Accurate finite sample inference for generalized linear models and models on overidentifying moment conditions

LÔ, Serigne Ndame

Abstract

Classical inference in statistic and econometric models is typically carried out by means of asymptotic approximations to the sampling distribution of estimators and test statistics. These approximations often do not provide accurate p-values and confidences intervals, especially when the sample size is small. Moreover, even if the sample size is large, the accuracy can be poor due to model misspecification (nonrobustness). Several alternative techniques have been proposed in the statistic and econometric literature to improve the accuracy of classical inference. In general, these alternatives address either the accuracy of the first-order approximations or the nonrobustness issue. However, the development of general procedures which are both robust and second order accurate is still an open question. In this thesis, we propose an alternative statistical test which has both robustness and small sample properties for two large and important classes of models: Generalized Linear Models (GLM) and models on overidentifying moments conditions.

Reference


URN : urn:nbn:ch:unige-4057
DOI : 10.13097/archive-ouverte/unige:405

Available at: http://archive-ouverte.unige.ch/unige:405

Disclaimer: layout of this document may differ from the published version.
ACCURATE FINITE SAMPLE INFERENCE FOR GENERALIZED LINEAR MODELS AND MODELS ON OVERIDENTIFYING MOMENT CONDITIONS

by Serigne Ndame Lô

Submitted for the degree of Ph.D in Econometrics and Statistics

Accepted on the recommendation of:

Prof. Christopher Field, Dalhousie University, Canada,
Prof. Jaya Krishnakumar, University of Geneva,
Prof. Elvezio Ronchetti, University of Geneva, thesis advisor,
Prof. Maria-Pia Victoria-Feser, University of Geneva.

Thesis number 607

Geneva, June 2006
La Faculté de sciences économiques et sociales, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par là, émettre aucune opinion sur les propositions qui s’y trouvent énoncées et qui n’engagent que la responsabilité de leur auteur.

Genève, le 30 juin 2006

Le Doyen:

Pierre ALLAN
To my family
Acknowledgments

First and foremost, I would like to express my gratitude to Professor Elvezio Ronchetti for his generous and precious guidance and his suggestions all along the way. My special thanks to him for having robustly taught me about statistical research.

I would also like to thank and express my sincere appreciation to Professor Maria-Pia Victoria-Feser and Dr. Eva Cantoni for their constant encouragement and support and their generous availability during the last five years. I gratefully acknowledge the trust they gave me during our collaboration.

My thanks as well to Professor Jaya Krishnakumar and Professor Christopher Field for the time they sacrificed to read the manuscript and for their invaluable assistance, their suggestions, and their comments which have greatly contributed in the clarity of the presentation.

I also would like to express my gratitude to Dr. Claude Tadonki without whom my work would not have seen the light, namely through his precious help in programming in Matlab. There is no doubt that his computing skills are as high as his generosity which I really appreciate.

My thanks also go to my friends Papa Diadji Gueye and Mostefa Bachtouti for their assistance in the correction of my english in the thesis.

Also, I am thankful to my professors and my friends and colleagues of the Faculty of Economics and Social Science of the University of Geneva especially Dominique Couturier, Ilir Roko and Silva Ohannessian for their useful discussions and comments.

Finally, I thank all my friends for their moral support throughout the project of my thesis in particular, Abdoul W. Dieng, Cheikh T. Dieye, Mbaye Dione, Djibril Fall and Mawa Thiam. Special thanks should go also to Mbathio Dieng.
Contents

1 Introduction 1
   1.1 Edgeworth Expansions ............................. 1
   1.2 Saddlepoint Approximations ........................ 3
   1.3 The Robustness Issue .............................. 5
   1.4 Combining Robustness and Second Order Accuracy .... 6
   1.5 Robust Second Order Accurate Inference for GLM .... 7
   1.6 Robust Second Order Accuracy for GMM .............. 9
   1.7 Outline ............................................ 11

2 Robust Second-order Accurate Inference for Generalized Linear Models 17
   2.1 Introduction ......................................... 20
   2.2 Classical and Robust Inference for Generalized Linear Models ... 22
   2.3 Second Order Accuracy and Robustness .............. 25
       2.3.1 Saddlepoint Test Statistic .................... 25
       2.3.2 Saddlepoint Test Statistic with Classical Score Function ... 27
       2.3.3 Saddlepoint Test Statistic with Robust Score Function ... 30
   2.4 Monte Carlo Study .................................. 33
   2.5 A Real Data Example ................................ 38
   2.6 Conclusion .......................................... 40
   2.7 APPENDIX ............................................. 41
       2.7.1 Classical Saddlepoint Test Statistic .......... 41
       2.7.2 Explicit formulas for $I_{ij}$, j=1,2,3 ............ 42
       2.7.3 Robust Saddlepoint Test Statistics .......... 44
       2.7.4 Some Special Robust Saddlepoint Test Statistics ... 48
       2.7.5 Code ........................................... 52

3 Robust Small Sample Accurate Inference in Moment Condition 61
   3.1 Introduction .......................................... 64
   3.2 Exponential tilting .................................. 67
   3.3 Robust Exponential Tilting .......................... 70
       3.3.1 Robust alternatives to the GMM .................. 70
       3.3.2 Robust exponential tilting estimator and test .... 71
   3.4 Monte Carlo Investigation ......................... 75
3.4.1 Model 1: Chi-squared Moments ......................... 76
3.4.2 Model 2: Hall-Horowitz (1996) .......................... 83
3.4.3 Model 3: Stochastic Lognormal Volatility Model ........ 87
3.5 Conclusion .............................................. 92
3.6 APPENDIX .................................................. 93
  3.6.1 The Algorithm ........................................ 93
  3.6.2 Computational aspects ................................. 94
  3.6.3 Code ............................................... 94
Abstract

Classical inference in statistic and econometric models is typically carried out by means of asymptotic approximations to the sampling distribution of estimators and test statistics. These approximations often do not provide accurate p-values and confidences intervals, especially when the sample size is small (first order approximations). Moreover, even if the sample size is large, the accuracy can be poor due to model misspecification (nonrobustness). Several alternative techniques have been proposed in the statistic and econometric literature to improve the accuracy of classical inference. In general, these alternatives address either the accuracy of the first-order approximations or the nonrobustness issue. However, the development of general procedures which are both robust and second order accurate is still an open question.

In this thesis, we propose an alternative statistical test which has both robustness and small sample properties for two large and important classes of models: Generalized Linear Models (GLM) and models on overidentifying moments conditions. The choice of these models is motivated by their wide use in statistics and econometrics. The former is very popular in many fields as various as psychology, marketing, economics and medical research. The latter, usually estimated by means of generalized method of moment (GMM), is commonly applied to complex non-linear models very often encountered in economics and finance.

As for the GLM, we will combine results from Cantoni and Ronchetti (2001), and Robinson, Ronchetti and Young (2003) to obtain a new robust second-order accurate test statistic for hypothesis testing and variable selection. Concerning the GMM, we will combine the boundedness of the orthogonality conditions (robustness) with information and entropy econometrics (or saddlepoint) techniques
(second-order) to define an alternative statistical test for overidentifying moment restrictions. In both cases, the alternative test is asymptotically $\chi^2$ distributed under the null hypothesis as the classical tests (likelihood ratio, score, Wald, and Hansen test) but with a relative error of order $O(n^{-1})$. Moreover, the accuracy of these new tests statistics is stable in a neighborhood of the model distribution and this leads to robust inference even in moderate to small samples. By a Monte Carlo investigation, we illustrate the performance of the new statistic in each model.
Résumé

Dans les modèles économétriques et statistiques, l’inférence classique est principalement basée sur une approximation asymptotique de la distribution empirique des estimateurs et des statistiques de test. Ces approximations conduisent souvent à des p-valeurs et des intervalles de confiance imprecis particulièrement quand l’échantillon est de petite taille (approximations de premier-ordre). En outre, dans le cas d’échantillon de moyenne à grande taille même, l’approximation de la distribution peut être inappropriée à cause du non respect des hypothèses sous-jacentes au modèle (non-robustesse). Plusieurs alternatives ont été proposées dans la littérature statistique et économétrique pour améliorer la précision de l’inférence classique. Généralement, ces alternatives s’occupent soit de l’approximation de premier-ordre, soit de la non-robustesse. Dès lors, le développement de procédures générales à la fois robustes et de second-ordre demeure un champs d’investigation.

Dans cette thèse, nous proposons un test statistique qui possède à la fois des propriétés de robustesse et une meilleure précision pour les échantillons de petites tailles, que nous appliquons à deux grandes classes de modèles caractérisées par leur large utilisation en statistique et en économie :

- les modèles linéaires généralisés (GLM), souvent utilisés pour analyser des données issues des recherches dans le domaine médical et des sciences humaines et sociales,

- les modèles aux moments suridentifiés (models on overidentifying moment condition), généralement estimés par le biais de la méthode des moments généralisés (GMM) et appliqués dans la spécification des modèles non-linéaires fréquemment rencontrés dans l’analyse quantitative économique et financière.
Le test développé dans le cadre des modèles GLMs est issu de la combinaison des résultats de Cantoni and Ronchetti (2001) et de Robinson, Ronchetti and Young (2003). Cette combinaison permet d'obtenir une nouvelle statistique robuste et de second-ordre qui peut être utilisée pour tester les hypothèses de test et pour la sélection de variables. Dans le cadre du GMM, nous procédons en bornant les conditions d'orthogonalité dans le cadre de la technique dite *information and entropy econometrics* (IEE) (ou dans celle du point de selle) pour fournir un nouveau test de spécification (test for overidentifying moment restriction) à la fois robuste et de second-ordre. Tout comme dans le cas des tests classiques (test du rapport de vraisemblance, du score, de Wald et de Hansen), les deux nouvelles statistiques de test que nous dérivons suivent une distribution asymptotique du $\chi^2$ sous l’hypothèse nulle mais avec une erreur relative d’ordre $O(n^{-1})$. La précision de ces nouveaux tests est stable au voisinage de la distribution du modèle et ceci nous mène à une inférence robuste même si la taille de l’échantillon est petite. Par une étude de Monte Carlo, nous illustrons la performance de la nouvelle statistique dans chaque modèle.
Chapter 1

Introduction

Classical inference in statistic and econometric models is typically carried out by means of asymptotic approximations to the sampling distribution of estimators and test statistics. These approximations often do not provide accurate p-values and confidence intervals, especially when the sample size is small. Moreover, even if the sample size is large, the accuracy can be poor in the far tails. Several alternatives have been proposed to improve the accuracy of asymptotic distribution. We restrict our attention to two important classes of techniques, Edgeworth and saddlepoint approximations.

1.1 Edgeworth Expansions

The idea of this technique goes back to Chebyshev (1890) and Edgeworth (1905). General references are James and Mayne (1962), Barndorff-Nielsen and Cox (1989), and Field and Ronchetti (1990, chap. 2). The basic idea is as follows. According to the Berry-Esseen theorem, the absolute error between the asymptotic normal distribution and the exact distribution of some statistic is of order $n^{-\frac{1}{2}}$, where $n$ is the sample size. This result suggests a way to improve the approximation of the exact sampling distribution by considering a complete asymptotic
expansion. In the case of the mean, Edgeworth proposed a series to approximate the exact distribution of the mean by considering a complete asymptotic expansion with terms in powers of \( n^{-\frac{1}{2}} \), where the leading term is the normal distribution.

Substantial work has been done on Edgeworth expansions. Theoretical results on its use and validity have been established in a very general setting. Some useful results and algorithms applying this method to statistics more general than the standardized mean have been published, see for instance Chibishov (1972), Bickel (1974) for nonparametric statistics, Bhattacharya and Rao (1976), Bhattacharya and Ghosh (1978), Phillips (1984) for the likelihood ratio test, Wald test and Lagrange multiplier test, Bickel, Götze, and van Zwet (1986) for U-statistics, and Taniguchi (1991) for higher order asymptotic results in time series analysis. In addition, Edgeworth expansions have been applied successfully to study the accuracy of the bootstrap. Hall (1992) developed and discussed the Edgeworth expansion as an approximation to distributions of estimators which can be written as smooth functions of means. Singh (1981), and DiCiccio and Romano (1988) used Edgeworth expansions to study the efficiency of the bootstrap method in classical finite parameter problems.

Edgeworth expansion approximations still have some drawbacks. The first one is that the resulting infinite series does not always converge. In fact, adding higher order terms does not necessarily improve the approximation and may make it worse. A second problem, encountered when the sample size is moderate, is that the approximation deteriorates in the tails, where it may even become negative.

Another disadvantage of this method is that the justification of its accuracy is based on the absolute error. Particularly in inferential statistics where we deal with very small values such as probabilities, the absolute error does not reflect the
level of accuracy. An alternative, more stringent tool to evaluate the accuracy is the relative error, which is defined by the difference between the approximation and the exact distribution divided by the exact. In practice, a bounded relative error implies a bounded absolute error, but the opposite is not guaranteed. This is the case of Edgeworth approximation where generally, only the absolute error is bounded. To address this problem, we devote the next section to a better approximation technique which has relative error of order $O(n^{-1})$.

1.2 Saddlepoint Approximations

A more accurate method was introduced by Daniels (1954) under the name of saddlepoint approximation. Daniels shed new light on the method of steepest descent by proving a very accurate approximation to the density of the mean of a sample of $n$ independent identically distributed (iid) random variables. General references are Barndorff-Nielsen (1978, chap. 5), Barndorff-Nielsen and Cox (1989, chap. 6), Field and Ronchetti (1990, chap. 3), and Jensen (1995). There are two basic steps to derive a saddlepoint approximation. First, the exact density of the mean can be expressed as the inversion integral of its characteristic function in the complex plane (see Field and Ronchetti (1990, chap. 3, section 3.3)). Secondly, applying the method of steepest descent to the integral, one can derive an expansion where the leading term is called the saddlepoint approximation to the density of the mean. An alternative way to derive this approximation is through the technique of a conjugate or associate distribution; see the references above.

The main property of this technique is that it leads to very accurate approximations of the exact distribution. That is, the saddlepoint approximation has relative error of order $O(n^{-1})$. Unlike Edgeworth expansions, saddlepoint tech-
niques gives very accurate approximations in the far tails of the distribution even for very small sample sizes. In addition, saddlepoint approximations cannot be negative and the error of the approximation tends to be locally stable.


Saddlepoint approximations are not limited to parametric statistics but their applications field has been extended to nonparametric second order analysis. Ronchetti and Welsh (1994) investigated the properties of empirical saddlepoint approximations. Davison and Hinkley (1988) showed that saddlepoint approximations can often replace simulation with excellent results in a variety of bootstrap problems. DiCiccio, Martin, and Young (1992) suggested techniques for constructing accurate approximate iterated bootstrap confidence intervals based on the saddlepoint method. Robinson et al. (2003) investigated the case where the underlying distribution is unknown and defined an empirical exponential likelihood test.
1.3 The Robustness Issue

Models are idealized approximations to reality and deviations from the assumed distribution can have important effects on classical estimators and tests based on these models (nonrobustness). In spite of their second order accuracy, standard classical inference based on saddlepoint technique can be drastically affected by small deviations from the underlying assumptions on the model. Robust statistics deals with such deviation and reliable tools, covering a variety of models, are now available; see for instance the books Huber (1981), and Hampel, Ronchetti, Rousseeuw, and Stahel (1986). While several approaches have been proposed in the literature, we focus here on the approach based on influence function.

Boundedness of the influence function implies that in a neighborhood of the model, the level of a test does not become arbitrarily close to 1 (robustness of validity) and the power does not become arbitrarily close to 0. Hence, a bounded influence function is a desirable local stability property of a test statistic. Then, for reliable inference, robust statistics should be used especially when the underlying distribution of the observations does not belong to the model but is assumed to lie in a neighborhood. In the framework of M-estimates, the influence function of the estimators is proportional to the underlying orthogonality function. Therefore, the stability of the level of tests is obtained by bounding the underlying orthogonality function on the model. This leads to robust versions of Wald, score and likelihood ratio tests based on M-estimators; see Heritier and Ronchetti (1994).
1.4 Combining Robustness and Second Order Accuracy

Several alternative techniques have been proposed in the statistics and econometric literature to improve the accuracy of inference. These alternatives address either the accuracy of the first order approximations or the nonrobustness issue. However, a procedure which is both robust and second order accurate is still an open question. In this thesis, we propose an alternative statistical test which has both robustness and small sample properties for two large and important classes of models: Generalized Linear Models (GLM) and models on overidentifying moments conditions. The choice of these estimating models is motivated by their wide use in statistics and econometrics. The former is very popular in many fields as various as psychology, marketing, economy and medical research. The latter, usually estimated by means of generalized method of moment (GMM), is commonly applied to complex non-linear models very often encountered for instance in finance.

As for the GLM, we will combine results from Cantoni and Ronchetti (2001) and Robinson et al. (2003) to obtain a new test statistic for hypothesis testing and variable selection which is asymptotically $\chi^2$ distributed as the three classical tests but with a relative error of order $O(n^{-1})$. Moreover, the accuracy of the new test statistic is stable in a neighborhood of the model distribution and this leads to robust inference even in moderate to small samples. Concerning the GMM, the boundedness of the orthogonality conditions will preserve the robustness properties of the estimator proposed by Ronchetti and Trojani (2001), while information and entropy econometrics (or saddlepoint) techniques will improve their finite sample performance.
1.5 Robust Second Order Accurate Inference for GLM

Generalized linear models (McCullagh and Nelder, 1989) have become the most commonly used class of regression models in the analysis of a large variety of data. This class includes classical linear models, logit and probit models for proportions, log-linear models for counts and some commonly used models for survival analysis. In particular, generalized linear model can be used to model the relationship between predictors and a function of the mean of a continuous or discrete response variable. Let $Y_1, \ldots, Y_n$ be $n$ independent observations of a response variable. Assume that the distribution of $Y_i$ belongs to the exponential family with $E[Y_i] = \mu_i$ and $Var[Y_i] = V(\mu_i)$, and

$$g(\mu_i) = \eta_i = x_i^T \beta, \quad i = 1, \ldots, n,$$

(1.1)

where $\beta \in \mathbb{R}^q$ is the vector of unknown parameters, $x_i \in \mathbb{R}^q$, and $g(.)$ is the link function.

The estimation of $\beta$ can be carried out by maximum likelihood or quasi-likelihood methods, which are equivalent if $g(.)$ is the canonical link, such as the logit function for logistic regression or the log for Poisson regression. Standard asymptotic inference based on likelihood ratio, Wald and score tests is then readily available for these models. However, there are two main problems which can potentially invalidate p-values and confidence intervals based on standard classical techniques.
First, the models are idealized approximations to reality and deviations from the assumed distribution can have important effects on classical estimators and tests for these models (nonrobustness). Second, even when the model is exact, standard classical inference is based on (first order) asymptotic theory and has absolute error of order $O(n^{-1/2})$. This can lead to inaccurate p-values and confidence intervals when the sample size is moderate to small or when probabilities in the extreme tails are required.

The nonrobustness of classical estimators and tests for parameters is a well known problem. Alternative methods have been suggested, see for instance Pregibon (1982), Stefanski, Carroll, and Ruppert (1986), Künsch, Stefanski, and Carroll (1989), Morgenthaler (1992), Bianco and Yohai (1996), Ruckstuhl and Welsh (2001), Cantoni and Ronchetti (2001), Victoria-Feser (2002), and Croux and Haesbroeck (2003). These methods are robust and can cope with deviations from the assumed distribution. However, they are based on first order asymptotic theory.

For these reasons, we would like to investigate alternative test statistics which combine robustness and good accuracy with small sample sizes. To build this new statistic test for hypothesis testing and variable selection, we combine two main results. The first one is the class of Mallows’s type M-estimators, developed by Cantoni and Ronchetti (2001). The second is the saddlepoint approximation to the density function of the M-estimates. To investigate the accuracy and the robustness of the new statistic, we will compare it with the classical test analytically and by the means of simulations.
1.6 Robust Second Order Accuracy for GMM

Procedures based on the generalized method of moments (GMM) (Hansen, 1982) are important tools in econometrics to estimate and to test the parameters of (highly) non-linear models. GMM is used to estimate the parameters of non-linear models and to test hypotheses about them in a computationally tractable way. In most cases, the theory available for conducting inference with these tools is asymptotic. It has been shown that GMM estimators converge asymptotically to a normal distribution and that the standard classical statistics for hypothesis testing converge asymptotically to a \( \chi^2 \) distribution. The main well-known problem is that the asymptotic distribution does not provide accurate p-values and confidence intervals. Moreover, if the underlying distribution of the observations does not belong to the model but is assumed to lie in a neighborhood, p-values and confidence intervals can be affected. To alleviate these problems, researchers have put forward solutions primarily in two different directions.

The first approach consists in improving classical GMM procedures. It is shown that imposing additional restrictions can appreciably improve inference based on GMM. An extended development was presented in the July 1996’s special issue of *Journal of Business and Economic Statistics*. In this issue, Hansen, Heaton, and Yaron, opted for continuous updating estimators. Other authors such as Christiano and Haan have found that imposing certain restrictions leads to substantial improvements in the small-sample properties of the statistical tests. Andersen and Sørenson stressed that it is generally not optimal to include many moments in the estimation procedure if the sample size is limited. In fact, the preferred number of moments is typically lower than the standard choice in the literature concerned with estimation on the basis of high-frequency financial data.
In the same line of thought, other authors have attempted to refine the asymptotic distribution of a GMM statistic by means of bootstrap techniques. Hall and Horowitz (1996) gave conditions under which the bootstrap provides asymptotic refinements to the critical values of $t$-tests and tests for overidentifying moment restrictions. In addition, Ronchetti and Trojani (2001) derived a robust GMM estimator that generates robust tests for a broad class of GMM test statistics. Though important improvements are noted in these procedures, they are all still based on first order asymptotic theory.

Other researchers have proposed to improve accuracy by developing a second order accurate inference with new statistical tests. Robinson et al. (2003) suggested parametric and non-parametric statistics based on saddlepoint approximations. Similarly, Imbens, Spady, and Johnson (1998) introduced a new class of tests based on the direct distance between the empirical distribution function and the nearest distribution function satisfying the moment restrictions. This technique, derived from information and entropy econometrics (IEE), was investigated in detail in the March 2002 special issue of the *Journal of Econometrics*. IEE can be used as an alternative method in the search of small sample statistics properties. IEE is also used in statistical inference for economic problems given incomplete knowledge or data, as well as in diagnostics and analysis of statistical properties of information measures.

Both procedures listed above give the same statistical test for overidentifying moment restrictions in the fully identified case (M-estimator), where the estimated parameter vector is the M-estimator defined by the orthogonality condition. Under the model, the alternative test has relative error of order $O(n^{-1})$. In spite of this second order accuracy, the statistic can be drastically affected by small deviations from the underlying assumptions on the model. Furthermore, when we are
not on the fully identified case, the equivalence between the statistic for testing overidentifying moment restrictions based on GMM and those based on IEE is not established.

Thus, we are motivated to find new statistics which combine both robustness properties and small sample performance. Specifically, the boundedness of the orthogonality conditions will preserve the robustness properties of the estimator proposed by Ronchetti and Trojani (2001), while information and entropy econometrics (or saddlepoint) techniques will improve the finite sample performance. Finally, we illustrate the accuracy of the new statistic by means of simulations for three classical models in the overidentifying moments condition.

1.7 Outline

This thesis is organized as a collection of two articles. The first concerns the robust second order accurate inference for Generalized Linear Models. The second focuses on the second order accuracy for Generalized Method of Moments. At the end of each article, we present detailed codes used for the computation of the new technique in the form of an appendix.
Bibliography


Chapter 2
Robust Second-order Accurate Inference for Generalized Linear Models

by

Serigne N. Lô and Elvezio Ronchetti
Department of Econometrics
University of Geneva
CH-1211 Geneva, Switzerland

May 2006
Abstract

In the framework of generalized linear models, the nonrobustness of classical estimators and tests for the parameters is a well known problem and alternative methods have been proposed in the literature. These methods are robust and can cope with deviations from the assumed distribution. However, they are based on first order asymptotic theory and their accuracy in moderate to small samples is still an open question. In this paper we propose a test statistic which combines robustness and good accuracy for moderate to small sample sizes. We combine results from Cantoni and Ronchetti (2001) and Robinson, Ronchetti and Young (2003) to obtain a robust test statistic for hypothesis testing and variable selection which is asymptotically $\chi^2$-distributed as the three classical tests but with a relative error of order $O(n^{-1})$. This leads to reliable inference in the presence of small deviations from the assumed model distribution and to accurate testing and variable selection even in moderate to small samples.

Keywords: M-estimators, Monte Carlo, Robust inference, Robust variable selection, Saddlepoint techniques.
2.1 Introduction

Generalized linear models (GLM) (McCullagh and Nelder, 1989) have become the most commonly used class of models in the analysis of a large variety of data. In particular, GLM can be used to model the relationship between predictors and a function of the mean of a continuous or discrete response variable. Let $Y_1, ..., Y_n$ be $n$ independent observations of a response variable. Assume that the distribution of $Y_i$ belongs to the exponential family with $E[Y_i] = \mu_i$ and $Var[Y_i] = V(\mu_i)$, and

$$g(\mu_i) = \eta_i = x_i^T \beta, \quad i = 1, ..., n, \quad (2.1)$$

where $\beta \in \mathbb{R}^q$ is a vector of unknown parameters, $x_i \in \mathbb{R}^q$, and $g(.)$ is the link function.

The estimation of $\beta$ can be carried out by maximum likelihood or quasi-likelihood methods, which are equivalent if $g(.)$ is the canonical link, such as the logit function for logistic regression or the log for Poisson regression. Standard asymptotic inference based on likelihood ratio, Wald and score test is then readily available for these models.

However, two main problems can potentially invalidate p-values and confidence intervals based on standard classical techniques. First of all, the models are ideal approximations to reality and deviations from the assumed distribution can have important effects on classical estimators and tests for these models (nonrobustness). Secondly, even when the model is exact, standard classical inference is based on (first order) asymptotic theory. This can lead to inaccurate p-values and confidence intervals when the sample size is moderate to small or when probabilities in the extreme tails are required (and in some cases both are required). Since
these tests are typically used for model comparison and variable selection, these problems can have important implications in the final choice of the explanatory variables. Consider for instance the data set discussed in section 2.5 where a Poisson regression is used to model adverse events of a drug on 117 patients affected by Crohn’s disease (a chronic inflammatory disease of the intestine) by means of 7 explanatory variables describing the characteristics of each patient. In this case a classical variable selection tends to include too many explanatory variables, while a deviance analysis obtained using our new test gives more reliable results; see section 2.5.

The nonrobustness of classical estimators and tests for $\beta$ is a well known problem and alternative methods have been proposed in the literature; see, for instance Pregibon (1982), Stefanski, Carroll, and Ruppert (1986), Künsch, Stefanski, and Carroll (1989), Morgenthaler (1992), Bianco and Yohai (1996), Ruckstuhl and Welsh (2001), Cantoni and Ronchetti (2001), Victoria-Feser (2002), and Croux and Haesbroeck (2003). These methods are robust and can cope with deviations from the assumed distribution. However, they are based on first order asymptotic theory and their accuracy in moderate to small samples is still an open question.

In this paper we propose a test statistic which combines robustness and good accuracy for small sample sizes. We combine results from Cantoni and Ronchetti (2001) and Robinson, Ronchetti, and Young (2003) to obtain a test statistic for hypothesis testing and variable selection in GLM which is asymptotically $\chi^2$-distributed as the three classical tests but with a relative error of order $O(n^{-1})$. This is in contrast with the absolute error of order $O(n^{-\frac{1}{2}})$ for the classical tests. Moreover, the accuracy of the new test statistic is stable in a neighborhood of the model distribution and this leads to robust inference even in moderate to small samples. The new test statistic is easily computed. Given a robust estimator for
\( \beta \), it has an explicit form in the case of a simple hypothesis and it requires a simple additional minimization in the case of a composite hypothesis.

The paper is organized as follows. Section 2.2 reviews the classical and robust estimators for GLM. In section 2.3 we provide a second order accurate test statistic based on saddlepoint approximations. We apply this statistic to GLM with a classical score function to derive a classical saddlepoint test statistic. Then, by replacing the classical score function by its robust version, we propose a robust saddlepoint test statistic. Section 2.4 presents a simulation study in the case of Poisson regression which shows the advantage of robust saddlepoint tests with respect to standard classical tests. The new procedure is applied on a real data example in section 2.5. Finally, section 2.6 concludes the article with some potential research directions.

### 2.2 Classical and Robust Inference for Generalized Linear Models

Let \( Y_1, \ldots, Y_n \) be \( n \) of independent random variables with density (or probability function) belonging to the exponential family

\[
 f_Y(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y; \phi) \right\}, \tag{2.2}
\]

for some specific functions \( a(\cdot), b(\cdot) \) and \( c(\cdot) \). Then \( E[Y_i] = \mu_i = b'(\theta_i) \) and \( Var[Y_i] = b''(\theta_i)a(\phi) \). Given \( n \) observations \( x_1, \ldots, x_n \) of a set of \( q \) explanatory variables \( (x_i \in \mathbb{R}^q) \), (2.1) defines the relationship between a linear predictor of the \( x_i \)'s and a function \( g(\mu_i) \) of the mean response \( \mu_i \). When \( g(\mu_i) \) is the canonical link, the maximum likelihood estimator and the quasi-likelihood estimator of \( \beta \) are the solution of the system of equations.
\[
\sum_{i=1}^{n} (y_i - \mu_i) \cdot x_{ij} = 0, \quad j = 1, \ldots, q, \quad (2.3)
\]
where \( \mu_i = g^{-1}(x_i^T \beta). \)

The maximum likelihood and the quasi-likelihood estimator defined by (2.3) can be viewed as an M-estimator (Huber, 1981) with score function

\[
\psi(y_i; \beta) = (y_i - \mu_i) \cdot x_i, \quad (2.4)
\]
where \( x_i = (x_{i1}, \ldots, x_{iq})^T. \)

Since \( \psi(y; \beta) \) is in general unbounded in \( x \) and \( y \), the influence function of the estimator defined by (2.3) is unbounded and the estimator is not robust; see Hampel, Ronchetti, Rousseeuw, and Stahel (1986). Several alternatives have been proposed. One of these methods is the class of M-estimators of Mallows’s type (Cantoni and Ronchetti 2001) defined by the score function:

\[
\psi(y_i; \beta) = \nu(y_i, \mu_i) w(x_i) \mu_i' - \tilde{a}(\beta), \quad (2.5)
\]
where \( \tilde{a}(\beta) = \frac{1}{n} \sum_{i=1}^{n} E[\nu(y_i, \mu_i)] w(x_i) \mu_i' \), \( \mu_i' = \frac{\partial \mu_i}{\partial \beta} \),
\(
\nu(y_i, \mu_i) = \psi_c(r_i) \frac{1}{V^{1/2}(\mu_i)}, \quad r_i = \frac{y_i - \mu_i}{V^{1/2}(\mu_i)} \)
are the Pearson residuals, \( V^{1/2}(\cdot) \) the square root of the variance function, and \( \psi_c \) is the Huber function defined by

\[
\psi_c(r) = \begin{cases} 
  r & |r| \leq c \\
  c \cdot \text{sign}(r) & |r| > c.
\end{cases}
\]

When \( w(x_i) = 1 \), we obtain the so-called Huber quasi-likelihood estimator.

The tuning constant \( c \) is typically chosen to ensure a given level of asymptotic efficiency and \( \tilde{a}(\beta) \) is a correction term to ensure Fisher consistency at the model.
(Notice that the tuning constant $c$ does not have anything to do with the function $c$ defined in (2.2).) $\tilde{a}(\beta)$ can be computed explicitly for binomial and Poisson models and does not require numerical integration. The advantage of this estimator is that standard inference based on robust quasi-deviances is available; see Cantoni and Ronchetti (2001). This will allow us to compare our new robust test with classical and robust tests based on first order asymptotic theory.
2.3 Second Order Accuracy and Robustness

2.3.1 Saddlepoint Test Statistic

Let \( Y_1, ..., Y_n \) be an independent, identically distributed sample of random vectors from a distribution \( F \) on some sample space \( \mathcal{Y} \). Define the M-functional \( \beta(F) \) to satisfy

\[
E[\psi(Y; \beta)] = 0, \tag{2.6}
\]

where \( \psi \) is assumed to be a smooth function from \( \mathcal{Y} \times \mathbb{R}^q \rightarrow \mathbb{R}^q \) with \( q = \text{dim}(\beta) \) and the expectation is taken with respect to \( F \). Suppose we wish to test the hypothesis \( u(\beta) = \eta_0 \), where \( u : \mathbb{R}^q \rightarrow \mathbb{R}^{q_1}, q_1 \leq q \) and consider test statistics based on \( u(T_n) \), where \( T_n \) is the M-estimate of \( \beta \) given by the solution of

\[
\sum_{i=1}^n \psi(Y_i; T_n) = 0. \tag{2.7}
\]

When \( q_1 = 1 \), saddlepoint approximations with relative error of order \( O(n^{-1}) \) for the p-value \( P[u(T_n) > u(t_n)] \), where \( t_n \) is the observed value of \( T_n \), are available; see for instance DiCiccio, Field, and Fraser (1990), Tingley and Field (1990), Daniels and Young (1991), Jing and Robinson (1994), Fan and Field (1995), Davidson, Hinkley, and Worton (1995), and Gatto and Ronchetti (1996). More generally, in the multidimensional case (\( q_1 > 1 \)), Robinson, Ronchetti, and Young (2003) proposed the one dimensional test statistic \( h(u(T_n)) \), where

\[
h(y) = \inf_{\{\beta : u(\beta) = y\}} \sup_{\lambda} \{-K_\psi(\lambda; \beta)\} \tag{2.8}
\]

and

\[
K_\psi(\lambda; \beta) = \log E[e^{\lambda^T \psi(Y; \beta)}] \tag{2.9}
\]

is the cumulant generating function of the score function \( \psi(Y; \beta) \) and the expectation is taken with respect to \( F \) under the null hypothesis.
They proved that under the null hypothesis, $2nh(u(T_n))$ is asymptotically $\chi^2_{q_1}$ with a relative error of order $O(n^{-1})$. Therefore, although this test is asymptotically (first order) equivalent to the three standard tests, it has better second order properties because the latter are asymptotically $\chi^2_{q_1}$ with an absolute error of order $O(n^{-1/2})$.

Notice that (2.8) can be rewritten as

$$h(y) = \inf_{\{\beta: u(\beta) = y\}} \{-K_\psi(\lambda(\beta); \beta)\},$$

(2.10)

where $K_\psi$ is defined by (2.9) and $\lambda(\beta)$ is the so-called saddlepoint satisfying

$$K_\psi'(\lambda; \beta) \equiv \frac{\partial}{\partial \lambda} K_\psi(\lambda; \beta) = 0.$$  

(2.11)

Moreover, in the case of a simple hypothesis, i.e. $u(\beta) = \beta$, (2.10) simply becomes $h(\beta) = -K_\psi(\lambda(\beta); \beta)$.

In order to apply the saddlepoint test statistic to GLM, we first need to generalize this result to the case when the observations $Y_1, \ldots, Y_n$ are independent but not identically distributed. In this case the formulae given above still hold with the cumulant generating function replaced by

$$K_\psi(\lambda; \beta) = \frac{1}{n} \sum_{i=1}^n K^i_\psi(\lambda; \beta),$$

(2.12)

where $K^i_\psi(\lambda; \beta) = \log E_{F^i}[e^{\lambda \psi(Y_i; \beta)}]$ and $F^i$ is the distribution of $Y_i$.

Notice that here the $F^i$ differ only by the parameter $\theta_i$ and this simplifies the computation of $K^i_\psi$; see Appendix 2.7.1.

Since the proof about the accuracy of the test is based on the saddlepoint approximation of the density of the M-estimator $T_n$, the result still holds in this case by replacing (1.3) and (2.1) in Robinson, Ronchetti and Young (2003) by the corresponding formulas at the bottom of p. 323 and at the top of p. 324 in
Ronchetti and Welsh (1994); see also Field and Ronchetti (1990), section 4.5.c.

In this modified form, the saddlepoint test statistic can be applied to GLM with different score functions $\psi$, such as those defined by (2.4) and (2.5). In the next section, we will exploit the structure of GLM to provide explicit formulas for the new test statistic.

### 2.3.2 Saddlepoint Test Statistic with Classical Score Function

The quasi-likelihood and the maximum likelihood estimators of $\beta$ are defined by the same score function. The solution of (2.3) is an M-estimator defined by the score function (2.4).

Let $K_\psi(\lambda; \beta) = \frac{1}{n} \sum_{i=1}^{n} K^i_\psi(\lambda; \beta)$, where $K^i_\psi(\lambda; \beta) = \log E_{F_0} [e^{\lambda^T \psi(Y_i; \beta)}]$ and $F_0$ is the distribution of $Y_i$ defined by the exponential family (2.2) with $\theta = \theta_0i$ and $b'(\theta_0i) = \mu_0i = g^{-1}(x_i^T \beta_0)$. Then by (2.4) we can write

$$K^i_\psi(\lambda; \beta) = \log \int e^{\lambda^T \psi(y; \beta)} f_{Y_i}(y; \theta_0i, \phi) \cdot dy$$

$$= \log \int e^{\lambda^T (y-\mu_i)x_i} \cdot e^{\frac{\mu_0i-b(\theta_0i)}{a(\phi)}} \cdot e^{c(y; \phi)} \cdot dy$$

$$= \log \int e^{-\mu_i^T x_i} \cdot e^{-\frac{b(\theta_0i)}{a(\phi)}} \cdot e^{\frac{\mu_0iT x_i a(\phi)}{a(\phi)}} \cdot e^{c(y; \phi)} \cdot dy$$

$$= \log \int e^{-\mu_i^T x_i} \cdot e^{-\frac{b(\theta_0i)}{a(\phi)}} \cdot e^{\frac{b(\theta_0i + \lambda^T x_i a(\phi)) - b(\theta_0i + \lambda^T x_i a(\phi))}{a(\phi)}} \cdot e^{c(y; \phi)} \cdot dy$$

$$= \log \left[ e^{-[\mu_i^T x_i + \frac{b(\theta_0i)}{a(\phi)}]} \cdot e^{\frac{b(\theta_0i + \lambda^T x_i a(\phi)) - b(\theta_0i + \lambda^T x_i a(\phi))}{a(\phi)}} \cdot \int e^{\frac{y(\theta_0i + \lambda^T x_i a(\phi)) - y(\theta_0i + \lambda^T x_i a(\phi))}{a(\phi)}} \cdot e^{c(y; \phi)} \cdot dy \right]$$

$$= \frac{b(\theta_0i + \lambda^T x_i a(\phi)) - b(\theta_0i)}{a(\phi)} - \mu_i^T x_i.$$

(2.13)
By taking into account the fact that \( \mu_i = b'(\theta_i) \), and that \( b'(\cdot) \) is injective, the solution \( \lambda(\beta) \) of (2.11) with \( K_\psi \) defined by (2.12) and (2.13) is unique and satisfies (see Appendix 2.7.1):

\[
\lambda^T(\beta)x_i = \frac{\theta_i - \theta_{0i}}{a(\phi)}.
\]

Therefore,

\[
h(\beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{b'(x_i^T\beta)x_i^T(\beta - \beta_0) - (b(x_i^T\beta) - b(x_i^T\beta_0))}{a(\phi)}.
\] (2.14)

The test statistic \( 2nh(\hat{\beta}) \) given by (2.14) where \( \hat{\beta} \) is MLE (the solution of (2.3)) is asymptotically \( \chi^2_q \) under the simple null hypothesis \( \beta = \beta_0 \) and can be used to test this null hypothesis.

Notice that \( 2nh(\hat{\beta}) \) can be rewritten using the classical likelihood ratio test statistics for (2.2) by replacing the \( y_i \) by \( \hat{\mu}_i \), i.e.

\[
2nh(\hat{\beta}) = 2[l(\hat{\beta} | \hat{\mu}_1, ..., \hat{\mu}_n) - l(\beta_0 | \hat{\mu}_1, ..., \hat{\mu}_n)],
\] (2.15)

where \( l(\beta | y_1, ..., y_n) = \sum_{i=1}^{n} \log f_Y(y_i; \theta_i, \phi) \) is the log-likelihood function of model (2.2) with \( \theta_i = x_i^T\beta \) and \( \theta_{0i} = x_i^T\beta_0 \) (canonical link).

To test the more general hypothesis \( u(\beta) = \eta_0 \), where \( u : \mathbb{R}^q \to \mathbb{R}^{q_1}, q_1 \leq q \), the test statistic is given by \( 2nh(u(\hat{\beta})) \), where \( h(y) \) is defined by (2.10) and

\[
-K_\psi(\lambda(\beta); \beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{b'(x_i^T\beta)x_i^T(\beta - \beta_0) - (b(x_i^T\beta) - b(x_i^T\beta_0))}{a(\phi)}.
\] (2.16)

and \( \beta_0 \) such that \( u(\beta_0) = \eta_0 \).

**Special cases**

(i) \( Y_i \sim N(\mu_i, \sigma^2) \)

\[
b(\theta) = \frac{\sigma^2}{2}, \quad a(\phi) = \sigma^2
\]

28
Then,

\[ h(\hat{\beta}) = \frac{1}{2n\sigma^2} (\hat{\beta} - \beta_0)^T \left[ \sum_{i=1}^{n} x_i x_i^T \right] (\hat{\beta} - \beta_0) \]

and \( 2nh(\hat{\beta}) \) is in this case exactly the classical Wald test statistic.

(ii) \( Y_i \sim P(\mu_i) \)
\[ b(\theta) = e^{\theta}, \quad a(\phi) = 1 \]

Then,
\[ h(\hat{\beta}) = \frac{1}{n} \left[ \sum_{i=1}^{n} e^{x_i^T \hat{\beta}} x_i^T (\hat{\beta} - \beta_0) - \sum_{i=1}^{n} (e^{x_i^T \hat{\beta}} - e^{x_i^T \beta_0}) \right]. \]

(iii) \( Y_i \sim Bin(m, \pi_i) \)
\[ b(\theta) = m \log(1 + e^{\theta}), \quad a(\phi) = 1 \]

Then,
\[ h(\hat{\beta}) = \frac{m}{n} \left[ \sum_{i=1}^{n} \frac{e^{x_i^T \hat{\beta}}}{1 + e^{x_i^T \hat{\beta}}} x_i^T (\hat{\beta} - \beta_0) - \sum_{i=1}^{n} \left[ \log(1 + e^{x_i^T \hat{\beta}}) - \log(1 + e^{x_i^T \beta_0}) \right] \right]. \]

The saddlepoint test defined by (2.14) will be more accurate than the standard classical test when the model is exact. However, both are based on the (unbounded) classical score function (2.4) and will be inaccurate (even for large \( n \)) in the presence of deviations from the model. In the next section, we construct a better robust test.
2.3.3 Saddlepoint Test Statistic with Robust Score Function

From (2.5), the robust score function is defined by  $	ilde{\psi}_R(y; \beta) = \psi_c(r) w(x) \frac{1}{\sqrt{1/2(\mu)}} \mu' - \tilde{a}(\beta)$ and the cumulant generating function of the robust score function by

$$K_{\tilde{\psi}_R}^i(\lambda; \beta) = \frac{1}{n} \sum_{i=1}^n K_{\tilde{\psi}_R}^i(\lambda; \beta),$$

where

$$K_{\tilde{\psi}_R}^i(\lambda; \beta) = \log E_F[ e^{\lambda^T \tilde{\psi}_R(Y_i; \beta)} ] .$$

As in the classical case, the robust cumulant generating function $K_{\tilde{\psi}_R}^i(\cdot)$ for each observation $i$ can be written as

$$K_{\tilde{\psi}_R}^i(\lambda; \beta) = \log \int e^{\lambda^T \tilde{\psi}_R(y; \beta)} f_Y(y; \theta_0, \phi) \cdot dy$$

$$= \log \left[ \int_{r_i < -c} e^{\lambda^T \tilde{\psi}_R(r_i; \mu) \frac{w(x)}{V^{1/2}(\mu)}} \mu' - \lambda^T \tilde{a}(\beta) \cdot e^{\frac{y \theta_0 - b(\theta_0)}{a(\phi)}} \cdot e^{c(y; \phi)} \cdot dy \right]$$

$$= \log \left[ \int_{r_i < c} e^{\lambda^T \tilde{\psi}_R(y; \beta) \frac{w(x)}{V^{1/2}(\mu)}} \mu' - \lambda^T \tilde{a}(\beta) \cdot e^{\frac{y \theta_0 - b(\theta_0)}{a(\phi)}} \cdot e^{c(y; \phi)} \cdot dy \right]$$

$$+ \int_{c < r_i < c} e^{\lambda^T \tilde{\psi}_R(y; \beta) \frac{w(x)}{V^{1/2}(\mu)}} \mu' - \lambda^T \tilde{a}(\beta) \cdot e^{\frac{y \theta_0 - b(\theta_0)}{a(\phi)}} \cdot e^{c(y; \phi)} \cdot dy$$

$$+ \int_{r_i > c} e^{\lambda^T \tilde{\psi}_R(y; \beta) \frac{w(x)}{V^{1/2}(\mu)}} \mu' - \lambda^T \tilde{a}(\beta) \cdot e^{\frac{y \theta_0 - b(\theta_0)}{a(\phi)}} \cdot e^{c(y; \phi)} \cdot dy$$

$$= \log [I_{i1} + I_{i2} + I_{i3}],$$

where

$$r_i = \frac{y - \mu}{V^{1/2}(\mu)} .$$

For the explicit calculations of $I_{ij}$ for $j = 1, 2, 3$, we refer to Appendix 2.7.2.
Finally, the cumulant generating function can be written as

\[ K_{\psi_R}^i(\lambda; \beta) = \log[I_{i1} + I_{i2} + I_{i3}] \]

\[ = \log\left[ e^{-\lambda^T c_{x_i} V_1(\mu_i) \mu_i^T \tilde{a}(\beta)} \cdot \mathbb{P}(Z_i \leq -cV^{1/2}(\mu_i) + \mu_i) \right. \]

\[ + e^{-\frac{\lambda^T c_{x_i} V_1(\mu_i) \mu_i^T \tilde{a}(\beta)}{V_1(\mu_i)} \cdot e^{b(\theta_0i + \frac{\lambda^T c_{x_i} \mu_i}{V_1(\mu_i)} + b(\theta_0i)) \cdot \mathbb{P}(Z_i \leq -cV^{1/2}(\mu_i) + \mu_i)^{-1} \cdot s(\lambda^0; \beta) \cdot s(\lambda^0, \beta) = 0. \] (2.19)

(2.19) can be easily solved numerically. Alternatively, we can approximate the solution of (2.19) by a one-step Newton’s algorithm, i.e.

\[ \tilde{\lambda}(\beta) \approx \lambda^0 - \left[ \frac{\partial s(\lambda; \beta)}{\partial \lambda} \right]^{-1} \cdot s(\lambda^0; \beta), \] (2.20)

where \( \lambda^0 = \hat{\beta}_R - \beta_0 \) and \( \hat{\beta}_R \) is the robust estimator defined by (2.7) and (2.5). The explicit computations of \( s(\lambda; \beta) \) and \( \frac{\partial s(\lambda; \beta)}{\partial \lambda} \) are provided in Appendix 2.7.3.
For a given distribution of $Y_i$ this leads to an analytical expression of the robust saddlepoint statistic $h_R(\cdot)$:

$$h_R(\beta) = \frac{1}{n} \sum_{i=1}^{n} K^i \tilde{\psi}_R (\tilde{\lambda}(\beta); \beta),$$  

(2.21)

where $\tilde{\lambda}(\beta) \cong (\hat{\beta}_R - \beta_0) - \left[ \sum_{i=1}^{n} x_i x_i^T A_i(\hat{\beta}_R - \beta_0) \right]^{-1} \cdot s(\hat{\beta}_R - \beta_0; \beta)$

and $A_i(\cdot)$ a scalar function defined by the distribution of $Y_i$. For the important cases of Normal, Poisson and Binomial distributions, we refer to Appendix 2.7.4.

The test statistic $2nh_R(\hat{\beta}_R)$ given by (2.21) where $\hat{\beta}_R$ is the robust M-estimator defined by (2.7) with the score function given by (2.5) is asymptotically $\chi^2_q$ under the simple null hypothesis $\beta = \beta_0$ and can be used to test this null hypothesis.

To test the more general hypothesis $u(\beta) = \eta_0$, where $u : \mathbb{R}^q \to \mathbb{R}^{q_1}$, $q_1 \leq q$, the robust test statistic is given by $2nh_R(u(\hat{\beta}_R))$, where $h_R(y)$ is defined by

$$h_R(y) = \inf_{\{\beta : u(\beta) = y\}} \{-K^\tilde{\psi}_R(\tilde{\lambda}(\beta); \beta)\},$$  

(2.22)
2.4 Monte Carlo Study

To illustrate and compare the different tests, we consider a Poisson regression with canonical link \( g(\mu) = \log(\mu) \) and 3 explanatory variables plus intercept \( (q = 4) \), i.e.

\[
\log(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} ,
\]

where \( x_{ij} \sim U[0, 1], j = 2, 3, 4 \). The \( Y_i' \)'s are generated according to the Poisson distribution \( P(\mu_i) \) and a perturbed distribution of the form \( (1 - \epsilon)P(\mu_i) + \epsilon P(\nu\mu_i) \), where \( \epsilon = 0.05, 0.10 \) and \( \nu = 2 \). The latter represents situations where the distribution of the data is not exactly the model but lies in a small neighborhood of the model. The null hypothesis is \( \beta_2 = \beta_3 = \beta_4 = 0 \) \( (q_1 = 3) \) and we choose two sample sizes \( n = 30, 100 \).

We consider four tests: the classical test, the robust quasi-deviance test developed in Cantoni and Ronchetti (2001), and the two saddlepoint tests derived from them in sections 3.2 and 3.3. The latter are defined by the new test statistics \( 2nh(\hat{\beta}) \) and \( 2nh_R(\hat{\beta}_R) \) respectively. The tuning constant \( c \) in the robust score function (2.5) is set to 1.345. Since the \( x \)-design is balanced and there are no leverage points, we set \( w(x_i) \equiv 1 \ \forall i \).

The computation of the new saddlepoint test statistics involves explicit expressions in the case of a simple hypothesis and an additional simple minimization in the case of a composite hypothesis. S-PLUS code is available from the authors upon request. The evaluation of the robust version of the saddlepoint test requires the computation of \( \hat{\beta}_R \), the robust estimator defined by (2.7) and (2.5). Code is available in R (rglm) and S-PLUS (http://www.unige.ch/ses/metri/cantoni/).
The results of the simulations are represented by PP-plots of p-values against $U[0, 1]$ probabilities. In Figures 2.1 to 2.3, PP-plots for the classical test (left) and the saddlepoint test based on the classical score function (right) are given in Panel (a). Panel (b) shows the corresponding PP-plots for their robust versions. The first row reports the simulation results for sample size $n = 30$ and the second one for $n = 100$.

Figures 2.1 shows the results when there are no deviations from the model. Even in this case the asymptotic approximation of the classical test statistic is inaccurate (deviation from the 45° line) both for $n = 30$ and 100 while the new test statistic clearly improves the accuracy. The robust quasi-deviance test is already doing better than its classical counterpart and the new robust saddlepoint test statistic provides a very high degree of accuracy. In the presence of small deviations from the model (Figures 2.2), the classical test is extremely inaccurate (even for $n = 100$), its saddlepoint version and robust quasi-deviance version are better but still inaccurate, while the robust saddlepoint test is very accurate even down to $n = 30$.

Finally, in the presence of larger deviations from the model (Figures 2.3), the robust saddlepoint test is not as accurate as in the previous cases but it is still useful. Notice however that this is an extreme scenario especially for $n = 30$.

To summarize: The new saddlepoint statistic clearly improves the accuracy of the test. When it is used with a robust score function, it can control the bias due to deviations from the model and the resulting test is very accurate in the presence of small deviations from the model and even down to small sample sizes.
Figure 2.1 (a): *Poisson model*

Figure 2.1 (b): *Poisson model*
Figure 2.2 (a): Contaminated model: $\epsilon = 0.05$, $\nu = 2$

Figure 2.2 (b): Contaminated model: $\epsilon = 0.05$, $\nu = 2$
Figure 2.3 (a): Contaminated model: $\epsilon = 0.10$, $\nu = 2$

Figure 2.3 (b): Contaminated model: $\epsilon = 0.10$, $\nu = 2$
2.5 A Real Data Example

We consider a data set issued from a study of the adverse events of a drug on 117 patients affected by Crohn’s disease (a chronic inflammatory disease of the intestines).

In addition to the response variable AE (number of adverse events), 7 explanatory variables were recorded for each patient: BMI (body mass index), HEIGHT, COUNTRY (one of the two countries where the patient lives), SEX, AGE, WEIGHT, and TREAT (the drug taken by the patient: placebo, drug 1, and drug 2). We consider a Poisson regression model. Table 2.1 presents a classical analysis of deviance table which selects the variables BMI, HEIGHT, COUNTRY, SEX, AGE. The same analysis by means of the (first-order) robust quasi-deviance test developed in Cantoni and Ronchetti (2001) provides strong evidence only for a smaller subset of variables, i.e. BMI, HEIGHT, and COUNTRY. Finally, a robust analysis of deviance based on the new robust saddlepoint test provides strong evidence only for the variable BMI. This can be explained on one side by looking at Figure 4 which shows the robust weight provided by the robust analysis: three cases (23, 49, and 51) have small weights and are clearly flagged as influential points. These points have a big influence on the classical analysis and lead to a wrong variable selection when using the classical test.

The second analysis provides a more reliable variable selection since it is robust and therefore not influenced by single points. However, in view of the robustness and better finite sample behavior of the new test, we recommend the result obtained by the third analysis.
Table 2.1: Analysis of deviance for classical, robust asymptotic (rob. as.) and robust saddlepoint (rob. sad.) methods

<table>
<thead>
<tr>
<th>Variable</th>
<th>P.val (class.)</th>
<th>P.val (rob. as.)</th>
<th>P.val (rob. sad.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BMI</td>
<td>0.0001</td>
<td>0.00001</td>
<td>0.0000</td>
</tr>
<tr>
<td>HEIGHT</td>
<td>0.0001</td>
<td>0.0029</td>
<td>0.0778</td>
</tr>
<tr>
<td>COUNTRY</td>
<td>0.0043</td>
<td>0.0079</td>
<td>0.0839</td>
</tr>
<tr>
<td>SEX</td>
<td>0.0015</td>
<td>0.1052</td>
<td>0.2616</td>
</tr>
<tr>
<td>AGE</td>
<td>0.0346</td>
<td>0.6555</td>
<td>0.2732</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>0.9651</td>
<td>0.9637</td>
<td>0.3958</td>
</tr>
<tr>
<td>TREAT</td>
<td>0.5855</td>
<td>0.4993</td>
<td>0.9288</td>
</tr>
</tbody>
</table>

Figure 4: Plot of the robust weights for each observation
2.6 Conclusion

We derived a robust test for GLM with second order accuracy. It keeps its level in the presence of small deviations from the assumed model and is accurate even down to small sample sizes. An open research direction is the study of its power. Moreover, since this test requires only a robust score function, similar test procedures can be developed for other models where such score functions are available.

Acknowledgments The authors would like to thank E. Cantoni, M.-P. Victoria-Feser, J. Krishnakumar, C. Field, J. Jureckova, M. Huskova, F. Peracchi and seminar participants at the University of Sydney (Australia), the Francqui Foundation Workshop (Brussels, Belgium) and Dalhousie University, Halifax (Canada) for useful remarks and suggestions on an early draft of the paper. They also thank Dr. Enrico Chavez (Genexion SA, Geneva) who kindly provided the data set analyzed in section 2.5.
2.7 APPENDIX

2.7.1 Classical Saddlepoint Test Statistic

To determine \( \lambda(\beta) \), we calculate

\[
-n \frac{\partial K_\psi(\lambda; \beta)}{\partial \lambda} = - \sum_{i=1}^{n} \frac{\partial K_i^i(\lambda; \beta)}{\partial \lambda} \\
= \sum_{i=1}^{n} \frac{\partial \left( \mu_i \lambda^T x_i + \frac{b(\theta_{0i}) - b(\theta_{0i} + \lambda^T x_i a(\phi))}{a(\phi)} \right)}{\partial \lambda} \\
= \sum_{i=1}^{n} \mu_i x_i - b'(\theta_{0i} + \lambda^T x_i a(\phi)) \cdot x_i \\
= 0
\]

and the solution \( \lambda(\beta) \) of this equation satisfies

\[
\lambda^T(\beta) x_i = \frac{\theta_i - \theta_{0i}}{a(\phi)}. 
\]

Then, by replacing \( \lambda^T(\beta) x_i \) in \( K_\psi \) and after simplification we obtain

\[
-K_i^i(\lambda(\beta); \beta) = \mu_i \left( \frac{\theta_i - \theta_{0i}}{x_i a(\phi)} \right) - \frac{b'(\theta_{0i}) - b(\theta_{0i})}{a(\phi)} \\
= \frac{(\theta_i - \theta_{0i}) \mu_i - (b(\theta_i) - b(\theta_{0i}))}{a(\phi)},
\]

and

\[
h(\beta) = -K_\psi(\lambda(\beta); \beta) \\
= \frac{1}{n} \sum_{i=1}^{n} -K_i^i(\lambda(\beta); \beta) \\
= \frac{1}{n} \sum_{i=1}^{n} \frac{(\theta_i - \theta_{0i}) \mu_i - (b(\theta_i) - b(\theta_{0i}))}{a(\phi)} \\
= \frac{1}{n} \sum_{i=1}^{n} \frac{b'(x_i^T \beta) x_i^T (\beta - \beta_0) - (b(x_i^T \beta) - b(x_i^T \beta_0))}{a(\phi)}. 
\]
2.7.2 Explicit formulas for $I_{ij}$, j=1,2,3

Calculation of the integrals $I_{i1}$, $I_{i2}$, $I_{i3}$

(i)

$$I_{i1} = \int_{r_i < -c} e^{-\lambda T c \frac{w(x_i)}{\sqrt{2}(\mu_i)}} \mu'_i - \lambda^T \tilde{a}(\beta) \cdot e^{\frac{\theta_{0i} - b(\theta_{0i})}{u(\phi)}} \cdot e^{c(y;\phi)} \cdot dy$$

$$= e^{-\lambda T c \frac{w(x_i)}{\sqrt{2}(\mu_i)}} \mu'_i - \lambda^T \tilde{a}(\beta) \cdot \int_{r_i < -c} e^{\frac{\theta_{0i} - b(\theta_{0i})}{u(\phi)}} \cdot e^{c(y;\phi)} \cdot dy$$

$$= e^{-\lambda T c \frac{w(x_i)}{\sqrt{2}(\mu_i)}} \mu'_i - \lambda^T \tilde{a}(\beta) \cdot \int_{y < -c V^{1/2}(\mu_i) + \mu_i} e^{\frac{\theta_{0i} - b(\theta_{0i})}{u(\phi)}} \cdot e^{c(y;\phi)} \cdot dy$$

$$= e^{-\lambda T c \frac{w(x_i)}{\sqrt{2}(\mu_i)}} \mu'_i - \lambda^T \tilde{a}(\beta) \cdot P(Z^i \leq -c V^{1/2}(\mu_i) + \mu_i)$$

where $Z^i$ is a random variable distributed according to the exponential family (2.2) with parameter $\theta_{0i}$.
(ii) 

\[ I_{i2} = \int_{|r_i| < c} \frac{y^{T} \mu_i' \cdot w(x_i)}{V(\mu_i)} \cdot \frac{-\lambda^{T} \mu_i' \cdot w(x_i)}{V(\mu_i)} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{-\frac{y_{\theta_0i} - b(\theta_0i)}{a(\phi)}} \cdot e^{c(y;\phi)} \cdot dy \]

\[ = \int_{|r_i| < c} \frac{y^{T} \mu_i' \cdot w(x_i)}{V(\mu_i)} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{-\frac{b(\theta_0i)}{a(\phi)}} \cdot e^{c(y;\phi)} \cdot dy \]

\[ = \int_{|r_i| < c} \frac{-\lambda^{T} \mu_i' \cdot w(x_i)}{V(\mu_i)} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{-\frac{b(\theta_0i)}{a(\phi)}} \cdot e^{c(y;\phi)} \cdot dy \]

\[ = e^{-\frac{y(\theta_0i) + \lambda^{T} \mu_i' \cdot w(x_i) \cdot a(\phi)}{V(\mu_i)}} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{-\frac{b(\theta_0i)}{a(\phi)}} \cdot e^{c(y;\phi)} \cdot dy \]

\[ = e^{-\frac{-\lambda^{T} \mu_i' \cdot w(x_i)}{V(\mu_i)}} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{-\frac{b(\theta_0i) + \lambda^{T} \mu_i' \cdot w(x_i) \cdot a(\phi)}{V(\mu_i)}} \cdot e^{c(y;\phi)} \cdot dy \]

\[ = e^{-\frac{-\lambda^{T} \mu_i' \cdot w(x_i)}{V(\mu_i)}} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{c(y;\phi)} \cdot dy \]

\[ = P(-cV^{1/2}(\mu_i) + \mu_i < Z_{\tilde{X}} < cV^{1/2}(\mu_i) + \mu_i) \]

Where \( Z_{\tilde{X}} \) is a random variable distributed according to the exponential family (2.2) with parameter \( \left[ \theta_0i + \frac{\lambda^{T} \mu_i' \cdot w(x_i) \cdot a(\phi)}{V(\mu_i)} \right] \).

(iii) This result can be easily derived as in (i).

We obtain:

\[ I_{i3} = e^{cV^{1/2}(\mu_i) - \lambda^{T} \tilde{a}(\beta)} \cdot P(Z_{i} \geq cV^{1/2}(\mu_i) + \mu_i) \]
2.7.3 Robust Saddlepoint Test Statistics

For $i = 1, ..., n$, we have from 2.7.2:

$$
\frac{\partial I_1}{\partial \lambda} + \frac{\partial I_2}{\partial \lambda} + \frac{\partial I_3}{\partial \lambda} = -\left[ cw(x_i) \mu'_i \right] \cdot I_1
$$

$$
- \left[ \frac{\mu_i w(x_i)}{V(\mu_i)} + \bar{a}(\beta) \right] - \frac{\mu'_i w(x_i)}{V(\mu_i)} b'(\theta_0i + \frac{\lambda^T \mu'_i w(x_i) a(\phi)}{V(\mu_i)}) \cdot I_2
$$

$$
+ e^{-\lambda^T \mu_i w(x_i)/V(\mu_i)} \cdot e^{-\lambda^T \bar{a}(\beta)} \cdot e^{b(\theta_0i + \frac{\lambda^T \mu'_i w(x_i) a(\phi)}{V(\mu_i)}) - b(\theta_0i)} \cdot \frac{\mu'_i w(x_i)}{V(\mu_i)} E_{|r_i|<c}[Y]
$$

$$
- \frac{\mu_i w(x_i)}{V(\mu_i)} b'(\theta_0i + \frac{\lambda^T \mu'_i w(x_i) a(\phi)}{V(\mu_i)}) \cdot I_2
$$

$$
+ \left[ \frac{cw(x_i) \mu'_i}{V^{1/2}(\mu_i)} - \bar{a}(\beta) \right] \cdot I_3
$$

$$
= -\left[ \frac{cw(x_i) \mu'_i}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right] \cdot I_1
$$

$$
- \left[ \frac{\mu_i w(x_i)}{V(\mu_i)} + \bar{a}(\beta) \right] \cdot I_2
$$

$$
+ e^{-\lambda^T \mu_i w(x_i)/V(\mu_i)} \cdot e^{-\lambda^T \bar{a}(\beta)} \cdot e^{b(\theta_0i + \frac{\lambda^T \mu'_i w(x_i) a(\phi)}{V(\mu_i)}) - b(\theta_0i)} \cdot \frac{\mu'_i w(x_i)}{V(\mu_i)} E_{|r_i|<c}[Y]
$$

$$
+ \left[ \frac{cw(x_i) \mu'_i}{V^{1/2}(\mu_i)} - \bar{a}(\beta) \right] \cdot I_3
$$

44
Furthermore,

\[
\frac{\partial s(\lambda; \beta)}{\partial \lambda} = \sum_{i=1}^{n} \frac{\partial}{\partial \lambda} \left[ \frac{\partial I_{i_{1}} + \partial I_{i_{2}} + \partial I_{i_{3}}}{\partial \lambda} \right] \\
= \sum_{i=1}^{n} \frac{\partial^{2}(I_{i_{1}} + I_{i_{2}} + I_{i_{3}})}{\partial \lambda \partial \lambda^T} \cdot (I_{i_{1}} + I_{i_{2}} + I_{i_{3}}) - \left[ \frac{\partial I_{i_{1}} + \partial I_{i_{2}} + \partial I_{i_{3}}}{\partial \lambda} \right] \cdot \left[ \frac{\partial I_{i_{1}} + \partial I_{i_{2}} + \partial I_{i_{3}}}{\partial \lambda} \right]^T \\
= \sum_{i=1}^{n} \frac{\partial^{2}(I_{i_{1}} + I_{i_{2}} + I_{i_{3}})}{\partial \lambda \partial \lambda^T} \cdot (I_{i_{1}} + I_{i_{2}} + I_{i_{3}})^2.
\]

Let \( S_{1_i} \) and \( S_{2_i} \) such that:

\[
S_{1_i} : = \frac{\partial^{2}(I_{i_{1}} + I_{i_{2}} + I_{i_{3}})}{\partial \lambda \partial \lambda^T} \cdot (I_{i_{1}} + I_{i_{2}} + I_{i_{3}}) = \frac{\partial}{\partial \lambda} \left[ \frac{\partial I_{i_{1}} + \partial I_{i_{2}} + \partial I_{i_{3}}}{\partial \lambda} \right] \cdot (I_{i_{1}} + I_{i_{2}} + I_{i_{3}}) \\
= (I_{i_{1}} + I_{i_{2}} + I_{i_{3}}) \cdot \left[ \frac{\partial I_{i_{1}} + \partial I_{i_{2}} + \partial I_{i_{3}}}{\partial \lambda} \right] \cdot (I_{i_{1}} + I_{i_{2}} + I_{i_{3}}) \\
+ \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right] \cdot \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right]^T \cdot I_{i_{1}} \\
- \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right] \cdot \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right]^T \cdot I_{i_{2}} \\
- \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right] \cdot \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right]^T \cdot E_{|r_i| < c}[Y] \\
\cdot e^{-\lambda^T \mu_i w(x_i)} \cdot e^{-\lambda^T \bar{a}(\beta)} \cdot \frac{e^{b(\theta_{Q_1}) - b(\theta_{Q_2})}}{a(\beta)} \\
- \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right] \cdot \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right]^T \cdot E_{|r_i| < c}[Y] \\
\cdot e^{-\lambda^T \mu_i w(x_i)} \cdot e^{-\lambda^T \bar{a}(\beta)} \cdot \frac{e^{b(\theta_{Q_1}) - b(\theta_{Q_2})}}{a(\beta)} \\
+ \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right] \cdot \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right]^T \cdot E_{|r_i| < c}[Y^2] \\
+ \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right] \cdot \left[ \frac{\mu_i w(x_i)}{V^{1/2}(\mu_i)} + \bar{a}(\beta) \right]^T \cdot I_{i_{3}} \right) \\
]
\]
and

$$\begin{align*}
S_{2i} &= \left[\frac{\partial I_{1}}{\partial \lambda} + \frac{\partial I_{2}}{\partial \lambda} + \frac{\partial I_{3}}{\partial \lambda}\right] \cdot \left[\frac{\partial I_{1}}{\partial \lambda} + \frac{\partial I_{2}}{\partial \lambda} + \frac{\partial I_{3}}{\partial \lambda}\right] \\
&= \left[\frac{cw(x_{i})\mu_{i}^{T}}{V^{1/2}(\mu_{i})} + \tilde{a}(\beta)\right] \cdot \left[\frac{cw(x_{i})\mu_{i}^{T}}{V^{1/2}(\mu_{i})} + \tilde{a}(\beta)\right]^{T} I_{1}^{2} \\
&+ \left[\frac{\mu_{i}^{T}w(x_{i})}{V(\mu_{i})} + \tilde{a}(\beta)\right] \cdot \left[\frac{\mu_{i}^{T}w(x_{i})}{V(\mu_{i})} + \tilde{a}(\beta)\right]^{T} I_{2}^{2} \\
&+ e^{-\lambda^{T} \mu_{i}^{T}w(x_{i})/V(\mu_{i})} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{\frac{2b(\theta_{0i}) + \lambda^{T} \mu_{i}^{T}w(x_{i})\eta(\theta)}{a(\theta)}} \\
&\cdot e^{-2\lambda^{T} \tilde{a}(\beta)} \cdot e^{\frac{2b(\theta_{0i}) - \lambda^{T} \mu_{i}^{T}w(x_{i})\eta(\theta)}{a(\theta)}} \\
&+ 2 \cdot \left[\frac{cw(x_{i})\mu_{i}^{T}}{V^{1/2}(\mu_{i})} + \tilde{a}(\beta)\right] \left[\frac{\mu_{i}^{T}w(x_{i})}{V(\mu_{i})} + \tilde{a}(\beta)\right]^{T} I_{1} I_{2} \\
&- 2 \cdot \left[\frac{cw(x_{i})\mu_{i}^{T}}{V^{1/2}(\mu_{i})} + \tilde{a}(\beta)\right] \left[\frac{\mu_{i}^{T}w(x_{i})}{V(\mu_{i})} + \tilde{a}(\beta)\right]^{T} I_{1} I_{3} \\
&- 2 \cdot \left[\frac{\mu_{i}^{T}w(x_{i})}{V(\mu_{i})} + \tilde{a}(\beta)\right] \left[\frac{\mu_{i}^{T}w(x_{i})}{V(\mu_{i})} + \tilde{a}(\beta)\right]^{T} I_{2} I_{3} \\
&+ e^{-\lambda^{T} \mu_{i}^{T}w(x_{i})/V(\mu_{i})} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{\frac{b(\theta_{0i}) + \lambda^{T} \mu_{i}^{T}w(x_{i})\eta(\theta)}{a(\theta)}} \\
&\cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{\frac{b(\theta_{0i}) - \lambda^{T} \mu_{i}^{T}w(x_{i})\eta(\theta)}{a(\theta)}} \\
&- 2 \cdot \left[\frac{\mu_{i}^{T}w(x_{i})}{V(\mu_{i})} + \tilde{a}(\beta)\right] \left[\frac{cw(x_{i})\mu_{i}^{T}}{V^{1/2}(\mu_{i})} + \tilde{a}(\beta)\right]^{T} I_{2} I_{3} \\
&+ 2 \cdot \left[\frac{cw(x_{i})\mu_{i}^{T}}{V^{1/2}(\mu_{i})} + \tilde{a}(\beta)\right] \left[\frac{\mu_{i}^{T}w(x_{i})}{V(\mu_{i})} + \tilde{a}(\beta)\right]^{T} I_{2} I_{3} \\
&+ e^{-\lambda^{T} \mu_{i}^{T}w(x_{i})/V(\mu_{i})} \cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{\frac{b(\theta_{0i}) + \lambda^{T} \mu_{i}^{T}w(x_{i})\eta(\theta)}{a(\theta)}} \\
&\cdot e^{-\lambda^{T} \tilde{a}(\beta)} \cdot e^{\frac{b(\theta_{0i}) - \lambda^{T} \mu_{i}^{T}w(x_{i})\eta(\theta)}{a(\theta)}} .
\end{align*}$$
Then,

\[
\frac{\partial s(\lambda; \beta)}{\partial \lambda} = \sum_{i=1}^{n} \frac{[S_1 i - S_2 i]}{(I_{i1} + I_{i2} + I_{i3})^2} \\
= \sum_{i=1}^{n} \frac{[\mu'_i \cdot \mu_i^T]}{(I_{i1} + I_{i2} + I_{i3})^2} w^2(x_i) \left\{ \left[ \frac{c}{V^{1/2}(\mu_i)} - \frac{\mu_i}{V(\mu_i)} \right]^2 \cdot I_{i1} I_{i2} + [2 \frac{c}{V^{1/2}(\mu_i)}]^2 I_{i1} I_{i3} \\
+ \left[ \frac{c}{V^{1/2}(\mu_i)} + \frac{\mu_i}{V(\mu_i)} \right]^2 \cdot I_{i2} I_{i3} \right\} \\
+ 2 \cdot e^{-\lambda^T \tilde{a}(\beta)} \cdot e^{b(\theta_{0i} + \mathbf{\lambda}^T \nu(w(x_i),a(\phi)) - b(\theta_{0i}))} \cdot \frac{1}{V(\mu_i)} \\
\cdot \left[ \left( \frac{c}{V^{1/2}(\mu_i)} - \frac{\mu_i}{V(\mu_i)} \right) \cdot I_{i1} - \left( \frac{c}{V^{1/2}(\mu_i)} + \frac{\mu_i}{V(\mu_i)} \right) \cdot I_{i3} \right] \cdot E_{|\theta| < c}[Y] \\
+ e^{-\lambda^T \tilde{a}(\beta)} \cdot e^{b(\theta_{0i} + \mathbf{\lambda}^T \nu(w(x_i),a(\phi)) - b(\theta_{0i}))} \cdot \frac{1}{V^2(\mu_i)} \\
\cdot \left[ I_{i1} + I_{i2} + I_{i3} \right] \cdot E_{|\theta| < c}[Y^2] \\
- \left[ e^{-\lambda^T \tilde{a}(\beta)} \cdot e^{b(\theta_{0i} + \mathbf{\lambda}^T \nu(w(x_i),a(\phi)) - b(\theta_{0i}))} \right]^2 \cdot \frac{1}{V^2(\mu_i)} \cdot \left[ E_{|\theta| < c}[Y] \right]^2 \right\}.
\]
2.7.4 Some Special Robust Saddlepoint Test Statistics

(i) $Y_i \sim N(\mu_i, \sigma^2)$

$b(\theta_i) = \frac{\theta_i^2}{2} \quad a(\phi) = \sigma^2$

and in this case $\tilde{a}(\beta) = 0$. Then, we have:

$$\frac{\partial s(\lambda; \beta)}{\partial \lambda} = \sum_{i=1}^{n} x_i x_i^T \cdot \frac{w^2(x_i)}{(I_{i1} + I_{i2} + I_{i3})^2} \left\{ (c - x_i^T \beta)^2 \cdot I_{i1} I_{i2} + (2c)^2 \cdot I_{i1} I_{i3} + (c + x_i^T \beta)^2 \cdot I_{i2} I_{i3} + 2 \cdot e^{x_i^T \lambda w(x_i) x_i^T (2\beta_0 - \beta) + (x_i^T \lambda w(x_i) \sigma)^2} \cdot (\sigma^{-1} x_i^T \lambda w(x_i) x_i^T \beta I_{i1} - (c + x_i^T \beta) \cdot I_{i3}) \cdot E_{|r_i| < c}[Y] ight\}$$

$$+ e^{x_i^T \lambda w(x_i) x_i^T (2\beta_0 - \beta) + (x_i^T \lambda w(x_i) \sigma)^2} \cdot [I_{i1} + I_{i2} + I_{i3}] \cdot E_{|r_i| < c}[Y^2]$$

$$- \left[ e^{x_i^T \lambda w(x_i) x_i^T (2\beta_0 - \beta) + (x_i^T \lambda w(x_i) \sigma)^2} \cdot \left( E_{|r_i| < c}[Y] \right)^2 \right]$$

$$= \sum_{i=1}^{n} x_i x_i^T \cdot A_i(\lambda),$$

where $A_i(\lambda)$ is scalar function defined by

$$A_i(\lambda) = \frac{w(x_i)}{(I_{i1} + I_{i2} + I_{i3})^2} \cdot \left\{ (c - x_i^T \beta)^2 \cdot I_{i1} I_{i2} + (2c)^2 \cdot I_{i1} I_{i3} + (c + x_i^T \beta)^2 \cdot I_{i2} I_{i3} + 2 \cdot e^{x_i^T \lambda w(x_i) x_i^T (2\beta_0 - \beta) + (x_i^T \lambda w(x_i) \sigma)^2} \cdot (\sigma^{-1} x_i^T \lambda w(x_i) x_i^T \beta I_{i1} - (c + x_i^T \beta) \cdot I_{i3}) \cdot E_{|r_i| < c}[Y] ight\}$$

$$+ e^{x_i^T \lambda w(x_i) x_i^T (2\beta_0 - \beta) + (x_i^T \lambda w(x_i) \sigma)^2} \cdot [I_{i1} + I_{i2} + I_{i3}] \cdot E_{|r_i| < c}[Y^2]$$

$$- \left[ e^{x_i^T \lambda w(x_i) x_i^T (2\beta_0 - \beta) + (x_i^T \lambda w(x_i) \sigma)^2} \cdot \left( E_{|r_i| < c}[Y] \right)^2 \right].$$
(ii) $Y_i \sim \mathcal{P}(\mu_i)$

$b(\theta) = e^\theta, \quad a(\phi) = 1$

Then, we have:

$$\frac{\partial s(\lambda; \beta)}{\partial \lambda} = \sum_{i=1}^{n} x_i x_i^T \cdot \frac{w^2(x_i) \cdot e^{2x_i^T \beta}}{(I_{i1} + I_{i2} + I_{i3})^2} \left\{ \left( ce^{-\frac{1}{2}x_i^T \beta} - 1 \right)^2 \cdot I_{i1}I_{i2} + \left( 2, ce^{-\frac{1}{2}x_i^T \beta} \right)^2 \cdot I_{i1}I_{i3} + \left( ce^{-\frac{1}{2}x_i^T \beta} + 1 \right)^2 \cdot I_{i2}I_{i3} \right\} + 2 \cdot e^{-x_i^T \lambda w e^T_i \beta - \lambda T \bar{a}(\beta)} \cdot e^{[x_i^T (\beta_0 + w(x_i)\lambda) - e^{x_i^T \beta_0}]} \cdot e^{-x_i^T \beta} \cdot E_{|\beta_i|<c}[Y]$$

$$\cdot \left[ \left( ce^{-\frac{1}{2}x_i^T \beta} - 1 \right) \cdot I_{i1} - \left( ce^{-\frac{1}{2}x_i^T \beta} + 1 \right) \cdot I_{i3} \right] + e^{-x_i^T \lambda w e^T_i \beta - \lambda T \bar{a}(\beta)} \cdot e^{[x_i^T (\beta_0 + w(x_i)\lambda) - e^{x_i^T \beta_0}]} \cdot e^{-2x_i^T \beta} \cdot E_{|\beta_i|<c}[Y]^2$$

$$- \left[ e^{-x_i^T \lambda w e^T_i \beta - \lambda T \bar{a}(\beta)} \cdot e^{[x_i^T (\beta_0 + w(x_i)\lambda) - e^{x_i^T \beta_0}]} \right] e^{-2x_i^T \beta} \cdot \left( E_{|\beta_i|<c}[Y] \right)^2 \right\}$$

$$= \sum_{i=1}^{n} x_i x_i^T \cdot A_i(\lambda),$$

where $A_i(\lambda)$ is scalar function defined by

$$A_i(\lambda) = \frac{w^2(x_i) \cdot e^{2x_i^T \beta}}{(I_{i1} + I_{i2} + I_{i3})^2} \left\{ \left( ce^{-\frac{1}{2}x_i^T \beta} - 1 \right)^2 \cdot I_{i1}I_{i2} + \left( 2, ce^{-\frac{1}{2}x_i^T \beta} \right)^2 \cdot I_{i1}I_{i3} + \left( ce^{-\frac{1}{2}x_i^T \beta} + 1 \right)^2 \cdot I_{i2}I_{i3} \right\} + 2 \cdot e^{-x_i^T \lambda w e^T_i \beta - \lambda T \bar{a}(\beta)} \cdot e^{[x_i^T (\beta_0 + w(x_i)\lambda) - e^{x_i^T \beta_0}]} \cdot e^{-x_i^T \beta} \cdot E_{|\beta_i|<c}[Y]$$

$$\cdot \left[ \left( ce^{-\frac{1}{2}x_i^T \beta} - 1 \right) \cdot I_{i1} - \left( ce^{-\frac{1}{2}x_i^T \beta} + 1 \right) \cdot I_{i3} \right] + e^{-x_i^T \lambda w e^T_i \beta - \lambda T \bar{a}(\beta)} \cdot e^{[x_i^T (\beta_0 + w(x_i)\lambda) - e^{x_i^T \beta_0}]} \cdot e^{-2x_i^T \beta} \cdot E_{|\beta_i|<c}[Y]^2$$

$$- \left[ e^{-x_i^T \lambda w e^T_i \beta - \lambda T \bar{a}(\beta)} \cdot e^{[x_i^T (\beta_0 + w(x_i)\lambda) - e^{x_i^T \beta_0}]} \right] e^{-2x_i^T \beta} \cdot \left( E_{|\beta_i|<c}[Y] \right)^2 \right\}.$$
(iii) $Y_i \sim \text{Bin}(m, \pi_i)$

\[
b(\theta) = m \cdot \log(1 + e^\theta), \quad a(\phi) = 1
\]

Then, we have:

\[
\frac{\partial s(\lambda; \beta)}{\partial \lambda} = \sum_{i=1}^{n} x_i x_i^T \frac{w^2(x_i)e^{2x_i^T\beta}}{(I_{i1} + I_{i2} + I_{i3})^2} \{(\frac{c - \sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}}{\sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}(1 + e^{x_i^T\beta})})^2 I_{i1}I_{i2}
+ \left(\frac{2c}{\sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}(1 + e^{x_i^T\beta})}\right)I_{i1}I_{i3} + \left(\frac{c + \sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}}{\sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}(1 + e^{x_i^T\beta})}\right)^2 I_{i2}I_{i3}
+ \frac{2(1 + x_i^T \beta_0 + x_i^T \lambda w(x_i))}{1 + \beta_0^T x_i} \cdot \frac{e^{-\gamma x_i^T \lambda_0 \mu_{w(x_i)} e^{x_i^T\beta}}}{1 + e^{x_i^T\beta}} \cdot e^{-\gamma x_i^T \hat{a}(\beta)} \cdot \frac{1}{m e^{x_i^T\beta}} \cdot \left[\frac{c - \sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}}{\sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}(1 + e^{x_i^T\beta})}\right] I_{i1} - \left(\frac{c + \sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}}{\sqrt{m} \cdot e^{1\sqrt{x_i^T\beta}}(1 + e^{x_i^T\beta})}\right) I_{i3} \cdot E[Z_{i1}^{Z_i}]
+ \frac{1 + x_i^T \beta_0 + x_i^T \lambda w(x_i)}{1 + x_i^T \beta_0} \cdot \frac{e^{-\gamma x_i^T \lambda_0 \mu_{w(x_i)} e^{x_i^T\beta}}}{1 + e^{x_i^T\beta}} \cdot e^{-\gamma x_i^T \hat{a}(\beta)} \cdot \frac{1}{m e^{x_i^T\beta}} \cdot \left[I_{i1} + I_{i2} + I_{i3}\right] \cdot E[Z_{i1}^{Z_i}] \cdot E[Y^2]
+ \frac{1 + x_i^T \beta_0 + x_i^T \lambda w(x_i)}{1 + x_i^T \beta_0} \cdot \frac{e^{-\gamma x_i^T \lambda_0 \mu_{w(x_i)} e^{x_i^T\beta}}}{1 + e^{x_i^T\beta}} \cdot e^{-\gamma x_i^T \hat{a}(\beta)} \cdot \frac{1}{m e^{x_i^T\beta}} \cdot \left[E[Z_{i1}^{Z_i}]\right]^2\}
= \sum_{i=1}^{n} x_i x_i^T \cdot A_i(\lambda),
\]

where $A_i(\lambda)$ is scalar function defined by
\[ A_i(\lambda) = \frac{w^2(x_i)e^{2x_i^T\beta}}{(I_{i1} + I_{i2} + I_{i3})^2}\left\{ \left( \frac{c - \sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}}{\sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}(1 + e^{x_i^T\beta})} \right)^2 I_{i1}I_{i2} + \left( \frac{c + \sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}}{\sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}(1 + e^{x_i^T\beta})} \right)^2 I_{i2}I_{i3} \right\}
\]

\[ + \left( \frac{2c}{\sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}(1 + e^{x_i^T\beta})} \right)^2 I_{i1}I_{i3} + \left( \frac{c + \sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}}{\sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}(1 + e^{x_i^T\beta})} \right)^2 I_{i2}I_{i3} \]

\[ + 2\left( \frac{1 + x_i^T\beta_0 + x_i^T\lambda w(x_i)}{1 + x_i^T\beta_0} \right)m \cdot e^{\frac{-m\lambda w(x_i)e^{x_i^T\beta}}{1 + e^{x_i^T\beta}}}.e^{-\lambda^T\tilde{a}(\beta)} \cdot \frac{1}{me^{x_i^T\beta}} \]

\[ \cdot \left[ \left( \frac{c - \sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}}{\sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}(1 + e^{x_i^T\beta})} \right)I_{i1} - \left( \frac{c + \sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}}{\sqrt{m} \cdot e^{\frac{1}{2}x_i^T\beta}(1 + e^{x_i^T\beta})} \right)I_{i3} \right] \cdot E^{Z_i}_{|r_i|<c}[Y] \]

\[ + \left( \frac{1 + x_i^T\beta_0 + x_i^T\lambda w(x_i)}{1 + x_i^T\beta_0} \right)m \cdot e^{\frac{-m\lambda w(x_i)e^{x_i^T\beta}}{1 + e^{x_i^T\beta}}}.e^{-\lambda^T\tilde{a}(\beta)} \cdot \frac{1}{m^2e^{2x_i^T\beta}} \]

\[ \cdot \left[ I_{i1} + I_{i2} + I_{i3} \right] \cdot E^{Z_i}_{|r_i|<c}[Y^2] \]

\[ - \left( \frac{1 + x_i^T\beta_0 + x_i^T\lambda w(x_i)}{1 + x_i^T\beta_0} \right)m \cdot e^{\frac{-m\lambda w(x_i)e^{x_i^T\beta}}{1 + e^{x_i^T\beta}}}.e^{-\lambda^T\tilde{a}(\beta)} \cdot \frac{1}{m^2e^{2x_i^T\beta}} \left[ E^{Z_i}_{|r_i|<c}[Y] \right]^2 \right\} \]
2.7.5 Code

simulcont.fun()

```r
simulcont.fun=function(N,contamin,lambda1,lambda2){
  ## random generation for Poisson distribution with contamination
  ## N := sample size
  ## contamin := percentage of contaminated data (ex > 0.05)
  ## lambda1 := (Nx1) vector of "normal" mean
  ## lambda2 := (Nx1) vector of "contaminated" mean
  ## written by S.N. Lo Jun 2005
  vect.unif=runif(N)
  vect.simul=vect.unif
  for(i in 1:N){
    if(vect.unif[i]>contamin){vect.simul[i]=rpois(1,lambda1[i])}
    else{vect.simul[i]=rpois(1,lambda2[i])}
  }
  return(vect.simul)
}
```

intP1()

```r
intP1=function(X,beta0,beta,lambda,w=1,c=1.345){
  ## to calculate the integral I.i1 (see P.9)
  ## X := matrix of exogenous variable
  ## beta0 := parameters vector under null hypothesis
  ## beta := estimated parameters vector
  ## c := the Huber constant set to 1.345
  a1=exp(-X%*%lambda*c*w*exp(1/2*(X%*%beta))-rep(Abeta%*%lambda,nrow(X)))
  P1= ppois(c*exp(1/2*(X%*%beta))+exp(X%*%beta),exp(X%*%beta0))
  int1=a1*P1 return(list(a1=a1,P1=P1,I1=int1)) }
```

intP2()

```r
intP2=function(X,beta0,beta,lambda,w=1,c=1.345){
  ## to calculate the integral I.i2 (see P.9)
  ## X := matrix of exogenous variable
  ## beta0 := parameters vector under null hypothesis
  ## beta := estimated parameters vector
  ## c := the Huber constant set to 1.345
  ABTA=rep(Abeta%*%lambda,nrow(X))
  a2=(exp(-X%*%lambda*exp(X%*%beta)*w-ABTA)*exp(exp(X%*%beta0+X%*%lambda*w)
    -exp(X%*%beta0)))
  P2=(ppois(c*exp(1/2*(X%*%beta))+exp(X%*%beta),exp(X%*%beta0+X%*%lambda*w))
    -ppois(-c*exp(1/2*(X%*%beta))+exp(X%*%beta),exp(X%*%beta0+X%*%lambda*w)))
```
int2=a2*P2
return(list(a2=a2,P2=P2,I2=int2))
}

intP3()
intP3=function(X,beta0,beta,lambda,w=1,c=1.345){
  ## to calculate the integral I.i3 (see P.9)
  ## X := matrix of exogenous variable
  ## beta0 := parameters vector under null hypothesis
  ## beta := estimated parameters vector
  ## c := the Huber constant set to 1.345
  a3=exp(X%*%lambda*c*w*exp(1/2*(X%*%beta))-rep(Abeta%*%lambda,nrow(X)))
  P3= 1-ppois(c*exp(1/2*(X%*%beta))+exp(X%*%beta),exp(X%*%beta0))
  int3=a3*P3
  return(list(a3=a3,P3=P3,I3=int3)) }

EspTrSimP()
EspTrSimP=function(X,beta0,beta,lambda,c=1.345,w=1,N.echant=50000){
  ## Estimation of truncated expectation by simulation
  ## X := matrix of exogenous variable
  ## beta0 := parameters vector under null hypothesis
  ## beta := estimated parameters vector
  ## lambda := represents the difference beta-beta0
  ## c := the Huber constant set to 1.345
  ## N.echant:= sample size for calculating the expectation set to 50'000
  Esp1=rep(NA,nrow(X))
  Esp2=rep(NA,nrow(X))
  for (i in 1:nrow(X)){
    Sim=rpois(N.echant,exp(X[i,]%*%beta0+X[i,]%*%lambda*w))
    Esp1[i]=mean(Sim[Sim>-c*exp(1/2*(X[i,]%*%beta))+exp(X[i,]%*%beta)&Sim<=c*exp(1/2*(X[i,]%*%beta))+exp(X[i,]%*%beta)])
    Esp2[i]=mean((Sim[Sim>-c*exp(1/2*(X[i,]%*%beta))+exp(X[i,]%*%beta)&Sim<=c*exp(1/2*(X[i,]%*%beta))+exp(X[i,]%*%beta]))^2)
  }
  return(list(Esp1=Esp1, Esp2=Esp2)) }

SimulK()
SimulK= function(n=30,N=50,c=1.345){
  ## Function to simulate (n) saddlepoint statistics when
  ## Y~Poisson(lambda) and construct two boxplots
  ## the first represents the Saddlepoint statistics vs chi2(4)
the second is the corresponding p-value vs Unif[0 , 1]
  n:= number samples
  N:= the size of each sample
  c:= the constant of Huber
  Kstat:= (nx1) vector of Saddlepoint statistics simulated
  Kstat:= (nx1) vector of corresponding p-value

written by S.N. Lo Jun 2005

Kstat=rep(NA,n) Pstat=rep(NA,n)
for (k in 1:n){
  # Simulation of endogenous / exogenous data
  const=rep(1,N) # the constant vector
  x1=runif(N,0,1)
  x2=runif(N,0,1)
  x3=runif(N,0,1)
  X=cbind(x1,x2,x3)
  X1=cbind(const,X) # (Nx4) matrix of exogenous variables
  beta0=rbinding(0,0,0,0) # (4x1) initial parameter vector
  mu=exp(X1%*%beta0) # (Nx1) vector of the expectations E[Y]
  y=simulcont.fun(N,.05,mu,2*mu) # (Nx1) vector of the endogenous variable
  # Estimation of the parameters of the logistic regression
  result.glm=glm.rob(X,y, choice="poisson",chuber=1.345)
  bet0=result.glm$coef[1]
  bet1=result.glm$coef[2]
  bet2=result.glm$coef[3]
  bet3=result.glm$coef[4]
  beta=c(bet0,bet1,bet2,bet3) # (4x1) Estimated parameters vector
  # calculation of a(beta)
  Abeta=result.glm$a.const
  # Initialization of lambda vector (4x1)
  lambda=beta-beta0
  # Calculation of components I_1, I_2, I_3, ### I_1.1
  I_1.1 = intP1(X1,beta0,beta,lambda,c=1.345)
  ### I_1.2
  I_1.2 = intP2(X1,beta0,beta,lambda,c=1.345)
  ### I_1.3
  I_1.3 = intP3(X1,beta0,beta,lambda,c=1.345)
  # Calculation of the truncated expectations
  EspTr=EspTrSimP(X1,beta0,beta,lambda,c=1.345)
Esp1 =EspTr$Esp1
Esp2 =EspTr$Esp2

# Determination of the diagonal matrix A
A=matrix(0,N,N)
for(i in 1:N){
  A[i,i]=(1/(I1.1*I1[i]+I1.2*I2[i]+I1.3*I3[i])^2)*(w*w*exp(2*(X1[i,]%*
  *beta)))*(w*w*exp(2*(X1[i,]%*beta)))*(c*exp(-1/2*(X1[i,]%*beta)*1)^2*Il.1*I1[i]...
  *Il.2*I2[i]+(2*c*exp(-1/2*(X1[i,]%*beta)))+1)^2*Il.1*I1[i]...
  *Il.3*I3[i]+2*exp(-X1[i,]%*lambda*w*exp(X1[i,]%*lambda)...) -Abeta%*lambda)*exp(exp(X1[i,]%*beta0+w*lambda)...
  -exp(X1[i,]%*beta0)*exp(-(X1[i,]%*beta0)*w*Esp1[i]... *((c*exp(-1/2*(X1[i,]%*beta))-1)*I1.1*I1[i]-(c*exp(-1/2*...
  (X1[i,]%*beta)))+1)*I1.3*I3[i]+exp(-X1[i,]%*lambda*w... *exp(X1[i,]%*beta)-Abeta%*lambda)*exp(exp(X1[i,]%...
  *(beta0+w*lambda)-exp(X1[i,]%*beta0))*exp(-2*(X1[i,]%... *%lambda-Abeta%*lambda)*exp(exp(X1[i,]%...*beta0+w*lambda))-(X1[i,]%*beta0)))^2*exp(-2*(X1[i,]%*beta))...
  *(Esp1[i])^2)
}

# Derivative of S() in relation to lambda
Dde.S=t(X1)%*%A%*%X1

# Determination the different components of S()
S.initial=matrix(NA,ncol=ncol(X1),nrow=N)
for (j in 1:N){
  S.initial[j,]=(1/N)*(1/(I1.1*I1[j]+I1.2*I2[j]+I1.3*I3[j]))*(-(c*w*X1[j,]%*
  %*lambda)*exp(exp(X1[j,]%*beta)-Abeta%*lambda)*exp(exp(X1[j,]%...
  *(beta0+w*lambda))-(X1[j,]%*beta0)))^2*exp(-2*(X1[j,]%*beta)) - Abeta)*I1.3*I3[j])
}
S=apply(S.initial,2,sum)

##### Estimation of Lambda.max by Newton method

Lambda.max=(beta-beta0) - ginverse(Dde.S)%*%S

Lambda.max
# Calculation of the components I_1, I_2, I_3 in relation to Lambda.max

### ILmax.1
ILmax.1= intP1(X1,beta0,beta,Lambda.max,c=1.345)
### ILmax.2
ILmax.2 = intP2(X1, beta0, beta, Lambda.max, c=1.345)

### ILmax.3
ILmax.3 = intP3(X1, beta0, beta, Lambda.max, c=1.345)
ABTA = rep(Abeta%*%Lambda.max, nrow(X))

# Calculation of K() for testing (β_1=β_2=β_3=0)
K = -X1%*%Lambda.max*w*exp(X1%*%beta) - ABTA + exp(X1%*%beta0 + X1%*%Lambda.max*w) ...
- exp(X1%*%beta0) + log((ILmax.1$A1/ILmax.2$A2)*ILmax.1$P1 + ILmax.2$P2 ... + (ILmax.3$A3/ILmax.2$A2)*ILmax.3$P3) # Calculation of h() the saddlepoint statistics Kstat[k] = -2*sum(K)

Pstat = 1 - pchisq(Kstat, 4) win.graph() par(mfrow=c(1,2))
plot(qchisq(ppoints(Kstat), 4), sort(Kstat), xlab="chi2(4)", ylab="Robust saddlepoint statistics")
abline(0,1) plot(qunif(ppoints(Pstat)), sort(Pstat), xlab="quantile of U[0,1]", ylab="P_value of Robust saddlepoint statistics")
abline(0,1) return(Kstat, Pstat)
Bibliography


Field, C. and Ronchetti, E. (1990), Small Sample Asymptotics, IMS Monograph Series: Hayward, California.


Chapter 3

Robust Small Sample Accurate Inference in Moment Condition Models

by

Serigne N. Lô and Elvezio Ronchetti

Department of Econometrics
University of Geneva
CH-1211 Geneva, Switzerland

June 2006
Abstract

Procedures based on the Generalized Method of Moments (GMM) are basic tools in modern econometrics. In most cases, the theory available for making inference in these models is based on first order asymptotic theory. It is well-known that the (first order) asymptotic distribution does not provide accurate p-values and confidence intervals in moderate to small samples. Moreover, in the presence of small deviations from the assumed model, p-values and confidence intervals based on classical GMM procedures can be drastically affected (nonrobustness). Several alternative techniques have been proposed in the literature to improve the accuracy of GMM procedures. These alternatives address either the first order accuracy of the approximations (information and entropy econometrics (IEE)) or the nonrobustness (Robust GMM estimators and tests). However, a procedure which is both robust and accurate in small samples is still an open question. In this paper, we propose a new alternative procedure which combines both robustness properties and small sample performance. Specifically, we combine IEE techniques as developed in Imbens, Spady, Johnson (1998) to obtain finite sample accuracy with robust methods obtained by bounding the original orthogonality function as proposed in Ronchetti and Trojani (2001). This leads to new robust estimators and tests in moment condition models with excellent finite sample accuracy. Finally, we illustrate the accuracy of the new statistic by means of some simulations for three models on overidentifying moment conditions.

Keywords: Exponential tilting, Generalized method of moments, Information and entropy econometrics, Robust tests, Saddlepoint techniques, Monte Carlo.
3.1 Introduction

Procedures based on the Generalized Method of Moments (GMM) (Hansen, 1982) are important tools in econometrics to estimate the parameters and make inference in moment condition models. In general, the inferential tools (p-values and confidence intervals) are based on first order asymptotic theory. More specifically, under appropriate regularity conditions, GMM estimators are asymptotically normal and the standard classical statistics for hypothesis testing are asymptotically \( \chi^2 \)-distributed. These results provide the tools used routinely in econometric analysis. However, there is evidence in the econometric literature that these asymptotic distributions do not provide accurate approximations to p-values and confidence intervals when the sample size is moderate to small; see for instance Altonji and Segal (1996), Burnside and Eichenbaum (1996), Hansen, Heaton, and Yaron (1996) among others in the July 1996’s special issue of the *Journal of Business and Economic Statistics*.

To alleviate this problem, several proposals have been put forward in the literature. An overview is presented in the July 1996’s special issue of *Journal of Business and Economic Statistics*. For instance, Hansen, Heaton, and Yaron (1996), opted for continuous updating estimators. Other authors such as Christiano and Haan (1996) found that imposing certain restrictions leads to substantial improvements in the small-sample properties of the statistical tests. Andersen and Sorensen (1996) stressed that it is generally not optimal to include many moments in the estimation procedure if the sample size is moderate to small. Bootstrap techniques have also been suggested to improve the approximation of the finite sample distribution of GMM statistics. Hall and Horowitz (1996) gave conditions under which the bootstrap provides asymptotic refinements to the critical values of \( t \)-tests and to the tests for overidentifying moment restrictions.

64
More recently, so-called information and entropy econometric (IEE) techniques have been used to improve the finite sample accuracy of GMM estimators and tests; see Imbens, Spady, and Johnson (1998) (ISJ thereafter) and for an overview, the March 2002 special issue of the *Journal of Econometrics*. The basic idea is to “tilt” the empirical distribution to the nearest distribution satisfying the moment conditions, where the distance is measured by a power divergence statistic (Cressie and Read, 1984) such as the Kullback-Leibler distance. These techniques are related to saddlepoint methods developed in the statistical literature for the fully identified case (M-estimators); see for instance Field and Ronchetti (1990), Spady (1991), Robinson, Ronchetti, and Young (2003).

In spite of their good finite sample accuracy when the model and the moment conditions are exactly satisfied, p-values and confidence intervals based on IEE techniques can be drastically affected as the original GMM procedures by small deviations from the underlying distribution of the model and from the corresponding moment conditions. Ronchetti and Trojani (2001) investigated this problem for the classical GMM procedures and derived robust alternatives to GMM estimators and tests. The goal of this paper is to extend these results to IEE techniques in order to obtain new estimators and tests which combine both robustness properties and good accuracy in moderate to small samples.

The paper is organized as follows. In section 3.2, we review IEE techniques by focusing in particular on exponential tilting (ET) techniques and provide a link with saddlepoint methods. Section 3.3 is devoted to the definition and the construction of a robust version of the exponential tilting estimator and corresponding test. In particular, we show that a necessary condition for the robustness of the ET estimator and test is the boundedness of the orthogonality function and its deriva-
tive with respect to the parameter. This implies a bounded influence function for the estimator and for the level of the corresponding test. When this condition is not satisfied by the original orthogonality function, we apply the technique developed in Ronchetti and Trojani (2001) to truncate the original orthogonality function and we use this modified orthogonality function in the ISJ procedure. This leads to new robust ET estimators and tests which are discussed in subsection 3.3.2. Section 3.4 presents a Monte Carlo study for three benchmark models which shows the excellent finite sample behavior of the new techniques both at the model and in the presence of small deviations from the model. Finally, section 3.5 provides some concluding remarks and suggestions for further research. The algorithm and the computational aspects are discussed in the Appendix.
3.2 Exponential tilting

Let $(Z_n)_{n \in \mathbb{N}}$ be a stationary ergodic sequence defined on an underlying probability space and taking values in $\mathbb{R}^N$ and let $\mathcal{P} = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^k\}$ be a family of distributions in $\mathbb{R}^N$ corresponding to the model distribution (or reference model). Further, let us define a function $h : \mathbb{R}^N \times \Theta \rightarrow \mathbb{R}^H$ that enforces a set of orthogonality conditions

$$E[h(Z; \theta_0)] = 0$$

(3.1)
on the structure of the underlying model. We assume that $\theta_0$ is the unique solution of (3.1) and we consider the case where the number of conditions $H$ is larger than the number of parameters $k$.

The GMM estimator $\hat{\theta}_{gmm}$ of $\theta_0$ (Hansen, 1982) is defined by

$$\hat{\theta}_{gmm} = \arg \min_{\theta} Q_W(\theta)$$

(3.2)
where $Q_W(\theta) = \left(\frac{1}{N} \sum_{i=1}^N h(Z_i; \theta)\right)' W^{-1} \left(\frac{1}{N} \sum_{i=1}^N h(Z_i; \theta)\right)$ for some positive semidefinite matrix $W$. Moreover, under (3.1) $N \cdot Q_W(\hat{\theta}_{gmm})$ is asymptotically $\chi^2_{H-k}$ distributed and can be used to test overidentifying conditions (Hansen’s test).

To improve the finite sample properties of the GMM estimator and Hansen’s test, ISJ proposed a class of alternative estimators based on the following idea. Given two discrete distributions $\tilde{\pi}$ and $\pi$ with common support and for a fixed scalar parameter $\lambda$, define the power-divergence statistic by (Cressie and Read, 1984)

$$I_\lambda(\tilde{\pi}; \pi) = \frac{1}{\lambda \cdot (1 + \lambda)} \sum_{i=1}^N \tilde{\pi}_i \left[\left(\frac{\tilde{\pi}_i}{\pi_i}\right)^\lambda - 1\right].$$

(3.3)
The estimator $\hat{\theta}$ of $\theta$, for a given $\lambda$, is then defined by the closest distribution to the empirical distribution, as measured by the Cressie-Read statistic, within the
set of distributions admitting a solution to the moment equations, i.e. \( \hat{\theta} \) is the solution of the problem

\[
\min_{\pi, \theta} I_\lambda(\tilde{\pi}; \pi), \quad \text{subject to} \quad \sum_{i=1}^{N} h(Z_i; \theta) \cdot \pi_i = 0 \quad \text{and} \quad \sum_{i=1}^{N} \pi_i = 1, \tag{3.4}
\]

where \( \tilde{\pi} \) is the vector of empirical frequencies \( \tilde{\pi}_i = \frac{1}{N} \) for \( i = 1, \ldots, N \).

Different values of \( \lambda \) lead to different estimators as discussed in ISJ. We focus on an important special case of this family of estimators, namely when \( \lambda \to -1 \). In this case, the optimization in (3.4) leads to the exponential tilted (ET) estimator \( \hat{\theta}_{et} \) which is defined as the minimizer of the Kullback-Leibler information criterion:

\[
\min_{\pi, \theta} \sum_{i=1}^{N} \pi_i \cdot \log(\pi_i) \quad \text{subject to} \quad \sum_{i=1}^{N} h(Z_i; \theta) \cdot \pi_i = 0 \quad \text{and} \quad \sum_{i=1}^{N} \pi_i = 1. \tag{3.5}
\]

It turns out that \( \pi_i \) is given by

\[
\pi_i = \frac{e^{t'h(Z_i; \theta)}}{\sum_{j=1}^{N} e^{t'h(Z_j; \theta)}}, \tag{3.6}
\]

and by defining the empirical cumulant generating function of \( h(Z_i; \theta) \),

\[
K(t; \theta) = \log \left( \frac{1}{N} \sum_{i=1}^{N} e^{t'h(Z_i; \theta)} \right), \tag{3.7}
\]

we obtain

\[-K(t, \theta) = \sum_{i=1}^{N} \pi_i \log(\pi_i) + \log(N). \tag{3.8}\]

Therefore (3.5) can be rewritten more compactly as

\[
\max_{t, \theta} K(t; \theta) \quad \text{subject to} \quad \frac{\partial}{\partial t} K(t; \theta) = 0, \tag{3.9}
\]

where \( \pi_i \) is defined by (3.6).

Under regularity conditions, the tilted estimator \( \hat{\theta}_{et} \) is asymptotically (first order) equivalent to the GMM estimator, i.e. \( \sqrt{N}(\hat{\theta}_{et} - \theta_0) \) has the same asymptotic
normal distribution as $\sqrt{N}(\hat{\theta}_{gmm} - \theta_0)$.

The corresponding test for overidentifying moment restrictions is based on the test statistic $-2 \cdot N \cdot K(t; \hat{\theta}^{et})$ ($= 2 \cdot N \cdot KLIC(\tilde{\pi}^{et}; \tilde{\pi})$ in ISJ, p. 342). Under the null hypothesis, this test statistic has the same asymptotic distribution as the classical Hansen test statistic, i.e. $\chi^2_d$, where $d = H - k$.

ISJ provide convincing evidence that $\hat{\theta}^{et}$ and the corresponding test have better finite sample properties than $\hat{\theta}_{gmm}$ and Hansen’s test. Furthermore, by (3.6), (3.7) and (3.9),

$$\frac{\partial}{\partial t} K(t; \theta) = e^{-K(t; \theta)} \cdot \frac{1}{N} \sum_{i=1}^{N} h(Z_i; \theta) e^{t' h(Z_i; \theta)}$$

$$= \sum_{i=1}^{N} h(Z_i; \theta) \pi_i(\theta) = E_{\pi} [h(Z; \theta)] = 0,$$

i.e. the empirical distribution $(\frac{1}{N}, ..., \frac{1}{N})$ is tilted to $(\pi_1, ..., \pi_N)$ in order to satisfy the orthogonality conditions under $(\pi_1, ..., \pi_N)$. This is the key procedure to obtain saddlepoint approximations of the distribution of estimators and test statistics which are well known to be highly accurate; cf. for instance Daniels (1954), Field and Ronchetti (1990), and Spady (1991) for the fully identified case (M-estimators). Indeed the empirical version used here corresponds to the so-called empirical saddlepoint approximation; see Ronchetti and Welsh (1994) and for a connection with empirical likelihood, Monti and Ronchetti (1993).
3.3 Robust Exponential Tilting

The tilted estimator $\hat{\theta}_{et}$ is an attractive alternative to the GMM estimator $\hat{\theta}_{gmm}$ when the moment conditions (3.1) are exactly specified. In this section, we want to investigate the behavior of the tilted estimator and the corresponding tests in the presence of slight misspecifications of the moment conditions. Let us first review these aspects for $\hat{\theta}_{gmm}$.

3.3.1 Robust alternatives to the GMM

The lack of robustness of the GMM estimator and tests in the presence of small deviations from the underlying distribution has already been studied extensively; see Ronchetti and Trojani (2001) and references therein. In particular, in that paper, it is shown that the influence function of the GMM estimator is proportional to the orthogonality function $h$. When $h(z; \theta)$ is unbounded in $z$, this leads to non-robust estimators. An alternative robust version was proposed as follows.

Consider the Huber function

\[ H_c : \mathbb{R}^H \rightarrow \mathbb{R}^H \]

\[ y \mapsto y \cdot w_c(y) = \begin{cases} 
  y & \text{if } \|y\| \leq c \\
  c \cdot \frac{y}{\|y\|} & \text{if } \|y\| > c ,
\end{cases} \quad (3.10) \]

where $w_c(y) = \min(1, \frac{c}{\|y\|})$ for $y \neq 0$ and $w_c(0) = 1$, and a new mapping $h^{A, \tau}_c : \mathbb{R}^N \times \Theta \rightarrow \mathbb{R}^H$ defined by

\[ h^{A, \tau}_c(z ; \theta) = H_c(A(\theta)[h(z ; \theta) - \tau(\theta)]) , \quad (3.11) \]

where the nonsingular matrix $A \in \mathbb{R}^{H \times H}$ and the vector $\tau \in \mathbb{R}^H$ are determined through the implicit equations:
\[
\begin{align*}
  & E_{\theta} \left[ h_{c}^{A,\tau}(Z ; \theta) \right] = 0 \\
  & \frac{1}{N} \sum_{i=1}^{N} \left[ h_{c}^{A,\tau}(Z_{i} ; \theta) \right] \cdot \left[ h_{c}^{A,\tau}(Z_{i} ; \theta) \right]' = I.
\end{align*}
\] (3.12)

Then, the GMM estimator \( \hat{\theta}^{gmm}_{c} \) and the corresponding tests defined by the modified bounded orthogonality conditions \( h_{c}^{A,\tau} \) have an influence function bounded by \( c \) (\( \geq \sqrt{H} \)) and are robust in the sense of Hampel, Ronchetti, Rousseeuw, and Stahel (1986). An iterative algorithm for the computation of \( A(\theta) \), \( \tau(\theta) \), and \( \hat{\theta}^{gmm} \) is provided by Ronchetti and Trojani (2001, p. 47); see also Appendix 3.6.1. The choice of the tuning constant \( c \) is discussed in Ronchetti and Trojani (2001).

### 3.3.2 Robust exponential tilting estimator and test

In view of section 3.2 and subsection 3.3.1, it seems natural at this point to try and derive an estimator (and the corresponding tests) with the good finite sample properties of \( \hat{\theta}^{et}_{c} \) and the robustness properties of \( \hat{\theta}^{gmm}_{c} \). This can be achieved by solving (3.9), with \( h(z ; \theta) = h_{c}^{A,\tau}(z ; \theta) \).

More specifically, by writing \( K_{c}(t ; \theta) = \log \left[ \frac{1}{N} \sum_{i=1}^{N} e^{t h_{c}(Z_{i} ; \theta)} \right] \) and \( h_{c}(\cdot ; \cdot) \) instead of \( h_{c}^{A,\tau}(\cdot ; \cdot) \) for simplicity, the new robust tilting estimator \( \hat{\theta}^{et}_{c} \) is defined by the optimization problem:

\[
\max_{t, \theta} K_{c}(t ; \theta),
\] (3.13)

subject to
\[
\sum_{i=1}^{N} h_c(Z_i; \theta) e^{t' h_c(Z_i; \theta)} = 0 \tag{3.13a}
\]

\[
E_{\theta} [h_c(Z; \theta)] = 0 \tag{3.13b}
\]

\[
\frac{1}{N} \sum_{i=1}^{N} h_c(Z_i; \theta) h'_c(Z_i; \theta) = I. \tag{3.13c}
\]

From a computational point of view, the same remarks as in ISJ, p. 339 apply.

\( \hat{\theta}^{et} \) is asymptotically equivalent to \( \hat{\theta}^{gmm} \), the robust GMM estimator defined by \( h_c(\cdot; \cdot) \). Moreover, when \( c \to \infty \), we recover the classical estimator \( \hat{\theta}^{et} \). Notice that even in the case where the cumulant generating function of \( h(Z; \theta) \) does not exist and \( \hat{\theta}^{et} \) is not defined, \( \hat{\theta}^{et} \) with a finite \( c \) exists and is an alternative to the classical estimator; see subsections 3.4.2 and 3.4.3. Finally, the corresponding robust test for overidentifying restrictions is defined by the test statistic \(-2 \cdot N \cdot K_c(t_c; \hat{\theta}^{et})\) which is asymptotically \( \chi^2_d \) under the null hypothesis, where \( d = H - k \).

Let us now investigate in more details the robustness properties of \( \hat{\theta}^{et} \). The tilting estimator can be viewed as an M-estimator (ISJ, p. 337) with estimating equations \( \sum_{i=1}^{N} \rho^{et}(Z_i; \hat{\theta}^{et}, \hat{t}^{et}) = 0 \), where

\[
\rho^{et}(z; \theta, t) = \begin{pmatrix}
    t' \frac{\partial h}{\partial \theta}(z; \theta) \cdot \exp(t' h(z; \theta)) \\
    h(z; \theta) \cdot \exp(t' h(z; \theta))
\end{pmatrix}. \tag{3.14}
\]

The influence function of estimators defined by estimating equations (M-estimators) is proportional to the estimating function (Huber, 1981), i.e.

\[
IF(z; \hat{\theta}^{et}, P_{\theta}) = E \left[ - \frac{\partial \rho^{et}}{\partial (\theta', t')} (z; \hat{\theta}^{et}, t) \right]^{-1} \rho^{et}(z; \hat{\theta}^{et}, t). \tag{3.15}
\]

72
The boundedness of the influence function implies a bounded bias of the estimator and of the level of the corresponding test when the underlying distribution lies in a neighborhood of the model (see Heritier and Ronchetti, 1994 and Ronchetti and Trojani, 2001). Here, the IF for ET estimators is bounded if and only if \( \rho^e_t \) is bounded with respect to \( z \). Therefore, we can focus our analysis on the function \( \rho^e_t \) to determine the robustness properties of the corresponding estimators and tests.

Generally, in the classical version, \( \rho^e_t \) is not bounded. In fact, both \( h(.,.) \) and \( \frac{\partial h}{\partial \theta} \) are not necessarily bounded. So the resulting estimators are not guaranteed to be robust. For \( \hat{\theta}^e_t, h_c(.,.) \) is bounded by construction, and therefore this estimator is robust if \( \frac{\partial h_c}{\partial \theta} \) is bounded.

Consider the robust ET estimator defined by the orthogonality function (3.11). It follows, with \( y = A(\theta)[h(z ; \theta) - \tau(\theta)] \),

\[
\frac{\partial}{\partial \theta} h_c(z ; \theta) = \frac{\partial}{\partial \theta} K_c(y) =
\begin{cases}
A' A^{-1} y + A(\frac{\partial h}{\partial \theta} h(z ; \theta) - \tau) & \text{if } \| y \| \leq c \\
c\{ I - \frac{y}{\| y \|} \cdot \frac{y^T}{\| y \|} \} \cdot \{ A' A^{-1} \frac{y}{\| y \|} + A \frac{1}{\| y \|} \left[ \frac{\partial h}{\partial \theta} h(z ; \theta) - \tau \right] \} & \text{if } \| y \| > c.
\end{cases}
\]

Thus \( \frac{\partial}{\partial \theta} h_c(z ; \theta) \) is bounded with respect to \( z \) if and only if

\[
\begin{cases}
\frac{\partial h}{\partial \theta} \text{ is bounded when } \| y \| \leq c \\
\frac{1}{\| y \|} \frac{\partial h}{\partial \theta} \text{ is bounded when } \| y \| > c.
\end{cases}
\]

73
The last two conditions are satisfied when $\frac{\partial h}{\partial \theta}$ is bounded everywhere. Then, for a given model, when the derivative of the moment vector with respect to the parameters is bounded, the robustness properties of the estimator $\hat{\theta}_{et}^c$ follow. If this is not the case, we have to check the boundedness of the conditions defined by (3.17) to determine the robustness properties of the estimator and of the test for the specific model.

Figure 3.1: Schematic illustration of the robustness properties of the exponential tilting estimator and test.

* This includes the case when the cumulant generating function of $h(\cdot; \cdot)$ is not defined.
3.4 Monte Carlo Investigation

To illustrate and compare the behavior of classical and robust ET estimators and tests, we perform a Monte Carlo experiment for three benchmark models (Chi-squared moments, Hall-Horowitz, stochastic lognormal volatility model). In each case we work with data generated from the model and from various slight perturbations of the model. We compute \( \hat{\theta}_{et} \) and \( \hat{\theta}_{et}^c \) and their corresponding tests for overidentifying moment restrictions based on the test statistics 

\[-2 \cdot N \cdot K(t ; \hat{\theta}_{et}) \]

and 

\[-2 \cdot N \cdot K_c(t_c ; \hat{\theta}_{et}^c) \]

respectively. Under the null hypothesis, these tests statistics are asymptotically distributed as \( \chi_d^2 \), where \( d = H - k \). We also report, where they are available, the best results obtained by ISJ by means of other tilted test statistics.

In each experiment, we simulate 5000 samples and we report the actual sizes \( P[T > v_\alpha] \) for each test based on a test statistic \( T \) corresponding to the nominal sizes \( \alpha = 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.2 \), where \( v_\alpha \) is the critical value of the test, i.e. \( P[\chi_d^2 > v_\alpha] = \alpha \). QQ-plots with respect to \( \chi_d^2 \) quantiles and relative errors \( (P[T > v_\alpha] - \alpha)/\alpha \) for each tests are also reported.
3.4.1 Model 1: Chi-squared Moments

The first Monte Carlo experiment focuses on a two moments, one parameter problem defined by the moment vector:

\[ h(Z; \theta) = \begin{pmatrix} Z - \theta \\ Z^2 - \theta^2 - 2\theta \end{pmatrix}. \]

The distribution of \( Z \) is \( \chi^2_1 \), \( \theta_0 = 1 \), and the data are generated from this model. Here \( h(z; \theta) \) is unbounded in \( z \) and \( \frac{\partial h}{\partial \theta}(z; \theta) = -\left( \frac{1}{2(\theta + 1)} \right) \) is constant with respect to \( z \). Therefore, we can use \( h_c(z; \theta) \) and we can expect good robustness and finite sample accuracy from \( \hat{\theta}^{et} \) and its corresponding tests.

The results of our simulations, for two sample sizes \( N = 500, 250 \), are presented in Table 3.1. In the first case (\( N = 500 \)), we can compare the results of our new robust test to those obtained by the test statistics in ISJ. The first column shows the actual size of the test based on the ET estimator of the Lagrange multipliers (\( ISJ.500 \)), cf. ISJ p.343. The following two columns give the results for the classical ET (\( classET.500 \)) and for the robust ET (\( robET.500 \)) respectively. We notice that the nominal sizes of \( robET.500 \) are the closest to the actual size. Notice that \( classET.500 \) test is very similar to the classical GMM specification test and shows a very liberal behavior on terms of size. The ISJ.500 is between the two ET statistics in terms of accuracy.

For a better evaluation of the small sample properties of the robust ET statistics, we also tested a reduced sample size of 250. The results of the classical and robust ET, \( classET.250 \) and \( robET.250 \), are reported in the last two columns of Table 3.1. Even with such a small sample size, the robust ET outperforms the classical test, and the corresponding nominal sizes are very close to the actual
sizes. These conclusions are confirmed by the graphical analysis in Figure 3.2.

Finally, we plot the relative errors, a much more stringent measure than absolute errors, for \( \text{robET.500} \) in Figure 3.3 (a) and \( \text{robET.250} \) in Figure 3.3 (b). Again, these plots demonstrate the very high accuracy of the robust ET test. In fact, the relative error in the tail for the robust ET test for \( N = 250 \) is smaller than 4\% down to \( \alpha = 0.02 \) and still reasonable for smaller sizes. The relative errors of the classical statistics are not reported because they exceed 100\% already for \( \alpha = 0.05 \). These results show that even in the case of no contamination, the robust ET test has a very high finite sample accuracy and is an interesting alternative to classical GMM and ET tests.

Table 3.1: Comparison of actual and nominal size of the statistics applied to \textit{Chi-Squared Moments} (Model 1) without contamination i.e. \( Z \sim \chi^2_1, H = 2, k = 1 \) and 5000 replications. \( ISJ.N \) = best statistics from ISJ; \( classET.N \) = classical ET test; \( robET.N \) = robust test. \( .N \) indicates the sample size. The tuning constant for the robust test was set to \( c = 2 \).

<table>
<thead>
<tr>
<th>\text{nom.size}</th>
<th>ISJ.500</th>
<th>\text{classET.500}</th>
<th>\text{robET.500}</th>
<th>\text{classET.250}</th>
<th>\text{robET.250}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>0.237</td>
<td>0.2552</td>
<td>0.2108</td>
<td>0.2772</td>
<td>0.1996</td>
</tr>
<tr>
<td>0.100</td>
<td>0.125</td>
<td>0.1554</td>
<td>0.1048</td>
<td>0.1820</td>
<td>0.0986</td>
</tr>
<tr>
<td>0.050</td>
<td>0.068</td>
<td>0.1044</td>
<td>0.0488</td>
<td>0.1266</td>
<td>0.0486</td>
</tr>
<tr>
<td>0.025</td>
<td>0.038</td>
<td>0.0712</td>
<td>0.0246</td>
<td>0.0930</td>
<td>0.0260</td>
</tr>
<tr>
<td>0.010</td>
<td>0.019</td>
<td>0.0474</td>
<td>0.0100</td>
<td>0.0642</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.005</td>
<td>0.010</td>
<td>0.0354</td>
<td>0.0048</td>
<td>0.0524</td>
<td>0.0040</td>
</tr>
<tr>
<td>0.001</td>
<td>0.003</td>
<td>0.0180</td>
<td>0.0004</td>
<td>0.0293</td>
<td>0.0006</td>
</tr>
</tbody>
</table>
Figure 3.2: QQ-plots of overidentifying ET statistics versus $\chi^2_1$.

Figure 3.3: Relative errors for robET.500 (a) and robET.250 (b)
Figure 3.4: Probability distribution functions (pdf) and QQ-plots versus $\chi^2_1$ for 100'000 simulated observations of two distributions (a) & (b) $0.95\chi^2_1 + 0.05\chi^2_{10}$ and (c) & (d) $\Gamma\left(\frac{1}{4}; \frac{1}{4}\right)$.

Let us now investigate the behavior of the classical procedure when the data follow a slightly perturbed model distribution. To illustrate the effects, we assume that $Z$ does not follow the model distribution $\chi^2_1$ but two contaminated distributions

$$Z \sim 0.95 \cdot \chi^2_1 + 0.05 \cdot \chi^2_{10}$$  \hspace{1cm} (3.18)

and

$$Z \sim \Gamma\left(\frac{1}{4}; \frac{1}{4}\right).$$  \hspace{1cm} (3.19)
In the first case, the Kolmogorov distance between the model and the contaminated distribution (i.e. the maximum difference between the distribution functions of the two distributions) is less than 0.05. In the second case, the Kolmogorov distance is 0.19. This means that (3.18) can be viewed as a small perturbation of the model distribution. On the other hand, (3.19) represents a larger perturbation. Figure 3.4 shows the QQ-plots of the distributions with respect to the $\chi^2_1$ quantiles. We do not argue that (3.18) or (3.19) should replace the original $\chi^2_1$. These are just illustrations of potential small deviations from the model. We still assume the original model with its moment conditions but we take into account the fact that in reality, the data might come from a slightly different unknown distribution with slightly different moment conditions. Thus, our goal is to have procedures based on the original model and moment conditions which still behave reasonably well in the presence of (unknown) small deviations.

The results of Tables 3.2, 3.3 and Figures 3.5, 3.6 show that the classical ET test is very inaccurate whereas the robust ET test is stable and very accurate even in the presence of small deviations from the underlying model.
Table 3.2: Comparison of actual and nominal size of the statistics applied to \textit{Chi-Squared Moments} (Model 1) with contaminated data (3.18), \( H = 2, \ k = 1 \) and 5000 replications. \textit{classET.N} = classical ET test; \textit{robET.N} = robust ET test. \( .N \) indicates the sample size.

The tuning constant for the robust test was set to \( c = 2 \).

<table>
<thead>
<tr>
<th>nom.size</th>
<th>classET.500</th>
<th>robET.500</th>
<th>classET.250</th>
<th>robET.250</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>0.2640</td>
<td>0.2028</td>
<td>0.2742</td>
<td>0.2074</td>
</tr>
<tr>
<td>0.100</td>
<td>0.1624</td>
<td>0.0952</td>
<td>0.1688</td>
<td>0.1118</td>
</tr>
<tr>
<td>0.050</td>
<td>0.1006</td>
<td>0.0466</td>
<td>0.1108</td>
<td>0.0530</td>
</tr>
<tr>
<td>0.025</td>
<td>0.0666</td>
<td>0.0240</td>
<td>0.0750</td>
<td>0.0270</td>
</tr>
<tr>
<td>0.010</td>
<td>0.0378</td>
<td>0.0084</td>
<td>0.0504</td>
<td>0.0110</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0280</td>
<td>0.0042</td>
<td>0.0352</td>
<td>0.0060</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0154</td>
<td>0.0006</td>
<td>0.0178</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

Table 3.3: Comparison of actual and nominal size of the statistics applied to \textit{Chi-Squared Moments} (Model 1) with contaminated data (3.19), \( H = 2, \ k = 1 \) and 5000 replications. \textit{classET.N} = classical ET test; \textit{robET.N} = robust ET test. \( .N \) indicates the sample size.

The tuning constant for the robust test was set to \( c = 2 \).

<table>
<thead>
<tr>
<th>nom.size</th>
<th>classET.500</th>
<th>robET.500</th>
<th>classET.250</th>
<th>robET.250</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>0.2870</td>
<td>0.1940</td>
<td>0.2830</td>
<td>0.2178</td>
</tr>
<tr>
<td>0.100</td>
<td>0.1748</td>
<td>0.0960</td>
<td>0.1764</td>
<td>0.1168</td>
</tr>
<tr>
<td>0.050</td>
<td>0.1054</td>
<td>0.0462</td>
<td>0.1138</td>
<td>0.0576</td>
</tr>
<tr>
<td>0.025</td>
<td>0.0728</td>
<td>0.0240</td>
<td>0.0812</td>
<td>0.0310</td>
</tr>
<tr>
<td>0.010</td>
<td>0.0436</td>
<td>0.0078</td>
<td>0.0568</td>
<td>0.0132</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0312</td>
<td>0.0044</td>
<td>0.0440</td>
<td>0.0076</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0148</td>
<td>0.0008</td>
<td>0.0244</td>
<td>0.0022</td>
</tr>
</tbody>
</table>
Figure 3.5: QQ-plots of overidentifying ET statistics versus $\chi^2_1$ with $Z \sim 0.95\chi^2_1 + 0.05\chi^2_{10}$.

Figure 3.6: QQ-plots of overidentifying ET statistics versus $\chi^2_1$ with $Z \sim \Gamma(\frac{1}{4}; \frac{1}{4})$. 

82
3.4.2 Model 2: Hall-Horowitz (1996)

In this experiment, we consider a design investigated by Hall and Horowitz (1996), where the moment vector has the form:

\[
h(Z; \theta) = \begin{pmatrix}
    e^{-0.72 - \theta \cdot (Z^{(1)} + Z^{(2)}) + 3 \cdot Z^{(2)}} - 1 \\
    Z^{(2)} \cdot [e^{-0.72 - \theta \cdot (Z^{(1)} + Z^{(2)}) + 3 \cdot Z^{(2)}} - 1]
\end{pmatrix}.
\]

The vector \((Z^{(1)}, Z^{(2)})'\) follows a bivariate normal distribution with means zero \((0, 0)'\), variances 0.16 and correlation coefficient zero. The true value of \(\theta\) is \(\theta_0 = 3\).

We follow the same approach as in the first model. The simulation results are reported in Table 3.4. In order to compare our analysis to the ISJ results, we simulate data with two different sample sizes, 200 and 100. The columns \(ISJ.N\) represent the closest nominal size from the ISJ investigation with sample sizes of 200 and 100 respectively (cf. ISJ p.345). The columns \(robET.N\) report the simulations result of the robust ET, where the constant \(c\) is fixed to 2.

Since, for this model, the cumulant generating function of the score vector does not exist, the classical ET test cannot be defined. However, we can “simulate” this case by means of our robust ET test with a large tuning constant \(c\) (for example \(c = 80\)). We call this test a “classical” ET test (“classET”). Notice however, that we do not recommend using this test, the accuracy of the robust ET test with \(c = 2\) being so much better.

Inspection of Table 3.4 reveals the high accuracy of the robust ET test and its better performance compared to the best statistics from ISJ for the two sample sizes. This results are confirmed by the QQ-plots in Figure 3.7 and the analysis of the relative error in Figure 3.8 (a) and (b).
Table 3.4: Comparison of actual and nominal size of the statistics applied to Hall-Horowitz’s design (Model 2) without contamination i.e.

\[
\begin{pmatrix}
Z^{(1)} \\
Z^{(2)}
\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .16 & 0 \\
0 & .16
\end{pmatrix}\right),
\]

\(H = 2, k = 1\) and 5000 replications. ISJ.\(N\) = best statistics for ISJ; “classET”.\(N\) = “classical” ET test; robET.\(N\) = robust ET test. \(N\) indicates the sample size. The tuning constant for the robust test was set to \(c = 2\).

<table>
<thead>
<tr>
<th>nom.size</th>
<th>ISJ.200</th>
<th>“classET”.200</th>
<th>robET.200</th>
<th>ISJ.100</th>
<th>“classET”.100</th>
<th>robET.100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>0.228</td>
<td>0.2486</td>
<td>0.2020</td>
<td>0.250</td>
<td>0.2807</td>
<td>0.2092</td>
</tr>
<tr>
<td>0.100</td>
<td>0.125</td>
<td>0.1459</td>
<td>0.0972</td>
<td>0.128</td>
<td>0.1776</td>
<td>0.1022</td>
</tr>
<tr>
<td>0.050</td>
<td>0.065</td>
<td>0.0923</td>
<td>0.0468</td>
<td>0.070</td>
<td>0.1175</td>
<td>0.0524</td>
</tr>
<tr>
<td>0.025</td>
<td>0.035</td>
<td>0.0582</td>
<td>0.0270</td>
<td>0.043</td>
<td>0.0800</td>
<td>0.0286</td>
</tr>
<tr>
<td>0.010</td>
<td>0.016</td>
<td>0.0338</td>
<td>0.0110</td>
<td>0.022</td>
<td>0.0509</td>
<td>0.0134</td>
</tr>
<tr>
<td>0.005</td>
<td>0.008</td>
<td>0.0231</td>
<td>0.0042</td>
<td>0.013</td>
<td>0.0376</td>
<td>0.0070</td>
</tr>
<tr>
<td>0.001</td>
<td>0.002</td>
<td>0.0012</td>
<td>0.0008</td>
<td>0.004</td>
<td>0.0194</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

Figure 3.7: QQ-plots of overidentifying ET statistics versus \(\chi^2_1\).
Similarly to Model 1, we studied the robustness of our new statistics for Hall-Horowitz’s model when the data are contaminated according to the following distribution

\[
\begin{pmatrix}
Z^{(1)}
\end{pmatrix}
\sim 0.95 \cdot \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .16 & 0 \\ 0 & .16 \end{pmatrix}\right) + 0.05 \cdot \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\right)
\]

In spite of this perturbation, the results for the robust ET statistics are only slightly modified compared to results from non-contaminated data. In contrast, the results for the “classical” ET tests are markedly worse cf. Table 3.5 and Figure 3.9.
Table 3.5: Comparison of actual and nominal size of the statistics applied to Hall-Horowitz’s design (Model 2) with contaminated data. $H = 2$, $k = 1$ and 5000 replications. “classET”.N= “classical” ET test; robET.N= robust ET test. .N indicates the sample size. The tuning constant for the robust test was set to $c = 2$.

<table>
<thead>
<tr>
<th>nom.size</th>
<th>“classET”.200</th>
<th>robET.200</th>
<th>“classET”.100</th>
<th>robET.100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>0.2692</td>
<td>0.2160</td>
<td>0.3010</td>
<td>0.1988</td>
</tr>
<tr>
<td>0.100</td>
<td>0.1620</td>
<td>0.1010</td>
<td>0.1908</td>
<td>0.1072</td>
</tr>
<tr>
<td>0.050</td>
<td>0.1022</td>
<td>0.0564</td>
<td>0.1236</td>
<td>0.0554</td>
</tr>
<tr>
<td>0.025</td>
<td>0.0678</td>
<td>0.0314</td>
<td>0.0902</td>
<td>0.0320</td>
</tr>
<tr>
<td>0.010</td>
<td>0.0412</td>
<td>0.0130</td>
<td>0.0630</td>
<td>0.0136</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0300</td>
<td>0.0078</td>
<td>0.0506</td>
<td>0.0080</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0122</td>
<td>0.0024</td>
<td>0.0298</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

Figure 3.9: QQ-plots of overidentifying ET statistics versus $\chi^2_1$ with perturbed data.
3.4.3 Model 3: Stochastic Lognormal Volatility Model

The stochastic lognormal volatility (SLV) model offers a powerful alternative to GARCH-type models to explain the well-documented time varying volatility. Moreover, the SLV model provides a reasonable first approximation to model the properties of most financial return series.

During the last ten years, a number of Monte Carlo studies have explored the small sample properties of these estimators. Since, the maximum likelihood approach is difficult to implement, this has left the field open to competition among alternative procedures such as GMM (Melino and Turnbull, 1990), maximum likelihood Monte Carlo (Sandmann and Koopman, 1996), quasi-maximum likelihood, Bayesian Markov Chain Monte Carlo (Jacquier, Polson, and Rossi, 1994), maximum likelihood through numerical integration (Fridman and Harris, 1998) and efficient method of moments (EMM) (Gallant and Tauchen, 1996). Andersen, Chung, and Sørenson (1999) have investigated the finite sample comparison of various methods for estimating SLV. Out of the six alternative methods mentioned above, they found that EMM completely overshadows the others with its flexibility and efficiency.

Consider the simple version of SLV model defined by:

\[
\begin{align*}
    y_t &= \sigma_t Z_t \\
    \log \sigma_t^2 &= w + \beta \log \sigma_{t-1}^2 + \sigma_u u_t
\end{align*}
\]

where \( t = 1, ..., N \), \( \theta = (w, \beta, \sigma_u) \) is the parameter vector, and \((Z_t, u_t)\) are iid \(N(0, I_2)\), that is, the error terms are mutually independent and distributed according to a standard normal distribution. In the model, returns display zero serial
correlation but the dependence in the higher-order moments is induced through
the stochastic volatility term, $\sigma_t$, the logarithm of which follows a first order
autoregressive $[AR(1)]$ model. The volatility persistence parameter, $\beta$, is estimated
to be less than unity, but quite close to it in most empirical studies. Finally, the
assumption of lognormality of the volatility process is a convenient parameteriza-
tion that allows for closed-form solutions of the moments and is consistent with the
evidence of excess kurtosis or “fat tails” in the unconditional return distribution.

When we impose the inequality constraints $0 < \beta < 1$ and $\sigma_u \geq 0$ to the model,
the return innovation series $y_t$ becomes strictly stationary and ergodic, and uncondi-
tional moments of any order exist. Throughout, we work with parameter values
that satisfy these additional inequalities. To implement the robust ET procedure,
we use 5 orthogonality conditions used by Andersen and Sørenson (1996). The
moment vector is defined by

$$h(y, \theta) = \begin{pmatrix}
|y_t| - \sqrt{\frac{2}{\pi}} \exp \left( \frac{\mu}{2} + \frac{\sigma^2}{8} \right) \\
y_t^2 - \exp \left( \mu + \frac{\sigma^2}{2} \right) \\
|y_t y_{t-1}| - \frac{2}{\pi} \exp \left( \mu + \frac{\sigma^2}{4} \right) \exp \left( \beta \frac{\sigma^2}{4} \right) \\
|y_t y_{t-3}| - \frac{2}{\pi} \exp \left( \mu + \frac{\sigma^2}{4} \right) \exp \left( \beta^3 \frac{\sigma^2}{4} \right) \\
|y_t y_{t-5}| - \frac{2}{\pi} \exp \left( \mu + \frac{\sigma^2}{4} \right) \exp \left( \beta^5 \frac{\sigma^2}{4} \right)
\end{pmatrix},$$

where $\mu = \frac{w}{1-\beta}$, $\sigma^2 = \frac{\sigma^2}{1-\beta^2}$ and $\theta = (w, \beta, \sigma_u)$.

We simulate 5000 samples of 500 observations from the SLV model. The vector
of parameters fixed for the simulation of the data vector $Z$ is $\theta_0 = (-.368, .95, .260)$. These values correspond to those used in the empirical study of the SLV by Jacquier
et al. (1994) and Andersen et al. (1999), among others.
The samples of size 500 are small by the standards of high-frequency financial time series analysis so the results presented here show the small sample properties of the ET method. Table 3.6 show in this case the good accuracy of the robust ET method for the overidentifying moments test. The nominal size of the robust ET method is close to the actual size even in the extreme tail. The QQ-plot in Figure 3.10 (a) confirms these results. Even when the data is contaminated (according to the configuration given in Table 3.7), the accuracy of the robust ET test is good. Figure 3.10 (b) confirms these results.

Finally, we estimate the SLV model (without contamination) by EMM and compare the results obtained with our new robust ET estimator. We choose EMM because of its flexibility and efficiency. The EMM computations are based on the procedure outlined in Gallant and Tauchen (2001), implemented in Finmetrics, with the optimal auxiliary model chosen automatically. Table 3.8 shows the bias, the variance and the associated root mean squared errors (RMSE) for each parameter and for both EMM and robust ET method. With respect to bias, variance, and RMSE, the robust ET method dominates EMM for all parameters except for $\sigma_u$, where the bias of EMM is smaller than that of the robust ET method. In particular, the reduction in RMSE is substantial.
Table 3.6: Comparison of actual and nominal size of the robust ET statistic applied to SLV’s design (Model 3) without contamination i.e.

\[
\begin{pmatrix}
Z_t \\
u_t
\end{pmatrix}
\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)
\]

\(H = 5, k = 3\) and 5000 replications. \(\text{robET.N} = \) robust ET test. \(N\) indicates the sample size. The tuning constant was set to \(c = 2.5\)

<table>
<thead>
<tr>
<th>nom.size</th>
<th>robET.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>0.2276</td>
</tr>
<tr>
<td>0.100</td>
<td>0.1186</td>
</tr>
<tr>
<td>0.050</td>
<td>0.0634</td>
</tr>
<tr>
<td>0.025</td>
<td>0.0318</td>
</tr>
<tr>
<td>0.010</td>
<td>0.0100</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0050</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0012</td>
</tr>
</tbody>
</table>
Table 3.7: Comparison of actual and nominal size of the robust ET statistic applied to SLV’s design (Model 3) with contaminated data i.e.

\[
\begin{pmatrix}
Z_t \\
u_t
\end{pmatrix}
\sim 0.95 \cdot \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) + 0.05 \cdot \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix} \right)
\]

\(H = 5, \ k = 3\) and 5000 replications. \(\text{robET}.N\) = robust ET test. \(N\) indicates the sample size. The tuning constant was set to \(c = 2.5\)

<table>
<thead>
<tr>
<th>nom.size</th>
<th>robET.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>0.2330</td>
</tr>
<tr>
<td>0.100</td>
<td>0.1186</td>
</tr>
<tr>
<td>0.050</td>
<td>0.0604</td>
</tr>
<tr>
<td>0.025</td>
<td>0.0300</td>
</tr>
<tr>
<td>0.010</td>
<td>0.0114</td>
</tr>
<tr>
<td>0.005</td>
<td>0.0068</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

Figure 3.10: QQ-plots of overidentifying ET statistics versus \(\chi^2(2)\) with normal (a) and contaminated data (b).
Table 3.8: Comparison of RMSEs of EMM and ET method with 5000 replications; true parameters \((w, \beta, \sigma_u) = (-.368, .95, .260)\).

<table>
<thead>
<tr>
<th>Method</th>
<th>(w)</th>
<th>(\beta)</th>
<th>(\sigma_u)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bias var RMSE</td>
<td>bias var RMSE</td>
<td>bias var RMSE</td>
</tr>
<tr>
<td>EMM</td>
<td>-.1640 .3367 .6030</td>
<td>-.0150 .0063 .0810</td>
<td>.0150 .0398 .2002</td>
</tr>
<tr>
<td>robust ET</td>
<td>.1020 .0380 .2220</td>
<td>.0144 .0006 .0299</td>
<td>-.0346 .0118 .1142</td>
</tr>
</tbody>
</table>

3.5 Conclusion

The Robust ET method is a useful procedure which provides attractive alternative estimators and tests to standard GMM methods. Our analysis shows that the new test statistic for overidentifying restrictions has excellent small sample properties for inference. Moreover, by its robustness, the procedure provides reliable estimators and tests when the model does not hold exactly. Furthermore, ET method is as flexible as GMM because it requires only a modified moments vector. This is an advantage e.g. with respect to EMM which requires an auxiliary model. Future research directions include the application of this method to other more complex models and the development of more efficient computational procedures.
3.6 APPENDIX

Here we provide the algorithm and the computational aspects for solving (3.13) under the constraints (3.13.a), (3.13.b) and (3.13.c). For the particular models studied in this paper, Matlab’s code is available from the authors upon request.

3.6.1 The Algorithm

To develop the algorithm for a general robust ET, we extend the procedure presented in Ronchetti and Trojani (2001, p. 47).

Specifically, for a given bound $c > \sqrt{H}$, the computation of the robust ET estimator can be performed by the following four steps:

i. Fix a starting value $\theta_0$ for $\theta$ and initial values $\tau_0 = 0$ and $A_0$ such that

$$A_0' A_0 = \left[ \frac{1}{N} \sum_{i=1}^{N} h(Z_i; \theta_0) h(Z_i; \theta_0)' \right]^{-1}$$

ii. Compute new values $\tau_1$ and $A_1$ for $\tau$ and $A$ defined by

$$\tau_1 = \frac{E_{\theta_0}[ h(Z ; \theta_0) w_c(A_0(h(Z ; \theta_0) - \tau_0))] }{ E_{\theta_0} w_c(A_0(h(Z ; \theta_0) - \tau_0))}$$

(3.20)

and

$$(A_1' A_1)^{-1} = \frac{1}{N} \sum_{i=1}^{N} [(h(Z_i ; \theta_0) - \tau_0))(h(Z_i ; \theta_0) - \tau_0)' \times w_c^2(A_0(h(Z_i ; \theta_0) - \tau_0))].$$

(3.21)

iii. Compute the optimal ET estimator $\theta_1$ associated to the orthogonality function $h_c^{\theta_1; \tau_1}$ by solving (3.13) subject to (3.13.a).

iv. Replace $\tau_0$ and $A_0$ by $\tau_1$ and $A_1$, respectively, and iterate the second and the third step described above until convergence.
3.6.2 Computational aspects

We used the \texttt{fmincon()} procedure for optimization in MATLAB 6.5. This algorithm is based on a Sequential Quadratic Programming (SQP) method, in which a Quadratic Programming (QP) subproblem is solved at each iteration. An estimate of the Hessian of the Lagrangian is updated at each iteration using the BFGS formula.

A particular point in this algorithm is the calculation of the vector $\tau$ defined by (3.20). The expectation in (3.20) is easily computed by simulating a sample of size 75000.

3.6.3 Code

\texttt{simulContLSV()}

\begin{verbatim}
1: function [R]=simulContLSV(n,epsilon,THETA0)
2: % Generate (nx1) temporal series R
3: % from simple Stochastic Lognormal Volatility (SLV) model
4: % \( R_t = \sigma_t Z_t \)
5: % \( \log(\sigma_t^2) = w + \beta \log(\sigma_{t-1}) + \sigma_u u_t \)
6: % THETA0: 3 - true parameter vector
7: % R: n - vector of the observations
8: % Written by: Serigne N. Lo, Department of Econometrics, University of Geneva CH-1211 Geneva, Switzerland
9: % n; % sample size
10: epsilon; %pourcentage of contamination
11: w=THETA0(1);
12: beta=THETA0(2);
13: sigma_u=THETA0(3);
14: n1=n+1;
15: sigma=zeros(n1+1,1);
16: R=rand(n,1);
17: Z=randn(n,1);
\end{verbatim}

94
U=randn(n,1);
ZZ=3*randn(n,1);  %normal with variance=9
UU=3*randn(n,1);  %normal with variance=9

% Initialization of sigma
sigma(1,1)=1;
% generating SLV vector R
for i=2:n1
  if (R(i-1)>epsilon)
    sigma(i,1)=exp(.5*(w+2*beta*log((sigma(i-1,1)))+sigma_u*U(i-1,1)));
    R(i-1,1)=sigma(i,1)*Z(i-1,1);
  else
    sigma(i,1)=exp(.5*(w+2*beta*log((sigma(i-1,1)))+sigma_u*UU(i-1,1)))...
    ;
    R(i-1,1)=sigma(i,1)*ZZ(i-1,1);
  end
end

function [A0] = ComputeLSVA0(R,a0)
% Function to estimate the matrix A0(5x5) (See Ronchetti and Trojani(2001) ... 
P.47)
% STOCHASTIC LOGNORMAL VOLATILITY (SLV) model
% R_t=sigma_t*Z_t
% log(sigma_t^2)=w + beta*log(sigma_(t-1)) + sigma_u*u_t
% a0: model parameters (3x1)
% R : n - vector of the observations
% Written by: Serigne N. Lo, Department of Econometrics, University of Geneva CH-1211 Geneva, Switzerland
r=5;  % maximum number of lag
n = length(R);  % sample size

% initialization of the parameter
w=a0(1);
beta=a0(2);
sigma_u=a0(3);
mu=w/(1-beta);
sigma2=sigma_u^2/(1-beta^2);
% Definition of the moments vectors
M1=abs(R(1:n-r)) - sqrt(2/pi)*exp(mu/2 + sigma2/8);
M2=R(1:n-r).^2 - exp(mu + sigma2/2);
M3=abs(R(1:n-r).*R(2:n-r+1))- (2/pi)*exp(mu + sigma2/4)*exp(beta*sigma2/4);
M4=abs(R(1:n-r).*R(4:n-r+3))- (2/pi)*exp(mu + sigma2/4)*exp((beta^3)*sigma2/4);
M5=abs(R(1:n-r).*R(6:n-r+5))- (2/pi)*exp(mu + sigma2/4)*exp((beta^5)*sigma2/4);

%%%M1 M2 M5 M7 M9
Phi=[M1 M2 M3 M4 M5];

W=Phi'*Phi/(n-r);
WW = inv(W);
A0 = chol(WW);

ComputeLSVA()
1: function [A] = ComputeLSVA(R,A,Tau,c,a0)
2:
3: % Function to estimate the matrix A_i(5x1) (See Ronchetti and ...
4: % Trojani(2001) P.47)
5: % R_t=sigma_t*Z_t
6: % log(sigma_t^2)=w + beta*log(sigma_(t-1)) + sigma_u*u_t
7: % a0=initial parameter vector (3x1)
8: %
9: % R: n - vector of the observations
10: % A: matrix (5x5) = A_(i-1)
11: % Tau0: initial vector (5x1) =Tau_(i-1) at first step
12: % c: Huber constant
13: % a0: model parameters (3x1)
14: %
15: % Written by: Serigne N. Lo, Department of Econometrics,
16: % University of Geneva CH-1211 Geneva, Switzerland
17:
18: n = length(R); % sample size
19: r=5; % maximum number of lag
20: % initialization of the parameter
21: w=a0(1);
22: beta=a0(2);
23: sigma_u=a0(3);
24:
25: mu=w/(1-beta);
26: sigma2=sigma_u^2/(1-beta^2);
27:
28: % Calculus of h(X,a0)
29: M1=abs(R(1:n-r)) - sqrt(2/pi)*exp(mu/2 + sigma2/8);
30: M2=R(1:n-r).^2 - exp(mu + sigma2/2);
31: M3=abs(R(1:n-r).*R(2:n-r+1))- (2/pi)*exp(mu + sigma2/4)*exp(beta*sigma2/4);
32: M4=abs(R(1:n-r).*R(4:n-r+3))- (2/pi)*exp(mu + sigma2/4)*exp((beta^3)... 
33: *sigma2/4); 
34: M5=abs(R(1:n-r).*R(6:n-r+5))- (2/pi)*exp(mu + sigma2/4)*exp((beta^5)... 
35: *sigma2/4); 
36: 
37: % Calculus of h(X,a0)-Tau 
38: M1_T=M1-Tau(1); 
39: M2_T=M2-Tau(2); 
40: M3_T=M3-Tau(3); 
41: M4_T=M4-Tau(4); 
42: M5_T=M5-Tau(5); 
43: Phi_Tau=[M1_T M2_T M3_T M4_T M5_T]; 
44: 
45: AH = A*Phi_Tau'; % matrix (5xN) 
46: NH = sqrt(sum(AH.*AH)); % AH.*AH calculate the square of each component ... 
47: % of AH 
48: % and the sum finish the calculus of the norm 
49: % NH N-vector 
50: I = find(NH == 0); 
51: if (~isempty(I)) 
52: NH(I) = 1; 
53: end 
54: U = c*ones(1,n-r); 
55: WC = U ./ NH; % WC N-vector composed by c/norm(NH) 
56: I = find(WC >1); 
57: if (~isempty(I)) 
58: WC(I) = 1; % All values WC > 1 are replaced by 1 
59: end 
60: 
61: W2 = WC.^2; 
62: AA=(Phi_Tau'.*(repmat(W2,5,1))*Phi_Tau)/(n-r); 
63: 
64: A = chol(inv(AA)); 

ComputeLSVTau() 
1: function [Tau] = ComputeLSVTau(Rr,A,Tau0,c,a0) 
2: 
3: % Function to estimate the vector Tau_i(5x1) (See Ronchetti and ... 
4: % Trojani(2001) P.47) 
5: % R_t=sigma_t*Z_t 
6: % log(sigma_t^2)=w + beta*log(sigma_(t-1)) + sigma_u*u_t 
7: % a0=initial parameter vector (3x1) 
8: % 
9: % Rr: 75000 - vector of simulated observations from SLV model 
10: % A: matrix (5x5) 
11: % Tau0: initial vector (5x1) =Tau_(i-1) at first step 
12: % c: Huber constant
% a0: model parameters (3x1)

% Written by: Serigne N. Lo, Department of Econometrics, University of Geneva CH-1211 Geneva, Switzerland

%load simulRr % Rr a simulated vector from a SLV model
r=5; % maximum number of lag
N = length(Rr); % sample size

% initialization of the parameter
w=a0(1);
beta=a0(2);
sigma_u=a0(3);

mu=w/(1-beta);
sigma2=sigma_u^2/(1-beta^2);

% Calculus of h(X,a0)
M1=abs(Rr(1:N-r)) - sqrt(2/pi)*exp(mu/2 + sigma2/8);
M2=Rr(1:N-r).^2 - exp(mu + sigma2/2);
M3=abs(Rr(1:N-r).*Rr(2:N-r+1))- (2/pi)*exp(mu + sigma2/4)...*
  *exp(beta*sigma2/4);
M4=abs(Rr(1:N-r).*Rr(4:N-r+3))- (2/pi)*exp(mu + sigma2/4)*exp((beta^3)...*
  *sigma2/4);
M5=abs(Rr(1:N-r).*Rr(6:N-r+5))- (2/pi)*exp(mu + sigma2/4)*exp((beta^5)...*
  *sigma2/4);

Phi=[M1 M2 M3 M4 M5];

% Calculus of h(X,a0)-Tau0
M1_T=M1-Tau0(1);
M2_T=M2-Tau0(2);
M3_T=M3-Tau0(3);
M4_T=M4-Tau0(4);
M5_T=M5-Tau0(5);

Phi_Tau=[M1_T M2_T M3_T M4_T M5_T];

AH = A*Phi_Tau'; % matrix (5xN)
NH = sqrt(sum(AH.*AH)); % AH.*AH calculate the square of each component of ...

% AH

if (~isempty(I))
59:   NH(I) = 1;
60: end
61: U = c*ones(1,N-r);
62: WC = U./ NH;  % N-vector composed by c/norm(NH)
63: I = find(WC >1);
64: if (~isempty(I))
65:   WC(I) = 1;  % All values WC > 1 are replaced by 1
66: end
67: Tau=sum(Phi'.*repmat(WC,5,1),2)/sum(WC);

myfunLSV()
1: function f=myfunLSV(V,R,A,Tau,c)
2: % Calculus of the cumulant generate function K(t,theta)
3: % R_t = sigma_t*Z_t
4: % log(sigma_t^2) = w + beta*log(sigma_(t-1)) + sigma_u*u_t
5: % a0 = initial parameter vector (3x1)
6: %
7:
8: % V: 8 - vector of estimating parameters (T+a0)
9: % with T=tilting parameters and a0= model parameters
10: % R: n - vector of the observations
11: % A: matrix (5x5) = A_(i-1)
12: % Tau0: initial vector (5x1) =Tau_(i-1) at first step
13: % c: Huber constant
14: %
15: % Written by: Serigne N. Lo, Department of Econometrics,
16: % University of Geneva CH-1211 Geneva, Switzerland
17: % initialization of the parameter (theta) and the tilting parameters (T)
18: m=8;
19: n=length(R);
20: T=V(1:m-3);  % tilting parameters
21: w=V(m-2);
22: beta=V(m-1);
23: sigma_u=V(m);
24:
25: n = length(R);  % sample size
26: r=5;  % maximum number of lag
27: mu=w/(1-beta);
28: sigma2=sigma_u^2/(1-beta^2);
29: % Calculus of h(X,a0)-Tau
30: M1=abs(R(1:n-r)) - sqrt(2/pi)*exp(mu/2 + sigma2/8);
31: M2=R(1:n-r).^2 - exp(mu + sigma2/2);
32: M3=abs(R(1:n-r).*R(2:n-r+1)) - (2/pi)*exp(mu + sigma2/4)*exp(beta*sigma2/4);

99
37: \[ M_4 = \text{abs}(R(1:n-r) \times R(4:n-r+3)) - \frac{2}{\pi} \exp(\mu + \sigma^2/4) \times \exp((\beta^3)\frac{\sigma^2}{4}); \]
38: \*\*\*\sigma^2/4; \]
39: \[ M_5 = \text{abs}(R(1:n-r) \times R(6:n-r+5)) - \frac{2}{\pi} \exp(\mu + \sigma^2/4) \times \exp((\beta^5)\frac{\sigma^2}{4}); \]
40: \*\*\*\sigma^2/4; \]
41: % Calculus of h(X,a0)-Tau
42: M1_T=M1-Tau(1);
43: M2_T=M2-Tau(2);
44: M3_T=M3-Tau(3);
45: M4_T=M4-Tau(4);
46: M5_T=M5-Tau(5);
47: Phi_Tau=[M1_T M2_T M3_T M4_T M5_T];
48: Y=A*Phi_Tau'; % matrix (5xN)
50: %
51: Ynorm=sqrt(sum(Y.*Y)); % Y.*Y calculate the square of each component of ...
52: % Y
53: % and the sum finish the calculus of the norm
54: % Ynorm N-vector
56: I=find(Ynorm > c );
57: h_y=zeros(5,n-r);
58: U=ones(1,n-r);
59: h=Y.*repmat((c*U./Ynorm),5,1); % WC N-vector composed by c/norm(NH)
60: if (~isempty(I))
61: h_y(:,I)=h(:,I);
62: end
63: I=find(Ynorm<=c);
64: if (~isempty(I))
65: h_y(:,I)=Y(:,I);
66: end
67: f=-log(sum(exp(T'*h_y))/(n-r)); % objective function value

mycontrLSV()
1: function [in_K_t,K_t]=mycontrLSV(V,R,A,Tau,c)
2: % in_K_t: Compute nonlinear inequalities
3: % K_t: Compute nonlinear equalities.
5: % R_t = \sigma_t^2 * Z_t
6: % log(\sigma_t^2) = w + beta*log(\sigma_(t-1)) + \sigma_u*u_t
7: % a0 = initial parameter vector (3x1)
8: %
9: % V: 8 - vector of estimating parameters (T(5x1)+a0(3x1))
10: % with T=tilting parameters and a0= model parameters
11: % R: n - vector of the observations
12: % A: matrix (5x5) = A_(i-1)
13: % Tau0: initial vector (5x1) = Tau_(i-1) at first step
14: % c: Huber constant

100
% initialization of the parameter (theta) and the tilting parameters (T)
m=8;
n=length(R);
T=V(1:m-3); % lagrangian T
w=V(m-2);
beta=V(m-1);
sigma_u=V(m);

n = length(R); % sample size
r=5; % maximum number of lag

mu=w/(1-beta);
sigma2=sigma_u^2/(1-beta^2);

% Calculus of h(X,a0)-Tau
M1=abs(R(1:n-r)) - sqrt(2/pi)*exp(mu/2 + sigma2/8);
M2=R(1:n-r).^2 - exp(mu + sigma2/2);
M3=abs(R(1:n-r).*R(2:n-r+1)) - (2/pi)*exp(mu + sigma2/4)*exp(beta*sigma2/4);
M4=abs(R(1:n-r).*R(4:n-r+3)) - (2/pi)*exp(mu + sigma2/4)*exp((beta^3)*sigma2/4);
M5=abs(R(1:n-r).*R(6:n-r+5)) - (2/pi)*exp(mu + sigma2/4)*exp((beta^5)*sigma2/4);

% Calculus of h(X,a0)-Tau
M1_T=M1-Tau(1);
M2_T=M2-Tau(2);
M3_T=M3-Tau(3);
M4_T=M4-Tau(4);
M5_T=M5-Tau(5);
Phi_Tau=[M1_T M2_T M3_T M4_T M5_T];

Y=A*Phi_Tau'; % matrix (5xN)

Ynorm=sqrt(sum(Y.*Y)); % Y.*Y calculate the square of each component of ...
% and the sum finish the calculus of the norm
I=find(Ynorm > c );
hc_y=zeros(5,n-r);
U=ones(1,n-r);
h=Y.*repmat((c*U./Ynorm),5,1); % WC N-vector composed by c/norm(NH)

if (~isempty(I))
61:     hc_y(:,I)=h(:,I);
62: end
63: I=find(Ynorm<=c);
64: if (~isempty(I))
65:     hc_y(:,I)=Y(:,I);
66: end
67:
68: in_K_t = [(beta-0.9999);(-sigma_u)]; % the inequalities constraint imposed
69: % to parameters beta and sigma_u
70: K_t=sum((hc_y.*repmat(exp(T'*hc_y),5,1)),2);

OptLSV()
1: function [V,fval,K_t]=OptLSV(T,THETA,R,A,Tau,c)
2:
3: % Minimization of -K() subject to
4: %   K_t() = 0
5: %       beta < 1
6: %       sigma_u > 0
7: %
8: %   T: Tilting parameters (5x1)
9: %   THETA: model parameters (3x1)
10: %   R: n - vector of the observations
11: %   A: matrix (5x5) = A_(i-1)
12: %   Tau: robustness vector (5x1)
13: %   c: Huber constant
14: %
15: % Written by: Serigne N. Lo, Department of Econometrics,
16: % University of Geneva CH-1211 Geneva, Switzerland
17: V0 = [T;THETA];
18: options = optimset('Display','iter');
19: [V,fval]=fmincon('myfunLSV',V0,[],[],[],[],[],[],...
20: 'mycontrLSV',options,R,A,Tau,c);
21: [in_K_t,K_t]=mycontrLSV(V,R,A,Tau,c);

IterateLSV()
1: function [Tau,A,V,THETA_k,fval]=IterateLSV(R,Rr,T,THETA,c,a0)
2:
3: % Minimization of -K() subject to non-linear constraint
4: % and estimation of A and Tau by iterations
5: %   K_t() = 0
6: %       beta < 1
7: %       sigma_u > 0
8: %
9: %   T: Tilting parameters (5x1)
10: %   THETA: model parameters (3x1)
% R: n - vector of the observations
% A: matrix (5x5) = A_{i-1}
% Tau: robusteness vector (5x1)
% c: Huber constant

% Written by: Serigne N. Lo, Department of Econometrics,
% University of Geneva CH-1211 Geneva, Switzerland

[A0] = ComputeLSVA0(R,a0);
% Iteration
[A_k] = A0; % initialization of A matrix
Tau_k = [0;0;0;0;0]; % initialization of Tau vector
THETA_k=THETA;
Delta1 = 1; % initialization for convergence condition for w
Delta2=1; % initialization for convergence condition for beta
Delta3=1; % initialization for convergence condition for sigma_u
MAX=50; % maximum number of iterations
I = 0; % Iteration number

while( Delta1>1e-3 & Delta2>1e-3 & Delta3>1e-3 & I<MAX); % convergence ...
% conditions
I = I + 1;
[Tau] = ComputeLSVTau(Rr,A_k,Tau_k,c,a0);
[A] = ComputeLSVA(R,A_k,Tau_k,c,THETA_k);
[V,fval,K_t]=OptLSV(T,THETA_k,R,A,Tau,c);

Delta1=abs(V(6)-THETA_k(1));
Delta2=abs(V(7)-THETA_k(2));
Delta3=abs(V(8)-THETA_k(3));
fprintf('--------------------------------------------------------
');
fprintf('Iteration N° %d Delta1=%f Delta2=%f ...
',I,Delta1,Delta2);
fprintf('Delta3=%f\n',I,Delta1,Delta2,Delta3);
fprintf('--------------------------------------------------------
');

THETA_k=V(6:8);
Tau_k = Tau;
A_k = A;
end

fprintf('--------------------------------------------------------
');
fprintf('Iteration N° %d Delta1=%f Delta2=%f ...
',I,Delta1,Delta2);
fprintf('Delta3=%f\n',I,Delta1,Delta2,Delta3);
fprintf('--------------------------------------------------------
');
Bibliography


