2*((muy2 - muy1)^2)/(2*(sy2^2 - sy1^2)*(mux2 - mux1))+(mux1 +
        mux2)/2 + sx^2*log(sy2/sy1)/(mux2 - mux1);
9
10  out = integral2(@(x,y) exp(-(((x - mux1).^2)/(2*sx^2) + ((y - muy1).^2)
        /(2*sy1^2))),...
11      max(b, mux1 - 4*sx),...
12      mux1 + 4*sign(k)*sx,...
13      @(x) max(a - sqrt((x - b)/k), muy1 - 4*sy1),...
14      @(x) min(a + sqrt((x - b)/k), muy1 + 4*sy1),...
15      'Method', 'iterated');
```

```
16
17 out = abs(out)/(2*pi*sx*sy1); % we take absolute value of out, because
       with
18                                % mux2>mux1 -> beta<0 and we have inverse
19                                % integration interval
20
21 %disp(out);
22
23 end
```

```
1  function out = W14(mux1,mux2,muy1,muy2,sx,sy1,sy2)
2  % calculates second part of the integral for the case 1.3.3
3
4  k = (sy2^2 - sy1^2)*sx^2/(2*sy1^2*sy2^2*(mux1 - mux2));
5
6  a = (muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2);
7
8  b = sx^2*((muy2 - muy1)^2)/(2*(sy2^2 - sy1^2)*(mux2 - mux1))+(mux1 +
       mux2)/2 + sx^2*log(sy2/sy1)/(mux2 - mux1);
9
10 out = integral2(@(x,y) exp(-(((x - mux2).^2)/(2*sx^2) + ((y - muy2).^2)
       /(2*sy2^2))) ,...
11     max(b, mux2 - 4*sx) ,...
12     mux2 + 4*sign(k)*sx ,...
13     @(x) max(a - sqrt((x - b)/k), muy2 - 4*sy2) ,...
14     @(x) min(a + sqrt((x - b)/k), muy2 + 4*sy2) ,...
15     'Method', 'iterated');
16
17 out = abs(out)/(2*pi*sx*sy2); % we take absolute value of out, because
       with
18                                % mux2>mux1 -> beta<0 and we have inverse
19                                % integration interval
20
21 %disp(out);
22
23 end
```

```
1  function out = W15(mux1,mux2,muy1,muy2,sx1,sx2,sy)
2  % calculates first part of the integral for the case 1.3.4
3
4  k = (sx2^2 - sx1^2)*sy^2/(2*sx1^2*sx2^2*(muy1 - muy2));
5
6  a = (mux1*sx2^2 - mux2*sx1^2)/(sx2^2 - sx1^2);
7
8  b = sy^2*((mux2 - mux1)^2)/(2*(sx2^2 - sx1^2)*(muy2 - muy1))+(muy1 +
       muy2)/2 + sy^2*log(sx2/sx1)/(muy2 - muy1);
9
10 out = integral2(@(x,y) exp(-(((x - mux1).^2)/(2*sx1^2) + ((y - muy1)
       .^2)/(2*sy^2))) ,...
11     mux1 - 4*sx1 ,...
12     mux1 + 4*sx1 ,...
13     @(x) min(k*(x - a).^2 + b, muy1 - 4*sy) ,...
14     @(x) min(k*(x - a).^2 + b, muy1 + 4*sy) ,...
15     'Method', 'iterated');
16
```

```
17 out = abs(out)/(2*pi*sx1*sy); % we take absolute value of out, because
       with
18                               % mux2>mux1 -> beta<0 and we have inverse
19                               % integration interval
20
21 %disp(out);
22
23 end
```

```
1  function out = W16(mux1,mux2,muy1,muy2,sx1,sx2,sy)
2  % calculates second part of the integral for the case 1.3.4
3
4  k = (sx2^2 - sx1^2)*sy^2/(2*sx1^2*sx2^2*(muy1 - muy2));
5
6  a = (mux1*sx2^2 - mux2*sx1^2)/(sx2^2 - sx1^2);
7
8  b = sy^2*((mux2 - mux1)^2)/(2*(sx2^2 - sx1^2)*(muy2 - muy1))+(muy1 +
       muy2)/2 + sy^2*log(sx2/sx1)/(muy2 - muy1);
9
10     out = integral2(@(x,y) exp(-(((x - mux2).^2)/(2*sx2^2) + ((y - muy2
           ).^2)/(2*sy^2))),...
11         mux2 - 4*sx2,...
12         mux2 + 4*sx2,...
13         @(x) min(k*(x - a).^2 + b, muy2 - 4*sy),...
14         @(x) min(k*(x - a).^2 + b, muy2 + 4*sy),...
15         'Method', 'iterated');
16
17 out = abs(out)/(2*pi*sx2*sy); % we take absolute value of out, because
       with
18                               % mux2>mux1 -> beta<0 and we have inverse
19                               % integration interval
20
21 %disp(out);
22
23 end
```

```
1  function out = W17(mux1,mux2,muy1,muy2,sx1,sx2,sy1,sy2)
2  % calculates first part of the integral for the case 1.4
3
4  ax = (sx2^2 - sx1^2)/(sx1^2*sx2^2);
5
6  ay = (sy2^2 - sy1^2)/(sy1^2*sy2^2);
7
8  bx = (mux1*sx2^2 - mux2*sx1^2)/(sx2^2 - sx1^2);
9
10 by = (muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2);
11
12 cx = ((mux2 - mux1)^2)/(sx2^2 - sx1^2);
13
14 cy = ((muy2 - muy1)^2)/(sy2^2 - sy1^2);
15
16 d = log((sx2^2*sy2^2)/(sx1^2*sy1^2));
17
18 out = integral2(@(x,y) exp(-(((x - mux1).^2)/(2*sx1^2) + ((y - muy1)
       .^2)/(2*sy1^2))),...
```

```
    mux1 - 4*sx1,...
    mux1 + 4*sx1,...
    @(x) max(by - sqrt((cx + cy + d)/ay - ax*((x - bx).^2)/ay).*...
        (((cx + cy + d)/ay - ax*((x - bx).^2)/ay)>0), muy1 - 4*sy1),...
    @(x) min(by + sqrt((cx + cy + d)/ay - ax*((x - bx).^2)/ay).*...
        (((cx + cy + d)/ay - ax*((x - bx).^2)/ay)>0), muy1 + 4*sy1),...
    'Method', 'iterated');

out = out/(2*pi*sx1*sy1);

%disp(out);

end
```

```
function out = W18(mux1,mux2,muy1,muy2,sx1,sx2,sy1,sy2)
% calculates second part of the integral for the case 1.4

ax = (sx2^2 - sx1^2)/(sx1^2*sx2^2);

ay = (sy2^2 - sy1^2)/(sy1^2*sy2^2);

bx = (mux1*sx2^2 - mux2*sx1^2)/(sx2^2 - sx1^2);

by = (muy1*sy2^2 - muy2*sy1^2)/(sy2^2 - sy1^2);

cx = ((mux2 - mux1)^2)/(sx2^2 - sx1^2);

cy = ((muy2 - muy1)^2)/(sy2^2 - sy1^2);

d = log((sx2^2*sy2^2)/(sx1^2*sy1^2));

out = integral2(@(x,y) exp(-(((x - mux2).^2)/(2*sx2^2) + ((y - muy2)
    .^2)/(2*sy2^2))),...
    mux2 - 4*sx2,...
    mux2 + 4*sx2,...
    @(x) max(by - sqrt((cx + cy + d)/ay - ax*((x - bx).^2)/ay).*...
        (((cx + cy + d)/ay - ax*((x - bx).^2)/ay)>0), muy2 - 4*sy2),...
    @(x) min(by + sqrt((cx + cy + d)/ay - ax*((x - bx).^2)/ay).*...
        (((cx + cy + d)/ay - ax*((x - bx).^2)/ay)>0), muy2 + 4*sy2),...
    'Method', 'iterated');

out = out/(2*pi*sx2*sy2);

%disp(out);

end
```

Subroutine for application of the CCR method in the SUR model:

```
% Example SUR model

clear all
% set up model parameters
G=5; % number of equations
k=3; % number of regressors
```

```
N = 300; % number of observations

% error term u~N(mu,Omega)
Omega = [1   0.5  -0.7  0.24  -0.33;
         0.5  1   -0.2  0.45  0.28;
        -0.7 -0.2  1    0.25  0.17;
         0.24 0.45 0.25 1     0.35;
        -0.33 0.28 0.17 -1.65 1];


% d = rand(G,1); % The diagonal values
% t = triu(bsxfun(@min,d,d.').*rand(G),1); % The upper trianglar random
    values
% Omega = diag(d)+t+t.'; % Put them together in a symmetric matrix
% for i=1:G
%     Omega(i,i) = 1;
% end


mu = zeros(G,1);
u = mvnrnd(mu,Omega,N);

% regressors for SUR
x_hat = rand(G*N,k-1);
x_hat = [ones(G*N,1) x_hat];
X = zeros(G*N,G*k);
for i=1:G
        X(N*(i-1)+1:N*i,k*(i-1)+1:k*i) = x_hat(N*(i-1)+1:N*i,:);
end

% % regressors for panel
% x_hat = rand(N,k-1);
% x_hat = [ones(N,1) x_hat];


% real values of the parameters
beta = round(rand(G*k,1)*10-5);

% vectorize error terms for the model
v = zeros(N*G,1); % error term in vector form
for i=1:G
    for j=1:N
        v(N*(i-1)+j) = u(j,i);
    end
end

% real model
y = X*beta+v;

% estimate using SOLS
b = (X'*X)\X'*y;

% estimate residuals
v_hat = y - X*b; % residuals in vector form
```

```
60
61 % separate residuals for each equation (reverse vectorization)
62 u_hat = zeros(N,G); % residuals in martix form
63 for i=1:G
64     for j=1:N
65         u_hat(j,i) = v_hat(N*(i-1)+j);
66     end
67 end
68
69 Omega_hat = u_hat'*u_hat/(N-G*k)    ;% estimated covariance matrix
70 mu_hat = mean(u_hat)';% estimated mean
71
72 % for i=1:G
73 %     for j=i+1:G
74 %         CCR(mu_hat([i,j]), Omega_hat([i,j],[i,j]), mu([i,j]), Omega([
       i,j],[i,j]), N)
75 %     end
76 % end
77
78 %disp('with zero correlation');
79 l=1;
80 p = zeros(size(Omega));
81 for i=1:G
82     for j=i+1:G
83         disp(['Case ' num2str(l) ': equation ' '<strong>' num2str(i) '
              </strong>' ' with equation ' '<strong>' num2str(j) '</strong
              >']);
84         p(i,j) = CCR(mu_hat([i,j]), Omega_hat([i,j],[i,j]), mu([i,j]),
              [Omega_hat(i,i), 0; 0, Omega_hat(j,j)], N);
85         l=l+1;
86     end
87 end
88
89
90 % disp('with estimated mean');
91 % for i=1:G
92 %     for j=i+1:G
93 %         CCR(mu_hat([i,j]), Omega_hat([i,j],[i,j]), mu_hat([i,j]),
       Omega([i,j],[i,j]), N)
94 %     end
95 % end
96 %
97 % disp('with estimated mean and covariance');
98 % for i=1:G
99 %     for j=i+1:G
100 %         CCR(mu_hat([i,j]), Omega_hat([i,j],[i,j]), mu_hat([i,j]), [
       Omega_hat(i,i), Omega(i,j); Omega(i,j), Omega_hat(j,j)], N)
101 %     end
102 % end
103
104
105 % estimate with GLS and FGLS
106
107 b_GLS = (X'*kron(inv(Omega),speye(N))*X)\(X'*kron(inv(Omega),speye(N))*
```

```
    y);
b_GLS = full(b_GLS);
v_hat_GLS = y - X*b_GLS;

b_FGLS = (X'*kron(inv(Omega_hat),speye(N))*X)\(X'*kron(inv(Omega_hat),
    speye(N))*y);
b_FGLS = full(b_FGLS);
v_hat_FGLS = y - X*b_FGLS;

% Ta to compare the quality of estimation by OLS, GLS and FGLS

VarNames = {'beta', 'b', 'b_GLS', 'b_FGLS', 'difference_OLS', '
    difference_GLS', 'difference_FGLS'};
T = table(beta, b, b_GLS, b_FGLS, abs(beta-b), abs(beta - b_GLS), abs(
    beta - b_FGLS), 'VariableNames', VarNames);
disp(T);

% Sum of  squared differences between real and estimated regressors

diff_OLS = (beta - b)'*(beta - b);
diff_GLS = (beta - b_GLS)'*(beta - b_GLS);
diff_FGLS  = (beta - b_FGLS)'*(beta - b_FGLS);

disp(['total deviation from real beta in OLS is: ' num2str(diff_OLS)]);
disp(['total deviation from real beta in GLS is: ' num2str(diff_GLS)]);
disp(['total deviation from real beta in FGLS is: ' num2str(diff_FGLS)
    ]);

disp(['GLS improved estimation by ' num2str((1 - diff_GLS/diff_OLS)
    *100) '%']);
disp(['FGLS improved estimation by ' num2str((1 - diff_FGLS/diff_OLS)
    *100) '%']);

% Sum of squared residuals and R^2 to compare

RSS_OLS = v_hat'*v_hat; % residual sum of squares
RSS_GLS = v_hat_GLS'*v_hat_GLS;
RSS_FGLS = v_hat_FGLS'*v_hat_FGLS;

ESS_OLS = (X*b - mean(y))'*(X*b - mean(y)); % explained sum of squares
ESS_GLS = (X*b_GLS - mean(y))'*(X*b_GLS - mean(y));
ESS_FGLS = (X*b_FGLS - mean(y))'*(X*b_FGLS - mean(y));

R2_OLS = 1 - RSS_OLS/ESS_OLS; % R squared
R2_GLS = 1 - RSS_GLS/ESS_GLS;
R2_FGLS = 1 - RSS_FGLS/ESS_FGLS;

disp(['Sum of squared residuals for OLS is: ' num2str(RSS_OLS) '; R
    squared is ' num2str(R2_OLS)]);
disp(['Sum of squared residuals for GLS is: ' num2str(RSS_GLS) '; R
    squared is ' num2str(R2_GLS)]);
disp(['Sum of squared residuals for FGLS is: ' num2str(RSS_FGLS) '; R
    squared is ' num2str(R2_FGLS)]);

```

```
%% For GLS I calculate the best measure of the goodness of fit - it is
%% a coefficient of correlation between the actual and the fitted
    values
%% of the dependent variable
%
% y_hat_OLS = X*b;
% y_hat_GLS = X*b_GLS;
% y_hat_FGLS = X*b_FGLS;
%
% GF_OLS = zeros(G,1);
% GF_GLS = zeros(G,1);
% GF_FGLS = zeros(G,1);
%
% for i=1:G
%     GF_OLS(i) = corr(y(N*(i-1)+1:N*i),y_hat_OLS(N*(i-1)+1:N*i));
%     GF_GLS(i) = corr(y(N*(i-1)+1:N*i),y_hat_GLS(N*(i-1)+1:N*i));
%     GF_FGLS(i) = corr(y(N*(i-1)+1:N*i),y_hat_FGLS(N*(i-1)+1:N*i));
% end
%
% disp(['Correlations between the actual and the fitted values for OLS
    are: ' num2str(GF_OLS(1)) ', ' num2str(GF_OLS(2)) ', ' num2str(
    GF_OLS(3))]);
% disp(['Correlations between the actual and the fitted values for GLS
    are: ' num2str(GF_GLS(1)) ', ' num2str(GF_GLS(2)) ', ' num2str(
    GF_GLS(3))]);
% disp(['Correlations between the actual and the fitted values for FGLS
     are: ' num2str(GF_FGLS(1)) ', ' num2str(GF_FGLS(2)) ', ' num2str(
    GF_FGLS(3))]);
```

Part 2 Approach for Testing Serial Correlation in Panel Data Models, Based on a Joint Mean-Variance CCR Test

2.1 Introduction

Popularity of panel data models has been increasing over the decades. Large amounts of data collected and progress in the field of IT have lead to the development of new methods, that can handle such types of data (see e.g. Baltagi (2008)). Some methods stem from the cross-section analysis, such as Pooled Ordinary Least Squares (POLS), and may be implemented in panels by extending the time dimension. Others were expanded from time series methods (see e.g. Baltagi & Li (1995), Okui (2009)) by adding new dimensions that deal not with one, but with many individuals. Different assumptions, required to be fulfilled for the each of the methods, mean that a researcher needs to analyze and sometimes transform the data before the model estimation. For example, strict exogeneity is one of the assumptions that should be fulfilled in order to apply widely used Fixed Effects and Random Effects models.
Serial correlation, i.e. correlation across different time periods, is one the problems that may occur in the data and influence the bias and consistency of estimator. In this Part I will present a new approach for testing the panel data for serial correlation. This approach is based on the cumulative distribution function confidence region (CCR) test from Part 1. It could be classified as a portmanteau test since it tests the total randomness of the data w.r.t. the time periods, but not the interaction between specific error terms.
Alternatives to be presented here to the CCR-based test for serial correlation are also considered. A vast amount of existing tests for serial correlation may be found in the literature. Starting from a portmanteau test by Box & Pierce (1970) with improvement in the paper of Ljung & Box (1978), it was further improved for application in panels by Okui (2009). Simultaneously, another branch of tests, based on the Lagrange Multiplier (LM) statistic, arose. First the LM-test for serial correlation was proposed in the paper of Breusch & Godfrey (1981) and then in the papers of Baltagi & Li (1991) and Baltagi & Li (1995). LM-based methods continue to be under view with more recent improvements by Born & Breitung (2016). More portmanteau tests were presented by Inoue & Solon (2006), Jochmans (2019) and Wooldridge (2002), p.283 with its modification for heteroscedasticity and autocorrelation consistent (HAC) standard errors in Drukker (2003). All the tests have different requirements for panel data. For example, normality of the error terms for one of the specifications of the test presented by Born & Breitung (2016). These alternative tests also possess different asymptotic properties. For example, a test by Breusch & Godfrey (1981) converges to the given limiting distribution only with a number of individuals and a number of time periods growing to infinity.
In this Part I will demonstrate and compare performances and real significance levels

of established tests and a new CCR-based test for serial correlation. The comparison will be based on the panel data with different set ups for the number of time periods, relations between error terms, including autoregressive processes that error terms will follow, and heteroscedasticity issues. This comparison will be performed with the help of Monte-Carlo generated panels and further, as an example, these tests will also be applied to the real-data panel describing health satisfaction in Germany.
The results, obtained from the comparisons of the testing techniques, demonstrated good performance of the CCR-based test on the panels with a small number of time periods T (mention that $T > 3$). This result is possible, because the CCR approach itself does not converge to any specific distribution, but already provides sufficiently good results for such a small sample size. Furthermore, the CCR-based test displayed robustness to the panels with heteroscedasticity across time in the error terms. Robustness to a heteroscedastic panel expands the scope of the CCR-based test. For example, microeconomic data does not usually possess homoscedastisity properties thus, demands special methods; one of which may be the CCR-based test. Finally, the CCR-based test for serial correlation is a good instrument to choose and improve the estimation technique based on the feasible data. It is also profitable for the panels with a small number of time periods available.

2.2 Existing Tests

2.2.1 Model Set Up

All the tests that will be discussed and compared in this Part start by establishing the same model. Consider panel data with an outcome y_{it} and a group of K regressors x_{it} with individuals $i = 1, \dots N$ and time time $t = 1, \dots T$. Next, I assume that there is no dependence across different individuals i, but there most likely is dependence within individuals, i.e. across different time points of each separate individual. The regression model in (2.1) is specified to analyze and estimate this data.

$$y_{it} = x'_{it}\beta + c_i + \varepsilon_{it} \tag{2.1}$$

In the model (2.1) β is a $K \times 1$ vector of unknown coefficients, c_i is a vector of time invariant individual effects. ε_{it} defines idiosyncratic error term, which is a point of interest. On the one hand, ε_{it} is assumed to be independent of individual effects c_i ($\varepsilon_{it} \perp c_i$) and across the individuals ($\varepsilon_{it} \perp \varepsilon_{js} \quad \forall i \neq j; \forall t, s$). On the other hand, error terms may be correlated across time points of one individual. And this introduces the general aim of all the tests for serial correlation: to check whether the idiosyncratic error term is correlated within individuals, i.e. if $Cov(\varepsilon_{it}, \varepsilon_{is}) = 0 \ \forall t, s$. More specifically, null hypotheses will be presented for each of the tests as they may differ.

2.2.2 Wooldridge-Drukker Test

The test, proposed by Wooldridge (2002), p.283, starts from the first difference of residuals $\Delta\hat{\varepsilon}_{it} = \hat{\varepsilon}_{it} - \hat{\varepsilon}_{i,t-1}$, estimated from equation (2.1) and using first difference estimator. In the paper Wooldridge constructs a regression equation (2.2):

$$\Delta\hat{\varepsilon}_{it} = \theta\Delta\hat{\varepsilon}_{i,t-1} + \eta_{it} \tag{2.2}$$

Based on the first difference equation for residuals (2.2), the null hypothesis for the Wooldridge test claims no first order serial correlation of error terms, i.e. $H_0 : Cov(\varepsilon_{it}, \varepsilon_{i,t-1}) = 0 \, \forall t = 2, \dots T$. Under the null hypothesis, ordinary least squares (OLS) estimation of the regression (2.2) gives parameter $\hat{\theta}$, which converges in probability to -0.5. Therefore, test statistic is (2.3):

$$WD = \frac{\hat{\theta} + 0.5}{\hat{s}_\theta} \tag{2.3}$$

$\hat{s}_\theta$ is a standard error of $\hat{\theta}$, estimated from the equation (2.2). To improve the test statistic, Drukker (2003) suggests using heteroscedasticity and robust consistent (HAC) standard errors, presented in the equation (2.4), as first differences of the residuals $\Delta\hat{\varepsilon}_{it}$ are naturally correlated.

$$\hat{s}_\theta = \frac{\sqrt{\sum_{i=1}^{N}\left(\sum_{t=3}^{T}\Delta\hat{\varepsilon}_{i,t-1}\hat{\eta}_{it}\right)^2}}{\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta\hat{\varepsilon}_{i,t-1}} \tag{2.4}$$

Wooldridge-Drukker WD statistic, calculated by plugging equation (2.4) into (2.2), follows t-distribution with $(T-3)$ degrees of freedom. Hence, a restriction $T > 3$ is natural. Additionally, the restriction on T is needed, as soon as equation (2.4) is treated with lagged differences, which leads to a loss of 2 time periods for every individual.

Furthermore, the Wooldridge-Drukker test is robust to data with heteroscedasticity across individuals but not across time points. Secondly, the Wooldridge-Drukker test is not applicable to unbalanced panels (see Born & Breitung (2016)).

2.2.3 Box-Pierce and Consecutive Tests

A portmanteau test for autocorrelations in time series, first presented in Box & Pierce (1970), has spawned a whole range of similar tests. All the tests were based on the idea of summing up the weighted sample autocorrelations for different lags. The next development of the test, by Ljung & Box (1978), provided a statistic, that was closer to asymptotic distribution. However, both of the tests demonstrated quite poor results for small T. On the other hand, both the Box-Pierce and Ljung-Box approaches gave an opportunity to test, not only for the autocorrelation of the first order, but for any order up to some p. For these reasons, Fu, Li, Fung (2002) and

Okui (2009) presented their own modifications of Box-Pierce test, designed for panel data. Additionally, Fu, Li, Fung (2002) proposed another modification for a small number of time periods T. Equation (2.5) presents the statistic obtained by Okui (2009) from the Box-Pierce test. And equation (2.6) presents another modification from the same paper for panel data, but with the Ljung-Box test as a framework. Null hypothesis is no autocorrelation up to order p.

$$Q_{BP} = NT \sum_{k=1}^{p} \hat{\rho}_k^2 \tag{2.5}$$

$$Q_{LB} = NT \sum_{k=1}^{p} \frac{T+2}{T-k} \left(\frac{T-k}{T} \hat{\rho}_k \right)^2 \tag{2.6}$$

Where $\hat{\rho}_k$ is asymptotically unbiased sample correlation between $\hat{\varepsilon}_{it}$ and $\hat{\varepsilon}_{i,t-k}$, calculated from the sample covariance matrix. The term $(T+2)/(T-k)$ in Q_{LB} statistic improves the size of the test, compared to Q_{BP} (see Ljung & Box (1978)). Both of the statistics (2.5) and (2.6) follow χ^2-distribution with p degrees of freedom. But Q_{LB} corrects the bias of Q_{BP} and therefore is usually closer to χ^2-distribution. In addition, there is a natural restriction on number of time periods: $T \geq p+1$ to test the autocorrelations up to order p ($T \geq 2$ for first order serial correlation).

2.2.4 LM-based Tests

Statistic, derived in Breusch & Godfrey (1981), was the first LM-based one to test for the first order serial correlation. Thus, null hypothesis is defined as no first order autocorrelation within individuals in the model (2.1). To compute this statistic, first define for convenience new variables A and B in equations (2.7) and (2.8):

$$A = \frac{\sum_{i=1}^{n} \left(\sum_{t=1}^{T} \tilde{\varepsilon}_{it} \right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\varepsilon}_{it}^2} - 1 \tag{2.7}$$

$$B = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \tilde{\varepsilon}_{it} \tilde{\varepsilon}_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=1}^{N} \tilde{\varepsilon}_{it}^2} \tag{2.8}$$

Where $\tilde{\varepsilon}_{it}$ are the residuals of model (2.1), estimated with Pooled OLS. Then the statistic of Breusch-Godfrey test is computed as (2.9):

$$LM = \frac{NT^2}{T-1} B^2 \tag{2.9}$$

The Breusch-Godfrey statistic LM follows χ^2 distribution with 1 degree of freedom, or similarly, $\sqrt{LM}$ follows standard normal distribution. However, this test has significant limitations. Firstly, error terms should be homoscedastic and normally distributed. Secondly, LM follows χ^2 distribution only for large T, N, and $\frac{T}{N} \to 0$. Born & Breitung (2016) state that otherwise LM statistic posses a Nickel bias of the OLS estimator, described by Nickell (1981). Nickel bias occurs due to demeaning of

panel data. This provokes correlations between regressors and error terms.
To eradicate the bias in LM-statistic (2.9), Baltagi & Li (1995) proposes adjusted test statistic LM^*, presented in equation (2.10):

$$LM^* = \frac{NT^2}{(T-1)(1-\frac{2}{T})}\left(B - \frac{A}{T}\right)^2 \tag{2.10}$$

LM^* statistic follows χ^2 distribution with 1 degree of freedom. Note that $T \geq 3$ to avoid division by zero. Moreover, LM^* in equation (2.10) corrects for the Nickel bias, but still demands quite large N and T. When T is fixed, both of the statistics produce huge errors. To avoid this limitation, Born & Breitung (2016) presented their own statistic (2.11), which improves LM statistic from Breusch & Godfrey (1981), presented in (2.9):

$$\widetilde{LM} = \sqrt{\frac{(T-1)^3}{(T+1)(T-2)^2}}\left(\sqrt{LM} + \sqrt{\frac{N}{T-1}}\right) \tag{2.11}$$

The test statistic $\widetilde{LM}$ follows standard normal distribution. Additionally, it may be used to test for serial correlation for fixed T but the restrictions for homoscedastic and normally distributed error terms still hold. Furthermore, note that $T \geq 3$ to avoid zero in the denominator.
Another development of the LM-based test was achieved by Baltagi (2008), p.97, where he showed that the similar LM statistic, as in (2.9), may be obtained using $\hat{\varepsilon}_{it}$ - residuals from the fixed effects (FE) estimation. Therefore, in Baltagi (2008), p.97 was defined another statistic (2.12):

$$LM_{FE} = \frac{NT^2}{T-1}\,\frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{N}\hat{\varepsilon}_{it}^2} \tag{2.12}$$

Presented statistic LM_{FE} follows χ^2-distribution with 1 degree of freedom. However, restrictions discussed for LM statistic, such as normality and homoscedastisity of error terms, also apply to LM_{FE}, as well as $T \geq 3$.

2.3 CCR Approach in Testing Panel Data Models

In this section I present a new approach that tests for serial correlation across time periods, which is based on the CCR test, presented in Part 1. Starting from the equation (2.1), assume normality of error terms ε_{it}. For convenience, I use first difference estimator with demeaning matrix Q_T, defined in (2.13):

$$Q_T = I_T - 1_T(1_T'1_T)^{-1}1_T' \tag{2.13}$$

With identity matrix $I_T \in \mathbb{R}^{T\times T}$ and vector of ones $1_T \in \mathbb{R}^{T\times 1}$ in (2.13). Note that

matrix Q_T has rank $(T-1)$. Multiplication of equation (2.1) by Q_T will delete all time invariant variables, including individual effects, and transform it into equation (2.14), with new demeaned variables defined in (2.15). x_i, y_i and ε_i denote the data for each individual i, collected through all time points $t = 1, \dots T$.

$$\ddot{y}_i = \ddot{x}_i'\beta + \ddot{\varepsilon}_i, \text{ with:} \tag{2.14}$$

$$\begin{cases} \ddot{y}_i = Q_T * y_i \\ \ddot{x}_i = Q_T * x_i \\ \ddot{\varepsilon}_i = Q_T * \varepsilon_i \end{cases} \tag{2.15}$$

Applying least squares estimator to demeaned model (2.14) produces parameter $\hat{\beta}_{FE}$, estimated with fixed effects. FE estimator is used for further calculation of residuals $\hat{\ddot{\varepsilon}}_i$ in (2.16). In the next step residuals are implemented to find $\hat{\ddot{\Omega}}$ - covariance matrix, defined in (2.17). This covariance matrix captures interactions between time points that are traced for all of the individuals.

$$\hat{\ddot{\varepsilon}}_i = \ddot{y}_i - \ddot{x}_i'\hat{\beta}_{FE} \tag{2.16}$$

$$\hat{\ddot{\Omega}} = \frac{1}{N}\sum_{i=1}^{N} \hat{\ddot{\varepsilon}}_i \hat{\ddot{\varepsilon}}_i' = Q_T * \hat{\Omega} * Q_T \tag{2.17}$$

The second part of equation (2.17) includes matrix $\hat{\Omega}$, which is the covariance matrix of non-demeaned model (2.1). Unless I do not calculate $\hat{\Omega}$, the right-hand side of equation (2.17) clearly demonstrates that matrix $\hat{\ddot{\Omega}}$ is obtained by multiplication of $\hat{\Omega}$ on the both sides with Q_T, which has a rank deficit. Thus, matrix $\hat{\Omega}$ is not restorable from $\hat{\ddot{\Omega}}$ in the general case. As a result, theoretical counterparts of covariance matrices (normal and demeaned) follow the same property and it is not possible to recover Ω from $\ddot{\Omega}$. However, demeaned covariance matrix $\hat{\ddot{\Omega}}$ cannot be used to test for serial correlation, as it includes relations between demeaned parameters (Nickel bias). Therefore, demeaned covariance matrix $\hat{\ddot{\Omega}}$ does not reflect the real covariances between time points. As a consequence, below I propose a 3-step procedure that develops a pseudo-covariance matrix $\check{\Omega}$ from $\hat{\ddot{\Omega}}$. Then, I demonstrate how matrices $\check{\Omega}$ and Ω are related to each other, so that pseudo-covariance matrix may be used to test for serial correlation.

3-Steps procedure to derive a pseudo-covariance matrix $\check{\Omega}$

Step 1 Denote the elements of desired matrix $\check{\Omega}$ as $\check{\sigma}_{ts}$, $\forall t, s = 1, \dots T$, with variances on the diagonal $\check{\sigma}_t^2 = \check{\sigma}_{tt}$. Similarly, $\hat{\ddot{\sigma}}_{ts}$, $\forall t, s = 1, \dots T$ - the elements of estimated with FE covariance matrix $\hat{\ddot{\Omega}}$, and variances on the diagonal are $\hat{\ddot{\sigma}}_t^2 = \hat{\ddot{\sigma}}_{tt}$. Construct the system of equations (2.18):

$$\check{\sigma}_t^2 + \check{\sigma}_s^2 = \hat{\check{\sigma}}_t^2 + \hat{\check{\sigma}}_s^2 - 2\hat{\check{\sigma}}_{ts}, \quad t, s = 1, \dots T \tag{2.18}$$

In (2.18) a system of linear equations is constructed. As soon as covariance matrix is symmetrical, the system (2.18) is invariant to the interchange of t and s. Thus, there are actually $T(T-1)/2$ unique equations in the system (2.18) and T unknowns (all the variances of the matrix $\check{\Omega}$). Therefore, with $T = 1$ or $T = 2$ the system is underidentified and does not have a solution. With $T = 3$ the system (2.18) is exactly identified and the solution always produces a diagonal matrix $\check{\Omega}$. When $T > 3$, the system (2.18) has more equations than unknowns and the solution is derived by minimizing the total distance to all of the equations. Hence, at this point a restriction $T > 3$ has to be established. This fact will be strictly proved further.

Step 2 Before the second step additional notation should be introduced in (2.19):

$$R_t = \sum_{s=1}^{T} \check{\sigma}_{ts} \tag{2.19}$$

According to formula (2.19), R_t is the sum of all the covariances in column t of the required covariance matrix $\check{\Omega}$. Using new variables R_t, I can set up a second system of equations (2.20):

$$\sum_{k=1}^{T} R_k - 2TR_t = T^2(\hat{\check{\sigma}}_t^2 - \check{\sigma}_t^2), \quad t = 1, \dots T \tag{2.20}$$

The system of linear equations (2.20) consists of T unique equations and T unknowns, $R_t, t = 1, \dots T$, and therefore has 1 solution.

Step 3 In the last step I use diagonal elements $\check{\sigma}_t^2$, acquired in step 1 and column sums R_t, obtained in the second step. Next, I calculate the missing non-diagonal elements $\check{\sigma}_{ts}$ from the system (2.21):

$$\check{\sigma}_{ts} = \frac{1}{T^2}\left(T^2\hat{\check{\sigma}}_{ts} + TR_s + TR_t - \sum_{k=1}^{T} R_k\right), \quad \forall t \neq s \tag{2.21}$$

Computed on the first step $\check{\sigma}_t^2$ together with computed in this step $\check{\sigma}_{st}$ form a pseudo-covariance matrix $\check{\Omega}$, presented in (2.22). Obtained matrix $\check{\Omega}$ is diagonal and positive semi-definite.

$$\check{\Omega} = \begin{pmatrix} \check{\sigma}_1^2 & \check{\sigma}_{12} & \dots & \check{\sigma}_{1T} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \check{\sigma}_{1T} & \check{\sigma}_{2T} & \dots & \check{\sigma}_T^2 \end{pmatrix} \tag{2.22}$$

Furthermore, an important property of matrix $\check{\Omega}$ is described in Proposition 1, that is introduced below. This property allows one to use matrix $\check{\Omega}$ to test for serial correlation.

Proposition 1 Pseudo-covariance matrix $\check{\Omega}$ will be diagonal (all the covariances are insignificant), if and only if the covariance matrix $\hat{\Omega}$ is diagonal (its covariances are insignificant).

Proof: The proof is illustrated in two parts: in the first part I assume the diagonal form of $\hat{\Omega}$ and show that pseudo-covariance matrix $\check{\Omega}$ is also diagonal. The second part deals with the matrix $\hat{\Omega}$, including at least one covariance that may not be rejected as insignificant. In this case matrix $\check{\Omega}$ also has significant non-diagonal elements. These two parts describe all possible cases and define the unambiguous relation between $\hat{\Omega}$ and $\check{\Omega}$.

Part 1: diagonal form of covariance matrix. In introduction to the first case, assume that $\hat{\Omega}$ is diagonal at this point and is defined in (2.23). Note that homoscedastisity is not required.

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\sigma}_T^2 \end{pmatrix} \tag{2.23}$$

Then the estimated covariance matrix from the FE model $\hat{\ddot{\Omega}}$ can be calculated, using equation (2.17) with matrix Q_T, defined in (2.13). Obtained matrix $\hat{\ddot{\Omega}}$ is presented elementwise in (2.24):

$$\begin{cases} \hat{\ddot{\sigma}}_t^2 = \frac{1}{T^2}\left(T(T-2)\hat{\sigma}_t^2 + tr(\hat{\Omega})\right) \\ \hat{\ddot{\sigma}}_{ts} = \frac{1}{T^2}\left(-T(\hat{\sigma}_t^2 + \hat{\sigma}_s^2) + tr(\hat{\Omega})\right) \end{cases} \tag{2.24}$$

Where $t, s = 1, \dots T; t \neq s$ and $tr(\hat{\Omega}) = \sum_{t=1}^{T} \hat{\sigma}_t^2$ is the trace of matrix $\hat{\Omega}$. Non-diagonal elements of matrix $\hat{\ddot{\Omega}}$, defined in (2.24), in general are not zeros. This fact, clearly demonstrates that even when there is no serial correlation between time points, covariance matrix of FE within estimation $\hat{\ddot{\Omega}}$ will not be diagonal. The elements of estimated matrix $\hat{\ddot{\Omega}}$, provided in (2.24), are used to run the 3-step procedure, presented above, and define the pseudo-covariance matrix $\check{\Omega}$. In the first step, one plugs the elements of the matrix from (2.24) into the first step system of equations

(2.18). After re-arranging the terms, this system transforms to (2.25):

$$\check{\sigma}_t^2 + \check{\sigma}_s^2 = \hat{\sigma}_t^2 + \hat{\sigma}_s^2,\ t, s = 1, \dots T \tag{2.25}$$

The obvious solution of the system (2.25) is elementwise equal terms from left and right-hand sides (2.26):

$$\check{\sigma}_t^2 = \hat{\sigma}_t^2, \quad t = 1, \dots T \tag{2.26}$$

To verify that (2.26) is the only solution, assume that there exits a second, different solution $\check{\sigma}_t^{2*}$, such that $\exists t : \check{\sigma}_t^{2*} \neq \check{\sigma}_t^2$. For clarity, I assume that the first term of the second solution $\check{\sigma}_1^{2*} = \hat{\sigma}_1^2 + \phi > \hat{\sigma}_1^2$ with $\phi > 0$. Then, from the equation $\check{\sigma}_1^{2*} + \check{\sigma}_2^{2*} = \hat{\sigma}_1^2 + \hat{\sigma}_2^2$ the new solution for the second term is $\check{\sigma}_2^{2*} = \hat{\sigma}_2^2 - \phi < \hat{\sigma}_2^2$. Next, from the equation $\check{\sigma}_2^{2*} + \check{\sigma}_3^{2*} = \hat{\sigma}_2^2 + \hat{\sigma}_3^2$ I conclude that $\check{\sigma}_3^{2*} = \hat{\sigma}_3^2 + \phi > \hat{\sigma}_3^2$. Finally, from the equation $\check{\sigma}_1^{2*} + \check{\sigma}_3^{2*} = \hat{\sigma}_1^2 + \hat{\sigma}_3^2$ I should infer that $\check{\sigma}_1^{2*} = \hat{\sigma}_1^2 - \phi < \hat{\sigma}_1^2$. This makes a contradiction to the assumption about the second solution $\check{\sigma}_1^{2*} > \hat{\sigma}_1^2$. Hence, there exists only one, unique solution to the system of equations (2.25) from step 1, given in statement (2.26).
For the second step, I plug the derived elements of $\hat{\check{\Omega}}$ (2.24) into the equation (2.20) and calculate (2.27)-(2.28):

$$\sum_{k=1}^{T} R_k - 2TR_t = T^2(\hat{\check{\sigma}}_t^2 - \check{\sigma}_t^2) = T(T-2)\hat{\sigma}_t^2 + tr(\hat{\Omega}) - T^2\check{\sigma}_t^2 \tag{2.27}$$

$$\sum_{k=1}^{T} R_k - 2TR_t = -2T\hat{\sigma}_t^2 + tr(\hat{\Omega}), \quad t = 1, \dots T \tag{2.28}$$

Sum up all the equations of the obtained system (2.28). Additionally, remember that $tr(\hat{\Omega}) = \sum_{t=1}^{T} \hat{\sigma}_t^2$. Then the solution for R_t is derived in (2.29)-(2.32):

$$T\sum_{k=1}^{T} R_k - 2T\sum_{t=1}^{T} R_t = -2T\sum_{t=1}^{T} \check{\sigma}_t^2 + Ttr(\hat{\Omega}) \tag{2.29}$$

$$\sum_{k=1}^{T} R_k = tr(\hat{\Omega}), \text{ plug this sum into (2.28):} \tag{2.30}$$

$$tr(\hat{\Omega}) - 2TR_t = -2T\hat{\sigma}_t^2 + tr(\hat{\Omega}) \tag{2.31}$$

$$R_t = \hat{\sigma}_t^2, \quad t = 1, \dots T \tag{2.32}$$

After two steps it may be identified that on the diagonal of matrix $\check{\Omega}$ stand the variances $\hat{\sigma}_t^2$, and the sums of the columns of matrix $\check{\Omega}$ are $\hat{\sigma}_t^2$ as well. However, this does not necessarily mean that all the other non-diagonal elements are zeros. They may have different signs and thus compensate each other. Hence, the third step is needed to investigate whether non-diagonal elements are truly zeros.

In the third step, I plug the already calculated in statement (2.24) variables $\hat{\hat{\sigma}}_{st}$, column sums R_t from (2.32) and the sum of R_t's (2.30) into the system, defined in (2.21). The computations are presented in (2.33)-(2.35):

$$\check{\sigma}_{ts} = \frac{1}{T^2}\left(T^2\hat{\hat{\sigma}}_{ts} + TR_s + TR_t - \sum_{k=1}^{T} R_k\right) \tag{2.33}$$

$$\check{\sigma}_{ts} = \frac{1}{T^2}\left(-T\hat{\sigma}_t^2 - T\hat{\sigma}_s^2 + tr(\hat{\Omega}) + T\hat{\sigma}_s^2 + T\hat{\sigma}_t^2 - tr(\Omega)\right) \tag{2.34}$$

$$\check{\sigma}_{ts} = 0, \quad \forall t \neq s \tag{2.35}$$

The result, obtained in (2.35) proves that the pseudo-covariance matrix $\check{\Omega}$ will be diagonal, if the initial matrix $\hat{\Omega}$ is diagonal. The end of part 1 of the proof.

Part 2: non-diagonal form of covariance matrix. The second part of the proof proceeds in the same way, but now assumes that matrix $\hat{\Omega}$ is not diagonal, i.e. there are covariances that are significantly different from zero. To simplify, I consider the case with one non-zero covariance $\hat{\sigma}_{12} \neq 0$. Then matrix $\hat{\Omega}$ will have the next form (2.36):

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \dots & 0 \\ \hat{\sigma}_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \hat{\sigma}_T^2 \end{pmatrix} \tag{2.36}$$

Using just outlined matrix $\hat{\Omega}$, I can further derive the FE covariance matrix from the equation (2.17). After the re-arrangement of the terms, the FE covariance matrix $\hat{\hat{\Omega}}$ is obtained elementwise in (2.37):

$$\begin{cases} \hat{\hat{\sigma}}_t^2 = \frac{1}{T^2}\left(T(T-2)\hat{\sigma}_t^2 + \kappa_2\hat{\sigma}_{12} + tr(\hat{\Omega})\right) \\ \hat{\hat{\sigma}}_{ts} = \frac{1}{T^2}\left(-T(\hat{\sigma}_t^2 + \hat{\sigma}_s^2) + \kappa_1\hat{\sigma}_{12} + tr(\hat{\Omega})\right) \end{cases} \tag{2.37}$$

Where κ_1 and κ_2 are parameters defined in (2.38) as functions of dummies (2.39):

$$\begin{cases} \kappa_1 = (1-T)(j_1 + j_2 + (1-T)j_1j_2) \\ \kappa_2 = 1 - T(j_1 + j_2) \end{cases} \tag{2.38}$$

$$\begin{cases} j_1 = 1, \text{ if at least one of indexes } (t,s) \text{ is equal to 1} \\ j_2 = 1, \text{ if at least one of indexes } (t,s) \text{ is equal to 2} \end{cases} \tag{2.39}$$

Next, the first step system of equations is calculated in (2.40)-(2.42) by plugging matrix elements from (2.37) into (2.18):

$$\check{\sigma}_t^2 + \check{\sigma}_s^2 = \hat{\check{\sigma}}_t^2 + \hat{\check{\sigma}}_s^2 - 2\hat{\check{\sigma}}_{ts} \tag{2.40}$$

$$\begin{aligned}\check{\sigma}_t^2 + \check{\sigma}_s^2 = \frac{1}{T^2}\Big(&T(T-2)\hat{\sigma}_t^2 + \kappa_2\hat{\sigma}_{12} + tr(\hat{\Omega}) + T(T-2)\hat{\sigma}_s^2 + \kappa_2\hat{\sigma}_{12} + \\ &tr(\hat{\Omega}) + 2T(\hat{\sigma}_t^2 + \hat{\sigma}_s^2) - 2\kappa_1\hat{\sigma}_{12} - 2tr(\hat{\Omega}\Big)\end{aligned} \tag{2.41}$$

$$\check{\sigma}_t^2 + \check{\sigma}_s^2 = \hat{\sigma}_t^2 + \hat{\sigma}_s^2 + \frac{2(\kappa_2 - \kappa_1)}{T^2}\hat{\sigma}_{12}, \quad t, s = 1, \dots T \tag{2.42}$$

The system of linear equations for $\hat{\sigma}_t^2$, displayed in (2.42), differs by the last term from the system, defined for the case of diagonal covariance matrix $\hat{\Omega}$. An investigation of this term showed that it will turn into zero if $\kappa_1 = \kappa_2$.

In the specific case with the 3 time periods, system (2.42) needs additional investigation. On one hand, the equality of kappas is possible only when at least one of the indexes (t, s) equals to 1 or 2. On the other hand, system (2.42) for $T = 3$ describes the equations for pairs of time points (1,2), (1,3) and (2,3). For all 3 equations, these pairs include either index 1 or 2. Thus, for $T = 3$ the system (2.42) drops out the last term and coincides with the system (2.25), which is defined for uncorrelated case. As a result, it is impossible to distinguish pseudo-covariance matrices with and without correlations between elements for $T = 3$. Therefore, it is also impossible to test for serial correlation on the pseudo-covariance matrix $\check{\Omega}$, as it will be always diagonal. This confirms the restriction, set on number of time periods $T > 3$.

The solution of the first step system (2.42) is presented in (2.43):

$$\begin{cases}\check{\sigma}_t^2 = \hat{\sigma}_t^2, \text{ for } t = 1, 2 \\ \check{\sigma}_t^2 = \hat{\sigma}_t^2 + \nu, \text{ for } t = 3, \dots T\end{cases} \tag{2.43}$$

Where a tail ν is defined as (2.44):

$$\nu = \frac{2(\kappa_2 - \kappa_1)}{T^2}(A'A)^{-1}A'1_{T(T-1)/2} \tag{2.44}$$

With A - matrix of all 2-permutations of T without repetitions. In terms of dummies, defined in (2.39), the solution of this system may be rewritten as in (2.45):

$$\check{\sigma}_t^2 = \hat{\sigma}_t^2 + (1 - j_1 - j_2)\nu \tag{2.45}$$

A calculation of the tail ν is complicated and not required for the proof. It stands as an additional non-zero term to show that the solution differs from the one in the uncorrelated case (2.26).

Applying the solution from the first step (2.45), the system of equations for the next step is constructed in (2.46)-(2.48):

$$\sum_{k=1}^{T} R_k - 2TR_t = T^2\hat{\check{\sigma}}_t^2 - T^2\check{\sigma}_t^2 \tag{2.46}$$

$$\sum_{k=1}^{T} R_k - 2TR_t = T(T-2)\hat{\sigma}_t^2 + \kappa_2\hat{\sigma}_{12} + tr(\hat{\Omega}) - T^2\hat{\sigma}_t^2 - T^2(1-j_1-j_2)\nu \tag{2.47}$$

$$\sum_{k=1}^{T} R_k - 2TR_t = -2T\hat{\sigma}_t^2 + \kappa_2\hat{\sigma}_{12} + tr(\hat{\Omega}) - T^2(1-j_1-j_2)\nu \tag{2.48}$$

To proceed as in the uncorrelated case, and sum up all the equations (2.48) for $t = 1, \dots T$, foremost one needs to find an expression for the sum of $(1 - j_1 - j_2)\nu$ (both dummies and tail ν depend on t). This expression is computed in (2.49)-(2.54) by summing up all the $T(T-1)/2$ equations of the system from the first step (2.42). Note that on the left-hand side (LHS) of this system, each variance $\check{\sigma}_t^2$ occurs $(T-1)$ times. Therefore LHS is $(T-1)$ times the sum of all variances. The same logic applied to the RHS gives $(T-1)$ times the sum of all variances, which is the trace of $\hat{\Omega}$.

$$\sum_{t,s}\left(\check{\sigma}_t^2 + \check{\sigma}_s^2\right) = \sum_{t,s}\left(\hat{\sigma}_t^2 + \hat{\sigma}_s^2\right) + \frac{T(T-1)}{2}\frac{2(\kappa_2-\kappa_1)}{T^2}\hat{\sigma}_{12} \tag{2.49}$$

$$(T-1)\sum_{t=1}^{T}\check{\sigma}_t^2 = (T-1)tr(\hat{\Omega}) + \frac{(T-1)(\kappa_2-\kappa_1)}{T}\hat{\sigma}_{12} \tag{2.50}$$

$$\sum_{t=1}^{T}\check{\sigma}_t^2 = tr(\hat{\Omega}) + \frac{(\kappa_2-\kappa_1)}{T}\hat{\sigma}_{12} \tag{2.51}$$

Remember the solution for $\check{\sigma}_t^2$, given in (2.45), and plug it into (2.51):

$$\sum_{t=1}^{T}\left(\hat{\sigma}_t^2 + (1-j_1-j_2)\nu\right) = tr(\hat{\Omega}) + \frac{(\kappa_2-\kappa_1)}{T}\hat{\sigma}_{12} \tag{2.52}$$

$$tr(\hat{\Omega}) + \sum_{t=1}^{T}(1-j_1-j_2)\nu = tr(\hat{\Omega}) + \frac{(\kappa_2-\kappa_1)}{T}\hat{\sigma}_{12} \tag{2.53}$$

$$\sum_{t=1}^{T}(1-j_1-j_2)\nu = \frac{(\kappa_2-\kappa_1)}{T}\hat{\sigma}_{12} \tag{2.54}$$

Additionally, compute the sum of $\kappa_2\hat{\sigma}_{12}$ in (2.55) using the definition of κ_2 from (2.38):

$$\sum_{t=1}^{T}\kappa_2\hat{\sigma}_{12} = \sum_{t=1}^{T}(1-T(j_1+j_2))\hat{\sigma}_{12} = (T-2T)\hat{\sigma}_{12} = -T\hat{\sigma}_{12} \tag{2.55}$$

After obtaining the expressions in (2.54) and (2.55), I use them to sum up all T second step equations (2.48) and receive the sum of all column sums R_t in (2.56):

$$T\sum_{k=1}^{T} R_k - 2T\sum_{t=1}^{T} R_t = Ttr(\hat{\Omega}) + \sum_{t=1}^{T}\left(-2T\hat{\sigma}_t^2 + \kappa_2\hat{\sigma}_{12} - T^2(1-j_1-j_2)\nu\right) \quad (2.56)$$

$$-T\sum_{t=1}^{T} R_t = Ttr(\hat{\Omega}) - 2Ttr(\hat{\Omega}) - T\hat{\sigma}_{12} - T^2\frac{(\kappa_2-\kappa_1)}{T}\hat{\sigma}_{12} \quad (2.57)$$

$$\sum_{t=1}^{T} R_t = tr(\hat{\Omega}) + (1-\kappa_1+\kappa_2)\hat{\sigma}_{12} \quad (2.58)$$

Plug the sum of R_t from equation (2.58) into each of the equations of the system (2.48) to obtain the solution for every R_t (2.59):

$$R_t = \hat{\sigma}_t^2 + \frac{1-\kappa_1}{2T}\hat{\sigma}_{12} + \frac{T(1-j_1-j_2)}{2}\nu \quad (2.59)$$

Equation (2.59) represents the solution of the second step system of equations. In the final step, I use already computed variables $\hat{\check{\sigma}}_{ts}$, R_t, as well as the sum of R_t to deal with the third step system (2.21) and figure out non-diagonal elements of the matrix $\check{\Omega}$. The solution is calculated in (2.60)-(2.62):

$$\check{\sigma}_{ts} = \frac{1}{T^2}\left(T^2\hat{\check{\sigma}}_{ts} + TR_s + TR_t - \sum_{k=1}^{T} R_k\right) \quad (2.60)$$

$$\begin{aligned}\check{\sigma}_{ts} = \frac{1}{T^2}\Bigg(&- T(\hat{\sigma}_t^2 + \hat{\sigma}_s^2) + \kappa_1\hat{\sigma}_{12} + tr(\hat{\Omega}) + T\hat{\sigma}_t^2 + \frac{1-\kappa_1}{2}\hat{\sigma}_{12} + \\ &\frac{T^2(1-j_1-j_2)}{2}\nu + T\hat{\sigma}_s^2 + \frac{1-\kappa_1}{2}\hat{\sigma}_{12} + \frac{T^2(1-j_1-j_2)}{2}\nu - \\ &tr(\Omega) - (1-\kappa_1+\kappa_2)\hat{\sigma}_{12}\Bigg)\end{aligned} \quad (2.61)$$

$$\check{\sigma}_{ts} = \frac{\kappa_1-\kappa_2}{T^2}\hat{\sigma}_{12} + (1-j_1-j_2)\nu \quad (2.62)$$

Now the elements of pseudo-covariance matrix $\check{\Omega}$ are defined in equations (2.45) and (2.62). Non-diagonal elements $\check{\sigma}_{ts}$ turn into zero only when one of the indexes (t, s) is 1 or 2. All the other elements of pseudo-covariance matrix are significantly different from zero, as $\check{\sigma}_{12}$ is significant by assumption. Thus, in this case the matrix $\check{\Omega}$ is not diagonal.

Finally, it was demonstrated that the diagonal matrix $\hat{\Omega}$ always generates the diagonal pseudo-matrix $\check{\Omega}$. And vice versa, if $\hat{\Omega}$ does not possess diagonality properties, then the pseudo-covariance matrix $\check{\Omega}$ is also non-diagonal.■

Transformation and testing on the pseudo-covariance matrix is possible even for small number of time periods T (starting from $T = 4$). For T going to infinity, the effect of the first term in RHS of (2.62) does not negotiate due to its denominator T^2, because the highest order of κ_1 is also T^2. The second term of RHS does not converge to zero as well, and the pseudo-covariance matrix preserves its significance of non-diagonal

elements, even with a large number of time periods.
Proof of proposition 1, provided above, allows the application of the CCR technique from Part 1 to test for serial correlation, using the pseudo-covariance matrix $\check{\Omega}$, computed with the 3-step procedure.
Testing for serial correlation, based on the CCR approach, proceeds as next. First of all, assume that error terms $\varepsilon_i \in \mathbb{R}^{T\times 1}$ from model (2.1) are independent across individuals and that they are normally distributed. Secondly, the CCR-based test is performed pairwisely across time points. Hence, I divide a panel into $T(T-1)/2$ groups of 2 cross-sections for 2 different time points. Denoting cross-sections of error terms as ε_t, I get groups in the form $(\varepsilon_t', \varepsilon_s')'$, for $t, s = 1, \dots T$, $t \neq s$. Each group of 2 cross-sections is a bivariate sample for the CCR test. It therefore means that each separate observation in a given bivariate sample has 2 error terms for a specific individual i from 2 fixed time points t and s (same for each individual). Then, after the calculation of the pseudo-covariance matrix $\check{\Omega}$, I use this matrix to test each of $T(T-1)/2$ groups of bivariate samples of error terms to follow uncorrelated joint normal distribution (2.63):

$$H_0 : \begin{pmatrix} \varepsilon_t \\ \varepsilon_s \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \check{\sigma}_t^2 & 0 \\ 0 & \check{\sigma}_s^2 \end{pmatrix} \right) \tag{2.63}$$

Note that in (2.63) I assume zero mean of error terms, which is their natural property. The CCR test calculates the p-values for each of the $T(T-1)/2$ tests, defined in (2.63).
An overall null hypothesis for the whole panel implies no serial correlation (up to order $T-1$) in the panel data. This null hypothesis may not be rejected, if all of the p-values from the pairwise testing (2.63) are larger than the predefined significance level γ: $min(\text{p-value}) > \gamma$. Otherwise, an overall null hypothesis is rejected, if p-value for at least one pair of error terms is smaller than the predefined significance level: $\exists$ p-value: p-value $< \gamma$.
As a result, in this section I derived and described the procedure for testing for serial correlation in panel data, based on the CCR test. Its main advantage, as stated, is the potential for its application in panel data models with a small number of time periods and on the heteroscedastic panels (since no restrictions are set on the variance). In the next section, I run and compare the properties of all of the serial correlation testing methods, represented in this Part, together with the CCR-based technique.

2.4 Monte-Carlo Simulations

In this section I will present the performance of the tests, discussed in sections 2.2 and 2.3. The tests for serial correlation will be analyzed for their finite sample properties, based on the panels with different parameters, simulated with the help of Monte-Carlo method. I will start with the model (2.1), which is used to generate panel data. The number of individuals is fixed to $N = 500$ and number of regressors to $K = 10$. The number of time periods T varies and for every case I consider panels to have 5, 10, 20, 30 or 50 time periods. Regression parameters β are fixed integers from 1 to 10.

Regressors x_{it} are taken from standard normal distribution. Individual effects c_i are taken from standard normal distribution as well and are independent across N. But individual effects are to be correlated with regressors. Therefore, I add a part of the individual effect c_i to each of the element of regressor vector x_i to have a clear correlation between them.

Table 2.1 displays the ratio of rejected panels, based on 10000 Monte-Carlo replications for each case. Altogether, it cover 8 statistics, described above: Wooldridge-Drukker WD (2.3), Ljung-Box Q_{LB} (2.6), Box-Pierce Q_{BP} (2.5), Breusch-Godfrey LM (2.9), Baltagi-Li LM^* (2.10), Born-Breitung $\widetilde{LM}$ (2.11), Baltagi Fixed Effects LM_{FE} (2.12), as well as the CCR-based one. Each method is applied in various situations. No serial correlation in the data is assumed in case 1. Cases 2 and 3 describe the behavior of the methods when the first order serial correlation is present in the data. In case 4, the methods are used in the data with correlation included only between first and second time points, while case 5 assumes a relatively small first order correlation and larger second order serial correlation. Cases 6 and 7 cover the situation when the panel contains heteroscedastic error terms across time without (case 6) or with serial correlation (case 7).

The first case I consider, illustrates a control for real significance levels. The theoretical significance level is set to $\gamma = 0.05$. As this case assumes no correlation in the data, each method should reject 5% of the simulated samples. The first part of the Table 2.1 shows that the statistics, mentioned to give poor results for small T (Q_{LB}, Q_{BP}, LM and LM^*), alter a lot from theoretical significance level $\gamma = 0.05$ for $T = 5$ and $T = 10$. Additionally, WD statistic also overestimates the significance level for $T = 5$. With increasing number of time periods, all the methods get close to the $\gamma = 0.05$.

Cases 2 and 3, in Table 2.1, present a situation where error terms follow an autoregressive process of order 1. A correlation between neighboring time periods equal to 0.05 in case 2 and 0.2 in case 3. Thus, a perfect test should reject all 100% of the simulated panels. With a smaller correlation in case 2, the same 4 methods (Q_{LB}, Q_{BP}, LM and LM^*), as in case 1, reject only a small fraction of the samples (around 22% for Q_{BP}, LM and LM^* and $52,2\%$ for Q_{LB}). With a stronger serial correlation $= 0.2$ in the case 3, all 8 methods reject almost 100% of the samples for T starting from 10. With 5 time periods in the case 3, only the LM test does not produce acceptable results thus rejecting only 7.2% of the samples.

Case 4 of Table 2.1 displays the reaction of the methods on only 1 correlation between first and second time periods. The tests should reject all the panels in this case. It appeared to be a problem for Q_{LB}, Q_{BP} and LM. However, the CCR, Wooldridge-Drukker and the adjusted LM-methods perform well and reject almost all of the samples.

In the fifth case, error terms follow the autoregressive process of order 2, as each error term correlates with error terms from the two previous periods. Both of the serial correlations are relatively small (0.01 for the first order and 0.05 for the second order). Rejection of the simulated panels is more complicated in this case, as most of the tests only check the existence of the first order serial correlation. Q_{LB}, Q_{BP}

Table 2.1: Performance of 8 Methods for Testing for Serial Correlation in Panels Subject to Predefined Correlations and Heteroscedasticity (Cases 6 and 7)

T	WD	Q_{LB}	Q_{BP}	LM	LM^*	$\widetilde{LM}$	LM_{FE}	CCR
1. No correlation across time								
5	0.075	0.159	0.011	0.016	0.013	0.055	0.055	0.052
10	0.052	0.095	0.077	0.028	0.021	0.051	0.052	0.054
20	0.046	0.047	0.045	0.049	0.049	0.05	0.052	0.048
30	0.048	0.047	0.044	0.043	0.045	0.045	0.044	0.042
50	0.051	0.053	0.055	0.052	0.054	0.053	0.054	0.052
2. First order serial correlation = 0.05								
5	0.996	0.522	0.222	0.21	0.228	1	0.999	0.912
10	0.999	0.744	0.691	0.706	0.725	1	1	0.956
20	0.999	0.974	0.957	0.869	0.887	1	1	0.995
30	1	0.999	0.982	0.958	0.963	1	1	0.999
50	1	1	0.996	0.997	0.999	1	1	0.999
3. First order serial correlation = 0.2								
5	1	0.999	0.978	0.072	0.998	0.992	0.886	0.914
10	1	0.999	0.989	0.552	0.999	1	0.991	0.999
20	1	1	0.999	0.999	0.999	1	0.999	1
30	1	1	1	0.999	1	1	1	1
50	1	1	1	1	1	1	1	1
4. One non-zero covariance $\sigma_{12} = 0.5$								
5	0.999	0.126	0.017	0.339	1	1	0.895	0.991
10	0.999	0.183	0.041	0.781	1	1	0.997	0.999
20	1	0.058	0.059	0.992	1	1	0.999	1
30	1	0.095	0.04	0.999	1	1	1	1
50	1	0.092	0.076	1	1	1	1	1
5. Serial correlation of the first order = 0.01, of the second order = 0.05								
5	1	0.679	0.211	0.081	1	1	0.999	0.234
10	1	0.967	0.785	0.343	1	1	0.999	0.849
20	1	0.999	0.974	0.981	1	1	1	0.998
30	1	1	0.999	0.998	1	1	1	0.999
50	1	1	1	0.999	1	1	1	1
6. Heteroscedasticity and no correlation across time								
5	0.991	0.142	0.019	0.081	1	1	1	0.047
10	0.868	1	0.996	0.07	1	1	0.992	0.067
20	0.55	1	1	0.134	0.998	1	0.901	0.069
30	0.45	1	1	0.153	0.95	1	0.72	0.059
50	0.299	1	1	0.189	0.468	1	0.512	0.062
7. Heteroscedasticity across time and first order serial correlation = 0.2								
5	1	0.817	0.54	0.073	1	1	0.986	0.841
10	1	1	1	0.083	1	1	0.069	0.916
20	1	1	1	0.015	1	1	0.996	1
30	1	1	0.459	0.006	1	1	1	1
50	1	0.696	0.72	0.005	1	1	1	1

Considered tests: Wooldridge-Drukker WD, Ljung-Box Q_{LB}, Box-Pierce Q_{BP}, Breusch-Godfrey LM, Baltagi-Li LM^*, Born-Breitung $\widetilde{LM}$, Baltagi Fixed Effects LM_{FE}, and CCR-based approach CCR

and LM struggle with this modification for small T. The CCR approach produces a big deviation when T is small, but starting from $T = 20$ it rejects almost all of the samples. This could be explained by small autocorrelation coefficients (0.01 and 0.05), that are not rejected individually, but combined together they are rejected for the samples with T larger than 20. A better performance shows WD and all 3 adjusted LM statistics.

For the last 2 cases, heteroscedasticity across time points is assumed. It results in a lot of deviation in the methods. For example, case 6 also assumes no correlation and ideally the rejection rate of the tests should be close to the theoretical significance level $\gamma = 0.05$. Only the CCR-based test demonstrates this and rejects around 5% of the samples. In case 7, when additional serial correlation is included, the compared tests should reject null hypothesis. CCR rejects almost 100% of the samples. Other methods, except of Box-Pierce Q_{BP} and Breusch-Godfrey LM, possesses similar rejection rates close to 100%. Still all the other methods, except CCR, cannot be used with heteroscedastic error terms, as they failed to test for serial correlation correctly in case 6.

In summary, during the simulation studies the CCR-based test produced adequate results in the situation of heteroscedastic error terms across time. However, the CCR-based test demonstrated worse rejection rates for autoregressive process of order 2. This may be explained by relatively small serial correlations that CCR-based test did not capture. Furthermore, the CCR-based test performs well in the case of only 1 correlation. This proves the fact, discussed above, that with increasing T significance of non-diagonal elements does not vanish (2.62). Additionally, the adjusted LM tests (LM^*, $\widetilde{LM}$ and LM_{FE}) together with Wooldridge-Drukker test performed well in situations with homoscedastic error terms (cases 1-5 of Table 2.1).

2.5 Example

In this section I will apply all 8 methods to the German Health Care Usage Data, taken from the Journal of Applied Econometrics Archive[2]. This is an unbalanced panel with 7293 individuals. To apply all the serial correlation tests, firstly I clean up the data. After acquiring a balanced panel, I observed 887 individuals in 7 time points (years 1984, 1985, 1986, 1987, 1988, 1991 and 1994), giving altogether 6209 observations. Based on this data I constructed the model (2.64):

$$\begin{aligned} newhsat_{it} = \beta_0 + \beta_1 income_{it} + \beta_2 doctor_{it} + \beta_3 female_i + \beta_4 handdum_{it} + \\ \beta_5 educ_{it} + \beta_6 married_{it} + c_i + \varepsilon_{it} \end{aligned} \tag{2.64}$$

With the variables

- $newhsat_{it}$ - recorded value of health satisfaction from 0 (low) to 10 (high)
- $income_{it}$ - log of hourly salary

[2] http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/

- $doctor_{it}$ - dummy variable that = 1, if there were doctor visits in last three months, otherwise = 0
- $female_i$ - dummy, that = 1, if a person is female, otherwise = 0
- $handdum_{it}$ - dummy, that = 1, if the person is handicapped, otherwise = 0
- $educ_{it}$ - years of schooling
- $married_{it}$ - dummy, that = 1, if the person is married, otherwise = 0
- c_i - unobserved individual effect
- ε_{it} - idiosyncratic error term

All the methods were applied to the data to test for serial correlation. Obtained p-values are presented in the Table 2.2.

Table 2.2: P-Values from Testing for Serial Correlation with 8 Methods in the German Health Care Usage Data

WD	Q_{LB}	Q_{BP}	LM	LM^*	$\widetilde{LM}$	LM_{FE}	CCR
4.4409e-16	2.0743e-04	7.9755e-04	0.3316	0	0	8.0486e-07	1.0159e-06

Considered tests: Wooldridge-Drukker WD, Ljung-Box Q_{LB}, Box-Pierce Q_{BP}, Breusch-Godfrey LM, Baltagi-Li LM^*, Born-Breitung $\widetilde{LM}$, Baltagi Fixed Effects LM_{FE}, and CCR-based approach CCR

According to Table 2.2, the only outlier is the Breusch-Godfrey LM-test, with p-value 0.3316, which did not reject the null hypothesis about absence of serial correlation. A reason for that may be not sufficient number of time periods T in the data for implementation of this specific test. All the other approaches produced much smaller p-values thus rejected the null hypothesis about the absence of serial correlation at a 1% significance level.

Furthermore, I tested all the possible combinations of the time points for the existence of serial correlation ($\binom{7}{6} = 7$ cases for $T = 6$, $\binom{7}{5} = 21$ cases for $T = 5$ and $\binom{7}{4} = 35$ cases for $T = 4$). This allowed me to discover all the possible relations between time points. The search among panels with 1 time period thrown away indicated serial correlation in each of the 7 possible cases. But the search among panels reduced to 5 time points allowed to pick out such a sub-panel that has only insignificant correlations across time. As a result, the largest reduced panel with uncorrelated error terms consists of the years 1984, 1985, 1986, 1987 and 1988. The rest of the panel (years 1991 and 1994) may be estimated separately. Table 2.3 compares the FE estimators for the full and reduced panels. T-statistics are presented in the brackets. Estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_4$ in Table 2.3 are significant at a 1% level for both of the models. The effect of education $\hat{\beta}_5$ on the health satisfaction is insignificant in both models, while the effect of being married on health satisfaction $\hat{\beta}_6$ becomes insignificant in the reduced model. Note that $\hat{\beta}_3$ is not estimated, as $female_i$ is time-invariant.

One can notice a large difference between estimators for the full and reduced panel in Table 2.3. For example, the effect of log of income on health satisfaction ($\hat{\beta}_1$)

Table 2.3: Parameters Obtained from the FE Estimation for the Full and Reduced Panels with 7 and 5 Years Respectively. T-Statistic is Given in Brackets

Estimated parameter	Full panel	Reduced panel
$\hat{\beta}_1$	-0.40068 (-2.8917)	0.20514 (2.7026)
$\hat{\beta}_2$	-0.65141 (-12.0515)	-0.5033 (-7.7493)
$\hat{\beta}_4$	-0.15367 (-2.8403)	-0.14804 (-2.5882)
$\hat{\beta}_5$	-0.13447 (-1.0142)	-0.069125 (-0.3409)
$\hat{\beta}_6$	0.32963 (2.7144)	0.14003 (0.7667)

in the full model is negative. In the reduced model this effect becomes positive, demonstrating that the full model possibly estimated the effect of log of income with the wrong sign. According to the reduced model, a 1% increase in the salary improves the health satisfaction by approximately 0.002 c.p. This is an expected result, as higher income for most of the individuals allows to improve quality of food, health care and consumption in total, which may positively affect the health satisfaction. Furthermore, the application of the FE estimator in the reduced panel decreased the estimated parameters $\hat{\beta}_5$ and $\hat{\beta}_6$ approximately by half (in absolute values). However, both of the parameters are not significant for the reduced model, according to the t-statics. Table 2.3 implies that the negative insignificant effect of education on health satisfaction, displayed in $\hat{\beta}_5$, was too strong and the reduced model has even weakened it. On the contrary, according to the Table 2.3, the effect of being married ($\hat{\beta}_6$) on the health satisfaction was positive and significant for the full panel. But it appeared that the model with all 7 time periods has overestimated the effect of being married, while reduced model has estimated it smaller by half and insignificant.
Finally, the study of the CCR-based method in comparison with the other approaches on the real data example from German Health Care gave an opportunity to improve the estimation (1 parameter decreased by half and become insignificant, another parameter has changed its sign). Because of serial correlation tests, the improvement was possible by only using the FE estimator, without implementation of more complicated estimation techniques.

2.6 Conclusion

In this Part a new approach for testing panels for serial correlation was introduced. It is a portmanteau test, based on the CCR technique for jointly testing means and covariances of normally distributed samples. Firstly, a 3-step transformation procedure was introduced to obtain a pseudo-covariance matrix from the FE covariance matrix. In Proposition 1 it was proved, that pseudo-covariance matrix will be diagonal, if and only if the covariance matrix (usually unknown) of the model is diagonal. However,

the pseudo-covariance matrix does not reflect the structure of the covariance matrix, therefore cannot provide information about the correlation between two specific time points. But it may state whether there is any correlation across time present in the data. Next, the CCR-based test is performed pairwisely on all possible pairs from the pseudo-covariance matrix. The final p-value of the test is defined as a minimum among the p-values from the pairwise comparison. Further, the CCR-based test appeared to perform well on the panels with small T. In addition, note the restriction $T > 3$, proven in the Proposition 1. Besides, this test is well-fitted for the panels with heteroscedasticity across time, as no restriction on the variances in different time periods is set.

Moreover, I compared the CCR-based method with the other existing tests for serial correlation, that were also described in this Part. Comparison proceeds with the help of Monte-Carlo simulations of the panels with autoregressive error terms, with correlation between specific time points and on the panels with heteroscedastic error terms. Together with CCR-based test, good performance indicated the Wooldridge-Drukker test and the adjusted LM-based tests. However, they are not adapted for panels with time-heteroscedastic error terms.

The introduced example of application on German Health Care Usage Data demonstrated how the testing methods may be applied in the real panel data models. Based on the testing results, reduced panel with no correlation across time was estimated and compared with the full panel, eventually improving the estimated parameters. Alternatively, if the test rejects null hypothesis about absence of correlation in the panel data, techniques like FE generalized least squares or dynamic panel data models may be used.

References

Baltagi, B., Qi, L., 1991. A joint test for serial correlation and random individual effects. *Statistics & Probability Letters*; 11(3); 277-280.

Baltagi, B., Qi, L., 1995. Testing AR(1) against MA(1) disturbances in an error component model. *Journal of Econometrics*; 68(1); 133-151.

Baltagi, B., 2008. Econometric Analysis of Panel Data. John Wiley & Sons Ltd; Chichester.

Born, B., Breitung, J., 2016. Testing for Serial Correlation in Fixed-Effects Panel Data Models. *Econometric Reviews*; 35; 1290-1316.

Box, G., Pierce, D., 1970. Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*; 65(332); 1509–1526.

Breusch, T., Godfrey, L., 1981. A review of recent work on testing for autocorrelation in dynamic simultaneous models. In Currie, D., Nobay R., Peel, D. (eds.), Macroeconomic analysis: Essays in macroeconomics and econometrics. Croom Helm; 1981; 63-105.

Drukker, D., 2003. Testing for Serial Correlation in Linear Panel-data Models. *The Stata Journal*; 3; 168-177.

Fu, B., Li, W., Fung, W., 2002. Testing Model Adequacy for Dynamic Panel Data with Intercorrelation. *Biometrika*; 89(3); 591-601.

Inoue, A., Solon, G., 2006. A portmanteau test for serially correlated errors in fixed effects models. *Econometric Theory*; 22; 835-851.

Jochmans, K., 2019. Testing for correlation in error-component models. *Cambridge Working Papers in Economics No. 1993*; University of Cambridge; Faculty of Economics, 2019.

Ljung, G., Box, G., 1978. On a measure of lack of fit in time series models. *Biometrika*; 65(2); 297–303.

Nickell, S., 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica*; 49(6); 1417-1426.

Okui, R., 2009. Testing serial correlation in fixed effects regression models based on asymptotically unbiased autocorrelation estimators. *Mathematics and Computers in Simulation*; 79(9); 2897-2909.

Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel Data, first edition. The MIT Press; Cambridge, Massachusetts; 2002.

Appendix 2: MatLab Code

Main subroutine for testing and estimation of panel models with the CCR-based method, with an input $X(i).x \in \mathbb{R}^{K\times T}$ - matrix of regressors collected for each individual $i = 1, \ldots N$, and $y \in \mathbb{R}^{T\times N}$ - dependent variable:

```
function p = CCR_Panel(X,y)

T = size(y,1);
N = size(y,2);
K = size(X(1).x,2);

% demeaning matrix

QT = eye(T) - ones(T,1)*(ones(1,T)*ones(T,1))^(-1)*ones(1,T);

% demean regression by taking out time invariant parameters

y_dots = zeros(T,N);

for i=1:N
    X(i).x_dots = round(QT * X(i).x,10);
    y_dots(:,i) = round(QT * y(:,i),10);
end

% sum up the matrices for FE estimation

XX = 0;
Xy= 0;

for i=1:N
    XX = XX + X(i).x_dots'*X(i).x_dots;
    Xy = Xy + X(i).x_dots'*y_dots(:,i);
end

% find time-invariant regressors and delete them

zer = 0;
z = 1;
for k=1:K
    if XX(k,:) == zeros(1,K)
        zer(z) = k;
        z = z + 1;
    end
end

if zer ~= 0
    XX(zer,:) = [];
    XX(:,zer) = [];
    Xy(zer) = [];
    for i=1:N
        X(i).x_dots(:,zer) = [];
    end
end

```

```
% calculate FE estimators and errors

b_FE = XX\Xy;

u_dots_hat = zeros(size(y_dots));

for i = 1:N
    u_dots_hat(:,i) = y_dots(:,i) - X(i).x_dots*b_FE;
end

mu_u_dots_hat = mean(u_dots_hat,2);

% Estimate Omega for FE

Omega_hat = 0;

for i=1:N
    Omega_hat = Omega_hat + u_dots_hat(:,i)*u_dots_hat(:,i)';
end

Omega_hat = Omega_hat/N;

% asympotic variance sug. by Arrelano(1987)

XuuX = 0;

for i=1:N
    XuuX = XuuX + X(i).x_dots'*u_dots_hat(:,i)*u_dots_hat(:,i)'*X(i).
        x_dots;
end

Avar_FE = (XX\XuuX)/XX;

% Fixed Effects Generalized Least Squares FEGLS

% delete last T observation from x, y, u to avoid rank failure

for i=1:N
    X(i).x_dots1 = X(i).x_dots(1:end-1,:);
end

y_dots1 = y_dots(1:end-1,:);
u_dots_hat1 = u_dots_hat(1:end-1,:);

% estimate Covariance matrix between time periods Omega

Omega_hat1 = 0;

for i=1:N
    Omega_hat1 = Omega_hat1 + u_dots_hat1(:,i)*u_dots_hat1(:,i)';
end

Omega_hat1 = Omega_hat1/N;

```

```
% calculate FEGLS estimators

XOX = 0;
XOy = 0;

for i=1:N
    XOX = XOX + (X(i).x_dots1'/Omega_hat1)*X(i).x_dots1;
    XOy = XOy + (X(i).x_dots1'/Omega_hat1)*y_dots1(:,i);
end

b_FEGLS = XOX\XOy;

% Preparations for testing CCR. Because of demeaning Omega_hat =
% QT*Omega*QT

% Assuming that there is no correlation, we solve a system of  for
    sigma^2 from
% 1 to T. Equation for each upper non-diagonal element in Omega(symm.)
% sigma^2_i + sigma^2_j = Omega(i,i) + Omega(j,j) - 2*Omega(i,j)
% Altogether we have T*(T-1)/2 equations (i=1,T; j=1,T; j>i) with T
% unknowns. This will work if T>2

% right-hand side

RHS = zeros(T);
for i=1:T
    for j=i:T
        RHS(i,j) = Omega_hat(i,i) + Omega_hat(j,j) - 2*Omega_hat(i,j);
    end
end

% form an equation A*sigma=c and solve it

A = zeros(T*(T-1)/2,T);
c = zeros(T,1);
k=1;
for i=1:T
    for j=1:T
        if RHS(i,j)~=0
            A(k,i)=1;
            A(k,j)=1;
            c(k) = RHS(i,j);
            k=k+1;
        end
    end
end
Sigma_diag = (A'*A)\A'*c;

% Construct system if T equations for diagonal elements, using that we
    know
% sigma^2_i for i=1,T, with unknowns R_i = sum(sigma(i,:)) - sum of all
% sigmas with index i:
% sum(R_k)-2*T*R_i = B_ii - T^2*sigma^2_i, for i from 1 to T

```

```
155 % right-hand side
156 B = Omega_hat*(T^2);
157 c = diag(B)- T^2*Sigma_diag;
158
159 % left-hand side
160 A = ones(T);
161 for t=1:T
162     A(t,t) = A(t,t) - 2*T;
163 end
164
165 R = A\c;
166
167 % with matrix B and vector R we can restore Matrix Omega before
        demeaning
168 % as sigma(i,j) = ((B(i,j) + T*R_i + T*R_j - sum(R_k))/(T^2)
169 % No correlation in unknown Omega <-> no correlation in Omega_rest
170 Omega_rest = zeros(T);
171 for i=1:T
172     for j=1:T
173         Omega_rest(i,j) = (B(i,j) + T*R(i) + T*R(j) - sum(R))/(T^2);
174     end
175 end
176
177 % Apply CCR technique to check whether observations are correlated in
        time
178 % and whether it makes sence to use FEGLS
179
180 l=1;
181 p = ones(size(Omega_hat));
182 for t1=1:T-1
183     dt = 1;
184     while 1
185         %disp(['Case ' num2str(l) ': period ' '<strong>' num2str(t1)
                '</strong>' ' with period ' '<strong>' num2str(t1+dt) '</
                strong>']);
186         p(t1,t1+dt) = CCR(mu_u_dots_hat([t1,t1+dt]), Omega_rest([t1,t1+
                dt],[t1,t1+dt]), mu_u_dots_hat([t1,t1+dt]), [Omega_rest(t1,
                t1), 0; 0, Omega_rest(t1+dt,t1+dt)], N);
187         p_vec(l) = p(t1,t1+dt);
188         l=l+1;
189         dt = dt + 1;
190         if (t1+dt)>T %|| p<significance
191             break
192         end
193     end
194 end
195
196 [p_total, ind] = min( min(p));
197 p = p_total;
198 disp('P-value for H0: no correlation between time points');
199 disp(p_total);
200
201 disp(ind);
202
```

```
% estimation results if beta is not known

% VarNames = {'b_FE', 'b_FEGLS', 'difference_FEGLS_and_FE'};
% Table = table(b_FE, b_FEGLS, abs(b_FEGLS - b_FE), 'VariableNames',
    VarNames);
% disp(Table);
%
% diff = abs(b_FEGLS - b_FE)'*ones(size(b_FE));
%
% disp(['total deviation of FE from FEGLS is: ' num2str(diff)]);
% disp(Omega_rest);
```

Subroutine for application of the CCR-based method on healthcare panel:

```
% Example panel model

clear all

load healthcare

for i=size(A,1):-1:1
    if A(i,3) ~= 1984 && A(i,3) ~= 1985 && A(i,3) ~= 1986 &&...
       A(i,3) ~= 1987 && A(i,3) ~= 1991
        A(i,:)=[];
    end
end

T = length(unique(A(:,3))); % time periods
N = size(A,1)/T; % individuals
significance = 0.05; % significance level for running CCR

y = reshape(A(:,41),T,N); % regressand as a Matrix(T*N) for estimation
y(isnan(y))=0; % delete NaN
y = double(y); % adjust formats

x =[ones(N*T,1), A(:,36), A(:,26), A(:,2), A(:,5), A(:,8) A(:,9)]; %
    regressors
x(isnan(x))=0; % delete NaN
x = double(x); % adjust formats
K = size(x,2)-1; % number of regressors

for i=1:N % x as a structure with with X(i) - individual, for each i we
          % have a X(i).x - (T*(K+1)) matrix with K regsessors and
              constant
          % in T time periods
    X(i).x = x((i-1)*T+1:i*T,:);
end

clearvars healthcare_description healthcare A


% demeaning matrix

QT = eye(T) - ones(T,1)*(ones(1,T)*ones(T,1))^(-1)*ones(1,T);

```

```
% demean regression by taking out time invariant parameters

y_dots = zeros(T,N);

for i=1:N
    X(i).x_dots = round(QT * X(i).x,10);
    y_dots(:,i) = round(QT * y(:,i),10);
end

% sum up the matrices for FE estimation

XX = 0;
Xy= 0;

for i=1:N
    XX = XX + X(i).x_dots'*X(i).x_dots;
    Xy = Xy + X(i).x_dots'*y_dots(:,i);
end

% find time-invariant regressors and delete them

zer = 0;
z = 1;
for k=1:K+1
    if XX(k,:) == zeros(1,K+1)
        zer(z) = k;
        z = z + 1;
    end
end
XX(zer,:) = [];
XX(:,zer) = [];
Xy(zer) = [];
for i=1:N
    X(i).x_dots(:,zer) = [];
end

% calculate FE estimators and errors

b_FE = XX\Xy;

u_dots_hat = zeros(size(y_dots));

for i = 1:N
    u_dots_hat(:,i) = y_dots(:,i) - X(i).x_dots*b_FE;
end

mu_u_dots_hat = mean(u_dots_hat,2);

% Estimate Omega for FE

Omega_hat = 0;

for i=1:N
    Omega_hat = Omega_hat + u_dots_hat(:,i)*u_dots_hat(:,i)';
```

```
94 end
95
96 Omega_hat = Omega_hat/N;
97
98 % asympotic variance sug. by Arrelano(1987)
99
100 XuuX = 0;
101
102 for i=1:N
103     XuuX = XuuX + X(i).x_dots'*u_dots_hat(:,i)*u_dots_hat(:,i)'*X(i).
        x_dots;
104 end
105
106 Avar_FE = (XX\XuuX)/XX;
107
108 % Fixed Effects Generalized Least Squares FEGLS
109
110 % delete last T observation from x, y, u to avoid rank failure
111
112 for i=1:N
113     X(i).x_dots1 = X(i).x_dots(1:end-1,:);
114 end
115
116 y_dots1 = y_dots(1:end-1,:);
117 u_dots_hat1 = u_dots_hat(1:end-1,:);
118
119 % estimate Covariance matrix between time periods Omega
120
121 Omega_hat1 = 0;
122
123 for i=1:N
124     Omega_hat1 = Omega_hat1 + u_dots_hat1(:,i)*u_dots_hat1(:,i)';
125 end
126
127 Omega_hat1 = Omega_hat1/N;
128
129 % calculate FEGLS estimators
130
131 XOX = 0;
132 XOy = 0;
133
134 for i=1:N
135     XOX = XOX + (X(i).x_dots1'/Omega_hat1)*X(i).x_dots1;
136     XOy = XOy + (X(i).x_dots1'/Omega_hat1)*y_dots1(:,i);
137 end
138
139 b_FEGLS = XOX\XOy;
140
141 % Preparations for testing CCR. Because of demeaning Omega_hat =
142 % QT*Omega*QT
143
144 % Assuming that there is no correlation, we solve a system of  for
      sigma^2 from
145 % 1 to T. Equation for each upper non-diagonal element in Omega(symm.)
```

```
% sigma^2_i + sigma^2_j = Omega(i,i) + Omega(j,j) - 2*Omega(i,j)
% Altogether we have T*(T-1)/2 equations (i=1,T; j=1,T; j>i) with T
% unknowns. This will work if T>2

% right-hand side

RHS = zeros(T);
for i=1:T
    for j=i:T
        RHS(i,j) = Omega_hat(i,i) + Omega_hat(j,j) - 2*Omega_hat(i,j);
    end
end

% form an equation A*sigma=c and solve it

A = zeros(T*(T-1)/2,T);
c = zeros(T,1);
k=1;
for i=1:T
    for j=1:T
        if RHS(i,j)~=0
            A(k,i)=1;
            A(k,j)=1;
            c(k) = RHS(i,j);
            k=k+1;
        end
    end
end
Sigma_diag = (A'*A)\A'*c;

% Construct system if T equations for diagonal elements, using that we
    know
% sigma^2_i for i=1,T, with unknowns R_i = sum(sigma(i,:)) - sum of all
% sigmas with index i:
% sum(R_k)-2*T*R_i = B_ii - T^2*sigma^2_i, for i from 1 to T

% right-hand side
B = Omega_hat*(T^2);
c = diag(B)- T^2*Sigma_diag;

% left-hand side
A = ones(T);
for t=1:T
    A(t,t) = A(t,t) - 2*T;
end

R = A\c;

% with matrix B and vector R we can restore Matrix Omega before
    demeaning
% as sigma(i,j) = ((B(i,j) + T*R_i + T*R_j - sum(R_k))/(T^2)
% No correlation in unknown Omega <-> no correlation in Omega_rest
Omega_rest = zeros(T);
for i=1:T
```

```
198     for j=1:T
199         Omega_rest(i,j) = (B(i,j) + T*R(i) + T*R(j) - sum(R))/(T^2);
200     end
201 end
202
203 % Apply CCR technique to check whether observations are correlated in
        time
204 % and whether it makes sence to use FEGLS
205
206 p = zeros(T);
207 l=1;
208 for t1=1:T-1
209     dt = 1;
210     while 1
211         disp(['Case ' num2str(l) ': period ' '<strong>' num2str(t1) '</
              strong>' ' with period ' '<strong>' num2str(t1+dt) '</strong
              >']);
212         p(t1,t1+dt) = CCR(mu_u_dots_hat([t1,t1+dt]), Omega_rest([t1,t1+
              dt],[t1,t1+dt]), mu_u_dots_hat([t1,t1+dt]), [Omega_rest(t1,
              t1), 0; 0, Omega_rest(t1+dt,t1+dt)], N);
213         l=l+1;
214         dt = dt + 1;
215         if (t1+dt)>T %|| p<significance
216             break
217         end
218     end
219 end
220
221 % estimation results if beta is not known
222
223 VarNames = {'b_FE', 'b_FEGLS', 'difference_FEGLS_and_FE'};
224 Table = table(b_FE, b_FEGLS, abs(b_FEGLS - b_FE), 'VariableNames',
        VarNames);
225 disp(Table);
226
227 diff = abs(b_FEGLS - b_FE)'*ones(size(b_FEGLS));
228
229 disp(['total deviation of FE from FEGLS is: ' num2str(diff)]);
```

Part 3 Confidence Regions for Bivariate Normal Distributions. Extension of Mood Approach

3.1 Introduction

Testing hypotheses and constructing confidence intervals are the essential tasks of any model estimation. Testing the significance of the parameters and their combinations together with testing joint significance of those parameters compose standard instruments of the data analysis. T-test, F-test and χ^2-tests are well-known to every economist and econometrician and start the extensive list of tests designed for different purposes.

In this Part I will describe a new approach for jointly testing means and variances/-covariances of a sample of random vectors, based on the idea of a test, presented by Mood (1950), p.227. This approach may be used as an additional or alternative instrument of analysis in modern portfolio theory. Starting from Markowitz (1952), who aimed to find the best ratio between expected return as a mean and variance of return. Another constructive implementation is testing for serial correlation in panels, as an alternative to approaches proposed by Ljung & Box (1978) or Born & Breitung (2016).

I will formally describe a technique of testing whether the given sample from $\mathbb{R}^m$ may follow multivariate normal distribution with the proposed parameters $N(\mu_0, \Sigma_0)$ and with a predetermined significance level γ. Firstly, assume an independent and identically distributed (iid) sample of random vectors $\{\mathbf{X}_i\}_{i=1}^n = \mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n \in \mathbb{R}^2$ from bivariate normal distribution with unknown parameters $N(\cdot, \cdot)$. Additionally, note that mean is defined as $\mu \in \mathbb{R}^2$ and covariance matrix is defined as $\Sigma \in \mathbb{R}^{2\times 2}$. Therefore I will test whether the given sample $\{\mathbf{X}_i\}_{i=1}^n$ may follow normal distribution with the proposed parameters $N(\mu_0, \Sigma_0)$ as defined in Null Hypothesis in (3.1):

$$H_0 : (\mu, \Sigma) = (\mu_0, \Sigma_0) \tag{3.1}$$

An extension of this approach from bivariate to multivariate case is possible by jointly testing each pair of dimensions of a sample $\{\mathbf{X}_i\}_{i=1}^n$ from $\mathbb{R}^m$. It will be also presented in this Part and applied to the SUR model.

Most of the existing techniques were designed only for a univariate case of the introduced test (3.1). For example, approaches described in Frey, Marrero, Norton (2009) and Arnold & Shavelle (1998) may be applied only in univariate case. The technique, outlined in Jensen (1995), is also constructed for a univariate case. It may be expanded to a multivariate case, but this test is based on Rao distance and therefore may not be expressed analytically for dimensionality $m > 1$. However, in Part 1 Cumulative Distribution Function Confidence Region (CCR) approach

for bivariate case of hypothesis (3.1) was introduced. Because of lack of equivalent methods, CCR was only examined for the specific cases, such as testing for temporal dependencies in Panel models. In Part 2 I applied CCR to find serial correlations in Panels and compared this approach with LM based Breush-Godfrey test, introduced by Godfrey (1978), with classic portmanteau test developed by Ljung & Box (1978), and with the most modern approaches as well, like the test developed by Jochmans (2019).

As a result of the analysis made above, I have designed a procedure of running a test (3.1) in the multivariate case, based on the technique, introduced by Mood (1950), p.227. Mood (1950) in his book takes the t-test for means and Wald test for variances and merges them into one joint mean-variance test for univariate normal distribution. Simultaneously Mood's approach forms a confidence region as a part of a parabola, restricted by two horizontal lines.

The procedure for testing normal distributions in the bivariate case uses the same logic, as the univariate one. It sets up as a combination of a test for means with a test for variances/covariances into one joint test. I have called this method the bivariate Mood test. An approach for means, used by bivariate Mood method, is a Hotelling t^2-test, introduced by Hotelling (1931). This is an extension of the usual t-test for vectors. The second is a Wishart test, which matches covariance matrices for equality. It is based on the definition of the Wishart distribution (see for example Gupta & Nagar (2000), p.87), and appears to be an extension of Gamma distribution. Wishart distribution, as a matrix-represented, describes asymptotic properties of each element of covariance matrix simultaneously. Merging the t^2 and Wishart test establishes a new joint Mood test for means and covariances of bivariate Normal distribution (3.1). As mentioned, further bivariate Mood test may be also expanded to a multivariate case by testing all m dimensions of the sample of random vectors $\{\mathbf{X}_i\}_{i=1}^n$ pairwisely. More technically, the distribution of Hotelling t^2 is given by equation (3.2), where new variable $\overline{\mathbf{X}}$ is a sample mean of $\{\mathbf{X}_i\}_{i=1}^n$, given in (3.3). The distribution of the Wishart statistic is defined by equation (3.4) with S - sample covariance matrix of $\{\mathbf{X}_i\}_{i=1}^n$, corrected for bias (3.5). And the task that bivariate Mood approach solves in this Part is connection of the both test (3.2) and (3.4) regarding the significance level γ and dependencies with each other.

$$n(\overline{\mathbf{X}} - \mu)\Sigma^{-1}(\overline{\mathbf{X}} - \mu) \sim \chi_2^2 \tag{3.2}$$

$$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{m} \mathbf{X}_i \tag{3.3}$$

$$(n-1)S \sim W_2(n-1, \Sigma) \tag{3.4}$$

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})' \tag{3.5}$$

Thereafter I demonstrate in this Part how the bivariate Mood test (3.2)-(3.5) may be decomposed into a system of 4 scalar random variables: z, h_{11}, h_{22} and h_{22}, which are transformations of Hotelling and Wishart statistics. Distributions of these new

random variables are given in (3.6)-(3.9) with $f_{h_{12}}$ - PDF of normal variance-mean mixture, that uses gamma distribution as a mixing one.
The next section of the Part 3 illustrates how the bivariate Mood confidence region, defined by (3.6)-(3.9), may be further simplified using affine transformations from Part 1. This section is crucial in analyzing the shape of the confidence region, as well as highlighting the subregions with a smaller probability density for further comparison with the CCR method.
After the decomposition into scalars, the CCR and Mood techniques will be compared. For these two tests, which may interchange with each other, the next section of the Part 3 aims to analyze them with respect to computation speed, shape of the confidence set (for bivariate case one has to deal with a 5-dimensional figures), correspondence of theoretical and real significance level, reaction on non-standard samples from the other distributions and samples with outliers. Analyses made in this Part demonstrated a better efficiency of the Mood method w.r.t. computation time, but worse behavior w.r.t. to the shape of confidence set, especially on larger significance levels $\gamma > 20\%$. Furthermore both the Mood and CCR approaches indicated good results while testing their real significance level.
Another substantial point is that the bivariate Mood test is an exact test, while the CCR is an approximate one. This gives real significance levels of the bivariate Mood approach, that are following the theoretical distribution derived. On the contrary, the CCR method only guarantees that the real significance level will be in some predefined neighborhood of the theoretical significance level γ, but not converging to it.
Analysis of the shapes of these two methods shows the CCR as the more effective one. The bivariate Mood approach produces a confidence region in the form of a generalized cylinder. The CCR method deals with a problem of edges of the cylinder, that have lower probability density, and smoothen confidence region to the form of so called 'pseudo-ellipse', mentioned in Part 1.

$$nz \sim \chi^2_2 \tag{3.6}$$
$$(n-1)h_{11} \sim \chi^2_{n-1} \tag{3.7}$$
$$(n-1)h_{22} \sim \chi^2_{n-1} \tag{3.8}$$
$$(n-1)h_{12} \sim f_{h_{12}}(h_{12}, n) \tag{3.9}$$

Finally, in this Part I have derived a new Mood approach for jointly testing means and variances/covariances of multivariate normal distribution. Then the Mood method was analyzed and compared with the CCR technique with respect to the most important parameters, including effectiveness, accuracy and robustness. Among them CCR has resulted in being more effective w.r.t. shape. While the bivariate Mood test, as an exact one, is more accurate on big sample sizes, as it is convergent to the theoretical significance level γ.

3.2 Model Set Up

The construction of the bivariate Mood approach starts with the sample of iid random vectors $\{\mathbf{X}_i\}_{i=1}^n = \mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n \in \mathbb{R}^2$. Each of the elements of the sample are drown from bivariate normal distribution, but with unknown parameters μ and Σ. One can also define sample mean $\overline{\mathbf{X}}$ in (3.3) and sample variance matrix S in (3.5). Null hypothesis, defined in (3.1) checks if the sample $\{\mathbf{X}_i\}_{i=1}^n$ may follow normal distribution with the proposed set of parameters (μ, Σ) and for the given significance level γ.
Therefore, the task is to define a value for each set of a points (μ, Σ) that describes how probable normal distribution with the proposed parameters $N(\mu, \Sigma)$ may be the one from which our original sample $\{\mathbf{X}_i\}_{i=1}^n$ was drawn. Based on this probability and critical value, one can reject or non-reject Null hypothesis (3.1). Similarly, one can construct confidence region $\mathcal{R}(\mathbf{X})$ in (3.10) as a locus of all points that will not be rejected with a probability larger than $1 - \gamma$ (probability of non-rejection of critical value).

$$P((\vec{\mu}, \Sigma)) \in \mathcal{R}(\mathbf{X})) = 1 - \gamma \tag{3.10}$$

3.2.1 Test for Mean

The first part of the bivariate Mood approach is based on the t^2-test, first presented by Hotelling (1931). In the notations introduced, it describes the joint relation of two sample means of our sample $\overline{\mathbf{X}} \in \mathbb{R}^2$ to the proposed values $\mu \in \mathbb{R}^2$ (3.11):

$$\frac{n(n-2)}{2(n-1)}(\overline{\mathbf{X}} - \mu)' S^{-1} (\overline{\mathbf{X}} - \mu) \sim F_{2,n-2} \tag{3.11}$$

In fact, Hotelling t^2 measures the nonlinear distance between two means: sample and population. The further population mean is located from the sample average, the higher value this statistic will take. Based on Hotelling t^2 'distance' (3.11) from Johnson & Wichern (2002), p.235 I introduce another statistic z, that uses theoretical variance matrix Σ instead of estimated matrix S (3.12):

$$n * z = n(\overline{\mathbf{X}} - \mu)' \Sigma^{-1} (\overline{\mathbf{X}} - \mu) \sim \chi^2_2 \tag{3.12}$$

According to (3.12), one still deals with a distance. And (3.12) is a pivotal quantity as well, since it does not depend on the unknown parameters (see Wooldridge (2010)), p.439. It is highly important that a one-sided χ^2-test should be used, rejecting only the points that are too far from each other, and therefore having a higher statistic. Points that are extremely close to each other should not be rejected and they are not rejected by one-sided test. Albeit two-sided χ^2 test claims to reject them. Therefore, a one-sided test will be used for (3.12).

3.2.2 Test for Variance

The second part of the bivariate Mood test is based on the Wishart distribution. First, assume that the sample of iid random variables $\{\mathbf{X}_i\}_{i=1}^n \in \mathbb{R}^2$ defined before follows a zero mean normal distribution $N(0, \Sigma)$. For this case Gupta & Nagar (2000), p.88 proved that the sum of squared elements of our sample will follow the Wishart distribution (3.13):

$$\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i' \sim W_2(n, \Sigma) \tag{3.13}$$

In the next step one can switch back to the first sample $\{\mathbf{X}_i\}_{i=1}^n$ drawn from a normal distribution $N(\mu, \Sigma)$ and demean it. Then, according to Muirhead (1982), p.86 estimation of the covariance matrix, defined in (3.5) will follow the Wishart distribution with $(n-1)$ degrees of freedom (3.14):

$$\sum_{i=1}^{n} (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})' = (n-1)S \sim W_2(n-1, \Sigma) \tag{3.14}$$

Many papers researching this topic claim that one may use both (3.13) and (3.14) as an estimation for covariance matrix. To exemplify, Gupta & Bodnar (2014) mentions that there is only a slight difference between (3.13) and (3.14). This Part uses an unbiased estimator (3.5). Therefore, equation (3.14), which describes the distribution of S, is preferable.

To use formula (3.14) as a statistic, one has to transform the left-hand side in the way it follows the Wishart distribution, independent of parameter Σ. For this reason, assume that $\Sigma^{-\frac{1}{2}}$ is existing and known. Then according to the properties of the Wishart distribution, given in Gupta & Nagar (2000), p.90 one can rewrite the distribution of S as (3.15), denoting the function H:

$$(n-1) * H = (n-1) * \Sigma^{-\frac{1}{2}} * S * \Sigma^{-\frac{1}{2}} \sim W_2(n-1, I_2) \tag{3.15}$$

I_2 in (3.15) is a 2×2 identity matrix. Thus, by achieving a function whose distribution does not depend on the unknown parameters, one gets a pivotal quantity in (3.15). What about the existence of $\Sigma^{-\frac{1}{2}}$? Firstly, Σ^{-1} exists as an inverse of covariance matrix with no degenerate variables included (by set up). Note that the inverse is positive definite (see Horn & Johnson (2013), p.438). Secondly, existence of the square root is in common case not obvious, but here helps fact that Σ^{-1} is positive definite. And Horn & Johnson (2013), p.439 proved the existence of a square root for such matrices. Hence, for any Σ there always exists a matrix $\Sigma^{-\frac{1}{2}}$.

Next, mention that in 2-dimensional case Σ is defined as (3.16). Here and further indexes x and y stand for the first and second dimensions of the sample $\{\mathbf{X}_i\}_{i=1}^n$ and its functions. To calculate the matrix $\Sigma^{-\frac{1}{2}}$ I use a Cayley–Hamilton theorem and a direct algorithm, derived in Levinger (1980). This method gives 4 roots, but two of

them are defined in complex plane, which is not the point of interest. The other two just have opposite signs. Since in (3.15) one multiplies by $\Sigma^{-\frac{1}{2}}$ twice, its sign does not matter and I will further use a positive root of Σ^{-1} (3.17), given as a matrix $\Sigma^{-\frac{1}{2}}$ in (3.18):

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \tag{3.16}$$

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{(1-\rho^2)\sigma_x^2} & \frac{-\rho}{(1-\rho^2)\sigma_x\sigma_y} \\ \frac{-\rho}{(1-\rho^2)\sigma_x\sigma_y} & \frac{1}{(1-\rho^2)\sigma_y^2} \end{pmatrix} \tag{3.17}$$

$$\Sigma^{-\frac{1}{2}} = \frac{\sqrt{1-\rho^2}\sigma_x\sigma_y}{\sqrt{\sigma_x^2+\sigma_y^2+2\sigma_x\sigma_y\sqrt{1-\rho^2}}}\left(\Sigma^{-1} + \frac{1}{\sqrt{1-\rho^2}\sigma_x\sigma_y} I_2\right) \tag{3.18}$$

At this point one has two pivotal quantities: (3.12) and (3.15). As pivots, they are also uncorrelated and independent, as follows from Greene (2002), p. 891. For clarification of the next step, I merge their confidence intervals (3.19)-(3.20) together and obtain (3.21):

$$P(n * z \in A_1) = P(n * z < a) = 1 - \gamma_1 \tag{3.19}$$

$$P((n-1)H \in A_2) = P(b < (n-1) * \Sigma^{-\frac{1}{2}} * S * \Sigma^{-\frac{1}{2}} < c) = 1 - \gamma_2 \tag{3.20}$$

$$\begin{aligned} (1-\gamma_1)(1-\gamma_2) &= P(n * z \in A_1) * P((n-1)H \in A_2) \\ &= P(n * z \in A_1, (n-1)H \in A_2) = 1 - \gamma \end{aligned} \tag{3.21}$$

Where A_1 and A_2 are some neighborhoods around $\overline{\mathbf{X}}$ and S respectively. Pivot $n * z$ in (3.19) is only bounded by a from the above, since it uses one-sided χ^2-test, as was mentioned before.

As a reminder, the Wishart distribution is a matrix one. Therefore, equation (3.20) describes a matrix inequality. And to work with this inequality, one has to convert it to a system of scalar inequalities by calculating (3.15) and its distributions element-wise. Thus, I define elements of the sample covariance matrix S in (3.22) and elements of the pivot H in (3.23). To mention both of the matrices are symmetrical.

$$S = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} = \frac{1}{n-1}\begin{pmatrix} (\mathbf{X}_x - \overline{\mathbf{X}}_x)'(\mathbf{X}_x - \overline{\mathbf{X}}_x) & (\mathbf{X}_x - \overline{\mathbf{X}}_x)'(\mathbf{X}_y - \overline{\mathbf{X}}_y) \\ (\mathbf{X}_x - \overline{\mathbf{X}}_x)'(\mathbf{X}_y - \overline{\mathbf{X}}_y) & (\mathbf{X}_y - \overline{\mathbf{X}}_y)'(\mathbf{X}_y - \overline{\mathbf{X}}_y) \end{pmatrix} \tag{3.22}$$

$$H = \Sigma^{-\frac{1}{2}} * S * \Sigma^{-\frac{1}{2}} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix} \tag{3.23}$$

Now, as $\Sigma^{-\frac{1}{2}}$ is calculated in equation (3.18) and S in (3.22), it is time to plug them in function of H in (3.23). The result is quite bulky, therefore I provide the matrix H element-wise in (3.24)-(3.26):

$$h_{11} = \frac{s_{11}(\sigma_y + \sqrt{1-\rho^2}\sigma_x)^2 - 2s_{12}\rho\sigma_x(\sigma_y + \sqrt{1-\rho^2}\sigma_x) + s_{22}\rho^2\sigma_x^2}{(1-\rho^2)\sigma_x^2(\sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y\sqrt{1-\rho^2})} \tag{3.24}$$

$$\begin{aligned} h_{12} = (s_{12}\rho^2\sigma_x\sigma_y - s_{11}\rho\sigma_y(\sigma_y + \sqrt{1-\rho^2}\sigma_x) + s_{12}(\sigma_y + \sqrt{1-\rho^2}\sigma_x) \\ (\sigma_x + \sqrt{1-\rho^2}\sigma_y) - s_{22}\rho\sigma_x(\sigma_x + \sqrt{1-\rho^2}\sigma_y)) \\ /((1-\rho^2)\sigma_x\sigma_y(\sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y\sqrt{1-\rho^2})) \end{aligned} \tag{3.25}$$

$$h_{22} = \frac{s_{22}(\sigma_x + \sqrt{1-\rho^2}\sigma_y)^2 - 2s_{12}\rho\sigma_y(\sigma_x + \sqrt{1-\rho^2}\sigma_y) + s_{11}\rho^2\sigma_y^2}{(1-\rho^2)\sigma_y^2(\sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y\sqrt{1-\rho^2})} \tag{3.26}$$

In the next step, marginal distributions of matrix elements h_{11}, h_{12} and h_{22} may be obtained by applying the Barlett decomposition technique for Wishart distribution, designed by Kshirsagar (1959). I apply this procedure to the bivariate Wishart distribution (3.27) that describes the behavior of matrix H. Following Barlett decomposition, matrices A and L are obtained in (3.28), where L is a Cholesky factorization of variance I_2, which is still I_2. Matrix A consists of two χ^2-distributed random variables with $(n-1)$ and $(n-2)$ degrees of freedom and standard normally distributed random variable n_{21}. All 3 random variables defined in A are jointly independent by construction. Finally, plugging A and L in equation (3.27) gives the element-wise distribution of H in equation (3.30).

$$(n-1)H = LAA'L' \sim W_2(n-1, I_2) \tag{3.27}$$

$$L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} c_1 & 0 \\ n_{21} & c_2 \end{pmatrix}, \tag{3.28}$$

$$c_1^2 \sim \chi^2_{n-1}, \quad c_2^2 \sim \chi^2_{n-2}, \quad n_{21} \sim N(0,1) \tag{3.29}$$

$$(n-1)H = \begin{pmatrix} c_1^2 & c_1 n_{21} \\ c_1 n_{21} & c_2^2 + n_{21}^2 \end{pmatrix} \tag{3.30}$$

For the bivariate case, Kshirsagar (1959) showed that the elements h_{11}, h_{22} and h_{12} of the matrix H are uncorrelated and independent. Diagonal elements h_{11} and h_{22} will follow χ^2-distribution, given in equations (3.31)-(3.32). Distribution of h_{11} is obviously the same as in c_1^2. Distribution of h_{22} develops from the definition of χ^2-distribution; adding one more independent normally distributed random variable increases degrees of freedom from $(n-2)$ to $(n-1)$. Non-diagonal element h_{12} will follow a normal variance-mean mixture distribution (3.33). The PDF of this distribution is given in (3.34), with gamma function $\Gamma(\cdot)$ and modified Bessel function of the second kind $K_\cdot(\cdot)$. Note that there is a discontinuity point $h_{12} = 0$, but it may be neglected as a part of a null set. For the PDF $f(h_{12})$ this irregularity is a removable discontinuity point, which may be eliminated by additionally setting $f(0) = \lim_{x_0 \to 0}(f(x_0))$.

$$(n-1)h_{11} \sim \chi^2_{n-1} \tag{3.31}$$

$$(n-1)h_{22} \sim \chi^2_{n-1} \tag{3.32}$$

$$(n-1)h_{12} \sim f_{h_{12}}(h_{12}, n) \tag{3.33}$$

$$f_{h_{12}}(h_{12}, n) = \frac{|h_{12}|^{\frac{n-2}{2}}}{\Gamma(\frac{n-1}{2})\sqrt{2^{n-2}\pi}} K_{\frac{n-2}{2}}(|h_{12}|), h_{12} \neq 0 \tag{3.34}$$

To give an understanding how h_{12} behaves, a short description of the derived pdf $f_{h_{12}}$ will be provided. Firstly, it is symmetrical around the y-axis. This property is obtained from normal distribution. Secondly, it follows the same bell-shaped form, but with thicker tails, as may be seen on the Figure 3.1. Additionally, all non-central moments of h_{12} may be calculated as the product of corresponding non-central moments of normally distributed n_{21} and c_1, which is χ-distributed, random variables.

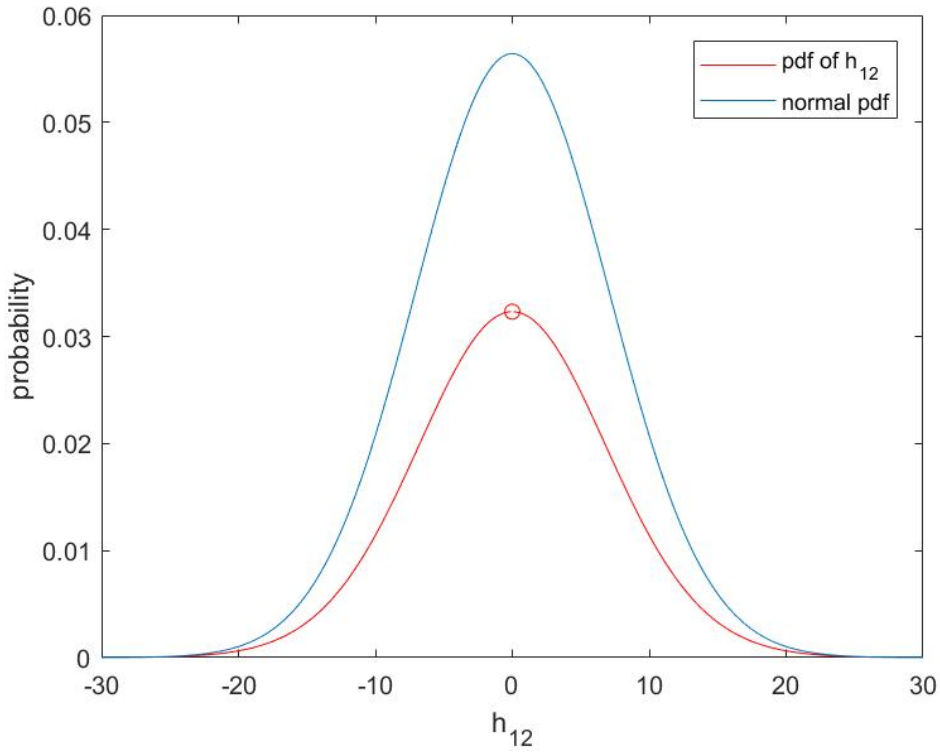

Figure 3.1: PDF of Non-Diagonal Element h_{12} from Wishart Distribution in Comparison with PDF of Normal Distribution for $\sigma^2 = 50$

3.2.3 Merging Mean and Variance Tests Together

The next step in the construction of the confidence set (3.10) is linking statistic z (3.12) with Wishart test statistics (3.31)-(3.33). Firstly, using the same notations as in the Wishart distribution part, I plug inverse covariance matrix (3.17) in equation (3.12) and get the scalar representation of statistic z in equation (3.35) with its distribution in (3.36):

$$z = \frac{\sigma_y^2 \left(\overline{\mathbf{X}}_x - \mu_x\right)^2 + \sigma_x^2 \left(\overline{\mathbf{X}}_y - \mu_y\right)^2 - 2\rho\sigma_x\sigma_y \left(\overline{\mathbf{X}}_x - \mu_x\right)\left(\overline{\mathbf{X}}_y - \mu_y\right)}{(1-\rho^2)\sigma_x^2\sigma_y^2} \tag{3.35}$$

$$nz \sim \chi_2^2 \tag{3.36}$$

After the disentangling of mean and variance inequalities, I can rewrite the confidence set (3.21) as a combination of scalar inequalities. To do this, remember that the elements of matrix H are jointly independent. z and H are also independent as pivots, hence h_{11}, h_{22}, h_{12} and z compose a set of jointly independent random variables. This allows me to write the confidence set in more detailed in (3.37). Note that h_{12} follows a distribution symmetrical around zero.

$$\begin{gathered} 1-\gamma = P((\vec{\mu}, \Sigma) \in \mathcal{R}(\mathbf{X})) = \\ P(nz < a, b < (n-1)h_{11} < c, d < (n-1)h_{22} < e, -f < (n-1)h_{12} < f) \\ = P(nz < a) * P(b < (n-1)h_{11} < c) * P(d < (n-1)h_{22} < e) * \\ P(-f < (n-1)h_{12} < f) = (1-\gamma_1)(1-\gamma_2)(1-\gamma_3)(1-\gamma_4) \end{gathered} \tag{3.37}$$

What about the selection of individual significance levels γ_i? Their choice and thus subsequent allocation to upper and lower tails does influence the volume and the amount of stretching and tightening of the confidence region. Furthermore, the influence of ratio between γ_i on the volume of confidence set will be discussed in subsection (3.2.6). However, the choice of γ_i in equation (3.37) does not influence the overall significance level γ, since all the gammas should be chosen in the way their confidence levels $(1-\gamma_i)$ multiply to the predefined confidence level $(1-\gamma)$. Hence, without loss of generality, I let $(1-\gamma_i) = \sqrt[4]{1-\gamma}$. The inverse statement for recovering γ from γ_i however is not true. To find the overall significance level γ, one should come up with the smallest possible confidence region that covers all the points given by γ_i, keeping chosen proportion of individual gammas unchanged. In the case considered all γ_i should be the same. Hence, a simple algorithm in (3.38) could be used to treat with gammas.

$$\begin{gathered} \gamma \text{ is known } \implies 1-\gamma_i = \sqrt[4]{1-\gamma} \\ \gamma_i \text{ is known } \implies 1-\gamma = (1-\min(\gamma_i))^4 \end{gathered} \tag{3.38}$$

At this point I can write down the confidence region $\mathcal{R}(\mathbf{X})$, derived in this subsection, as (3.39), and therefore set up a a procedure for joint (μ, Σ) test, defined in (3.1): a sample $\{\mathbf{X}_i\}_{i=1}^n$ is rejected by null hypothesis H_0, if its sample mean $\overline{\mathbf{X}}$ and sample covariance S are not a part of the confidence set $\mathcal{R}(\mathbf{X})$.

$$\mathcal{R}(\mathbf{X}) = \left\{ \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \right) : \begin{array}{l} z < \frac{a}{n} \\ \frac{b}{n-1} < h_{11} < \frac{c}{n-1} \\ \frac{d}{n-1} < h_{22} < \frac{e}{n-1} \\ -\frac{f}{n-1} < h_{12} < \frac{f}{n-1} \end{array} \right\} \tag{3.39}$$

As a result, the confidence set derived in (3.39) represents a range of points $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ in 5-dimensional space, such that each of the points may not be rejected by null hypothesis (3.1).

3.2.4 Limiting Properties of Pivots z, h_{11}, h_{22} and h_{12}

To verify the correctness of functional forms of z, h_{11}, h_{22} and h_{12}, I examine the limiting properties of these statistics, established in the previous subsections. The aim of this subsection is to show that whenever sample parameters get closer to theoretical values, statistics, determined earlier, will also get closer to the distributions these statistics follow. This means that significance γ_i will tend to its maximum of 100% when sample parameters get closer to theoretical values. In degenerate case when sample and theoretical moments are exactly the same, statistics should be equal to the mean of their limiting distributions.

I start with nz, which follows χ^2_2-distribution (equations (3.35)-(3.36)). In this case, only sample means are included in its formula and we limit only them to the theoretical means. The result obtained in (3.40) is expected, since one-sided test on χ^2 achieves its maximum significance at zero point, when all the distribution is in the right tail.

$$\lim_{\substack{\overline{\mathbf{X}}_x \to \mu_x \\ \overline{\mathbf{X}}_y \to \mu_y}} \left(\frac{\sigma_y^2 \left(\overline{\mathbf{X}}_x - \mu_x\right)^2 + \sigma_x^2 \left(\overline{\mathbf{X}}_y - \mu_y\right)^2 - 2\rho\sigma_x\sigma_y \left(\overline{\mathbf{X}}_x - \mu_x\right)\left(\overline{\mathbf{X}}_y - \mu_y\right)}{(1-\rho^2)\sigma_x^2\sigma_y^2} \right)$$
$$*n = n\frac{\sigma_y^2 * 0 + \sigma_x^2 * 0 - 2\rho\sigma_x\sigma_y * 0}{(1-\rho^2)\sigma_x^2\sigma_y^2} = 0 \tag{3.40}$$

Behavior of h_{11} and h_{22} is exactly the same as far as they may be obtained from each other by simple substitution of σ_x and σ_y. Therefore, I examine them together in (3.41) by tending elements of sample covariance matrix S to respective elements of covariance matrix Σ.

$$\begin{gathered} \lim((n-1)h_{11}) = \lim((n-1)h_{22}) \\ = (n-1)\frac{\sigma_x^2(\sigma_y + \sqrt{1-\rho^2}\sigma_x)^2 - 2\rho^2\sigma_x^2\sigma_y(\sigma_y + \sqrt{1-\rho^2}\sigma_x) + \rho^2\sigma_x^2\sigma_y^2}{(1-\rho^2)\sigma_x^2(\sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y\sqrt{1-\rho^2})} \\ = (n-1)\frac{\sigma_y^2(1-\rho^2) + \sigma_x^2(1-\rho^2) + 2\sigma_x\sigma_y(1-\rho^2)\sqrt{1-\rho^2}}{(1-\rho^2)(\sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y\sqrt{1-\rho^2})} = n-1 \end{gathered} \tag{3.41}$$

According to the calculated result above, diagonal elements of matrix H tend to $n-1$, which is the mean of χ^2_{n-1}. And this is the point where significance obtains its maximum, because all the points are allocated to the left and right tails of the distribution.

The limit of the non-diagonal element h_{12} is obtained in the same way as the diagonal elements by tending sample variance to its theoretical counterpart. The result is

$\lim((n-1)(h_{12})) \to 0$ as may be anticipated, because the distribution h_{12} follows (3.34) is symmetrical around zero and maximum of significance value sets naturally in this point.

The results of the marginal analysis made in this subsection shows that approaching the theoretical values of the distribution yields to the increase of significance value γ and approaching its maximum of 100% when the sample and theoretical values coincide.

3.2.5 Expanding the Bivariate Mood Technique to Multivariate Case

The bivariate Mood method, designed in this article, may be easily expanded to the case of any dimensionality. This extension was not possible for a univariate Mood test, but including correlation ρ in the bivariate test allows to apply Mood approach on any two elements of the tested sample of random vectors.

The procedure is next. Assume that the considered sample $\{\mathbf{X}_i\}_{i=1}^n$ has $m > 2$ dimensions. Splitting dimensions into pairs brings each pair to the case of bivariate sample $\{\mathbf{X}_i^{k,l}\}_{i=1}^n$ with $k, l = 1, \ldots, m$, $k \neq l$. And each of these sub-samples may treated with a bivariate Mood test giving $C_2^m = \frac{m(m-1)}{2}$ tests to run in total. After evaluating all the pairs of dimensions null hypothesis $H_0 : (\mu, \Sigma) = (\mu_0, \Sigma_0)$ should be rejected if for all of them $H_0^{k,l} : (\mu^{k,l}, \Sigma^{k,l}) = (\mu_0^{k,l}, \Sigma_0^{k,l})$ is rejected. Otherwise null hypothesis may not be rejected.

Obtaining a technique that may determine whether the sample follows normal distribution with given parameters for any dimension size is quite valuable. Applications in panel models, e.g. testing for serial correlation, is one of the use-cases (see Part 2). Another example, following Arnold & Shavelle (1998), is construction of confidence region for functions of mean and variance, as $\mu + 2\sigma$, $\frac{\mu}{\sigma}$. These functions are widely used in modern portfolio theory (e.g. the Markowitz model, see Markowitz (1952)) and the introduced technique may be used to construct confidence regions that deliver significance levels for functions of μ and σ.

3.2.6 Transformation and Shaping Properties of Bivariate Mood Confidence Set

In this subsection I will describe how bivariate Mood confidence sets may be simplified with transformation, used in Part 1. Thereafter, I shall exhibit the shape and geometrical properties of the bivariate Mood method. Additionally, in this subsection I apply transformation to demonstrate the optimal choice of significance levels γ_i with respect to the volume of confidence set.

The transformation starts from the sample $\{\mathbf{X}_i\}_{i=1}^n$. In previous subsection I presented how this sample may be tested on whether it follows normal distribution with the given parameters $N(\mu_0, \Sigma_0)$. Using the transformation matrix V, defined in Part 1, equations (1.17)-(1.18), one can reduce the hypothesis tested (3.1) to the one with a simpler structure in (3.42). The right-hand side of the hypothesis is mean and covariance matrix of the transformed sample defined in (3.43). Note that firstly, multiplication by matrix V also eliminates correlation between dimensions in

transformed sample $\tilde{\mathbf{X}}$. As a result, matrix $\tilde{\Sigma}$ is diagonal. Secondly, sample means of $\{\tilde{\mathbf{X}}_i\}_{i=1}^n$ are zeros, as the sample was demeaned in (3.43).

$$H_0 : \left(\begin{pmatrix} \tilde{\mu}_x \\ \tilde{\mu}_y \end{pmatrix}, \begin{pmatrix} \tilde{\sigma_x}^2 & 0 \\ 0 & \tilde{\sigma_y}^2 \end{pmatrix} \right) \sim (0, I_2) \tag{3.42}$$

$$\begin{pmatrix} \tilde{\mathbf{X}}_x \\ \tilde{\mathbf{X}}_y \end{pmatrix} = V \left(\begin{pmatrix} \mathbf{X}_x \\ \mathbf{X}_y \end{pmatrix} - \begin{pmatrix} \overline{\mathbf{X}}_x \\ \overline{\mathbf{X}}_y \end{pmatrix} \right) \tag{3.43}$$

Matrix V is defined from the system (3.44), where the elements of this matrix are v_{ij}. Full analytical solution is described in detail in Part 1, equations (1.13)-(1.16) thus not reported here.

$$\begin{cases} \sum_{i=1}^2 \sum_{j=1}^2 v_{1i} v_{2j} S(i,j) = 0; \\ \sum_{i=1}^2 \sum_{j=1}^2 v_{1i} v_{2j} \Sigma(i,j) = 0; \\ \sum_{i=1}^2 \sum_{j=1}^2 v_{1i} v_{1j} \Sigma(i,j) = 1; \\ \sum_{i=1}^2 \sum_{j=1}^2 v_{2i} v_{2j} S(i,j) = 1 \end{cases} \tag{3.44}$$

The obtained matrix V, most importantly, provides the results of transformation that reduces number of variables by 1 by fixing $\rho = 0$. This fact is used to simplify pivots z, h_{11}, h_{22} and h_{12}. New functions are shown in (3.45) together with their distributions, obtained earlier.

$$\begin{aligned} nz &= \frac{\tilde{\mu_x}^2}{\tilde{\sigma_x}^2} + \frac{\tilde{\mu_y}^2}{\tilde{\sigma_y}^2} \sim \chi^2_2 \\ (n-1)h_{11} &= (n-1)\frac{\tilde{s_{11}}}{\tilde{\sigma_x}^2} \sim \chi^2_{n-1} \\ (n-1)h_{22} &= (n-1)\frac{\tilde{s_{22}}}{\tilde{\sigma_y}^2} \sim \chi^2_{n-1} \\ (n-1)h_{12} &= (n-1)\frac{\tilde{s_{12}}}{\tilde{\sigma_x}\tilde{\sigma_y}} \sim f_{h_{12}} \end{aligned} \tag{3.45}$$

The form of pivots in (3.45) is much simpler for analysis of confidence region in terms of $(\tilde{\mu_x}, \tilde{\mu_y}, \tilde{\sigma_x}, \tilde{\sigma_y})$. Using critical values t_{11}, t_{12}, t_{21}, t_{22} and t_3 of the distributions that pivots follow, one can obtain a system of inequalities (3.46) that identifies shape and volume of the confidence set, as well as the location of the tested points inside or outside of the region. Note that sigmas are positive, therefore the use of square roots does not lead to loss of the solutions or any of their parts. What's more, inequality for h_{12} was skipped, because sample covariance matrix is diagonal after the transformation, $\tilde{s_{12}}$ becomes zero and this inequality always holds.

$$\mathcal{R}(\tilde{\mathbf{X}}) = \begin{cases} \frac{\tilde{\mu_x}^2}{\tilde{\sigma_x}^2} + \frac{\tilde{\mu_y}^2}{\tilde{\sigma_y}^2} < \frac{t_3}{n} \\ \sqrt{\frac{(n-1)\tilde{s_{11}}}{t_{12}}} < \tilde{\sigma_x} < \sqrt{\frac{(n-1)\tilde{s_{11}}}{t_{11}}} \\ \sqrt{\frac{(n-1)\tilde{s_{22}}}{t_{22}}} < \tilde{\sigma_y} < \sqrt{\frac{(n-1)\tilde{s_{22}}}{t_{21}}} \end{cases} \tag{3.46}$$

The confidence set, defined by inequalities in (3.46) is a 4-D figure with rectangle in

$(\tilde{\sigma_x}, \tilde{\sigma_y})$ plane and ellipse in $(\tilde{\mu_x}, \tilde{\mu_y})$ plane. An example of such a confidence region is demonstrated in Figures 3.2 and 3.3. I generate a sample of 200 random vectors from bivariate normal distribution with means 4 and 5, unit variances and covariance equal 0.7, transform it and construct a confidence set for this sample at 1% significance level. Stars on the plots stand for the theoretical parameters after the transformation, from which vectors were sampled. Clearly seen from the graphs that these theoretical values fall into the constructed confidence set and therefore cannot be rejected.
Another important question that transformation helps to solve is the choice of individual significance levels γ_i. Till this point I assumed them to be the same $1 - \gamma_i = \sqrt[4]{1-\gamma}$. On the other hand, shares of each significance level do influence volume of the confidence set. At the same time transformation introduced in this subsection allows for straightforward calculation of the volume (4-D) in equation (3.47):

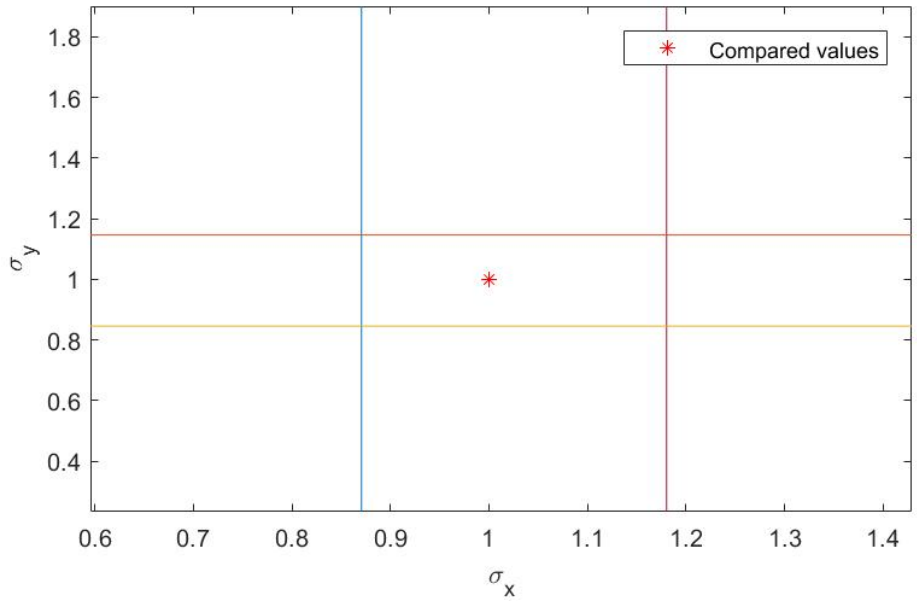

Figure 3.2: Projection of the Mood Confidence Region on (σ_x, σ_y)-Plane (Inner Rectangle) for Significance Level $\gamma = 1\%$ and Sample Size $n = 200$

$$\begin{aligned} V &= \iiiint_{\mathcal{R}(\tilde{\mathbf{X}})} d\tilde{\mu_x} d\tilde{\mu_y} d\tilde{\sigma_x} d\tilde{\sigma_y} \\ &= \frac{\pi(n-1)^2 \tilde{s_{11}} \tilde{s_{22}} t_3 (t_{12} - t_{11})(t_{22} - t_{21})}{4nt_{11}t_{12}t_{21}t_{22}} \end{aligned} \tag{3.47}$$

The optimal choice of γ_i is performed next by simple search on the grid and selecting the ratio of γ_i with the smallest generated volume. To mention, γ_2, γ_3 and γ_4 should be equal, as far as they are taken from the same Wishart distribution. Therefore, significance is distributed among γ_1 and all the other γ_i's. The result of the grid search is illustrated in Table 3.1. For each pair of sample sizes and total confidence level (first two columns of Table 3.1) I provide optimal significance γ_1 for pivot z and optimal significances γ_{2-4} for all 3 pivots from Wishart distribution, h_{11}, h_{12} and h_{22} in columns 3 and 4 of Table 3.1 respectively. Column 5 shows an aggregated significance level of Wishart pivots and column 5 is the volume of the constructed

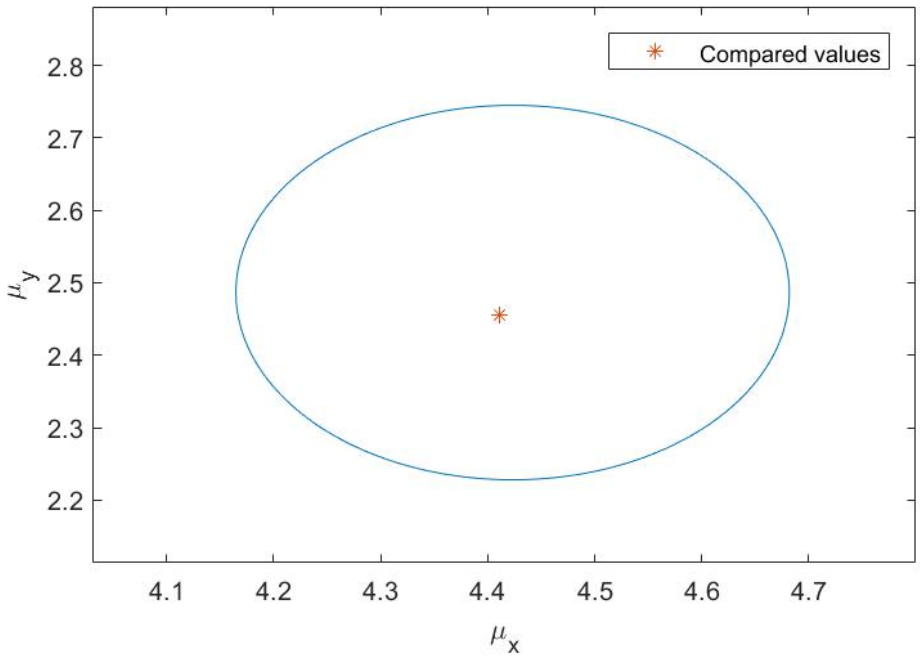

Figure 3.3: Projection of the Mood Confidence Region for Significance Level $\gamma = 1\%$ and Sample Size $n = 200$ on (μ_x, μ_y)-Plane

confidence region. The first thing to notice is the enormous size of the region for sample size $n = 10$. This may happen due to difficulties in estimation of variance from small samples. An increase of sample size quickly reduces confidence set, as already may be seen for $n = 50$.

The next and the most important point is the allocation of significance to gammas. When the sample size is small ($n = 10$), a huge share of significance is distributed to γ_{2-4}, producing optimal shares of significance that are close to equal. This supports the earlier choice $1 - \gamma_i = \sqrt[4]{1-\gamma}$. However, for larger sample sizes allocation of significances changes and optimal choice becomes close to $1 - \gamma_1 = (1 - \gamma_2)^3$. Hence, the proportion of significance captured by Hotelling t^2 and Wishart part becomes the same. Columns 3 and 5 provide close results starting from $n = 50$ and justify this allocation of significance. The fact that optimal shares of gamma change with increasing sample size may be explained with the distributions pivots follow (see (3.45)). While function z follows χ^2 distribution with fixed degrees of freedom, h_{11} and h_{22} follow χ^2 distribution that increases its degrees of freedom with increasing sample size n, getting thinner tails and approaching normal distribution.

Therefore, to obtain minimum-volume bivariate Mood confidence set for sample size $n < 50$, according to the analysis, set equal gammas $\gamma_i = 1 - \sqrt[4]{1-\gamma}$. With $n \geq 50$ minimum-volume significance levels are $\gamma_1 = 1 - \sqrt{1-\gamma}$ and $\gamma_{2-4} = 1 - \sqrt[6]{1-\gamma}$.

Concluding this subsection, transformation applied to the bivariate Mood method may be very useful as a simplification instrument. Using it, the shape and volume of confidence sets were demonstrated and optimal allocation of individual significances γ_i was found. Application of this transformation technique also allows one to avoid the distribution of sample covariance $f_{h_{12}}$ (3.34) in construction of confidence set, which is quite resource-intensive for large sample sizes because of the modified Bessel

function in its functional form.

Table 3.1: Allocations of Individual Significance Levels γ_1 and γ_{2-4} to Produce the Smallest Mood Confidence Set for Sample Sizes n from 10 to 1000 and Overall Confidence Level $1-\gamma$ from 0.8 to 0.99

n	$1-\gamma$	γ_1	γ_{2-4}	$1-(1-\gamma_2)^3$	Volume
	0.80	0.0720	0.0483	0.1379	3.5060
	0.90	0.0320	0.0240	0.0702	7.3343
10	0.95	0.0140	0.0123	0.0365	13.8889
	0.975	0.0065	0.0062	0.0186	24.2495
	0.99	0.0024	0.0025	0.0076	45.9204
	0.80	0.1040	0.0371	0.1071	0.0593
	0.90	0.0480	0.0185	0.0546	0.1041
50	0.95	0.0230	0.0093	0.0276	0.1652
	0.975	0.0110	0.0047	0.0142	0.2427
	0.99	0.0042	0.0019	0.0058	0.3697
	0.80	0.1080	0.0356	0.1031	0.0136
	0.90	0.0500	0.0179	0.0526	0.0236
100	0.95	0.0240	0.0090	0.0266	0.0363
	0.975	0.0120	0.0044	0.0132	0.0522
	0.99	0.0046	0.0018	0.0054	0.0778
	0.80	0.1120	0.0342	0.0991	0.0005
	0.90	0.0520	0.0172	0.0506	0.0009
500	0.95	0.0260	0.0083	0.0246	0.0013
	0.975	0.0125	0.0042	0.0127	0.0019
	0.99	0.0050	0.0017	0.0050	0.0027
	0.80	0.1120	0.0342	0.0991	0.0001
	0.90	0.0540	0.0165	0.0486	0.0002
1000	0.95	0.0260	0.0083	0.0246	0.0003
	0.975	0.0125	0.0042	0.0127	0.0005
	0.99	0.0050	0.0017	0.0050	0.0007

3.3 Comparison of Bivariate Mood and CCR Confidence Regions

The CCR technique, introduced in Part 1 together with the bivariate Mood approach, derived in this Part, are aimed to test the same joint hypothesis for mean and variance of two-dimensional random sample. Assumptions for both methods are also the same, requiring the sample to be iid from Normal distribution. Therefore, on one hand these tests look same from the point of view of a person, who wants to apply them to a specific problem. On the other hand, these tests differ in concept of construction of confidence region. While the bivariate Mood test is an exact one and pivots follow given distributions, CCR test is not. It can only get closer to the exact distribution with a predefined precision. Thus, to give the possibility of objective choice, in this section I analyze the CCR and Mood approaches with respect to their efficiency of calculation speed, accuracy of significance level, shape of the confidence set and robustness to non-normal samples and samples with outliers.

3.3.1 Significance Level

One of the most important qualities of the test is to reject hypothesis with the probability that meets the predefined significance level. Because of simulated data used, the theoretical mean and covariance matrix are known. Therefore, it is possible to check if the confidence region includes this point and calculate a fraction of samples rejected to the total number of repetitions. Predefined (theoretical) significance level should correspond with the calculated (real) one. To check how good theoretical and real significances coincide, I replicated 100,000 samples for each of the different sample sizes n from 10 to 50000 from standard bivariate normal distribution. For each of the samples p-value, based on the known mean and covariance, was calculated and compared with significance level γ. The fraction of samples with p-value$< \gamma$ to the total number of replications defines real significance level. The resulted shares for the Mood and CCR techniques are displayed in Table 3.2. More detailed tables may be found in Appendix 3.I.

Table 3.2: Comparison of Theoretical Significance Levels with the Real Ones Obtained from the Mood and CCR Approaches for Sample Sizes n from 10 to 1000 and Overall Significance Level γ from 0.01 to 0.2

n	γ	CCR γ_{real}	Mood γ_{real}
10	0.20	0.2050	0.1819
	0.10	0.0979	0.0881
	0.05	0.0526	0.0453
	0.025	0.0268	0.0225
	0.01	0.0091	0.0097
50	0.20	0.2001	0.1897
	0.10	0.0982	0.0957
	0.05	0.0505	0.0498
	0.025	0.0260	0.0258
	0.01	0.0116	0.0123
100	0.20	0.2030	0.1992
	0.1	0.0996	0.0974
	0.05	0.0512	0.0490
	0.025	0.0258	0.0246
	0.01	0.0107	0.0102
500	0.20	0.1961	0.1954
	0.10	0.0963	0.1027
	0.05	0.0457	0.0554
	0.025	0.0249	0.0266
	0.01	0.0100	0.0114
1000	0.20	0.1986	0.2030
	0.10	0.0996	0.1010
	0.05	0.0504	0.0537
	0.025	0.0257	0.0253
	0.01	0.0104	0.0106

Both of the methods give real significance levels that are pretty close to the theoretical ones. Even for small sample sizes, both the CCR and Mood techniques are still rejecting correctly. For sample size $n = 10$ the largest deviation of the real significance from theoretical one is a little bit less than 2%, which is a reasonable fluctuation. For

the large sample sizes the fluctuation decreases, e.g. when $n = 500$ the maximum difference from the theoretical significance is 0.54%, made by the Mood approach for $\gamma = 0.05$. However, on larger significance levels one can observe that the Mood approach systematically rejects a little bit less samples than the CCR. Figure 3.4 clearly demonstrates this phenomenon. In the upper right corner the CCR and Mood graphs almost coincide, while starting from roughly $1 - \gamma < 0.8$ Mood accepts distinctly more samples than the CCR method. This fact may be explained by the form of the confidence set of Mood method, that covers regions with lower probability density. It will be discussed in more detailed in the next subsection. Nevertheless, this fact does not play a significant role for most of the applications, since they work with small significance levels. In the case of extremely small significance level required ($\gamma < 0.0001$) I would propose to use the Mood method, because the significance level of the CCR technique in this case may get close to the over mentioned precision of CCR and bring additional noise to the test.

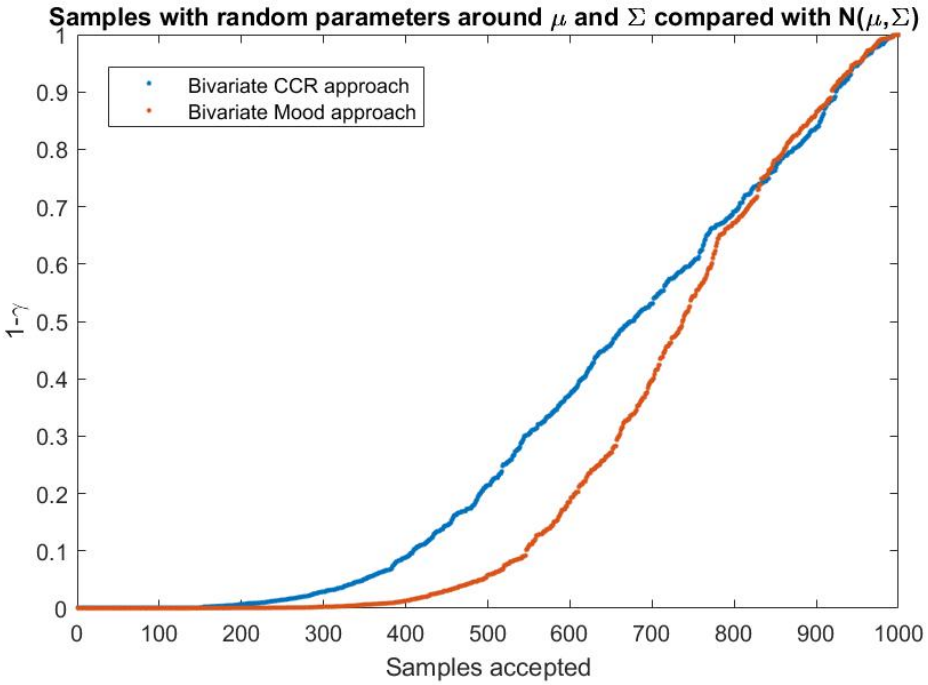

Figure 3.4: Acceptance of 1000 Randomly Generated Samples from the Mood and CCR Approaches as a Function of Confidence Level $1 - \gamma$

3.3.2 Shape and Volume

Earlier in equation (3.39) I constructed a bivariate Mood confidence set for a random normally distributed iid sample $\{\mathbf{X}_i\}_{i=1}^n$. This set represents a collection of vectors $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ in 5-dimensional space, so that every vector may not be rejected as a set of parameters of normal distribution, from which the sample $\{\mathbf{X}_i\}_{i=1}^n$ was generated. In this section I will describe the Mood confidence set and compare it to the CCR confidence set for both equal and optimal individual gammas w.r.t. the volume of the region. The first point is a form of the set, considered in Proposition 1.

Proposition 1: Mood confidence region $\mathcal{R}(\mathbf{X})$ is a convex set.

Proof: Assume that confidence set $\mathcal{R}(\mathbf{X})$ is not convex. It means that there exists 2 points $A_1(\mu_{x1}, mu_{y1}, \sigma_{x1}, \sigma_{y1}, \rho_1) \in \mathcal{R}(\mathbf{X})$ and $A_2(\mu_{x2}, mu_{y2}, \sigma_{x2}, \sigma_{y2}, \rho_2) \in \mathcal{R}(\mathbf{X})$, such that on the interval A_1A_2 one can find a point $B = \alpha A_1 + (1-\alpha)A_2$, $\alpha \in (0,1)$, that does not belong to confidence set $B \notin \mathcal{R}(\mathbf{X})$. Therefore, at least one of the pivots z, h_{11}, h_{22}, h_{12} is rejected at point B, since it lies outside the of critical values interval. Say pivot h_{22} is rejected. But this pivot at points $h_{22}(A_1)$ and $h_{22}(A_2)$ may not be rejected, as A_1 and A_2 are from the confidence set. As far as critical values define a continuous set, all the points between $h_{22}(A_1)$ and $h_{22}(A_2)$, among which is point $h_{22}(B)$ also belong to the confidence set. This gives a contradiction. Hence, for any two points of the confidence set $\mathcal{R}(\mathbf{X})$ their line segment also belongs to $\mathcal{R}(\mathbf{X})$, which is the definition of convex set.■

Proposition 1 defines the Mood confidence set as a convex one, which is confirmed by Figures 3.2 and 3.3. The transformation made in previous section showed the simpler form of confidence set, but gave a good starting point for understanding the shape of the primary confidence set. It will be a trapezoidal form in variances plane with ellipse in means plane for every variance point.

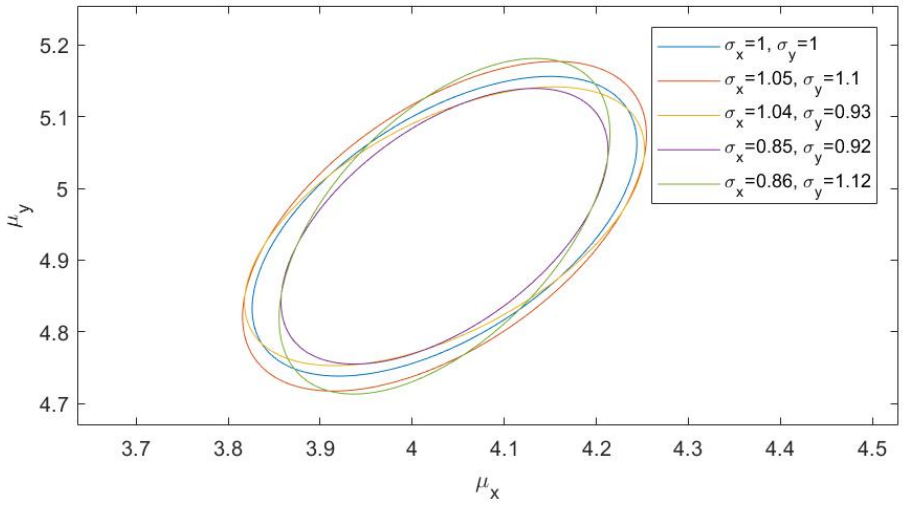

Figure 3.5: Mood Confidence Sets, Projected on (μ_x, μ_y)-Plane for Different Values of the Variances Entering the Same Confidence Set for $\gamma = 0.1$ and $n = 500$

Figure 3.5 demonstrates those ellipses for 5 different values of variances from the confidence region. Figure 3.5 also indicates the benefit from testing means and variances jointly: ellipses have a large common subset, but each one determines new points that may not be rejected and throws away other points that should be rejected. Even for small fluctuations of variances inside of the confidence set, one obtains new specific confidence set for means.

To compare further Mood with CCR confidence region, I imposed them one on another. Note that CCR approach produces ellipsoidal set, while Mood gives a cylindrical set in 5-D. See Figure 3.6 for means and Figure 3.7 for variances. Sample parameters are marked with center circles on the graphs.

Figure 3.6 shows that both methods produce ellipsoidal confidence sets for means and for the CCR method it is slightly larger. However, as seen on Figure 3.7, CCR

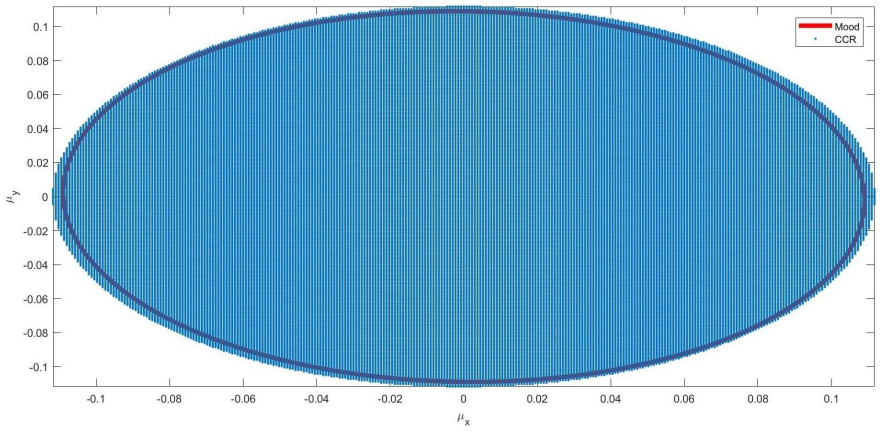

Figure 3.6: Transformed Mood (Bounded by Solid Line) and CCR (Dotted Region) Confidence Sets Projected on (μ_x, μ_y)-Plane for $\gamma = 0.1$ and $n = 500$

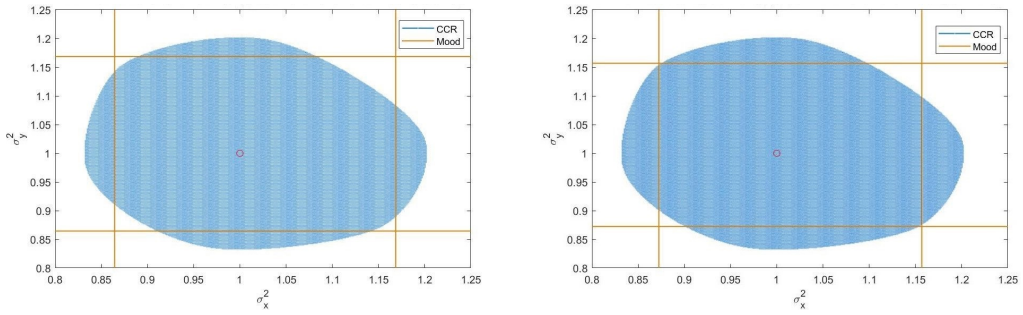

Figure 3.7: Transformed Mood (Inner Square) and CCR (Dotted Region) Confidence Sets Projected on (σ_x^2, σ_y^2)-Plane for $\gamma = 0.1$ and $n = 500$ (Left - Equal γ_i, Right - Optimal γ_i)

and Mood approaches definitely produce different confidence sets for variances. The CCR technique cuts the edges of the Mood confidence set, that are less likely to occur, because they are further from the center of the region, marked with a red circle. Due to the cutting edges of the Mood confidence set, the CCR approach can add additional subsets from up, down, left and right that are more likely to appear. Firstly, this infers that the CCR confidence set is more effective w.r.t. the shape of the confidence set, unless it is not the exact one. Secondly, it infers why the Mood method systematically rejected more samples than the CCR, as it was demonstrated on Figure 3.4.

Additionally, the left and right graphs on Figure 3.7 show how the Mood confidence set may be reduced by using optimal γ_i, derived in section 3.2.6. It makes the corners slightly smaller, but does not get rid of them. Note that at the same time transferring from the left to right graph increases the confidence set for means to keep the same significance level. It implies that choice of γ_i has only an effect on the volume of confidence set, but insignificant for total γ, and in applications one can simply use $\gamma_i = 1 - \sqrt[4]{1-\gamma}$.

3.3.3 Computation Speed

The time that computers spend on running a certain algorithm decreases each year with exponential speed. On the contrary, in the last decades analyses of big data

problems arises, with tremendously increasing datasets. Therefore, it is logical to compare the time that was spent checking the null hypothesis (3.1) with both the Mood and CCR techniques. My analysis includes time that computer[3] calculated p-values for 100 replications of a randomly generated samples of different sizes. The computation time will differ from computer to computer, but the ratios between techniques will be invariant. The results (in seconds) are presented in Table 3.3.

Table 3.3: Computation Time from 100 Repetitions of the Mood and CCR Tests

n	10	50	100	500	1000
CCR (sec)	27,177	26,385	27,737	31,928	25,580
Mood (sec)	0,403	0,659	0,732	3,336	3,244

The first thing to be noticed from Table 3.3 is that the Mood approach computes around 50 times faster than the CRR for small sample sizes $n < 500$. This fact is mainly due to double integral that the CCR technique needs to compute every time. One more factor that slows down the CCR method is transformation that is also performed each time.
The second thing to mention from Table 3.3 is the increase of computation time for the Mood method when sample size $n > 500$. Increase in sample size makes more effective to use bivariate Mood method with transformation, which I do in Table 3.3. This is due to Bessel function that is calculated for pivot h_{12}. Even with transformation the Mood approach still computes around 10 times faster than CCR.

3.3.4 Robustness

In this section I compare the Mood and CCR methods with respect to their stability properties. Robustness checks are done in two parts: reaction on non-normal samples and reaction on samples with outliers.
Non-normal samples were generated from two distributions. Bivariate t-distribution is the first one, as it is the one that is close to normal. It has thicker tails, but converges to normal distribution with increasing degrees of freedom. The second distribution used for generating samples is bivariate Poisson. This is a discrete distribution, but it also has a bell-shaped form of PDF. Among these distributions I also include crucial cases: t distribution with 3 degrees of freedom (extremely thick tails) and Poisson with parameter $\lambda = 3$ (skewed). Table 3.4 demonstrates real significance levels, calculated on 100000 repetitions for the Mood and CCR approaches applied on the non-normal samples. The sample size is fixed and equals to 250.
As expected, Table 3.4 shows too high real significance levels for t-distribution with 3 degrees of freedom (columns 2 and 3). Too thick tails force both techniques to reject high number of samples. Even for t-distribution with 50 degrees of freedom (columns 4-5) CCR still rejects more than 30% of the samples instead of 20%. But the Mood approach is much better in this case; rejecting only 2.2% more of samples, thus making it usable for t_{50}-distribution. Columns 6-7 of Table 3.4, with deviation from the real significance level of around 1%, show results that are as good as for normally distributed samples. The last 4 columns demonstrate reactions of both approaches

[3]Computer specifications: CPU Intel Core i5-6200U with 2.3 Ghz and 8 GB RAM type DDR4

Table 3.4: Real Significance Levels Produced by the Mood and CCR Approaches for the Samples from t-Distribution and Poisson Distribution with Different Number of Degrees of Freedom

γ	t_3 distr.		t_{50} distr.		t_{100} distr.		$Poiss(3)$ distr.		$Poiss(10)$ distr.	
	CCR	Mood	CCR	Mood	CCR	Mood	CCR	Mood	CCR	Mood
0.200	0.889	0.892	0.312	0.222	0.212	0.209	0.235	0.232	0.217	0.214
0.100	0.833	0.834	0.115	0.119	0.109	0.106	0.127	0.125	0.110	0.110
0.050	0.779	0.776	0.057	0.061	0.051	0.054	0.073	0.068	0.055	0.057
0.025	0.726	0.713	0.030	0.031	0.024	0.027	0.041	0.033	0.028	0.028
0.010	0.661	0.649	0.012	0.014	0.011	0.012	0.018	0.016	0.010	0.011

on Poisson distributed samples. Even with parameter $\lambda = 3$, both distributions show reasonably good results with deviation that does not exceed 3.5%, although the samples are generated from skewed distribution. With parameter λ increased to 10, skewness almost disappears and real significance levels become comparable to those from the normal distribution (see Table 3.2).

For comparison of the CCR and Mood techniques on samples with outliers, I generated normally distributed samples and randomly added fluctuations to part of the elements of the sample. This gave samples with 5%, 10% and 20% of outliers. Table 3.5 presents real significance levels calculated on such samples. Adding fluctuations was randomly assigned to elements of the sample for not to provoke serial correlations in data.

Table 3.5: Real Significance Levels Produced by the Mood and CCR Approaches from the Samples with Outliers

γ	5% outliers		10% outliers		20% outliers	
	CCR	Mood	CCR	Mood	CCR	Mood
0.200	0.201	0.193	0.259	0.253	0.731	0.662
0.100	0.095	0.098	0.143	0.138	0.592	0.506
0.050	0.046	0.053	0.076	0.076	0.457	0.375
0.025	0.021	0.026	0.041	0.041	0.344	0.275
0.010	0.013	0.009	0.017	0.019	0.232	0.180

Including random outliers to samples may seriously harm real significance levels. Already 20% of outliers make it impossible to use both the Mood and CCR techniques, as illustrated in the last 2 columns of Table 3.5. But with 5% of outliers, both methods still return real significance levels that differ from the theoretical ones by less than 1%(columns 2-3).

To conclude this subsection, both Mood and CCR showed satisfactory results in robustness checks with slightly better stability in the Mood approach. But still both tests may be used on samples close to normal and with a moderate number of outliers (5%-10%).

3.4 Applications

The application of the bivariate Mood technique includes estimation, testing hypothesis and construction of confidence regions. An extension of the Mood method to a multivariate case provides new scopes of application. For example, an approach for

testing for serial correlation in Panel Data Models, designed in Part 2 is based on the CCR technique and may be also used with the Mood method. According to a comparison made in the previous section, the Mood approach performs better than CCR with respect to the computation speed. Faster calculations may be crucial for large datasets, as the number of tests to be performed growth exponentially with number of time periods in panel.

Another application of the multivariate Mood technique is a SUR (seemingly unrelated regressions) model. I simulated a SUR model in MatLab with 5 equations. Each equation includes 5 unique regressors. Next, 300 observations were generated for each regressor in each equation. And using regressors, dependent variables were calculated for every case. Additionally, I randomly included correlations between error terms of the equations. Starting from OLS estimation, Table 3.6 displays estimated covariance matrix $\hat{\Omega}$. Estimated covariance matrix describes cross-equation connections and based on it I run a Mood test on each pair of equations. With Mood test I compare whether the error terms of each two equations may follow normal distribution with zero correlation. Obtained result is presented in Table 3.7. At 10% significance level 3 pairs are rejected: equation 3 and 4 with p-value 9.18%, equations 3 and 5 with p-value 1.1%, equations 4 and 5 with p-value 4.62%. Therefore, for all 3 mentioned pairs covariance is significantly different from zero, and considered SUR system of 5 equations has covariance matrix that shows cross-equation correlations. Further this system may be estimated by parts, as far as first two equations have insignificant correlations with the other equations. Thus, they form an independent subsystem, which may be estimated separately with OLS, while last three equations should be estimated with method that captures correlations between equations, for example feasible generalized least squares (FGLS). Computing described SUR model separately for parts with and without correlation between equations helped to improve estimators by roughly 20%, comparing to FGLS estimation of the whole SUR model.

Table 3.6: Estimated Covariance Matrix $\hat{\Omega}$ from the SUR Model to be Used in the Mood Method

	1	2	3	4	5
1	0.8766	0.1025	0.1139	0.0249	-0.0023
2	0.1025	1.1323	0.0186	-0.0265	-0.0537
3	0.1139	0.0186	0.9211	0.1206	0.1754
4	0.0249	-0.0265	0.1206	0.9214	0.1473
5	-0.0023	-0.0537	0.1754	0.1473	1.1035

Table 3.7: P-Values for each Pair of Equations Obtained Using the Estimated Matrix $\hat{\Omega}$ with the Mood Method

	1	2	3	4	5
1	0	0.2685	0.1099	0.9814	1.0000
2	0	0	0.9963	0.9856	0.8752
3	0	0	0	0.0918*	0.0110**
4	0	0	0	0	0.0462**
5	0	0	0	0	0

* p<.1; ** p<.05; *** p<.001

3.5 Conclusion

The bivariate Mood method, derived in this Part, is aimed to jointly test means and variances and to construct confidence set for samples from bivariate normal distribution. This is an exact test, based on linking Hotelling t^2 and Wishart distribution. Test statistics, derived for the Mood test, were examined for limiting properties to check the convergence of method. Using derived joint statistic, confidence region was arranged, examined and optimized with respect to its volume. It appeared that for small sample sizes minimum-volume confidence set is obtained by equalizing all individual gammas: $\gamma_i = 1 - \sqrt[4]{1-\gamma}$. But for larger sample sizes $n > 50$ equal shares of significance for t^2 and Wishart part produce minimum-volume confidence set: $\gamma_1 = 1 - \sqrt{1-\gamma}$ and $\gamma_{2-4} = 1 - \sqrt[6]{1-\gamma}$. Further analysis of the shape of confidence region showed that the Mood test can additionally be used for improvement of confidence sets constructed for means only. Furthermore, the Mood approach was expanded to multivariate case, allowing one to test samples of any dimensionality.

In the next section Mood was also compared with an existing technique, called CCR. The results showed that on the one hand, the Mood approach is much more effective than the CCR with respect to computation time as well as performing slightly better with non-normal samples and samples with outliers. On the other hand, because of the angular shape of the confidence region, the Mood method slightly overestimates real significance level for theoretical significance $\gamma > 20\%$. As a result, the Mood approach was demonstrated to be a good alternative to the CCR test, outperforming it in terms of computation time.

Application of the Mood technique was demonstrated on SUR model. Because of detected cross-equation relations with the Mood method, the whole system was divided into 2 parts and estimated using OLS and FGLS estimators. Implementation of this procedure helped to improve the estimation of the SUR model by roughly 20%.

Further development and application of the bivariate Mood approach are possible in any brunch of economics, where estimation and testing is used. For example, portfolio analysis or panel econometrics.

References

Arnold, B., Shavelle, R., 1998. Joint confidence sets for the mean and variance of a Normal distribution. *The American Statistician*; 52; 133-140.

Born, B., Breitung, J., 2016. Testing for serial correlation in fixed effects panel data models. *Econometric Reviews*; 35; 1290-1316.

Frey, J., Marrero, O., Norton, D., 2009. Minimum-area confidence sets for a Normal distribution. *Journal of Statistical Planning and Inference*; 139; 1023-1032.

Greene, W., 2002. Econometric Analysis. 5th Edition. Prentice Hall; New Jersey; 2002.

Godfrey, L., 1978. Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*; 46; 1293–1301.

Gupta, A., Nagar, D., 2000. Matrix Variate Distributions. Chapman & Hall; London; 2000.

Gupta, A., Bodnar, T., 2014. An exact test about the covariance matrix. *Journal of Multivariate Analysis*; 125; 176-189.

Horn, R., Johnson, C., 2013. Matrix Analysis. Cambridge University Press; Cambridge; 2013.

Hotelling, H., 1931. The generalization of Student's ratio. *Annals of Mathematical Statistics*; 2; 360–378.

Jensen, U., 1995. A review of the derivation and calculation of Rao distances with an application to portfolio theory. In Maddala, Phillips, Srinivasan (eds.), Advances in econometrics and quantitative economics. Blackwell; 2008; 413-462.

Jochmans, K., 2019. Testing for correlation in error-component models. *Cambridge Working Papers in Economics No. 1993*; University of Cambridge; Faculty of Economics, 2019.

Johnson, R., Wichern, D., 2002. Applied multivariate statistical analysis. Prentice hall; New Jersey; 2002.

Kshirsagar, A., 1959. Bartlett Decomposition and Wishart Distribution. *Ann. Math. Statist.*; 30; 239-241.

Levinger, B., 1980. The Square Root of a 2×2 Matrix. *Mathematics Magazine*; 53; 222-224.

Ljung, G., Box, G., 1978. On a measure of lack of fit in time series models. *Biometrika*; 65(2); 297–303.

Markowitz, H., 1952. Portfolio Selection. *The Journal of Finance*; 7(1); 77-91.

Mood, A., 1950. Introduction to the Theory of Statistics. McGraw-Hill; New York; 1950.

Muirhead, R., 1982. Aspects of Multivariate Statistical Theory. Wiley; New York; 1982.

Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data, second edition. The MIT Press; Cambridge, Massachusetts; 2010.

Appendix 3.I: Tables

Table 3.8: Theoretical Significance Levels γ vs. Real Significance Levels for the Mood and CCR Approaches. Sample Sizes $n = 10, 25, 50$

γ	CCR	Mood	CCR	Mood	CCR	Mood
	n=10		n=25		n=50	
0.99	0.9915	0.9913	0.991	0.9901	0.9903	0.9915
0.95	0.9505	0.9493	0.9522	0.9516	0.9543	0.9536
0.9	0.8991	0.9009	0.9034	0.8999	0.9019	0.9027
0.75	0.7516	0.7489	0.7535	0.7454	0.7502	0.7484
0.5	0.5114	0.4882	0.5018	0.4903	0.499	0.494
0.3	0.3106	0.2816	0.2981	0.294	0.3028	0.2888
0.2	0.205	0.1819	0.1987	0.1982	0.2001	0.1897
0.1	0.0979	0.0881	0.1033	0.0986	0.0982	0.0957
0.05	0.0526	0.0453	0.0537	0.0499	0.0505	0.0498
0.025	0.0268	0.0225	0.0279	0.0255	0.026	0.0258
0.01	0.0091	0.0097	0.0105	0.0106	0.0116	0.0123

Table 3.9: Theoretical Significance Levels γ vs. Real Significance Levels for the Mood and CCR Approaches. Sample Sizes $n = 100, 150, 200$

γ	CCR	Mood	CCR	Mood	CCR	Mood
	n=100		n=150		n=200	
0.99	0.9918	0.9907	0.9883	0.9886	0.9905	0.991
0.95	0.9521	0.954	0.9458	0.9449	0.9495	0.95
0.9	0.9005	0.9013	0.8913	0.8946	0.8997	0.9024
0.75	0.7486	0.7477	0.7408	0.7487	0.7487	0.7477
0.5	0.5012	0.5001	0.4796	0.5011	0.4991	0.4984
0.3	0.3081	0.2982	0.2834	0.3007	0.2922	0.2906
0.2	0.203	0.1992	0.1887	0.2014	0.1944	0.1942
0.1	0.0996	0.0974	0.0919	0.0993	0.0981	0.0967
0.05	0.0512	0.049	0.0449	0.0496	0.049	0.0479
0.025	0.0258	0.0246	0.0204	0.0241	0.0233	0.0251
0.01	0.0107	0.0102	0.0089	0.0098	0.0102	0.0104

Table 3.10: Theoretical Significance Levels γ vs. Real Significance Levels for the Mood and CCR Approaches. Sample Sizes $n = 250, 300, 350$

γ	CCR	Mood	CCR	Mood	CCR	Mood
	n=250		n=300		n=350	
0.99	0.9909	0.9905	0.9909	0.9901	0.9899	0.9916
0.95	0.95	0.9466	0.9488	0.9505	0.9488	0.9505
0.9	0.8982	0.8963	0.897	0.9013	0.9046	0.8996
0.75	0.7454	0.7518	0.7495	0.7463	0.7567	0.7500
0.5	0.4995	0.497	0.4955	0.4994	0.5114	0.4996
0.3	0.2999	0.3024	0.2993	0.3046	0.308	0.2975
0.2	0.1998	0.2014	0.2024	0.2034	0.2074	0.1996
0.1	0.1002	0.1027	0.0996	0.1052	0.1023	0.0988
0.05	0.0515	0.0554	0.0515	0.0533	0.0495	0.0486
0.025	0.0244	0.0266	0.0279	0.0286	0.0242	0.0233
0.01	0.0099	0.0114	0.011	0.0108	0.0096	0.0092

Table 3.11: Theoretical Significance Levels γ vs. Real Significance Levels for the Mood and CCR Approaches. Sample Sizes $n = 400, 450, 500$

γ	CCR	Mood	CCR	Mood	CCR	Mood
	n=400		n=450		n=500	
0.99	0.9901	0.9873	0.9905	0.9893	0.9905	0.9884
0.95	0.9484	0.9513	0.9539	0.9496	0.9499	0.9473
0.9	0.8989	0.9007	0.9000	0.8989	0.8974	0.8983
0.75	0.7467	0.7500	0.7509	0.7486	0.7558	0.7504
0.5	0.4972	0.5000	0.5096	0.5032	0.4995	0.5005
0.3	0.3023	0.2989	0.3096	0.3021	0.2920	0.3005
0.2	0.2018	0.2014	0.2073	0.2004	0.1961	0.1954
0.1	0.1044	0.0998	0.1027	0.0984	0.0963	0.1027
0.05	0.0501	0.0490	0.0531	0.0489	0.0457	0.0554
0.025	0.0259	0.0268	0.0272	0.0248	0.0249	0.0266
0.01	0.0107	0.0097	0.0111	0.0110	0.0100	0.0114

Table 3.12: Theoretical Significance Levels γ vs. Real Significance Levels for the Mood and CCR Approaches. Sample Sizes $n = 750, 1000, 2000$

γ	CCR	Mood	CCR	Mood	CCR	Mood
	n=750		n=1000		n=2000	
0.99	0.9901	0.9905	0.9885	0.9897	0.9893	0.9903
0.95	0.9515	0.9491	0.9507	0.9500	0.9507	0.9501
0.9	0.9072	0.8996	0.9001	0.9001	0.8957	0.9005
0.75	0.7646	0.7506	0.7551	0.7509	0.7463	0.7503
0.5	0.5321	0.5012	0.5003	0.5016	0.5030	0.4991
0.3	0.3295	0.2988	0.2978	0.3005	0.2961	0.2998
0.2	0.2249	0.2014	0.1986	0.2030	0.1972	0.1999
0.1	0.1203	0.1007	0.0996	0.1010	0.0991	0.0995
0.05	0.0614	0.0499	0.0504	0.0537	0.0529	0.0517
0.025	0.0329	0.0248	0.0257	0.0253	0.0265	0.0241
0.01	0.0129	0.0098	0.0104	0.0106	0.0099	0.0095

Table 3.13: Theoretical Significance Levels γ vs. Real Significance Levels for the Mood and CCR Approaches. Sample Sizes $n = 5000$ to 50000

γ	CCR	Mood	CCR	Mood	CCR	Mood	CCR	Mood
	n=5000		n=10000		n=20000		n=50000	
0.99	0.9902	0.9893	0.9889	0.9896	0.9885	0.9900	0.9913	0.9913
0.95	0.9497	0.9489	0.9501	0.9499	0.9474	0.9474	0.9521	0.9501
0.9	0.8938	0.8998	0.9017	0.9003	0.9027	0.8996	0.9011	0.9000
0.75	0.7445	0.7497	0.7458	0.7498	0.7558	0.7511	0.7478	0.7494
0.5	0.4968	0.5015	0.4883	0.5004	0.5100	0.4991	0.4991	0.5000
0.3	0.2915	0.2998	0.2937	0.3004	0.3065	0.3008	0.3003	0.3002
0.2	0.1962	0.1990	0.1961	0.1989	0.2028	0.2003	0.1991	0.2004
0.1	0.0953	0.1015	0.0993	0.0991	0.1085	0.1000	0.1001	0.0997
0.05	0.0480	0.0512	0.0480	0.0493	0.0516	0.0502	0.0476	0.0498
0.025	0.0231	0.0248	0.0233	0.0245	0.0248	0.0236	0.0257	0.0267
0.01	0.0087	0.0086	0.0091	0.0097	0.0079	0.0099	0.0093	0.0081

Appendix 3.II: MatLab Code

Main subroutine:

```
function p = Mood_mult(meanX, varX, mu, C, n)

% function that performs multivariate Mood test with H0:
% (muX,SigmaX)=(mu,C) for X - sample of 2-dimentional random vectors
% second part - sample mean and variance of X as an input (to coinside
% with CCR function)
% this is part 2

% input:
% meanX - sample mean of X
% varX - sample variance-covariance Matrix of X
% mu - (2*1) vector of means which we want to compare with sample means
% C - (2*2) variance matrix (symmetrical and positive definite) which
    we
% want to compare with sample variance
% n - sample size

if C(1,2)~=C(2,1) || det(C)<0 || C(1,1)<0 || C(2,2)<0
    error('incorrect data');
end


if n<=300 % to avoid modified Bessel function for large n,
    transformation is used for n>300

    % define statistics
    % define averages
    x_av = meanX(1);
    y_av = meanX(2);

    % define estiamted variance as statistic
    S1 = varX;

    % define values that are tested as a scalars
    mux = mu(1);
    muy = mu(2);
    sigmax = sqrt(C(1,1));
    sigmay = sqrt(C(2,2));
    rho = C(1,2)/(sigmax*sigmay);

    % define functions of statistic that we use to run test

    h11 = (S1(1,1)*(sigmay + sqrt(1-rho^2)*sigmax)^2 - 2*S1(1,2)*rho*
        sigmax*(sigmay + sqrt(1-rho^2)*sigmax) + S1(2,2)*rho^2*sigmax^2)
        /...
        ((1-rho^2)*sigmax^2*(sigmax^2 + sigmay^2 + 2*sigmax*sigmay*sqrt
            (1-rho^2)));

    h12 = (S1(1,2)*rho^2*sigmax*sigmay - S1(1,1)*rho*sigmay*(sigmay +
        sqrt(1-rho^2)*sigmax) + S1(1,2)*(sigmay + sqrt(1-rho^2)*sigmax)
        *(sigmax + sqrt(1-rho^2)*sigmay) - S1(2,2)*rho*sigmax*(sigmax +
```

```
        sqrt(1-rho^2)*sigmay))/...
45          ((1-rho^2)*sigmax*sigmay*(sigmax^2 + sigmay^2 + 2*sigmax*sigmay
                *sqrt(1-rho^2)));
46
47      h22 = (S1(1,1)*rho^2*sigmay^2 - 2*S1(1,2)*rho*sigmay*(sigmax + sqrt
            (1-rho^2)*sigmay) + S1(2,2)*(sigmax + sqrt(1-rho^2)*sigmay)^2)
            /...
48          ((1-rho^2)*sigmay^2*(sigmax^2 + sigmay^2 + 2*sigmax*sigmay*sqrt
                (1-rho^2)));
49
50      z = (sigmay^2*(mux - x_av)^2 + sigmax^2*(muy - y_av)^2 - 2*rho*
            sigmax*sigmay*(mux - x_av)*(muy - y_av))/...
51          ((1-rho^2)*sigmax^2*sigmay^2);
52
53      % calculate p values for each function defined previously
54
55      p1 = 1 - chi2cdf((n-1)*z,2);
56
57      p2 = 2*min(chi2cdf((n-1)*h11,n-1),1 - chi2cdf((n-1)*h11,n-1));
58
59      p3 = 2*min(chi2cdf((n-1)*h22,n-1),1 - chi2cdf((n-1)*h22,n-1));
60
61      p4 = 2*min(cov_cdf((n-1)*h12,n-1), 1 - cov_cdf((n-1)*h12,n-1));
62
63      p_min = min([p1,p2,p3,p4]); % find the point that is the furtherst
            from sample values. This point defines total critical level
64
65      p = 1 - (1 - p_min)^4;
66
67      %p = [p1; p2; p3; p4; p; z; h11; h22; h12];
68
69  else
70      [mu, C, meanX,varX] = Distr_new( mu, C, meanX, varX); % gives new
            values that need to be compared:
71                                                  % H0:(nu1,O1)=(nu2,O2);
72                                                  % with O1 and O2 diagonal
73      % define statistics
74      % define averages
75      x_av = meanX(1);
76      y_av = meanX(2);
77
78      % define estiamted variance as statistic
79      S1 = varX;
80
81      % define values that are tested as a scalars
82      mux = mu(1);
83      muy = mu(2);
84      sigmax = sqrt(C(1,1));
85      sigmay = sqrt(C(2,2));
86
87      h11 = S1(1,1)/(sigmax^2);
88
89      h22 = S1(2,2)/(sigmay^2);
90
```

```
91      z = (sigmay^2*(mux-x_av)^2 + sigmax^2*(muy-y_av).^2)./(sigmax^2*
            sigmay^2);
92
93      p1 = 1 - chi2cdf((n-1)*z,2);
94
95      p2 = 2*min(chi2cdf((n-1)*h11,n-1),1 - chi2cdf((n-1)*h11,n-1));
96
97      p3 = 2*min(chi2cdf((n-1)*h22,n-1),1 - chi2cdf((n-1)*h22,n-1));
98
99      p_min = min([p1,p2,p3]); % find the point that is the furtherst
            from sample values. This point defines total critical level
100
101     p = 1 - (1 - p_min)^3; % to the power of 4, as far as h_12 part
            always holds, but it still takes part in obtaining gamma
102
103 end
104
105 end
```

Transformation subroutines for bivariate Mood approach:

```
1  function [nu1,O1,nu2,O2] = Distr_new(mu1,C1,mu2,C2)
2
3  % derives new parameters of the distributions after the transformation
       of
4  % Random Variables N(mu1, C1) and N(mu2, C2) to N(nu1,O1) and N(nu2,O2)
       .
5  % Idea of the transforamtion is to get O1 - diagonal, O2 = I, nu2 = 0.
       Uses
6  % Trans_2 to calculate transformation coefficients and constant
7
8  [v,d_mu] = Trans2(mu1,C1,mu2,C2);
9
10 nu1 = v*mu1 - d_mu;
11 nu2 = v*mu2 - d_mu;
12 O1 = v*C1*v';
13 O2 = v*C2*v';
14 end
```

```
1  function [v,d_mu] = Trans2(~, C1, mu2, C2)
2
3  % Calculates the transformation coefficients v and constant d_mu to
       switch
4  % from the case N(mu1, C1) and N(mu2, C2) to the case of non-corretaled
5  % Random Vectors with second variable N(0,I)
6  % Input data as two vectors (2*1) of means mu1, mu2 and two covariance
7  % matrices (2*2) C1, C2
8  % check the uncorr by v*C1*v' and v*C2*v'
9  %tic
10
11 sx1 = sqrt(C1(1,1));
12 sx2 = sqrt(C2(1,1));
13 sy1 = sqrt(C1(2,2));
14 sy2 = sqrt(C2(2,2));
15 rho1 = C1(1,2)/(sx1*sy1);
```

```
rho2 = C2(1,2)/(sx2*sy2);

eps = 0.00001;

if abs(sy1^2*rho2*sx2*sy2 - rho1*sx1*sy1*sy2^2)>eps

    % Discriminant
    D = (sy1.^2.*sx2.^2 - sx1.^2.*sy2.^2).^2 - ...
        4*(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2)...
        .*(rho1.*sx1.*sy1.*sx2.^2 - rho2.*sx1.^2.*sx2.*sy2);

    % Auxiliary variables (case 1)

    phi1 = 1;
    phi2 = (rho1.*sx1.*sy1.*sx2.^2 - rho2.*sx1.^2.*sx2.*sy2)...
        ./(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2);
    phi3 = (sx1.^2.*sy2.^2 - sy1.^2.*sx2.^2 - sqrt(D))...
        ./(2*(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2));
    phi4 = (sx1.^2.*sy2.^2 - sy1.^2.*sx2.^2 + sqrt(D))...
        ./(2*(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2));


    %% Auxiliary variables (case 2)
    %
    % phi1 = 1;
    % phi2 = (rho1.*sx1.*sy1.*sx2.^2 - rho2.*sx1.^2.*sx2.*sy2)...
    %     ./(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2^2);
    % phi3 = (sx1.^2.*sy2.^2 - sy1.^2.*sx2.^2 + sqrt(D))...
    %     ./(2*(rho2.*sy1.^2.*sx2.*sy2-rho1.*sx1.*sy1.*sy2.^2));
    % phi4 = (sx1.^2.*sy2.^2 - sy1.^2.*sx2.^2 - sqrt(D))...
    %     ./(2*(rho2.*sy1.^2.*sx2.*sy2 - rho1.*sx1.*sy1.*sy2.^2));

    % Transformation coefficients (case 1)

    b = phi4;
    a = ones(size(b,1),size(b,2));
    c = ones(size(b,1),size(b,2));
    d = phi3./phi1;

    %% Transformation coefficients (case 2)
    %
    % a = phi1./phi2;
    % b = phi1./phi3;
    % c = phi2;
    % d = phi2.*phi3./phi1;

elseif abs(sy1^2*sx2^2 - sx1^2*sy2^2)<eps

    a = 1;
    b = 1;
    c = 1;
    d =  -(sx1.^2 + rho1.*sx1.*sy1)./(sy1.^2 + rho1.*sx1.*sy1);

else
```

```
70
71      sx1 = sx1+eps;
72      a = 1;
73      b = 1;
74      c = 1;
75      d =  -(sx1.^2 + rho1.*sx1.*sy1)./(sy1.^2 + rho1.*sx1.*sy1);
76
77  end
78
79  v = [a,b;c,d];
80
81  % additionally transform the solution so that new covariance matrix C2
        will
82  % be identity matrix and second vector standard normally distributed
83
84  C2_1 = v*C2*v';
85  k1 = sqrt(1/C2_1(1,1));
86  k2 = sqrt(1/C2_1(2,2));
87  r = [k1, 0; 0, k2];
88  v = r*v;
89
90  % calculate the constant term to make means of the second Random
        Variable
91  % zeros
92
93  d_mu = v*mu2;
94
95  %toc
96  end
```

Subroutine for computing volume of the bivariate Mood confidence region. Inputs are the random sample X and critical values t_{ij}:

```
1  function V = Mood_volume(X, t11, t12, t21, t22, t3)
2
3  S = cov(X);
4  n = size(X,1);
5
6  V = pi*t3*(n-1)^2*S(1,1)*S(2,2)*(t12-t11)*(t22-t21)/(4*n*t11*t12*t21*
       t22);
7  end
```

Subroutine for computing the optimal ratio of individual significance levels γ_i w.r.t. the volume of the confidence region:

```
1  %subroutine that computes optimal allocation of gamma w.r.t. the
       smallest
2  %volume
3  N = [10 50 100 500 1000];
4  G = [0.2 0.1 0.05 0.025 0.01];
5  O = zeros(length(N)*length(G),3);
6
7  q = 1; % counter
8  for j1 = 1:length(N)
9       for j2 = 1:length(G)
10 % calculate average volume on 1000 iterations and effective gammas for
```

```
% given sample size N and gamma G
opt = zeros(1000,3);
for  r = 1:1000
n = N(j1); % size of a sample
X = mvnrnd([0;0], [1  0; 0  1], n);
k = 50; % pieces we brake gamma

gamma = G(j2);
for i=1:k-1
    gamma1(i) = gamma - i*gamma/k;
    gamma2(i) = 1 - ((1-gamma)/(1-gamma1(i)))^(1/3);
    t11 = chi2inv(gamma2(i)/2,n-1);
    t12 = chi2inv(1 - gamma2(i)/2,n-1);
    t21 = t11;
    t22 = t12;
    t3 = chi2inv(1-gamma1(i),2);
    V(i) = Mood_volume(X,t11,t12,t21,t22,t3);
end
[~,a] = min(V);
opt(r,:) = [gamma1(a),gamma2(a), V(a)];
end

O(q,:) = mean(opt);
q = q + 1;
    end
end
```

Auxiliary subroutine that calculates the inverse marginal CDF of the covariance of Wishart distribution, presented in equation (3.34):

```
function [P,y_min,y_max] = cov_cdf(y,n)

% CDF of the distribution of covariance in Wishard distribution

f = @(x)(abs(x)).^((n-1)/2)./(gamma(n/2).*sqrt(2.^(n-1).*pi)).*besselk
    ((n-1)/2,abs(x));

% search for the boundary max, strating from which pdf is not defined
    with
% copmputer precision and is = 0
y_max1 = 1/n;
y_max2 = n^2;
eps = 0.0001;

while 1
    y_mid = (y_max1 + y_max2)/2;
    if isnan(f(y_mid)) || isinf(f(y_mid))
        y_max2=y_mid;
    else
        y_max1=y_mid;
    end
    if (abs(y_max2 - y_max1)<eps)
        break;
    end
end
```

```
y_max = y_max1;

% search for the boundary min, strating from which pdf is infinity with
% computer precision

y_min1 = 0;
y_min2 = y_max;
eps = eps^2;

while 1
    y_mid = (y_min1 + y_min2)/2;
    if isnan(f(y_mid)) || isinf(f(y_mid))
        y_min1 = y_mid;
    else
        y_min2 = y_mid;
    end
    if(abs(y_min2 - y_min1)<eps)
        break;
    end
end
y_min = y_min2;

% define CDF for each point
if y>=y_max
    P=1;
elseif (y<y_max) && (y>y_min)
    P = 1 - integral(f,y,y_max);
elseif (y<=y_min) && (y>=-y_min)
    P = y*(1 - 2*integral(f,y_min,y_max))/(2*y_min) +1/2;
elseif (y>-y_max) && (y<-y_min)
    P = integral(f,-y_max,y);
elseif y<=-y_max
    P=0;
end

if isnan(P) || isinf(P)
    P = 0.5;
end

end
```

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

Datum: 08.03.2021

Unterschrift: **Valerii Dashuk**