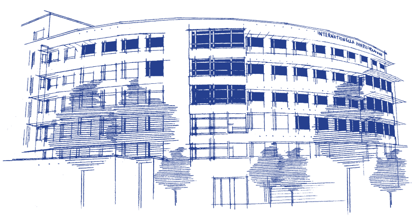
DS

Assessing Distributional Properties of High-Dimensional Data

RASHID MANSOOR



Jönköping International Business School
Jönköping University

JIBS Dissertation Series No. 092 • 2013

RASHID MANSOOR

Assessing Distributional Properties of High-Dimensional Data

JIBS Dissertation Series No. 092

JIBS

Assessing Distributional Properties of High-Dimensional Data

RASHID MANSOOR

This doctoral thesis consists of five papers in the field of multivariate statistical analysis of high-dimensional data. Because of the wide application and methodological scope, the individual papers in the thesis necessarily target a number of different statistical issues.

In the first paper, Monte Carlo simulations are used to investigate a number of tests of multivariate non-normality with respect to their increasing dimension asymptotic (IDA) properties as the dimension p grows proportionally with the number of observations n such that $p/n \rightarrow c$ where c is a constant.

In the second paper a new test for non-normality that utilizes principal components is proposed for cases when $p/n \rightarrow c$. The power and size of the test are examined through Monte Carlo simulations where different combinations of n and p are used.

The third paper treats the problem of the relation between the second central moment of a distribution to its first raw moment. In order to make inference of the systematic relationship between mean and standard deviation, a model that captures this relationship by a slope parameter (β) is proposed and three different estimators of this parameter are developed and their consistency proven in the context where the number of variables increases proportionally to the number of observations. In the fourth paper, a Bayesian regression approach has been taken to model the relationship between the mean and standard deviation of the excess return and to test hypotheses regarding the β parameter. An empirical example involving Stockholm exchange market data is included. Then finally in the fifth paper three new methods to test for panel cointegration of high-dimensional data in the error correction framework are introduced.

JÖNKÖPING INTERNATIONAL BUSINESS SCHOOL
JÖNKÖPING UNIVERSITY

ISSN 1403-0470 ISBN 978-91-86345-46-4

Assessing Distributional Properties of High-Dimensional Data

RASHID MANSOOR

JÖNKÖPING INTERNATIONAL
BUSINESS SCHOOL
JÖNKÖPING UNIVERSITY

Jönköping International Business School
P.O. Box 1026
SE-551 11 Jönköping
Tel.: +46 36 10 10 00
E-mail: info@jibs.hj.se
www.jibs.se

Assessing Distributional Properties of High-Dimensional Data
JIBS Dissertation Series No. 092

ISSN 1403-0470
ISBN 978-91-86345-46-4

Printed by ARK Tryckaren AB, 2013

This thesis, is dedicated to the memory
of my dear father Muhammad Ismail
who passed away before this thesis
was completed.

Acknowledgments

It was a long journey from Government Primary School Hajizai to Lund University and then finally Jönköping International business School (JIBS), Jönköping University. This journey has not been a lonely one. Several people have contributed in different ways. It has been a great privilege to spend several years in the Department of Economics, Finance and Statistics at Jönköping University, and its members will always remain dear to me.

I would especially like to express my thanks and gratitude to my head supervisor Professor Thomas Holgersson for dedicating his time to the supervision of this dissertation work over the years. I very much appreciate the informal discussions that we had throughout the length of the thesis and his never-ending patience reading my papers. I will always be in debt to him, for his encouragement and willingness to share his research skills over these years. Without his help this thesis would not have seen the light of day. I thank him for his patience and for all his support and help from day one until today. Professor Holgersson is truly a very good supervisor and teacher with a broad knowledge in multiple areas of statistics. He is always willing to share his knowledge with students and does so in a humble and kind way. I am thankful for all that he has taught me and for the example that he has set.

I would also like to thank my deputy supervisor Professor Ghazi Shukur for his support and guidance. He patiently provided the encouragement and advice necessary for me to proceed through the doctoral program and complete my dissertation. His deep and broad knowledge in econometrics has been inspiring to many doctoral students including myself. I thank Professor Shukur for his help, encouragement and support during my stay at JIBS. Professor Shukur is a person one instantly loves and never forgets.

Special thanks to professor Björn Holmquist from Lund University for his precious comments and productive suggestions on the earlier version of my papers in my final seminar. He was also my teacher in the master program in statistics at Lund University and taught me advanced statistical inference and multivariate methods. I am very grateful that he provided me with some of his books during many courses, which I greatly appreciate. I want to thank Professor Holmquist very much for everything from August 22, 2007 until now. Professor Holmquist likes to make people happy around him.

I would also like to thank my teachers at Lund University, Professor Krzysztof Nowicki, Professor Björn Holmquist, Associate Professor Panagiotis Mantalos, Associate Professor Mats Hagnell, and Peter Gustafsson, for their valuable help and cooperation during my stay at Lund. They will always remain dear to me, and I have many fond memories from Lund.

I am also grateful to Associate Professor Abdullah Almasri for teaching me wavelet analysis and Dr. Shakir Hussain for his valuable and interesting lectures on survival analysis, multilevel modeling and Bayesian multilevel modeling.

I also wish to express my deep appreciation to Professor Börje Johansson for his support and encouragement during the periodic follow-up postgraduate study meetings with PhD students. I always found a smile on his face and I really appreciate his positive attitude to students. A special word of thanks goes to Kerstin Ferroukhi for her help and advice. I could always knock at her office door whenever I had a problem. I will remember all her help throughout my life.

I would also like to express my gratitude to Professor Åke E. Andersson, Professor Charlie Karlsson, Professor Per-Olof Bjuggren, Professor Scott Hacker, Professor Charlotta Mellander, Associate Professor Johan Klaesson, and Assistant Professor Johan Eklund for their support and feedback during the periodic follow-up postgraduate study meetings with PhD students. I appreciate their efforts in helping the students to achieve their goals. I am grateful to Professor Andreas Stephan and Assistant Professor Agostino Manduchi for their patience in listening. I would also like to thank Research Coordinator Monica Bartels and administrators Susanne Hansson, Maria Farkas, and Katarina Blåman for their help.

It is time to express my deepest gratitude to my colleagues and friends Dr. Kristofer Månsson, Dr. Peter Karlsson, Dr. Zangin Zeebari, Associate Professor Pär Sjölander, and Dr. Hyunjoo Kim, who always helped me, all of them always there to share my moments of joy and sorrow. My first paper was with Dr. Peter Karlsson and my last paper was with Dr. Kristofer Månsson. I really enjoyed that time and learned a great deal from them and want to thank them all for their help and support.

All the former and current PhD students at the Department of Economics, Finance and Statistics made my work very enjoyable. I thank them all for their kindness and optimistic attitude. My gratitude also goes to all my friends outside the department. I am lucky to have lovely Pakistani friends in Sweden whose presence makes my stay more enjoyable; without their company, my stay in Sweden would never have been so wonderful.

A special word of thanks goes to the members of my family, especially my parents and my brothers Khalid Mansoor, Muhammad Adeel and Dr. Muhammad Aqeel for their love and support.

This dissertation is dedicated to the loving memory of my dear father Muhammad Ismail (may Allah rest his soul in peace), who believed in my academic abilities at a very young age and also was looking forward to the quick completion of this dissertation. Unfortunately, he is not here with us to see the whole thing finished.

Jönköping October 22, 2013
Rashid Mansoor

Abstract

This doctoral thesis consists of five papers in the field of multivariate statistical analysis of high-dimensional data. Because of the wide application and methodological scope, the individual papers in the thesis necessarily target a number of different statistical issues. In the first paper, Monte Carlo simulations are used to investigate a number of tests of multivariate non-normality with respect to their increasing dimension asymptotic (IDA) properties as the dimension p grows proportionally with the number of observations n such that $p/n \to c$ where c is a constant. In the second paper a new test for non-normality that utilizes principal components is proposed for cases when $p/n \to c$. The power and size of the test are examined through Monte Carlo simulations where different combinations of n and p are used.

The third paper treats the problem of the relation between the second central moment of a distribution to its first raw moment. In order to make inference of the systematic relationship between mean and standard deviation, a model that captures this relationship by a slope parameter (β) is proposed and three different estimators of this parameter are developed and their consistency proven in the context where the number of variables increases proportionally to the number of observations. In the fourth paper, a Bayesian regression approach has been taken to model the relationship between the mean and standard deviation of the excess return and to test hypotheses regarding the β parameter. An empirical example involving Stockholm exchange market data is included. Then finally in the fifth paper three new methods to test for panel cointegration of high-dimensional data in the error correction framework are introduced.

Table of Content

1 Introduction 11
2. High-Dimensional Data 12
3. Distributional Properties 13
Multivariate Normality of High-Dimensional Data 13
Mean-standard Deviation Ratio of Large Data Sets 15
Cointegration of High-Dimensional Data 16
4. Summary of the Articles 19
Article 1: Assessing Normality of High-Dimensional Data 19
Article 2: Using Principal Components to Test Normality of High-Dimensional Data 19
Article 3: Estimating Mean-Standard Deviation Ratios of Financial Data 20
Article 4: A Bayesian Approach for Estimating Mean-Standard Deviation Ratios of Financial Data 20
Article 5: Testing for Panel Cointegration in High-Dimensional Data in the Presence of Cross-Sectional Dependency 20
References 22

Articles 29
Article 1 Assessing Normality of High- Dimensional Data 31
Article 2 Using Principal Components to Test Normality of High- Dimensional Data 43
Article 3 Estimating Mean-Standard Deviation Ratios of Financial Data 57
Article 4 A Bayesian Approach for Estimating Mean-Standard Deviation Ratios of Financial Data 75
Article 5 Testing for Panel Cointegration in High-Dimensional Data in the Presence of Cross-Sectional Dependency 85

JIBS Dissertation Series 99

1 Introduction

Multivariate statistical analysis deals with the statistical analysis of data that comprise several measurements (variables) of a number of individuals or objects (observations). It is obvious that observing several variables gives us more information than observing only a single one. The multivariate approach allows us to explore the joint performance of the variables at hand and to assess the influence of each variable in the presence of others. There are many studies such as marketing research, where most business problems are multidimensional, studies on production processes, where a number of different measurements are taken on the same unit or studies on health and economy, etc. While the vast majority of the statistical research in the first half of the 1900s was set on univariate problems, the importance of developing methods for analysing multidimensional data was slowly becoming recognized. Scientists such as Pearson (1901), Hotelling (1933), and Fisher (1936) pioneered research on how to analyse multivariate data, in particular within classification problems. Other contributions include those by Wishart (1928), who derived the distribution of sample variances and covariances for multivariate normal samples, Wilks (1946), who provided procedures for additional tests of hypotheses on means, variances, and covariances, and Mahalanobis (1936) who developed distance measures with application areas such as outlier detection. From the mid-1950s and onwards, multivariate methods were recognized as a common and important part of the statistician's toolbox. Multivariate techniques were developed for applications in quality control, time series analysis, and regression analysis, which were previously restricted to univariate problems. While most of these methods are recognized as being well behaved for data of low dimensions, such as joint analysis of people's height, weight, and body mass index observed on a thousand individuals, recent attention has been devoted to cases where the numbers of variables and observations are proportional. In such cases the standard multivariate methods are usually not well behaved, which in turn produces biased estimates, and wrong conclusions are drawn about the data at hand.

Furthermore many of the statistical models depend on standard assumptions (e.g. normality, serial independence, homoscedasticity, moment restrictions, etc.) about the data; an appropriate statistical model is one whose major assumptions are suitable for the data analysed. When the distributional assumptions are violated, subsequent analyses will differ from those expected and can produce misleading inferences. However, while the problem has not suffered from lack of attention in cases where n is large relative to p, little, if any, attention has been given to cases when n is fixed and p increases or p/n increases simultaneously towards a constant c where $0 \leq c < 1$ and $c > 1$.

This thesis is concerned with the intersection of these two problems.

2. High-Dimensional Data

While it was almost impossible to analyse data of higher dimension than, say, five, before the introduction of the PC, we nowadays often face problems involving the analysis of very large datasets. Some examples include econometric panel datasets where one may have weekly observations of 50-100 companies over a 5-year period, the investigation of risk and return across a large number of heterogeneous assets in finance, or an analysis of surveys where budget constraints frequently will restrict the sample size n to be low compared to p. The traditional multivariate methods are, however, usually not well behaved in cases where p is large compared to n. A common requirement of an estimator is that of consistency, i.e. that an estimated parameter should converge to the unknown true value as the number of observations increase. But in situations where p is comparable to n the asymptotical properties are more realistically described through the mode where the ratio between p and n tends towards a constant, i.e. $p/n \to c$, $0 < c < 1$. In such cases the traditional assumptions are violated and the statistic at hand will usually not converge to its true parameter. For example, the standard estimator of the inverse covariance matrix is well known to degenerate when p grows proportionally with n (Girko, 1995; Serdobolskii, 1985). Other examples include methods for hypothesis tests of linear restrictions, which typically depend on fixed dimension asymptotics (Muirhead, 1982; Rao, 1973; Fujikoshi, 1970, 1988) and will hence not be well behaved under conditions where p increases. Hence there is a need to develop new statistical methods that remain well behaved also under more general asymptotic conditions. Although the influence of high-dimensional data has been noticed in many different directions of multivariate statistical inference, the problem has not yet been clearly cited in the literature, and the development of descriptive and inferential methods for such data is still a great challenge to research in statistics.

The development of increasing dimension analysis (IDA) was pioneered by Kolmogorov in the early 1970s to be followed by a number of important contributions, including problems of estimating the inverse covariance matrix (Girko 1975, 1990, 1995; Serdobolskii 1985, 2000, 2008), estimation of high-dimensional expected value vectors (Serdobolskii, 2008), hypothesis testing (Srivastava, 2007), discriminant analysis (Serdobolskii, 2008; Pavlenko and Von Rosen, 2001). Some other contributions include Yin et al. (1988), Jonsson (1982), and Silverstein and Bai (1995), who derived limits for eigenvalues.

A somewhat different problem arises when p is larger than the number of observations n such that $p/n \to c$ where $c > 1$. Such cases are seen in many statistical applications and are common in fields such as biological sciences and financial studies. For example, data from DNA micro arrays may involve thousands of gene expression levels measured in relatively few subjects. The methodological problem in this setting is how to extract the useful information from the data when the dimension of the data p is larger than the number of observations n. In such situations there is an option of using data reduction methods and present a system having many degrees of freedom by imposing restrictions on the covariance structures. Factor analysis, principal components analysis, and discriminant analysis are some of the techniques that can be used for dimensionality reduction.

3. Distributional Properties

Knowledge of the theoretical distribution that best approximates the underlying distribution of a random variable is crucial for deciding what statistical models are appropriate to use for those specific data. An appropriate statistical model is one whose fundamental assumptions are suitable for the data analysed. When the underlying assumptions are inappropriate, the actual distribution of the common point estimators or test statistics will differ from that expected. For example, most of the economic data consist of time series and there are very often temporal correlations in the model error terms corresponding to successive time periods. This is known as autocorrelation and the reason might be omitted variables or misspecification of the dynamic process. There is a large amount of literature on autocorrelation tests where several single-wise and system-wise autocorrelation tests have been developed. For example, tests for autocorrelated errors of univariate models were proposed by Ljung and Box (1978) and the BG test by Breusch (1978) and Godfrey (1978). Further development was made by Hosking (1980), Edgerton and Shukur (1999), and Holgersson (2004), where multivariate versions of autocorrelation tests were suggested. Similarly, another standard assumption is that of homoscedasticity. Homoscedasticity means that the disturbance variance should be constant at each observation. Violation of this assumption is called heteroscedasticity. A number of methods have been suggested for testing heteroscedasticity. For example, Doornik (1996) examined certain properties of a test for multivariate heteroscedasticity suggested by Kelejian (1982), while Holgersson and Shukur (2004) proposed a system-wise test for heteroscedasticity. Another common test was developed by White (1980). An additional way to characterize a distribution is through its third and fourth central moments, i.e. the coefficients of skewness and kurtosis. A somewhat different way to characterize the distributional properties of a multivariate dataset is through the link between different moments. In case there exists, for example, a linear relationship between the mean values and the standard deviations, then the distribution of the multivariate random variable may be characterized by a lower number of parameters than otherwise.

While significance tests for assessing distributional properties such as those discussed above are well established in cases where n is large relative to p, little attention has been given to cases when n is fixed and p increases or p/n increases simultaneously towards a constant c where $0 \le c < 1$. This thesis considers assessing distributional properties of different types in high-dimensional settings.

Multivariate Normality of High-Dimensional Data

"*It is not enough to know that a sample could have come from a normal population; we must be clear that it is at the same time improbable that it has come from a population differing so much from the normal as to invalidate the use of 'normal theory' tests in further handling of the material*." (*E. S. Pearson, 1930*)

Normality is one of the most common assumptions in the development and use of statistical methodology. The importance of normality is due to the fact that many of the statistical methods are strongly dependent on this assumption and require an effective test procedure to detect possible non-normality. The violation of the assumption of normality and its effect on standard statistical procedures date back before Bartlett's (1935) paper on the *t*-test. A number of well-known tests can be found in D'Agostino (1982) and Srivastava and Mudholkar (2003). These are usually based on (i) sample skewness and kurtosis measure, (ii) the characteristic function, or (iii) other characterizations of the normal distribution.

Also multivariate techniques frequently assume multivariate normality. Hopkins and Clay (1963), Mardia (1970), and Conover and Iman (1980) in their simulation studies highlight the importance of the multivariate normality assumptions, demonstrating that many of them lack robustness when they are applied to non-multivariate data. In recent times new approaches for testing multivariate normality have been recommended by Bowman and Foster (1993), Henze and Wagner (1997), and Romeu and Ozturk (1993), and multivariate extensions of tests by Shapiro and Wilk (1965) using order statistics was suggested by Royston (1983) and Srivastava and Hui (1987). Other authors, for example Mardia (1970), Srivastava (1984), and Lütkepohl and Theilen (1991), developed tests using multivariate skewness and kurtosis for detecting multivariate normality. These are of particular interest partly since they are simple to conduct and partly because they are commonly used and included in standard software.

The measures of multivariate skewness and kurtosis proposed by Mardia (1970) are defined as follows:

$$\beta_{1,M} = E\left[(\mathbf{X}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_j - \boldsymbol{\mu})\right]^3 \text{ and } \beta_{2,M} = E\left[(\mathbf{X}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu})\right]^2, \qquad (1)$$

where $\mathbf{X}_i$ and $\mathbf{X}_j$ are identically and independently distributed. The sample analogies are given by

$$b_{1,M} = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^3 \text{ and } b_{2,M} = n^{-1}\sum_{i=1}^{n} r_i^4 \text{, where}$$

$$r_i^2 = \left[\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)' \mathbf{S}^{-1}\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)\right]^2,\ r_{ij} = \left[\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)' \mathbf{S}^{-1}\left(\mathbf{X}_j - \bar{\mathbf{X}}\right)\right],$$

$$\bar{\mathbf{X}} = n^{-1}\sum\nolimits_{i=1}^{n}\mathbf{X}_i \text{ and } \mathbf{S} = n^{-1}\sum\nolimits_{i=1}^{n}\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)'.$$

Jarque and McKenzie (1995) suggested a combination of these two measures by

$$D_M = D_{1,M} + D_{2,M}^2 \text{ where } D_M \xrightarrow{L} \chi^2_{(p(p+1)(p+2)/6+1)}. \qquad (2)$$

A second class of measures of multivariate skewness and kurtosis was proposed by Lütkepohl and Theilen (1991). Let $\mathbf{L}$ denote the Cholesky decomposition of $\mathbf{S}$ such that

$\mathbf{S} = \mathbf{L}'\mathbf{L}$. Then define $b_{1,L} = (n/6)\sum_{j=1}^{p}\left(n^{-1}\sum_{i=1}^{n} z_{ij}^{3}\right)^{2}$ and $b_{2,L} = (n/24)\sum_{j=1}^{p}\left(n^{-1}\sum_{i=1}^{n} z_{ij}^{4} - 3\right)^{2}$ for $j = 1,..., p$, where $\mathbf{Z}_i = (z_{1i},...,z_{Pi})' = \mathbf{L}^{-1}\left(\mathbf{X}_i - \bar{\mathbf{X}}\right)$. Under the null hypothesis of normality the following asymptotic distribution holds:

$$D_L = b_{1,L} + b_{2,L} \xrightarrow{L} \chi^2_{2p}. \qquad (3)$$

The composite statistics D_M and D_L are frequently used in applied statistics to test for non-normality and are also included in many statistical software. The small sample properties of the tests have been investigated through simulation studies by, among others, Holgersson and Shukur (2001) and Doornik and Hansen (2008). While these investigations involved cases of fixed *p*, we here investigate the size and power properties of these tests with respect to their IDA properties, i.e. when *p* increases together with *n*, say $p/n \to c$, $0 < c < 1$. It has been exposed in this thesis that Mardia's (1970) and Srivastava's (1984) tests are inconsistent for some distributions, while Lütkepohl and Theilen (1991) in contrast are consistent for all the non-normal settings used in this thesis. One reason may be that the degrees of freedom of Mardia's (1970) tests increase quickly with the increase in dimension and tests based on a chi-square distribution with many degrees of freedom often have low power. The other reason might be that Mardia's (1970) measures of multivariate skewness and kurtosis both require the inversion of sample covariance matrices. For high-dimensional data the sample covariance matrices may be degenerate and their inversion may be unstable, which in turn may lead to biased analyses. If the dimension is larger than the number of observations, inversion of the sample covariance matrices will be impossible. Based on the above discussion, the Lütkepohl and Theilen (1991) test is used as a benchmark, and a test that utilizes principal components is proposed to test for non-normality when the number of dimensions is larger than the number of observations $(p > n)$, and also when the number of observations is larger than the dimension $(n > p)$ and p increases asymptotically along with n. This is further discussed in Paper 2.

Mean-standard Deviation Ratio of Large Data Sets

Datasets involving a large number of variables frequently involve parameter restrictions, in the sense that the distribution of the data is expressed using a parameter of lower dimension than the data. In such cases one may either use the information of possible parameter restrictions to reduce the dimension of data or to make inference on the parameter restrictions themselves. Relevant questions may involve the functional form of parameter restrictions or their numerical value. An important case involves financial economic theories in which it is stated that the higher the risk an investor is willing to take, the higher is the potential return from the investment. Hence there is an explicit relation between the expected values and standard deviations across a large number of variables. References include Osteryoung et al. (1977), Lopes et al. (1993), and Sharpe (1966), who investigated the ratio of mean value to standard deviation. In order to make inferences of the systematic relationship between mean and standard deviation values, a model that captures the relationship between standard deviations and the mean values may be applied. A link between the standard deviation and mean of a set of variables $X_1,...,X_n$ may be defined as follows:

$$\mu_i = \beta\sigma_i, \tag{4}$$

where $\mu_i = E[X_{it}]$, $\sigma_i = \sqrt{E[X_{it} - \mu_i]^2}$, $i = 1,...,n,\ t = 1,...,T$ and $-\infty < \mu_i < \infty$, $0 < \sigma_i < \infty$. Methods to consistently estimate the beta parameter in (4) are proposed in the thesis. Specifically, if $\{\bar{X}_j\}_{j=1}^n$ and $\{S_j\}_{j=1}^n$ denote the set of sample means and standard deviations, it may be shown that the following estimators are consistent:

$$\hat{\beta}_k = \frac{\sum_{i=1}^{n} w_i^k \bar{x}_i s_i}{\sum_{i=1}^{k} w_i^k \bar{x}_i s_i^2}, \quad k = 1,\ 2,\ 3.$$

Apart from consistency, there is also an issue of how to make inference on the beta parameter, e.g. to test whether or not it is equal to zero. While this is not straightforward to do in the standard frequentist way, Bayesian methods will be of great potential. Specifically, it may be shown that the joint posterior pdf β and σ_δ, which is given by

$$p(\beta, \sigma_\delta \mid \bar{X}_i, \sigma_i) \propto p(\beta, \sigma_\delta)\, p(\bar{X}_i \mid \sigma_i, \beta, \sigma_\delta), \tag{5}$$

facilitates estimation of the marginal posterior pdf of β. This is defined as follows:

$$p(\beta \mid \bar{X}_i, \sigma_i) = \int_0^\infty p(\beta, \sigma_\delta \mid \bar{X}_i, \sigma_i)\, d\sigma_\delta. \tag{6}$$

Of particular interest is to test if the β coefficients for two independent segments, or sub-samples, are equal, i.e. $H_0 : \beta_d = 0$ vs $H_1 : \beta_d \neq 0$. Details are given in Papers 3 and 4.

Cointegration of High-Dimensional Data

It was an important improvement in econometrics when it was established that if the ordinary least square (OLS) residuals from a regression between integrated variables of the same order are stationary, this may lead to superconsistent estimates of the coefficients. Granger (1981) named the concept cointegration; it was further developed by Engle and

Granger (1987). Several methods have been developed to test for cointegration. The most popular tests for cointegration are the residual-based cointegration tests by Engle and Granger (1987), Johansen (1988), and Stock and Watson (1988). All these univariate procedures lack statistical power and therefore the analysis of cointegration in the panel data setting has been a rich area of study in recent years and there has been an emergent interest in testing for panel cointegration in economic and financial variables. Panel cointegration tests are categorized into a first and a second generation of procedures, where the first generation assumes independence across the units in the panel. Along these lines, Pedroni (2004) and Kao (1999) generalized the residual-based cointegration tests by Engle and Granger (1987) and Phillips and Ouliaris (1990). Larsson et al. (2001) adapted the system approach of Johansen (1988), which in contrast to the above-mentioned residual-based testing procedure is able to detect multiple cointegration in the panel. McCoskey and Kao (1998) suggested a test for the null of panel cointegration. All these aforementioned tests do not take account of cross-sectional dependency. However the assumption of zero-independent cross-sectional dependency is in many contexts a highly restrictive one. Therefore a second generation of approaches takes account of cross-sectional dependency in the panel cointegration testing. These include, for example, the tests proposed by Bai and Kao (2006), Banerjee and Carrion-i-Silvestre (2006), Gengenbach et al. (2006), Groen and Kleibergen (2003), Nelson et al. (2005), Pedroni and Vogelsang (2005), and Westerlund (2005). However, these methods are only considered for the case when the number of cross-sectional units (N) is much smaller than the number of time period (T). Furthermore many studies fail to reject the null hypothesis of no cointegration, even in cases where the theory suggests strong cointegration. One reason for this failure may be that most residual-based cointegration tests require that the short-run adjustment process is equal to the long-run adjustment process. Banerjee et al. (1998), Kremers et al. (1992), and Mantalos and Shukur (1998) showed that its failure could cause a significant loss of power for residual-based cointegration tests.

In this thesis a different approach to test for panel cointegration in high-dimensional data is developed using the combining *p*-value approach, which was introduced to the panel unit root literature by Maddala and Wu (1999) and Choi (2001). A Fisher (1932) type of test is proposed, defined as

$$P = -2\sum_{i=1}^{N}\ln(p_i) \sim \chi^2_{(2N)}, \tag{7}$$

where p_i denotes the *p*-value of the test statistic for each cross-sectional unit *i*. Similarly Stouffer et al. (1949) proposed a test known as the inverse normal test defined as:

$$Z = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Phi^{-1}(p_i) \overset{L}{\rightarrow} N(0,1), \tag{8}$$

where $0 \le p_i \le 1$ and $\Phi^{-1}(p_i)$ is a standard normal variable. Yet another test proposed by Choi (2001) is defined as:

$$P_m = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\ln(p_i) + 1 \right) \xrightarrow{L} N\left(0,1\right). \tag{9}$$

The key interest is to investigate the asymptotic properties of these three tests as N grows proportionally to T, i.e. we investigate the size and power properties of these tests with respect to their IDA properties. This is discussed in detail in the fifth paper.

4. Summary of the Articles

Article 1: Assessing Normality of High-Dimensional Data

In the first paper Monte Carlo simulations are used to investigate the properties of three tests of multivariate non-normality proposed by Mardia (1970), Srivastava (1984), and Lütkepohl and Theilen (1991) with respect to their increasing dimension analysis (IDA) properties. Since the effects of non-normality within normal-theory based inference of multivariate data may be serious if the dimension p of data is proportional to the sample size n, it is important that tests of the normality assumption remain well behaved in such settings. It is argued that a good test should be consistent in the sense that the likelihood to reject a false H_0 increases as the number of observations increases in all kinds of asymptotic settings including cases when (i) p is fixed and n increases, (ii) n is fixed and p increases, and (iii) n and p increase simultaneously, since all three cases concern an increasing amount of information (i.e. $np \to \infty$). The Mardia and Srivastava tests are shown to be inconsistent against some distributions whereas the test of Lütkepohl and Theilen in contrast is consistent against all non-normal distributions used in the paper. The final conclusions are that when testing for multivariate non-normality in high-dimensional data, the Lütkepohl and Theilen test with simulated critical values should be applied, while the tests by Mardia and Srivastava should be used with caution unless $n \gg p$.

Article 2: Using Principal Components to Test Normality of High-Dimensional Data

In this paper a test that utilizes principal components to test for non-normality is developed. A simulation study was conducted to investigate the size and power properties of the test in the high-dimensional setting, i.e. (i) when the number of dimension p is greater than the number of observations n and (ii) when n is greater than p and in both cases p increases asymptotically along with n. In this study principal component analysis is used to reduce the dimension of the data. The size and power of the test using Monte Carlo critical values and chi-square critical values are investigated. Based on the result of a simulation study, the test is seen to have good overall size and power properties and detects non-normality in all the cases, i.e. when p is fix and n increases and when p and n increase simultaneously. Hence the proposed D_L^* test is recommended to be used for non-normality in high-dimensional settings.

Article 3: Estimating Mean-Standard Deviation Ratios of Financial Data

This paper treats the problem of the relation between the second central moment of a distribution to its first raw moment. The key interest is the reciprocal of the coefficient of variation, where, according to the theory of finance, an investor is compensated by increased expected returns for taking higher risks (i.e. variance of the investments returns), which is represented by the ratio of the mean value to the standard deviation. In most cases these applications have involved the problem of making inference of the relation between the mean and the standard deviation of a single variable. In this context, however, there is a systematic relation between the standard deviations and mean values of a large number of heterogeneous variables. In order to make inference of the systematic relationship between the mean values and standard deviations, a model that captures this relationship by a slope parameter (β) is proposed and three different estimators of this parameter developed. Their consistencies are proven in the context where the number of variables is proportional to the number of observations. An empirical study is conducted on the stocks of Stockholm Stock Exchange between June 1995 and June 2010 and the population divided into three segments depending on the market capitalization value. It is seen that the beta parameter is heterogeneous across the three segments.

Article 4: A Bayesian Approach for Estimating Mean-Standard Deviation Ratios of Financial Data

In this paper a Bayesian regression approach has been taken to model the relationship between the mean and the standard deviation of the excess return. This paper derives procedures for statistical inference, in particular interval estimation and hence hypothesis testing, for assessing whether two population segments, e.g. Large Cap and Small Cap stocks, have equal risk/return ratios. It is shown that hypothesis testing of the possible difference between return-to-risk ratios of two market segments can easily be conducted by Bayesian credible intervals. Specifically, a vague prior is used, which facilitates the use of previously proposed frequentist results with the important enhancement of inference methods. Hence a simple but yet comprehensive tool is available for investors to quantify and compare risk rewards of different markets. The use of the method is demonstrated through an empirical investigation of the Small Cap and Large Cap segments of the Stockholm Stock Exchange.

Article 5: Testing for Panel Cointegration in High-Dimensional Data in the Presence of Cross-Sectional Dependency

This paper proposes some new methods to test for panel cointegration in the error correction framework in high-dimensional data. These new panel tests are based on different *p*-value combinations (used in, for example, panel unit root testing in Choi, 2001) of the single equation cointegration test in the error-correction framework suggested by Kanioura and Turner (2005). The paper uses a Wald statistic that is defined through the combination of individual *p*-values, and by means of Monte Carlo simulations the size and power properties of the tests are investigated when the number of cross-sectional units

(N) and the number of time periods (T) increase simultaneously, i.e. $N/T \to c$, where c is a constant. Moreover, the effect of cross-sectional dependency caused by a single common factor is investigated and it is shown that by applying a simple demeaning procedure any common factor over the cross- sectional dimension will be removed through mathematical cancellation. Based on the simulated results the Z-test first proposed by Stouffer et al. (1949) is the one recommended to be used in real applications. The demeaned version of this test is robust to cross-sectional dependency caused by a common factor. Furthermore it has the best power properties among the suggested *p*-value combinations and we recommend this test to be used in the high- dimensional setting.

References

Bai, J. and Kao, C. (2006). On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence. In B. H. Baltagi (Ed.), *Panel Data Econometrics: Theoretical Contributions and Empirical Applications*, Amsterdam: Elsevier Science, pp. 3-30.

Banerjee, A. and Carrion-i-Silvestre, J. L. (2006). Cointegration in panel data with breaks and cross-section dependence. *ECB Working Paper No.* 0591.

Banerjee, A., Dolado, J. and Mestre, R. (1998). Error-correction Mechanism Tests for Cointegration in a Single-equation Framework. *Journal of Time Series Analysis*, 19(3), 267-283.

Bartlett, M. S. (1935). The Effect of Non-Normality on the t Distribution. *Proceedings of the Cambridge Philosophical Society*, 31, 223-231.

Bowman, A. W. and Foster, P. J. (1993). Adaptive Smoothing and Density Based Tests of Multivariate Normality. *Journal of the American Statistical Association*, 88(422), 529-537.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17(31), 334–355.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, 20(2), 249-272.

Conover, W. J. and Iman, R. L. (1980). The rank transformation as a method of discrimination with some examples. *Communications in Statistics – Theory and Methods*, 9(5), 465-487.

D'Agostino, R. B. (1982). Tests for departure from normality. In S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Science*, 2, pp. 315-324

Doornik, J. A. (1996). Testing Error Autocorrelation and Heteroscedasticity, revised version of paper presented at the 7th World Congress of the Econometric Society, Tokyo, Japan, August 22-29, 1995.

Doornik, J. A. and H. Hansen. (2008). An Omnibus Test for Univariate and Multivariate Normality. *Oxford Bulletin of Economics and Statistics*, 70(Supplement s1), 927-939.

Edgerton, D. and Shukur, G. (1999). Testing autocorrelation in a system perspective testing autocorrelation. *Econometric Reviews*, 18(4), 343-386.

Engle, R. F. and Granger, C. W. J. (1987). Cointegration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2), 251-276.

Fisher, R. A. (1936). The use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2), 179-188.

Fujikoshi, Y. (1970). Asymptotic Expansions of the Distributions of Test Statistics in Multivariate Analysis. *Journal of Science of the Hiroshima University Series A-I*, 34(1), 73-144.

Fujikoshi, Y. (1988). Comparison of powers of a class of tests for multivariate linear hypothesis and independence. *Journal of Multivariate Analysis*, 26(1), 48-58.

Gengenbach, C., Palm, F. C. and Urbain, J.-P. (2006). Cointegration Testing in Panels with Common Factors. *Oxford Bulletin of Economics and Statistics*, 68(Supplement s1), 683-719.

Girko, V.L. (1975). *Spectral Theory of Random Matrices*. Cambridge, UK: Elsevier Science.

Girko, V.L. (1990). *Theory of Random Determinants*. Doordrecht: Kluwer Academic Publishers.

Girko, V.L. (1995). *Statistical Analysis of Observations of Increasing Dimension*. Doordrecht: Kluwer Academic Publishers.

Godfrey, L. G. (1978). Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*, 46(6), 1293-1301.

Granger, C. W. J. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*, 16(1), 121-130.

Groen, J. J. J. and Kleibergen, F. (2003). Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models. *Journal of Business & Economic Statistics*, 21(2), 295-318.

Henze, N. and Wagner, T. (1997). A New Approach to the BHEP Tests For Multivariate Normality. *Journal of Multivariate Analysis*, 62(1), 1-23.

Holgersson, H. E. T. (2004). Testing for Multivariate Autocorrelation. *Journal of Applied Statistics*, 31(4), 379-395.

Holgersson, H. E. T. and Shukur, G. (2001). Some Aspects of Non-Normality Tests in Systems of Regression Equations. *Communications in Statistics – Computation and Simulation*, 30(2), 291-310.

Holgersson, H. E. T. and Shukur, G. (2004). Testing for Multivariate Heteroscedasticity, *Journal of Statistical Computation and Simulation*, 74(12), 879-896.

Hopkins, J. W. and Clay, P. P. F. (1963). Some Empirical Distributions of Bivariate T^2 and Homoscedasticity Criterion *M* Under Unequal Variance and Leptokurtosis. *Journal of the American Statistical Association*, 58(304), 1048-1053.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of American Statistical Association*, 75, 602-608.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(7), 498-520.

Jarque, C. M. and McKenzie, C. R. (1995). Testing for multivariate normality in simultaneous equations models. *Mathematics and Computers in Simulation*, 39(3-4), 323-328.

Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, 12(1), 1-38.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Econometric Dynamics and Control*, 12(2-3), 231-254.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, 90(1), 1-44.

Kelejian, H. H. (1982). An extension of a standard test for heteroskedasticity to a systems framework. *Journal of Econometrics*, 20(2), 325-333.

Kremers, J. J. M., Ericson, N. R., Dolado, J. J. (1992). The power of cointegration tests. *Oxford Bulletin of Economics and Statistics*, 54(3), 325-348.

Kanioura, A. and Turner, P. (2005). Critical values for an F-test for cointegration in a multivariate model. *Applied Economics*, 37(3), 265-270.

Larsson, R., Lyhagen, J and Löthgren, M. (2001). Likelihood-based cointegration tests in heterogeneous panels. *Econometrics Journal*, 4(1), 109-142.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303.

Lopes, A., Nezry, E., Touzi, R. and Laur, H. (1993). Structure detection and statistical adaptive speckle filtering in SAR images. *International Journal of Remote Sensing*, 14(9), 1735-1758.

Lütkepohl, H. and Theilen, B. (1991). Measures of multivariate skewness and kurtosis for tests of nonnormality. *Statistical Papers*, 32(1), 179-193.

Maddala, G. S. and Wu, S. (1999). A Comparative Study of Unit Root Test with Panel Data and a New Simple Test. *Oxford Bulletin of Economics and Statistics*, 61(S1), 631-652.

Mahalanobis, P. C. (1936). On the Generalised Distance in Statistics. *Proceedings of the National Institute of Science of India*, 2(1), 49-55.

Mantalos, P. and Shukur, G. (1998). Size and Power of the Error Correction Model Cointegration Test. A Bootstrap Approach. *Oxford Bulletin of Economics and Statistics*, 60(2), 249-255.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519-530.

McCoskey, S. and Kao, C. (1998). A residual-based test of the null of cointegration in panel data. *Econometric Reviews*, 17(1), 57–84.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. New York, NY: Wiley & Sons.

Nelson, C. M., Ogaki, M. and Donggyu, S. (2005). Dynamic Seemingly Unrelated Cointegrating Regressions. *Review of Economic Studies*, 72(3), 797-820.

Osteryoung, J. S., Scott, E. and Roberts, G. S. (1977). Selecting Capital Projects with the Coefficient of Variation. *Financial Management*, 6(2), 65-70.

Pavlenko, T. and Von Rosen, D. (2001). Effect of dimensionality on discrimination. *Statistics*, 35(3), 191-213.

Pearson, E.S. (1930). A further development of tests for normality. *Biometrika*, 22, 239-249.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11), 559-572.

Pedroni, P. (2004). Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis. *Econometric Theory, 20*(3), 597-625.

Pedroni, P. and Vogelsang, T. (2005). Robust Unit Root and Cointegration Rank Tests for Panels and Large Systems. *Department of Economics Working Papers*, Williams College, Williamstown, MA.

Phillips, P. C. B. and Ouliaris, S. (1990). Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica*, 58(1), 165-193.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). New York, NY: John Wiley & Sons.

Romeu, J. L. and Ozturk, A. (1993). A Comparative Study of Goodness-of-Fit Tests for Multivariate Normality. *Journal of Multivariate Analysis,* 46(2), 309-334.

Royston, J. P. (1983). Some Techniques For Assessing Multivariate Normality Based on the Shapiro-Wilk W. *Applied Statistics*, 32(2), 121-133.

Serdobolskii, V. I. (2008). *Multiparametric Statistics.* Amsterdam: Elsevier.

Serdobolskii, V. I. (2000). *Multivariate Statistical Analysis: A High-Dimensional Approach.* Doordrecht: Springer.

Serdobolskii, V. I. (1985). The resolvent and the spectral functions of sample covariance matrices of increasing dimension. *Russian Mathematical Surveys*, 40(2), 232-233.

Shapiro, S.S. and Wilk, M.B. (1965). An Analysis of variance Test For Normality (complete samples). *Biometrika*, 52, 591-611.

Sharpe, W. F. (1966). Mutual Fund Performance. *Journal of Business*, 39(1, Part 2), 119-138.

Silverstein, J. W. and Bai, Z. D. (1995). On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices. *Journal of Multivariate Analysis*, 54(2), 175-192.

Srivastava, M. S. and Hui, T.K. (1987). On assessing multivariate normality based on shapiro-wilk *W* statistic. *Statistics & Probability Letters*, 5(1), 15-18.

Srivastava, D. K. and Mudholkar, G. S. (2003). Goodness-of-fit Tests for Univariate and Multivariate Normal Models. In R. Khattree and C. R. Rao (Eds.), *Handbook of Statistics. Vol. 22: Statistics in Industry*, Elsevier, Amsterdam: Elsevier, pp. 869-906.

Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics & Probability Letters,* 2(5), 263-267.

Srivastava, M. S. (2007). Multivariate Theory For Analyzing High Dimensional Data. *Journal of the Japanese Statistical Society,* 37(1), 53-86.

Stock, J. H. and Watson, M. W. (1988). Testing for Common Trends. *Journal of the American Statistical Association*, 83(404), 1097-1107.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A. and Williams Jr., R. M. (1949). *Studies in Social Psychology in World War II: The American Soldier, Vol. 1, Adjustment During Army Life*. Princeton, NJ: Princeton University Press.

Westerlund, J. (2005). New Simple Tests for Panel Cointegration. *Econometric Reviews*, 24(3), 297-316.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48 (4), 817-838.

Wilks, S. S. (1946). Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality of Covariances in a Normal Multivariate Distribution. *Annals of Mathematical Statistics*, 17(3), 257-281.

Wilks, S. S. (1963). Multivariate Statistical Outliers. *Sankhya: The Indian Journal of Statistics, Series A*, 25(4), 407-426.

Wishart, J. (1928). The generalised product moment distribution in samples from a normal population. *Biometrika*, 20A(1-2), 32-52.

Yin, Y. Q., Bai, Z. D. and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields,* 78(4), 509-521.

Articles

Article 1
Assessing Normality of High-Dimensional Data
H. E. T. Holgersson and Rashid Mansoor
Published in Communications in Statistics—Simulation and Computation®, 42: 360–369, 2013

Article 2
Using Principal Components to Test Normality of High-Dimensional Data
Rashid Mansoor

Article 3
Estimating Mean-Standard Deviation Ratios of Financial Data
H. E. T. Holgersson, Peter S. Karlsson and Rashid Mansoor
Published in Journal of Applied Statistics, 39(3): 657-671, 2012

Article 4
A Bayesian Approach for Estimating Mean-Standard Deviation Ratios of Financial Data
Rashid Mansoor
Forthcoming in Journal of Statistics.

Article 5
Testing for Panel Cointegration in High-Dimensional Data in the Presence of Cross-Sectional Dependency
Rashid Mansoor and Kristofer Månsson

JIBS Dissertation Series

No. 001 Melander, Anders (1997): “*Industrial wisdom and strategic change – The Swedish pulp and paper industry 1945-1990*”, (Business Administration).

No. 002 Marmefelt, Thomas (1998): “*Bank-industry networks and economic evolution – An institutional-evolutionary approach*”, (Economics).

No. 003 Wiklund, Johan (1998): “*Small firm growth and performance – Entrepreneurship and beyond*”, (Business Administration).

No. 004 Braunerhjelm, Pontus (1999): “*Knowledge capital, firm performance and network production*”, (Economics).

No. 005 Frankelius, Per (1999): “*Företagande över tid – Kontextuellt perspektiv och visuellt beskrivningsspråk*”, (Business Administration).

No. 006 Klaesson, Johan (2001): “*A study of localisation economies and the transport sector*”, (Economics).

No. 007 Hatemi-J, Abdulnasser (2001): “*Time-series Econometrics Applied to Macroeconomic Issues*”, (Economics).

No. 008 Alhager, Eleonor (2001): “*Mervärdesskatt vid omstruktureringar, Iustus förlag AB*”, (Law).

No. 009 Hugoson, Peter (2001): “*Interregional Business Travel and the Economics of Business Interaction*”, (Economics).

No. 010 Pettersson, Lars (2002): “*Location, Housing and Premises in a Dynamic Perspective*”, (Economics).

No. 011 Paulsson, Thomas (2002): “*Decision Making, Risk, and Utility – Assessments and Applications of Alternative Decision Models*”, (Economics).

No. 012 Brundin, Ethel (2002): “*Emotions in Motion – The Strategic Leader in a Radical Change Process*”, (Business Administration).

No. 013 Wiklund, Hans (2002): “*Arenas for Democratic Deliberation – Decision-making in an Infrastructure Project in Sweden*”, (Political Science).

No. 014 Florin Samuelsson, Emilia (2002): “*Accountability and Family Business Contexts - An Interpretive Approach to Accounting and Control Practices*”, (Business Administration).

No. 015 Ahl, Helene J. (2002): “*The Making of the Female Entrepreneur – A Discourse Analysis of Research Texts on Women’s Entrepreneurship*”, (Business Administration).

No. 016 Olsson, Michael (2002): “*Studies of Commuting and Labour Market Integration*”, (Economics).

No. 017 Wigren, Caroline (2003): “*The Spirit of Gnosjö – The Grand Narrative and Beyond*”, (Business Administration).

No. 018 Hall, Annika (2003): “*Strategising in the context of genuine relations: An interpretative study of strategic renewal through family interactions*”, (Business Administration).

No. 019 Nilsson, Ulf (2003): *“Product costing in interorganizational relationships – A supplier’s perspective”*, (Business Administration).

No. 020 Samuelsson, Mikael (2004): *“Creating new ventures: A longitudinal investigation of the nascent venturing process”*, (Business Administration).

No. 021 Bruns, Volker (2004): *“Who receives bank loans? A study of lending officers’ assessments of loans to growing small and medium-sized enterprises”*, (Business Administration).

No. 022 Gustafsson, Veronica (2004): *“Entrepreneurial Decision-Making: Individuals, tasks and cognitions”*, (Business Administration).

No. 023 Agndal, Henrik (2004): *“Internationalisation as a Process of Strategy and Change – A Study of 16 Swedish Industrial SMEs”*, (Business Administration).

No. 024 Ejermo, Olof (2004): *“Perspectives on Regional and Industrial Dynamics of Innovation”*, (Economics).

No. 025 Barenfeld, Jesper (2005): *“Taxation of Cross-Border Partnerships: Double-Tax Relief in Hybrid and Reverse Hybrid Situations”*, (Law).

No. 026 Hilling, Maria (2005): *“Free Movement and Tax Treaties in the Internal Market”*, (Law).

No. 027 Brunninge, Olof (2005): *“Organisational self-understanding and the strategy process”*, (Business Administration).

No. 028 Blombäck, Anna (2005): *“Supplier brand image – a catalyst for choice: Expanding the B2B brand discourse by studying the role corporate brand image plays in the selection of subcontractors”*, (Business Administration).

No. 029 Nordqvist, Mattias (2005): *“Understanding the role of ownership in strategizing: a study of family firms”*, (Business Administration).

No. 030 Karlsson, Tomas (2005): *“Business Plans in New Ventures: An Institutional Perspective”*, (Business Administration).

No. 031 Johnson, Andreas (2005): *“Host Country Effects of Foreign Direct Investment: The Case of Developing and Transition Economies”*, (Economics).

No. 032 Nyström, Kristina (2006): *“Entry and Exit in Swedish Industrial Sectors”*, (Economics).

No. 033 Salvato, Carlo (2006): *“Micro-Foundations of Organizational Adaptation. A Field Study in the Evolution of Product Development Capabilities in a Design Firm”*, (Business Administration).

No. 034 Gråsjö, Urban (2006): *“Spatial Spillovers of Knowledge Production – An Accessibility Approach”*, (Economics).

No. 035 Dahlqvist, Jonas (2007): *“Assessing New Economic Activity – Process and Performance in New Ventures”*, (Business Administration).

No. 036 Andersson, Martin (2007): *“Disentangling Trade Flows – firms, geography and technology”*, (Economics).

No. 037 Nilsson, Désirée (2007): *“Essays on Trade Flows, Demand Structure and Income Distribution”*, (Economics).

No. 038 McKelvie, Alexander (2007): *“Innovation in New Firms: Examining the role of knowledge and growth willingness”*, (Business Administration).

No. 039 Garvi, Miriam (2007): "*Venture Capital for the Future - Implications of Founding Visions in the Venture Capital Setting*", (Business Administration).

No. 040 Rosander, Ulrika (2007): "*Generalklausul mot skatteflykt*", (Law).

No. 041 Hultman, Jens (2007): "*Rethinking adoption – Information and communications technology interaction processes within the Swedish automobile industry*", (Business Administration).

No. 042 Hilling, Axel (2007): "*Income Taxation of Derivatives and other Financial Instruments – Economic Substance versus Legal Form: A study focusing on Swedish non-financial companies*", (Law).

No. 043 Sjölander, Pär (2007): "*Simulation-Based Approaches in Financial Econometrics*", (Economics).

No. 044 Hang, Min (2007): "*Media Business Venturing: A Study on the Choice of Organizational Mode*", (Business Administration).

No. 045 Lövstål, Eva (2008): "*Management Control Systems in Entrepreneurial Organisations – A Balancing Challenge*", (Business Administration).

No. 046 Fridriksson, Helgi-Valur (2008): "*Learning processes in an inter-organizational context – A study of krAft project*", (Business Administration).

No. 047 Naldi, Lucia (2008): "*Growth through Internationalization: a Knowledge Perspective on SMEs*", (Business Administration).

No. 048 Wiberg, Daniel (2008): "*Institutional Ownership - the Anonymous Capital: Corporate Governance and Investment Performance*", (Economics).

No. 049 Eklund Johan E. (2008): "*Corporate Governance, Private Property and Investment*", (Economics).

No. 050 Glemdal, Michael (2008): "*Gubben på kullen - Om den smärtsamma skillnaden mellan politiska intentioner och praktiska resultat*", (Political Science).

No. 051 Bohman, Helena (2008): "*Trade, Knowledge and Income Distribution*", (Economics).

No. 052 Rendahl Pernilla (2008): "*Cross-Border Consumption Taxation of Digital Supplies - A Comparative Study of Double Taxation and Unintentional Non-Taxation of B2C E-Commerce*", (Law).

No. 053 Mellander, Charlotta (2008): "*The Wealth of Urban Regions - On the Location of Creative Individuals and Firms*", (Economics).

No. 054 Johansson, Monica (2008): "*Organizing Policy - A Policy Analysis starting from SMEs in Tuscany and the County of Jönköping*", (Political Science).

No. 055 van Weezel, Aldo (2009): "*Entrepreneurial Strategy-Making Mode and Performance: A Study of the Newspaper Industry*", (Business Administration)

No. 056 Hellerstedt, Karin (2009): "*The Composition of New Venture Teams: Its Dynamics and Consequences*", (Business Administration).

No. 057 Hunter, Erik (2009): "*Celebrity Entrepreneurship and Celebrity Endorsement: Similarities, differences and the effect of deeper engagement*", (Business Administration).

No. 058 Gerson, Anna (2009): "*Compensation of Losses in Foreign Subsidiaries within the EU: A Comparative Study of the Unilateral Loss-Compensation Mechanisms in Austria and Denmark*", (Law).

No. 059 Dahlström, Tobias (2009): "*Causes of corruption*", (Economics).

No. 060 Languilaire, Jean-Charles (2009): "*Experiencing work/non-work - Theorising individuals' process of integrating and segmenting work, family, social and private*", (Business Administration).

No. 061 Nguyen Tan, Phat (2009): "*Transfer Pricing: The Vietnamese System in the Light of the OECD Guidelines and the Systems in certain Developed and Developing Countries*", (Law).

No. 062 Karlsson, Ann Britt (2009): "*Institutionalisering av ansvar i kommunal revision – Lärande organisering*", (Political Science).

No. 063 Johansson, Sara (2010): "*Knowledge, Product Differentiation and Trade,* (Economics).

No. 064 Ots, Mart (2010): "*Understanding value formation - A study of marketing communications practices at the food retailer ICA*", (Business Administration).

No. 065 Raviola, Elena (2010): "*Paper meets Web: How the institution of news production works on paper and online*", (Business Administration).

No. 066 Palmberg, Johanna (2010): "*Family Ownership and Investment Performance*", (Economics).

No. 067 Borgström, Benedikte (2010): "*Supply chain strategising: Integration in practice*", (Business Administration).

No. 068 Wikner, Sarah (2010): "*Value co-creation as practice: On a supplier's capabilities in the value generation process*", (Business Administration).

No. 069 Karlsson, Peter (2011): "*Issues of Incompleteness, Outliers and Asymptotics in High-Dimensional Data*", (Statistics).

No. 070 Helin, Jenny (2011): "*Living moments in family meetings: A process study in the family business context*", (Business Administration).

No. 071 Markowska, Magdalena (2011): "*Entrepreneurial Competence Development: Triggers, Processes & Consequences*", (Business Administration).

No. 072 Cui, Lianguang (2011): "*Innovation and network development of logistics firms*", (Business Administration).

No. 073 Norbäck, Maria (2011): "*Making Public Service Television: A study of institutional work in collaborative TV production*", (Business Administration).

No. 074 Dzansi, James (2011): "*Essays on Financing and Returns on Investment*", (Economics).

No. 075 Månsson, Kristofer (2012): "*Issues of multicollinearity and conditional heteroscedasticy in time series econometrics*", (Statistics).

No. 076 Balkow, Jenny (2012): "*In the Middle: On Sourcing from China and the Role of the Intermediary*", (Business Administration).

No. 077 Karlsson, Hyunjoo Kim (2012): "*Dynamics of macroeconomic and financial variables in different time horizons*", (Economics).

No. 078 Bjerke, Lina (2012): "*Knowledge flows across space and firms*", (Economics).

No. 079 Högberg, Andreas (2012): "*Corporate Governance, Legal Origin and Firm Performance: An Asian Perspective*", (Economics).

No. 080 Wictor, Ingemar (2012): *"Born Globals: Rapid International Growth in New Ventures"*, (Business Administration).

No. 081 Skoglund, Per (2012): "*Sourcing decisions for military logistics in Peace Support Operations: A case study of the Swedish armed forces"*, (Business Administration).

No. 082 Haag, Kajsa (2012): "*Rethinking family business succession: From a problem to solve to an ongoing practice"*, (Business Administration).

No. 083 Zeebari, Zangin (2012): "*On Median and Ridge Estimation of SURE Models"*, (Statistics).

No. 084 Jenkins, Anna (2012): *"After firm failure: Emotions, learning and re-entry"*, (Business Administration).

No. 085 Ghazawneh, Ahmad (2012): *"Towards a Boundary Resources Theory of Software Platforms"*, (Informatics).

No. 086 Backman, Mikaela (2013): *"Regions, Human Capital and New Firm Formation"*, (Economics).

No. 087 Casanovas, Inés (2013): "*Online Education in Universities: Moving from Individual Adoption to Institutionalisation of an Information Technology Innovation"*, (Informatics).

No. 088 Nilsson, Pia (2013): *"Price Formation in Real Estate Markets"*, (Economics).

No. 089 Sallander, Ann-Sophie (2013): "*Ömsesidiga överenskommelser enligt skatteavtal"*, (Law).

No. 090 Trenta, Cristina (2013): "*VAT in Peer-to-peer Content Distribution – Towards a Tax Proposal for Decentralized Networks"*, (Law).

No. 091 Boers, Börje (2013): "*Organizational identity construction in family businesses a dualities perspective"*, (Business Administration).

No. 092 Mansoor, Rashid (2013): "Assessing Distributional Properties of High-Dimensional Data", (Statistics).