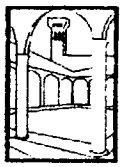


EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Robust Estimation of Panel Data; with an Application to Investment Equations

J.L.M. Wagenvoort

Thesis submitted for assessment with a view to obtaining the Degree of Doctor of the European University Institute

Florence, April 1998

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Robust Estimation of Panel Data; with an Application to Investment Equations

J.L.M. Wagenvoort

The Thesis Committee consists of:

Prof. Giampiero Gallo, University of Florence
" Luigi Guiso, Ente Einaudi, Rome
" Teun Kloek, Erasmus University Rotterdam
" Costas Meghir, University College London
" Robert Waldmann, EUI and IGIER, Milan, Supervisor

Acknowledgements

The present study could not have been accomplished without the help of a number of people and institutions.

I especially wish to thank my supervisor, Professor Robert Waldmann, who provided continuous intellectual support, excellent suggestions and "great" comments. Being his research assistant in the last stage of my dissertation was a very enjoyable and valuable experience. This gave me the excellent opportunity to work with him on a number of issues, especially in the field of the research themes of Chapter 3.

The first chapter of this thesis is inspired by joint work with my friend and colleague, Jeroen Hinloopen, which began in 1992 at the Department of Econometrics at the Erasmus University Rotterdam (EUR). Besides several scientific publications, our joint efforts also resulted in the foundation of the consultancy firm *Estimatica* last Spring. In the near future *precise estimation* will remain high on the agenda! Undertaking projects with Jeroen is allways interesting, challenging and encouraging. Evidently, he holds an honoured place in this vote of thanks.

I also thank Paul Schure with whom I had stimulating discussions about the recent ideas concerning financial intermediation presented in Chapter 4.

This work benefited from useful comments made by participants in seminars at the European University Institute and the European Investment Bank (EIB) in Luxembourg. My stay at the EIB provided the perfect stimulus to complete this work.

I owe special thanks to Melanie Roberts who carefully accomplished the laborious job of correcting the final version for grammar and spelling mistakes. Of course, the usual disclaimer applies.

Finally, the financial support of *Stichting Noorthey* is gratefully acknowledged.

Rien Wagenvoort
Firenze, August 1998

Contents

General Introduction and Summary. 1

Part I Robust Estimation

Chapter 1 An Example of a High Breakdown Point Generalized M-Estimator. 9

1.1 Efficiency Versus Robustness
1.2 Construction of a High Breakdown Point Generalized M-Estimator
    1.2.1 Mallow's First Order Condition
    1.2.2 Weights Based on Leverage Points
    1.2.3 Weights Based on Vertical Outliers
    1.2.4 Data and Model Suitability
1.3 The MVE Estimator of Location and Scatter
    1.3.1 Projection
    1.3.2 Resampling
    1.3.3 Testing the MVE-Approximations
1.4 Statistical Properties of the HBP GM Estimator
    1.4.1 t-values, Adjusted R^2 and Scale Estimate
    1.4.2 Breakdown Point
    1.4.3 Consistency
    1.4.4 Efficiency
1.5 Stars, Light and Heat
1.6 Conclusions
Appendix A1 Least median of Squares Regression

Chapter 2 B-Robust Estimation and Testing of the SUR Model With Fixed Efects and Heteroskedasticity. 47

2.1 Introduction
2.2 Model Assumptions
2.3 Testing For Correlated Fixed Effects
2.4 Testing For Cross-sectional Independence
2.5 Heteroskedasticity
    2.5.1 The Pagan and Vella Diagnostic Test on Heteroskedasticity
    2.5.2 Correction For Multiplicative Heteroskedasticity
    2.5.3 Simulation Results
2.6 Autocorrelation
    2.6.1 The Generalized Durbin-Watson Test
    2.6.2 Robust Seemingly Unrelated Regression
    2.6.3 A Monte Carlo Study
2.7 Conclusions

Chapter 3 Two Stage GM and Robust Generalized Method of Moments 71

3.1 Introduction
3.2 Two Stage Generalized M-Estimation
3.3 Robust Generalized Method of Moments Estimation
3.4 An Examination of the Performance of the Robust GMM Estimator in Simulated Panel Data Sets With Measurement Errors in the Explanatory Variables
3.5 Summary and Outlook
Appendix A3 The Choice of Instruments

Part II Investment and Financial Policy

Chapter 4 A Literature Review of the Relevance of Financial Policy 97

4.1 Modigliani and Miller's Proposition I
4.2 Incomplete Capital Markets
4.3 Credit Constraints
4.3.1 A Theoretical Explanation For Credit Rationing Equilibria
4.3.2 Implications of Credit Rationing
4.3.3 Empirical Evidence Regarding Credit Rationing
4.4 Tax Shields
4.5 Concluding Remarks

Chapter 5 Empirical Evidence Regarding the Sensitivity of Dutch Corporate Investment to the Financial Environment 109

5.1 Introduction
5.2 Neoclassical Investment Theory
5.2.1 Derivation of the Euler Equation
5.2.2 Derivation of the Q-model of Investment
5.3 The Data Sources
5.4 Financial Regimes
5.5 The Estimation Procedure
5.6 Empiricism
5.6.1 Estimation Results, Euler Equation
5.6.2 Estimation Results, Q-model
5.7 Conclusion
Appendix A5.1 Data and Variable Construction
Appendix A5.2 Tables Containing the Estimation Results

References 143

GENERAL INTRODUCTION AND SUMMARY

Testing economic theory on the basis of available data sets through econometric modelling is only legitimate if each condition of the entire collection of underlying assumptions is satisfied. This, at first glance obvious, statement has more implications than many of the practitioners of econometrics in the field are aware of. Even on a theoretical level, econometricians seem to show little agreement about the recommended modelling strategies or estimation procedures. This is imputed to the highly non-experimental nature of economic data and the fact that in many cases the theoretical variables do not coincide one-to-one with the available observed data. Disciples of both Fisher's experimental design paradigm or Gauss's theory of errors have to assume that the mechanism which generates the data is a nearly isolated system. See Spanos (1993) for an exposition of these different methodologies. This means that the systematic explanatory part of the model is somehow fixed or predetermined and the error terms can be viewed as non-systematic effects of omitted influencing factors or mismeasurement. In sharp conflict with this assumption, economic data is held to be generated by a fickle process where explanatory variables are often endogenous and influential factors show up and disappear. Usually, a specific economic theory only considers a sub set of the systematic explanatory components and makes predictions under the assumption that the influential factors which have been put aside are constant. These *ceteris paribus* clauses often do not hold when analyzing real data sets.

The *textbook* approach (a hybrid of Fisher's paradigm and the law of errors, see Spanos (1986), Fig. 1.1) towards econometric modelling consists of the following sequence of steps: First, the theoretical model is translated into an econometric model by appending an error term which has to obey certain conditions. Second, the unknown parameters are estimated with a classical estimation procedure such as OLS. Third, specification tests are performed. Most of the time, the applied econometrician only tests for homoskedasticity, independence and normality of the error terms. Fourth, if the specification tests give rise to rejection of the initial model assumptions then either another estimator is applied (for instance GLS or 2SLS), which is thought to be appropriate under new model assumptions or the theoretical model is redefined (by adding or removing explanatory variables). Followers of the textbook approach implicitly assume that the theoretical model coincides with the statistical model which describes the probabilistic structure of the data. Friction between the statistical model and the underlying generating process of the data, however, may cause serious misleading conclusions about empirical support for theoretical relationships. The following two examples, chosen among many others, put a question mark on the standard textbook approach: Mizon (1995) finds that autocorrelation correctors, as a response to rejection of the independence assumption by traditional autocorrelation tests, can produce inconsistent parameter estimates if the statistical model is misspecified. Furthermore, in applied econometric studies it is common to include the square of the explanatory variable in the design in order to model non-

linearities. Hinloopen and Wagenvoort (1995) show that this can amplify the exorbitant influence of outlying observations on the unreliable parameter estimates and t-values if the data is polluted.[1] Indeed, sometimes, juggling with models has earned the applied econometrician a bad reputation.

It is to the credit of the London School of Economics (LSE) tradition in econometric modelling that it calls attention to the proper specification of the statistical model. In this light, Spanos (1993) introduces his *Probabilistic Reduction* approach where the theoretical model is explicitly distinguished from the statistical one.[2] The first task of the empirical researcher is now to develop a well specified statistical framework which corresponds to the generating process of the observed data (specification). Data plots and simple statistics (covariances, mean, median etc.) determine the structure of the statistical model. Note that this data snooping is strictly forbidden in the textbook approach since it will invalidate statistical inference. In the second round of the probabilistic reduction approach the parameters of the statistical model are estimated and the underlying assumptions are tested (misspecification testing). The first and second steps are repeated until a well specified statistical model is found (respecification). Finally, the bridge is built between the statistical and theoretical model by reparameterizing or restricting the first (identification). Economic theory (represented by the restricted model) can be tested against the specific alternative, namely the full well specified statistical model.

Although the latter approach to econometric modelling is less dependent on *a priori* specification assumptions than the textbook methodology, it may suffer from many practical problems. Since we can only speculate when considering the unknown generating mechanism of economic data, it remains unclear how the appropriate statistical model has to be found. If the dimension of the joint distribution of the dependent and explanatory variables exceeds three then one can not rely on data plots. More importantly, what should one do if the statistical model is miles away from the theoretical structure (i.e. it can not be identified) or if two different statistical models both rendering the probabilistic information of the data lead to different economic hypotheses? Another difficulty, which also applies to the *textbook* approach and which will bring us to the next modelling strategy, is the occurence of observations which do not follow the pattern of the bulk of the data. It may happen that, although the variables of interest have population moments corresponding to a joint normal distribution, having relatively few outlying observations can cause the rejection of the normality assumption. Even if in this case the modeller chooses to abandon the normality assumption by choosing, for example, a t-distribution as the proper probabilistic structure, it cannot be guaranteed that the outliers will not spoil the statistical inference.

In reply to the susceptibility of the classical estimators towards outliers, new so-called robust estimators were developed in the research area of *robust statistics*. The aim of robust

[1] Outliers are defined as observations which do not follow the pattern of the majority of the data.

[2] The probabilistic reduction approach belongs to the class of *general to specific* modelling strategies (see e.g. Hendry (1987)).

statistics is to trace outliers and to diminish their influence on statistical inference. This is achieved by relaxing the usual idealized assumptions in econometrics. The crux of the robust theory is the approximate character of the strict parametric model. An econometrician of the *probabilistic reduction* approach searches for the specified distribution function $G(x,\theta)$ which describes all the data x at hand through the unknown parameters θ, whereas in robust statistics we assume that a fraction $1-\varepsilon$ of the data obeys G but the rest of the observations emerge from an unknown distribution $H(x,\theta)$. Thus, the distribution of the complete set of data points is now represented by $F(x,\theta) = (1-\varepsilon)G(x,\theta) + \varepsilon H(x,\theta)$. Huber (1964) was the first statistician who considers such a "full neighborhood" of a parametric model in his minimax strategy and introduces M-estimators which are defined by a generalization of the criterion function of the maximum likelihood estimator. Hampel et al. (1986, p.21) discuss a number of reasons for deviation from the strict parametric model. The occurence of gross errors (due to typing, copying or computation errors etc.) and/or rounding and grouping of the data can invalidate the distributional assumptions of the model. The theoretical model may be only appropriate for a sub set of the data or only coincide with the probabilistic structure of the data in the limit, by virtue of the central limit theorem. Furthermore, the independence and homoskedasticity conditions may also only be approximately fulfilled.

To provide a global measure of the sensitivity of an estimator to outlying observations Donoho and Huber (1983) introduce the idea of a breakdown point. The maximum fraction of data contamination which leaves the estimator determinate defines its breakdown point. Classical estimators such as OLS, GLS, 2SLS, GMM etc. have a breakdown point of zero, i.e. a single observation can cause the estimator to produce any estimates and any standard errors. Another important concept related to the robustness of an estimator is the influence function which measures the effect of an additional observation in any local point x on the estimator, given a (large) sample with distribution F (see Hampel et al. (1986), p.41). Estimators with bounded influence functions are called B-robust.

Data snooping is permitted in robust statistics. In fact, semi-parametric or non-parametric tools can be used to discover outlying observations before the parametric estimation procedure is started. For example, distances based on the Minimum Volume Ellipsoid (MVE) estimates of location and scatter of the design are reliable indicators of leverage points (see Rousseeuw (1985)).[3] These robust distances measure the extent to which an observation on the regressors deviates from the center of the independent variables. However, one should not use the regression residuals or design distances which are computed with classical estimators when unmasking outliers. These residuals and distances are likely to be based on wrong estimates, therefore outliers can be left unnoticed.

In the thesis in hand we adopt the robust estimation methodology by considering neighbourhoods of parametric models. In the first part of this monologue we develop new robust estimators for panel data models. The second part contains an application; the new estimators are

[3] Data points for which the independent variable lies far from the bulk of the explanatory observations are called leverage points (see Rousseeuw and van Zomeren (1990) and Chapter 1).

employed when estimating a neoclassical investment model in order to test for imperfections in the financial markets.

We start in Chapter 1 with the presentation of a High Breakdown Point Generalized M-estimator (HBP GM) as a B-robust alternative to the well-known ordinary least squares method for the standard linear model. Joint research with Jeroen Hinloopen, encapsulated in two papers (see Hinloopen and Wagenvoort (1995, 1997)), provides a good background for this part of the thesis.[4] The HBP GM estimator is constructed according to the rules outlined in Simpson, Ruppert and Carroll (1992) and combines the quality of being able to handle up to 50% data pollution without taking arbitrary values. Further, it exhibits the desirable asymptotic properties of consistency and normality under weak conditions. The Gauss-Markov theorem ensures that OLS provides the Best Linear Unbiased Estimator. The price we have to pay for the robustness of our estimator is an alleged loss in efficiency since the HBP GM method downweighs (good) leverage points. However, by means of a simulation study we demonstrate that the HBP GM estimator is only slightly less efficient than OLS for clean data but, in most cases, is much more efficient than both the robust Least Median of Squares (LMS) estimator (see Rouseeuw (1984)) and OLS if the data is contaminated. An example from the astronomy is included with a view to highlighting the successful performance of the HBP GM estimator with respect to LMS and OLS when anomalous data is present and to introduce graphical tools for the discovery of outliers. Simulated t-statistics, under clean and polluted generated data sets, are reported in order to demonstrate the reliability of the robust statistical inference.

The identification of leverage points, within our specific HBP GM method, is based on the MVE estimates of the centre and dispersion of the regressors. Since no closed form solution can be obtained for these MVE estimators, we have to rely on approximations. For this purpose, the projection algorithm and the resampling algorithm have been invented to approximate robust distances (see Rouseeuw and Van Zomeren (1990, 1991)). Our simulation results, however, indicate that both algorithms are inadequate and can be improved by a specific correction factor.

The aim of the second chapter is to show how the HBP GM technique can be extended in order to produce more efficient estimates for models where the homogeneity and/or independence conditions are not fulfilled. We present the robust counterpart of Pagan and Vella's classical test on multiplicative heteroskedasticity associated with the HBP GM estimator of Chapter 1 (see Pagan and Vella (1989)) and robustify Harvey's (1976) two-step method. Multiplicative heteroskedasticity and the occurence of outlying observations are, for some real data sets, two sides of the same coin. In this case, application of the classical Harvey procedure is not an adequate solution to the outlier problem since this procedure suffers from the masking effect mentioned earlier in the introduction. Furthermore, a bounded influence estimator for the Seemingly Unrelated Regression (SUR) model is derived. This type of regression model is especially relevant for panel data which give cause to correlated errors across different time-periods but zero cross-sectional correlation. Monte Carlo

[4] In particular, Section 1.5 corresponds with Hinloopen and Wagenvoort (1995).

results affirm that substantial efficiency gains can be achieved for various experiments by consulting the robust proposals of Chapter 2.

Up to now we have ignored important aspects of non-experimental data. Economic variables often fluctuate simultaneously and the explanatory factors of the statistical model are usually measured with error. Therefore, in Chapter 3, we relax another assumption of the standard linear model. It so happens that the design matrix is not necessarily orthogonal to the model errors. In this case, instrumental variable techniques such as Two Stage Least Squares (2SLS) provide consistent estimates given that outliers do not corrupt the data. In imitation of Krasker (1986), we draw a parallel with robust regression and introduce the Two Stage Generalized M (2SGM) estimator.

Although the 2SGM estimator produces heteroskedastic consistent estimates and standard errors, it fails to fully exploit all the moment restrictions if the error terms exhibit a non-diagonal covariance matrix or heteroskedasticity. In response, we develop a more efficient estimator than 2SGM in the guise of Robust Generalized Method of Moments (RGMM) for models with correlated and/or heteroskedastic errors. The usual GMM unbounded influence estimator is defined by equating the empirical moment restrictions, corresponding with the theoretical moment conditions, to zero. For example, the sample mean is the method of moments estimator of the population mean of a random variable. This statistic, however, is vulnerable to outliers since it does not discriminate between different observations in the sample, i.e. each of them is given a weight equal to one over the number of observations. For that reason, the median of the observations can be a much better predictor of the first theoretical moment than the standard empirical moment (the sample mean). This gave us the idea of replacing the usual GMM defining sample moment condition by a robust empirical moment restriction. The latter is obtained by means of inserting weights for each observation in the moment condition depending on both the leverage of the instruments and the position of the observation with respect to the bulk of the data from the viewpoint of the relationship between the dependent variable and the predicted explanatory factors. These weights are computed with the help of the Minimum Volume Ellipsoid estimator and the 2SGM method. Conditions are provided for which the RGMM estimator is consistent, asymptotic normally distributed and possesses a bounded influence function.

An illustrative example is given by employing the RGMM estimator on a generated panel data model with erroneously measured explanatory variables which is extensively studied by Griliches and Hausman (1986). They derive the optimal set of instruments under different assumptions imposed on the stationarity and covariance structure of the measurement errors. From our Monte Carlo results we conclude that the RGMM estimator is consistent and almost as efficient as the standard GMM approach if the data are free from any outliers. The HBP GM estimates are downward biased in the cases both of first difference regression and when estimating in the within dimension of the data by calculating deviations around individual means, to eliminate the individual fixed effects. This result is expected since it has been shown that OLS under-estimates the parameters too when there are errors in the regressors. Furthermore, the Wu-Hausman (see Wu (1973) and Hausman (1978)) test statistic reveals that the first difference and within

estimates are significantly different for the first experiment. This is also explained by the occurence of measurement errors in the explanatory factors. The standard GMM estimator breaks down when the data are contaminated. However, the robust version of the generalized method of moments still produces reliable relatively efficient estimates when the orthogonality condition is rejected and the data are polluted. To our knowledge, Chapter 3 presents the first example of an estimator which combines these properties.

A brief literature survey of the conclusive arguments for an active corporate financial policy heralds the second part of the thesis. In Chapter 4 we discuss the consequences of incompleteness of the capital markets, credit constraints and taxes for Modigliani and Miller's famous Proposition I of 1958 on the irrelevance of the capital structure. The conlusion to be drawn from this review is that, in general, financial policy is relevant.

Keeping this result at the back of our mind, we explicitly take into account the financing side of investment projects when modelling the optimal capital choice of a production firm in Chapter 5. Following Bond and Meghir (1992), a neoclassical investment model is introduced with convex capital adjustment costs where the managers of the firm simultaneously decide on the level of investment, the financial structure and the input factors of the production process. Both the Euler equation and Tobin's Q-specification (see Tobin (1969) and Hayashi (1982)) of the model are derived and estimated, using a panel data set of Dutch manufacturing companies which are quoted on the Amsterdam stock exchange. Due to several very large outliers among the observations, in most of the cases, the classical estimation methods fail to reveal the direction and signifiance of the relationship, implied by the majority of the data, between the investment capital ratio and its explanatory factors. In constrast with our estimation results associated with classic GMM and most of the empirical studies on the Q-theory presented so far, we find that the empirical performance of the Q-model is satisfactory. The empirical Euler equation is less appropriate for describing the investment behaviour of the Dutch manufacturing companies. The new robust GMM estimator is successful in securing consistent relatively efficient estimates of the parameters of our panel data model with endogenous regressors, autocorrelated errors, fixed effects and heterogenous observations including severe outliers.

PART I
ROBUST ESTIMATION

CHAPTER 1

An Example of a High Breakdown Point Generalized M-Estimator

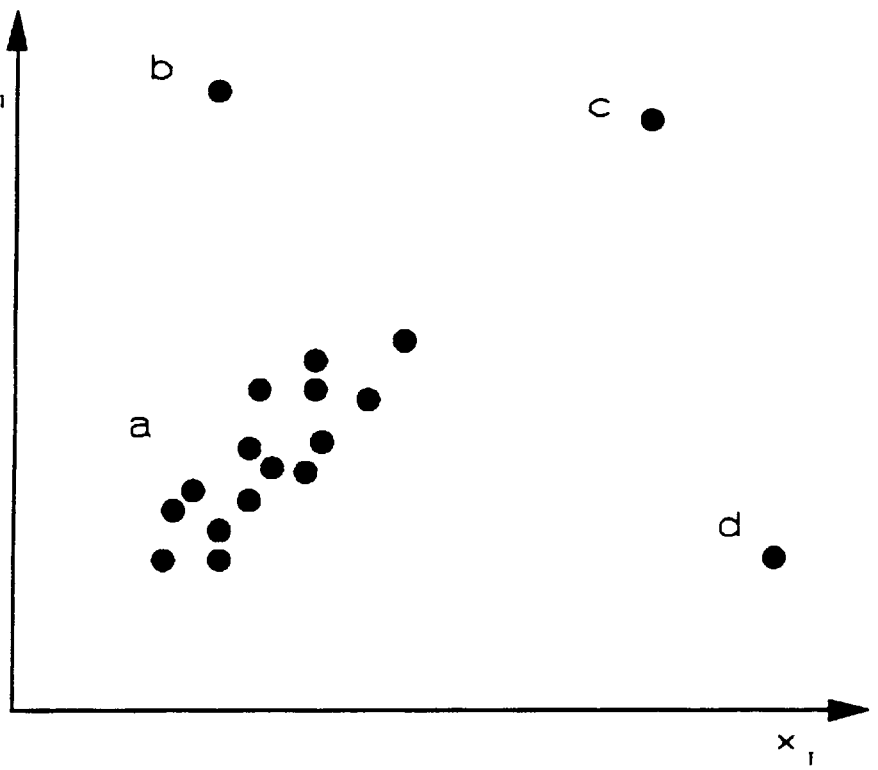
Section 1.1 Efficiency Versus Robustness

At the beginning of last century, the widely known Ordinary Least Squares (OLS) method was introduced by the two celebrated statisticians Legendre and Gauss. To this day, empirical researchers use OLS and its generalizations since the Gauss-Markov theorem asserts that OLS provides the Best Linear Unbiased Estimator (BLUE) for the parameters of the standard linear model. The estimator is best in the sense that it is the most efficient one among the linear estimators when the data are smooth, i.e. do not contain outlying observations. Following Rousseeuw and Van Zomeren (1990) we distinguish leverage points and/or vertical outliers (see Figure 1.1) when defining these anomalous observations. Data items for which the independent variable lies far from the majority of the explanatory observations are called leverage points. These observations deserve special attention from the econometrician since leverage points may invalidate classical statistical inference. On the other hand, they may substantially lower the standard errors if the variation in the design matrix reveals the statistical pattern contained in the data set. Observations which do not follow this relationship which is observed by the majority of the data, are defined as vertical outliers. Note that an observation can be both a vertical outlier and a leverage point (point d in Figure 1) or can be horizontally outlying without being a vertical outlier when it perfectly fits the relation between the independent variable and the explanatory factors (point c). Outliers are present in almost any real data set for a number of reasons: they may emerge from typos or measurement errors, the model error term may come from a fat and/or long tailed distribution, or the theoretical model may be only appropriate for a sub set of the data. If the data is contaminated with vertical outliers and leverage points then OLS can not only be a very inefficient regression technique but also produce dramatically different estimates when a single outlying observation is added or deleted. The survival of OLS for about two centuries in empirical studies is not justified by its performance on contaminated data.

Ever since the discovery of the least squares criterium statisticians were aware of these shortcomings as can be seen from the following two citations:[1]

[1] Cited by Rousseeuw and Leroy (1987).

gure 1.1 Simple Regression Example with (a) Regular Observations, (b) Vertical Outlier, (c) Leverage Point and (d) Vertical Outlier and Leverage Point



Legendre in 1805:

If among these errors are some which appear too large to be admissible, then those observations which produced these errors will be rejected, as coming from too faulty experiments, and the unknowns will be determined by means of the other observations, which will then give much smaller errors.

Edgeworth in 1887:

The method of Least Squares is seen to be our best course when we have thrown overboard a certain portion of our data--a sort of sacrifice which has often to be made by those who sail upon the stormy seas of probability.

endre and Edgeworth point here at the two core issues of robust estimation: which observations e excessive influence on the estimation results and how can the distortions be circumvented. It ns rather drastic to completely ignore part of the data as is suggested by the two scientists above. his case, good leverage points, i.e. a horizontal outlier which is not vertically outlying, can no

longer contribute to the efficiency of the estimator. Evidently, one needs to have prior beliefs about the true parameters of the model to determine whether a leverage point is good or bad. There exists a trade-off between efficiency and outlier robustness. The aim of Part I of this thesis is to find the best terms of exchange for various settings of the linear model. We search for relatively efficient estimators, in comparison with the classical proposals such as OLS, GLS, 2SLS and GMM, subject to the restriction that any observation has limited influence on the parameter estimates, i.e. the influence function associated with the estimator has to be bounded.[2] In Chapter 1 we discuss a bounded influence or so-called B-robust estimator for the standard linear model which is almost as efficient as OLS when it is applied to clean data sets. Chapter 2 and 3 provide B-robust relatively efficient solutions to the estimation of the linear covariance model and linear simultaneous equations respectively. Beside the efficiency argument, other reasons can motivate the inclusion of the outliers. In some studies the researcher has special interest in analysing and predicting the outlying observations. Hinloopen (1997) discusses in his Ph.D. thesis an entertaining example about engineers who have to decide on the height of a dyke. Throwing the outliers overboard would in this case cause the water to flood over the dyke at the Dutch village *Borgharen dorp* during stormy weather.

An early attempt to generate robust estimates was made in 1887 by Edgeworth who proposes to replace the objective of OLS to minimize the sum of the squared errors by the criterium: minimize the sum of the absolute errors. The resulting least absolute values estimator or so-called L_1-estimator is less sensitive than OLS to certain types of outliers (see Judge (1988, p.899)), but it can still be tricked by one outlying observation. Using the terminology developed in the research area of robust statistics, we say that its breakdown point, defined as the smallest fraction of data contamination which causes the estimator to take on arbitrary values[3], is equal to zero. The notion of breakdown point provides a global measure of reliability whereas the influence function is developed to indicate the effect of a single (or cluster of) outlying observation(s). An example of a high breakdown point estimator is provided by Rousseeuw's (1984) Least Median of Squares (LMS) estimator which minimizes the median of the squared residuals instead of the sum. The LMS estimator has the appealing property that its breakdown point is 50%. By definition this is the highest percentage achievable, since

---

[2] The influence function is formally defined by the change in the estimates returned by the estimator β when the portion t of the initial data distribution F is replaced by the distribution G which puts mass 1 at the point x:

$$IF(x;\beta,F) = \lim_{t\downarrow 0} \frac{\beta((1-t)F+tG(x)) - \beta(F)}{t}$$

for those x where this limit exists (Hampel et al. (1986), p. 84). The gross-error sensitivity γ^* of the estimator to small data contamination is then obtained by evaluating the upper bound on the influence function:

$$\gamma^* = \sup_x |IF(x;\beta,F)|.$$

See Hampel et al. (1986) for a survey of robust statistics based on influence functions.

[3] In Section 1.4 we give a formal definition of this concept.

beyond this limit the distinction between *good* and *bad* data becomes arbitrary.[4] The price of the robustness of the LMS estimator in terms of efficiency loss is substantial. Therefore, generalizations of the Least Median of Squares estimator are presented by Rousseeuw and Yohai (1984) in the form of robust S estimators[5]. These S estimators exhibit the desirable properties of having both a breakdown point of 50% and good asymptotic performance in terms of efficiency.

Generalized M-estimators belong to the class of bounded influence estimators but, on their turn, exhibit the deficiency of having a low breakdown point in multiple regression models. Maronna, Bustos and Yohai (1979) prove that the breakdown point of the GM estimator is at most $1/(p+1)$, where p is the number of explanatory variables. M-estimators are defined by minimizing the sum of transformed residuals, where the symmetric transformation function ρ has a unique minimum at zero (i.e. the quadratic objective function of OLS is replaced by another functional form), see Huber (1973). The shape of the transformation function determines the relative influence of differently sized residuals on the parameter estimates. The robust M-estimator assigns less weight to vertical outliers than OLS. In addition to bounding the influence of observations which have relatively large residuals, GM estimators also bound the influence of leverage points irrespective of their residuals emanating from the regression. Krasker and Welsch (1982) derive the most efficient B-robust estimator given a bound on the gross-error sensitivity (see footnote 2).

The breakdown point of the General M estimator can be driven up by resorting to a two-step procedure suggested by Simpson, Ruppert and Carroll (1992), henceforth SRC. In this chapter we work out a specific example of a High Breakdown Point General M-estimator (HBP GM) according to the rules set down in SRC (1992). The idea of the HBP GM technique is to initiate the Newton-Raphson (NR) iterations, which are executed to solve the GM defining equation, with a preliminary HBP estimator. We have chosen Rousseeuw's LMS method to serve as preliminary estimator. SRC show that under mild conditions the breakdown point of the preliminary estimator(s) carries over to the final HBP GM estimator. They also demonstrate that the HBP GM estimator yields consistent asymptotically normal estimates. In addition, we show in this chapter, by means of a simulation experiment, that the HBP GM estimator we use is more efficient than it's preliminary LMS estimator when different kinds of data contamination are involved. Moreover, these simulations reveal that the efficiency loss incurred when using the HBP GM estimator instead of OLS in experiments with non-polluted data is small.

The HBP GM estimator we employ, involves computation of Minimum Volume Ellipsoid (MVE) location and scale estimates (see Rousseeuw and Leroy (1987)). Since there is no analytical expression for this estimator we must rely on numerical approximations. In the literature two algorithms are developed for this purpose (see Rousseeuw and Van Zomeren (1990)): the resampling

[4] Strictly speaking 50% is the highest breakdown point achievable for regression equivariant estimators.

[5] S estimators are defined by minimizing a robust measure of the scale of the residuals (see e.g. Hampel et al. (1986), p.114).

and the projection algorithm. We derive a correction factor which improves both algorithms for computing the Minimum Volume Ellipsoid estimator.

The set up of Chapter 1 is as follows. Section 1.2 contains an analytical derivation of the HBP GM estimator. Here we also indicate for which models and what kind of data this estimator is suitable. In the subsequent section we provide the definition and the computational procedure of the MVE estimator and test the above mentioned correction factor. The statistical properties (breakdown point, consistency and efficiency) of the HBP GM estimator together with its covariance matrix and t-values are discussed in Section 1.4. In this section we also show how to compute the adjusted coefficient of determination and how to obtain a robust estimate of the variance of the HBP GM regression residuals. Section 1.5 presents a real data example while conclusions are stated in Section 1.6. In Appendix A1 we go into the details of the computation of the LMS estimator.

Section 1.2 Construction of a High Breakdown Point Generalized M-Estimator

The aim of Chapter 1 is to provide a B-robust relatively efficient estimator of the unknown parameter vector β of the linear model

$$y_i = x_i\beta + \varepsilon_i \quad i=1,...,n, \tag{1.1}$$

where the errors $\varepsilon_1,...,\varepsilon_n$ are i.i.d. with distribution function F and finite variance σ^2. F is assumed to have a density symmetric around zero which is not necessarily equal to the normal. As usual, y_i is the regressand and x_i is a row vector of length p of observable explanatory variables. The regressors are assumed to be orthogonal to the model errors.

1.2.1 Mallow's First Order Condition

The M-estimator of β is defined by the objective function

$$\min_{\beta} \sum_{i=1}^{n} \rho(r_i), \tag{1.2}$$

where $r_i = y_i - x_i\beta$. Here we assume that the transformation function ρ is twice continously differentiable on the domain [-c,c], $\rho \in C^2[-c,c]$, $\rho(\cdot):\mathbb{R}\rightarrow\mathbb{R}_+$, $c\in\mathbb{R}$. The OLS and the maximum likelihood estimator (MLE) are members of the class of M-estimators as can be seen from choosing ρ equal to $\rho(r_i) = r_i^2$ and $\rho(r_i) = -\ln[f(r_i)]$ respectively, where ln denotes the natural logarithm and f the probability density function. In the case of Huber's (1973) proposal, the first order condition associated with criterium (1.2) can be written as

$$\sum_{i=1}^{n} x_i^T r_i w_r(r_i) = 0. \tag{1.3}$$

$w_r(\cdot): \mathbb{R} \rightarrow \mathbb{R}_+$ is a weight function which controls for vertical outliers. x_i^T is the transpose of x_i. Huber suggests downweighing observations whenever the absolute value of their residuals exceeds a certain positive constant c:

$$w_r(r_i) = \min\left(1, \frac{c}{|r_i|}\right) \tag{1.4}$$

Leverage points are fully taken into account if the corresponding residuals are less than or equal to c. Consider a cloud of data points without outlying observations generated by a simple regression model where the regressand is independent of the regressor. It suffices to add one leverage point to this data set to let the M-estimator return any estimate. In fact, the fitted regression line will connect the leverage point with the cloud containing the rest of the observations. The breakdown point of M-estimators is thus 1/n which approaches zero as the number of observations increases. In this example, not so much the estimate of β but especially the estimate of its standard error may induce spurious regression results. In order to bound the influence of leverage points, Mallows (1975) generalizes the first order condition of M-estimators to

$$\sum_{i=1}^{n} x_i^T w_x(x_i) r_i w_r(r_i) = 0, \tag{1.5}$$

where $w_x(\cdot): \mathbb{R}^p \rightarrow \mathbb{R}_+$ is the weight function based on the identification of horizontal outliers. The solution to equation (1.5) results in a bounded influence or so-called Generalized M-estimator for appropriate choices of the weight functions w_x and w_r which downweigh both leverage points and vertical outliers.

The Newton-Raphson (NR) algorithm is employed to solve (1.5). For this purpose we compute the Hessian by differentiating the LHS of (1.5) (the gradient) with respect to β,

$$H = -\sum_{i=1}^{n} x_i^T w_x(x_i)(x_i w_r(r_i) - r_i w_r'(r_i)), \tag{1.6}$$

where

$$w_r'(r_i) = \frac{\partial w_r(r_i)}{\partial r_i}\frac{\partial r_i}{\partial \beta} = -\frac{\partial w_r(r_i)}{\partial r_i} x_i. \tag{1.7}$$

For each step j of the NR iterations

$$\beta_{j+1} = \beta_j - H^{-1}\left(\sum_{i=1}^{n} x_i^T w_x(x_i) r_{ij} w_r(r_{ij})\right) \tag{1.8}$$

we repeatly determine which observations are considered to be vertically outlying since the size of the ith residual r_{ij} and hence also $w_r(r_{ij})$ depend on the parameter estimates β_j. On the other hand, leverage points are distinguished independently of the statistical pattern contained in the data. Thus, $w_r(r_{ij})$ are updated for every iteration whereas $w_x(x_i)$ have to be computed only once.

Simpson, Ruppert and Carroll (1992) show that the final GM estimator (after convergence of procedure (1.8)) inherits the breakdown point of the preliminary estimators that are used to (i) compute w_x and (ii) to obtain an initial estimate of β, provided that certain conditions on w_r are fulfilled. With this end in view, we base the identification of leverage points on the MVE-estimates of location and scale and employ Rousseeuw's LMS technique to deliver a starting value of $\hat{\beta}$ for the NR algorithm. Both the LMS and MVE estimators have the comfortable breakdown point of 50%. In Section 1.4 we verify whether the SRC conditions necessary for the transfer of the breakdown point apply to our choice of w_r. To compute the LMS estimator we have used the resampling algorithm with the intercept updating rule, both as described in Rousseeuw and Leroy (1987, p.197-202) and Appendix A1. The computation of the MVE is extensively discussed in Section 1.3.

1.2.2 Weights Based On Leverage Points

What we need in order to detect leverage points and to determine weights correspondingly is to measure the extent to which an observation deviates from the center of the explanatory variables. A classical measure is the diagonal of the hat matrix (see e.g. Judge et al. (1988, p.892)),

$$h_i = x_i(X^TX)^{-1}x_i^T \quad i=1,...,n, \tag{1.9}$$

where $X = (x_1^T,...,x_n^T)^T$. Related to (1.9) is the Mahalanobis distance[6]

$$MD_i = \sqrt{(x_i - T(X))C(X)^{-1}(x_i - T(X))^T} \quad i=1,...,n, \tag{1.10}$$

where $T(\cdot)$ is the arithmetic mean and $C(\cdot)$ the sample covariance matrix. These classical indicators show how far an observation is located from the sample average. Although the use of (1.9) and (1.10) is to trace the leverage points, in seeming contradiction, both can however be corrupted by them. Consequently, mild leverage points or a cluster of huge horizontal outliers can escape easily from being discovered when these classical non-robust criteria are used (For a prominent exposition of this phenomenon see Rousseeuw and Van Zomeren (1990). See also the example of Section 1.5). To overcome this *masking effect* we insert robust Minimum Volume Ellipsoid (MVE) estimates of the location $T(X)$ and the dispersion $C(X)$ in (1.10). This results in robust distances RD_i since

[6] Rousseeuw and Van Zomeren (1990) mention that $h_i = MD_i^2/(n-1) + 1/n$.

the MVE estimator has a breakdown point of 50% (see Rousseeuw and Leroy (1987)).

The squared Mahalanobis distances follow a chi-squared distribution with p degrees of freedom if the p explanatory factors are drawn from a joint normal distribution. If we are willing to assume that the majority of the explanatory observations of the data represent the normal then weights based on leverage points can be calculated as:

$$w_x(x_i) = \min\left(1, \frac{\sqrt{\chi^2_{0.975}(p)}}{RD_i}\right) \quad i=1,...,n. \tag{1.11}$$

Notice that all observations with a robust distance exceeding the 97.5th percentile of the $\chi^2(p)$ distribution are identified as leverage points and receive weights less than unity.

1.2.3 Weights Based On Vertical Outliers

Our next goal is to determine weights based on vertical outliers. To grasp a better understanding of the sensitivity of OLS to outliers, we differentiate the objective function $\rho(\cdot)$ with respect to r_i. Although the resulting function, $\psi(r_i) = \partial\rho(r_i)/\partial r_i$, is somewhat different from the influence function defined in footnote 2, they are closely related (see Hampel et al. (1986), p.316). $\psi(r_i)$ measures the rate of change of the objective function (1.2) due to an infinitesimal change in the residual associated with the ith data point. In the case of OLS, $\psi(r_i) = 2r_i$. As a consequence, large positive or negative residuals have a stronger impact on (1.2) than small absolute residuals given a certain accompanying value of x. Thus, OLS is extremely vulnerable to vertical outliers.

In order to bound the influence of vertical outliers we force the ψ-function to be redescending in its argument. In particular, Tukey's bi-square function, depicted in Figure 1.2,

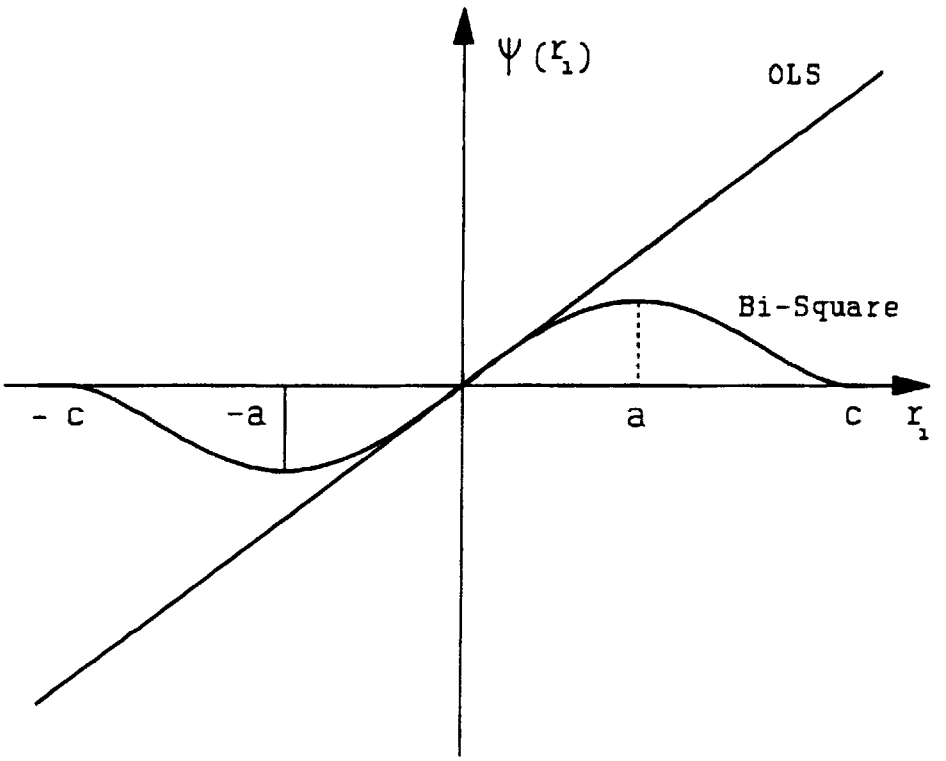
$$\begin{aligned} \psi(r_i) &= r_i(1-(r_i/(\sigma c))^2)^2 \quad |r_i/\sigma| < c \\ &= 0 \qquad\qquad\qquad\quad |r_i/\sigma| \geq c, \end{aligned} \tag{1.12}$$

is used when constructing the weights w_r:

$$\begin{aligned} w_r(r_i) &= \frac{\psi(r_i)}{r_i} \quad r_i \neq 0 \\ &= 1 \qquad r_i = 0. \end{aligned} \tag{1.13}$$

Specification of c and σ is required to compute (1.12). The LMS scale estimate (see Appendix A1) is used as an approximation for the standard deviation whereas a common choice for c is 4.685 (see Beaton and Tukey (1974)); c reflects the tradeoff between efficiency and outlier robustness of the HBP GM estimator as mentioned in the introduction to Chapter 1. Choosing values lower than 4.685 means that more observations are downweighted and lower weights apply to the vertical outliers. Thus,

Figure 1.2 Tukey's Bi-square Function and The OLS Psi-Function



efficiency losses are sustained if the data is free from outliers. On the other hand, the efficient maximum likelihood estimator (MLE) is obtained when c grows to infinity but then the estimates are unreliable if the data is contaminated. The chosen number c = 4.685 is found by simulation. The M-estimator (i.e. $w_i(x_i)$ = 1, i=1,..,n) based on (1.12) with c = 4.685 reaches an efficiency level of 95% in comparison with MLE if the large sample of model errors are standard normally distributed. Notice that our HBP GM estimator also downweighs leverage points which are not vertically outlying. This creates another channel that reduces efficiency. Monte Carlo experiments are performed and reported in Section 1.4 to examine the efficiency of our HBP GM estimator.

Two differences between Huber's proposal (1.4) and Tukey's suggestion for the computation of the weights w_r stand out: (i) The bi-square function downweighs the observations to zero if their standardized residuals step over the threshold value c whereas the most excessive vertical outlier still has influence on the estimates if Huber's weights are employed, (ii) Huber does not discriminate between observations with small and large residuals as long as the absolute value of the residual is less than c. Therefore, among the observations which are not identified as vertically outlying, the ones exhibiting relatively small residuals have less influence on the objective function (1.2) than the others. On the other hand, all vertical outliers have the same impact if the explanatory part of these

observations is equal.[7] In contrast with (1.4), the Tukey weights based on (1.12) do not only distinguish a group of huge vertical outliers (which are downweigted to zero) but also divide the other group between observations which receive more importance when the absolute value of the residual increases and the mild vertical outliers which receive less and less influence the further they are away from their concomitant predicted value. The standardized residuals of the latter group of mild outlying observations take an absolute value between a and c in Figure 1.2. In support of the bi-square function one could argue the following: It is not void of reason to ignore those observations with extremely large residuals[8] since they by no means reveal the predicted theoretical relationship between the independent variable and the explanatory factors. The theoretical (economic) model is misspecified for these outliers. Of course, the applied econometrician should always report how many and which observations are put aside. In this way, insight is gained into what extent the data suits the theory.

1.2.4 Data and Model Suitability

After convergence of the NR algorithm (1.8), the HBP GM estimator can be written as

$$\beta_{GM} = \frac{\sum_{i=1}^{n} x_i^T w_x(x_i) w_r(r_i) y_i}{\sum_{i=1}^{n} x_i^T w_x(x_i) w_r(r_i) x_i}, \tag{1.14}$$

where $w_x(x_i)$ is given by (1.11) and $w_r(r_i)$ by (1.13).[9] In obvious matrix notation (1.14) reads

$$\beta_{GM} = (X^T W_x(X) W_r(r) X)^{-1} (X^T W_x(X) W_r(r) y), \tag{1.15}$$

where $W_x(X)$ is a diagonal matrix comprising $w_x(x_i)$, $i=1,..,n$, and $W_r(r)$ is a diagonal matrix with entries $w_r(r_i)$, $i=1,..,n$.

When should the HBP GM estimator (1.15) be applied? In economics, data can be qualitative or quantitative and be measured either at micro (e.g. firm or household) or macro (e.g. aggregate) level. In what follows we will discuss each of these cases and indicate when the HBP GM estimator is to be used.

If the response variable is qualitative (giving rise to probit, logit, etc. models), the HPB GM

---

[7] In the Huber case, $\psi(r_i) = r_i$ *if* $r_i \leq c$, $\psi(r_i) = 1$ otherwise.

[8] Note that if the model errors are normally distributed then more than 99.75% of the domain of the standardized errors is covered by the interval [-c,c]. In the normal case, standardized absolute residuals which exceed c are considered to belong to the group of extreme outliers.

[9] Our GAUSS computer program returns a HBP GM estimate which slightly differs from (1.14) since NR assumes that the second order term of the Taylor expansion is negligible (see formula (1.39) of sub-section 1.4.1).

estimation technique is inappropriate. In this case, the bi-square function does not act as the right sounding board for the measurement of outlying residuals while correcting for leverage points is not at all straightforward. On the other hand, if (some of the) explanatory variables are dichotomous (i.e. dummies) and the dependent variable is not, the HBP GM estimator can be applied.

Although time series often contain outlying observations, the HBP GM estimation procedure is not developed to estimate ARIMA type models. In particular, outliers present in time series are of a fundamentally different nature (see e.g. Gómez and Maravall (1994)) than the ones considered in this chapter. Using pooled cross-section-time-series data (so called panel or longitudinal models) creates specific problems which are analyzed in Chapter 2 and Chapter 3. For a discussion on joint estimation of model parameters and outlier effects in time series see Chen and Liu (1993).

Micro data are notorious for their erratic behaviour. Here especially robust estimation techniques are called for. Aggregation of micro data can remove some of the irregularities, but also in this case robust estimation is still desired since classic and robust estimates are almost identical when data are smooth. On the other hand, if the aggregate data remain erratic, the HBP GM estimator yields more reliable estimates compared to classic techniques. Moreover, cross-country (macro economic) surveys are likely to contain (severe) outliers despite their aggregate nature.

To summarize, the HBP GM estimator described in this section is appropriate for estimating a cross-sectional data model in which the response variable is not dichotomous.

1.3 The MVE Estimator of Location and Scatter

The Minimum Volume Ellipsoid (MVE) estimator (Rousseeuw (1985)) is based on the hyper ellipsoid of minimum hyper volume containing at least half of the observations. The estimate of location corresponds to the centre of this hyper ellipsoid while the corresponding covariance estimate is the hyper ellipsoid multiplied by some factor to obtain consistency.

The determinant of a scatter matrix (that is, a positive definite symmetric matrix) is proportional to the squared volume of the corresponding tolerance ellipsoid (see Rousseeuw and Leroy (1987), p.259). The MVE estimator is defined as the pair (T,C) which

$$\begin{aligned} &\textit{minimizes}\ \det(C) \\ &(T,C) \\ &\text{subject to} \\ &\#\{i;[x_i-T]C^{-1}[x_i-T]^T \le \delta^2\} \ge h, \end{aligned} \tag{1.16}$$

where $h = [(n+p+1)/2]$. Recall that p refers to the number of explanatory variables. The notation [x] stands for the largest integer less than or equal to x. If it is assumed that the majority of the data comes from a normal distribution, δ^2 is set equal to the 50th percentile of the $\chi^2(p)$ distribution. We then denote

$$RD(x_i) = \sqrt{[x_i - T]C^{-1}[x_i - T]^T} \ , \tag{1.17}$$

as the robust distance of case x_i. Observe that T and C are defined such that the h-th order statistic of the robust distances is

$$\{RD(x_i)\}_h = \sqrt{\chi^2_{0.50}(p)} \ . \tag{1.18}$$

Two procedures are reported in the literature to approximate the solution to (1.16): the *resampling* algorithm and the *projection* algorithm. However, those algoritms do not yield robust distances satisfying (1.18), as will become clear below. We therefore propose to correct the *approximated* MVE-distances $RD^*(\cdot)$, computed by either algorithm, by multiplying each computed robust distance with the correction factor (this factor was originally conceived by Kloek (1996))

$$\frac{\sqrt{\chi^2_{0.50}(p)}}{\{RD^*(x_i)\}_h} \ . \tag{1.19}$$

In this way the median (or more precisely, the h-th order statistic) of the $RD^*(\cdot)$ is always equal to the root of the 50-th percentile of the χ^2-distribution with p degrees of freedom.

Further, the breakdown point of the MVE estimator is $[(n-p+1)/2]/n$, which is 50% as the number of observations, n, goes to infinity (see Rousseeuw and Leroy (1987, p.264)).

To assess the correction factor we have run some simulations, the results of which are presented below, after brief descriptions of the two algorithms used to approximate the MVE-distances.

1.3.1 Projection

The need for a correction factor like (1.19) is made clear by considering the *projection* algorithm of Rousseeuw and Van Zomeren (1990), which is based on Gasko and Donoho (1982). This algorithm is built around

$$u_i = \max_v \frac{|x_i v^T - L(x_1 v^T, \ldots, x_n v^T)|}{S(x_1 v^T, \ldots, x_n v^T)} \ , \tag{1.20}$$

which is the exact one-dimensional version of (1.17) applied to the projections $x_i v^T$ if $L(\cdot)$ and $S(\cdot)$ are the MVE-estimates of location and scatter respectively. The latter are defined as

$$L(z_1,\ldots,z_n) = (z_j + z_{j-h+1})/2 \tag{1.21}$$

and

$$S(z_1,\ldots,z_n) = \gamma(z_j - z_{j-h+1}) \tag{1.22}$$

where $z_1 \le z_2 \le \ldots \le z_n$ is any set of numbers, where $z_j - z_{j-h+1}$ is the smallest of the differences

$$z_h - z_1, z_{h+1} - z_2, \ldots, z_n - z_{n-h+1} \tag{1.23}$$

and where γ is some constant used to achieve consistent estimates of the unknown scale parameter.[10] If the projections $x_i v^T$ are (standard) normally distributed then one can verify that γ must equal

$$\gamma = \frac{1}{2\sqrt{\chi^2_{0.50}(1)}}, \tag{1.24}$$

for large samples.

In principle all directions v should be considered to compute (1.16). Because in practise this is impossible, Rousseeuw and Van Zomeren (1990) propose to restrict the set of directions to all v defined as $x_l - M$, $l = 1,..,n$, with $M = \{ \underset{j}{med}(x_{j1}), \ldots, \underset{j}{med}(x_{jp}) \}$.

Observe however that the u_i in (1.20) do not necessarily satisfy (1.18) unless there is only one explanatory variable. Indeed, in the special case of one dimension we have

$$u_i = \frac{|x_i - (x_j + x_{j-h+1})/2|}{\gamma(x_j - x_{j-h+1})} \tag{1.25}$$

where $x_1 \le x_2 \le \ldots \le x_n$. In this case we can compute the h-th order statistic of the u_i exactly. As a first step, we find the remarkable identity

$$\{|x_i - T|\}_h = \underset{j-h+1 \le i \le j}{maximum} |x_i - T| = (x_j - x_{j-h+1})/2, \tag{1.26}$$

because h of the x_i's are in between x_{j-h+1} and x_j. Hence,

[10] Rousseeuw and Van Zomeren (1990) argue that the factor $\gamma(n)$ depends on the sample size but do not provide an explanation or a specification of γ.

$$\{u_i\}_h = \left\{\frac{|x_i - T|}{S(\cdot)}\right\}_h = \frac{1}{2\gamma} = \sqrt{\chi^2_{0.50}(1)}. \tag{1.27}$$

Given that γ is fixed independently of the number of explanatory variables, the robust distances $RD^*(x_i)$ based on the projection algorithm in general do not satisfy (1.18). For higher dimensions we can restore equation (1.18) by multiplying the u_i's by (1.19). Robust distances are then given by

$$RD^*(x_i) = \frac{\sqrt{\chi^2_{0.50}(p)}}{\{u_i\}_h} u_i. \tag{1.28}$$

The following steps summarize the algorithm as we have used it in our simulations:

Algorithm P (Projection Algorithm). This algorithm calculates robust distances for each case x_i using approximated MVE-estimates of location and scale obtained by *projecting* x_i on a one-dimensional subspace. The MVE-estimates of location and scale are approximated as being the midpoint and length respectively of the shortest half of all projections considered.

P1. [Projection of each case x_i on a one-dimensional subspace] Compute the projections $z_i = x_i v_l^T$ for all cases x_i in the direction $v_l = x_l - M$, with $M = \{\underset{j}{med}(x_{j1}), \ldots, \underset{j}{med}(x_{jp})\}$.

P2. [Shortest half of the projections] Order the numbers z_i such that $z_1 \le z_2 \le \ldots \le z_n$. Compute the smallest of the differences

$$z_h - z_1, z_{h+1} - z_2, \ldots, z_n - z_{n-h+1},$$

where $h = [(n+p+1)/2]$. Denote this shortest half by $z_j - z_{j-h-1}$.

P3. [Midpoint and the length of the shortest half] Compute the location and scale estimate according to (1.21) and (1.22) respectively (observe that because the final robust distances are multiplied by (1.19), in this step set γ equal to 1 (see (1.22)).

P4. [Repetition] Repeat step P1 through P3 for all projection directions v_l, $l = 1, .., n$.

P5. [Robust distances] Compute robust distances u_i.

P6. [Correction factor] Compute correction factor (1.19) as

$$\frac{\sqrt{\chi^2_{0.50}(p)}}{\underset{i=1,..,n}{med}\{u_i\}}.$$

P7. [Final correction] Multiply each robust distance u_i by the factor of step p6.

1.3.2 Resampling

The oldest algorithm to approximate the MVE is the *resampling* algorithm proposed by Rousseeuw and Leroy (1987, pp. 259-60); see also Rousseeuw and Van Zomeren (1990, 1991). The idea is to compute a sufficient number of ellipsoids containing half of all observations. Among these, the one with lowest volume is taken as the approximation of (1.16).

In particular, let

$$\bar{x}_K = \frac{1}{p+1}\sum_{i\in K} x_i \tag{1.29}$$

and

$$C_K = \frac{1}{p}\sum_{i\in K} (x_i-\bar{x}_K)^T(x_i-\bar{x}_K) \tag{1.30}$$

be respectively the arithmetic mean and corresponding covariance matrix of a subsample of $(p+1)$ different observations, indexed by $K = \{i_1,\ldots,i_{p+1}\}$. To assure that exactly h observations are contained in the corresponding ellipsoid, C_K should be inflated or deflated by the h-th order statistic

$$m_K^2 = \left\{(x_i-\bar{x}_K)C_K^{-1}(x_i-\bar{x}_K)^T\right\}_h. \tag{1.31}$$

Indeed, the ellipsoid containing h observations corresponding to subsample K is given by $m_K^2 C_K$, with a concomitant volume proportional to

$$\det(m_K^2 C_K). \tag{1.32}$$

The algorithm consists of computing the value of (1.32) for many different subsamples, after which the subsample K^*, corresponding to the lowest value of the objective function, is retained. The MVE-approximation of location and scatter are then respectively given by

$$T^* = \bar{x}_{K^*}. \tag{1.33}$$

and

$$C^* = \frac{m_{K^*}^2}{\chi^2_{0.50}(p)} C_{K^*}. \tag{1.34}$$

Robust distances are then obtained by inserting (1.33) and (1.34) into (1.17). Observe that up to this point (1.18) always holds even if (T^*, C^*) are not based on the ellipsoid of minimum volume.

Recall that according to formula (1.11) observations are suspicious when their robust distances exceed $\sqrt{\chi^2_{0.975}(p)}$. By generating normally distributed series and application of the resampling algorithm one can determine the extent to which the computed robust distances RD^* clash with the χ^2-distribution. With such an experiment, the cutoff value and weights (1.11) can be approved or not. In this way, an improvement in the resampling algorithm is found, due to Rousseeuw and Van Zomeren (1991). They determine an empirical factor c, which depends both on n and p, such that on average the 97.5-th percentile of the $RD(x_i)$ approaches the square root of the χ^2-distribution with the appropriate degrees of freedom. In particular, (1.34) is multiplied by

$$c(n,p)^2 = (1+15/(n-p))^2. \tag{1.35}$$

Our simulations show that (1.35) indeed improves the MVE estimates (see Section 1.3.3) but still even more accurate MVE estimates can be obtained.

Another improvement had already been proposed by Rousseeuw and Leroy (1987, p.260). They assign to each observation a weight w_i according to

$$\begin{aligned} w_i &= 1 \quad \textit{if} \quad (x_i - T^*)C^{*-1}(x_i - T^*)^T \le \kappa \\ &= 0 \quad \textit{otherwise,} \end{aligned} \tag{1.36}$$

where, for instance, $\kappa = \chi^2_{0.975}(p)$. More efficient (reweighted) estimators for location and scatter are then suggested as being respectively

$$T^{**} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{1.37}$$

and

$$C^{**} = \frac{\sum_{i=1}^{n} w_i (x_i - T^{**})^T (x_i - T^{**})}{\sum_{i=1}^{n} w_i - 1}. \tag{1.38}$$

The h-th order statistic of the robust distances RD^* based on the MVE-estimates (1.37) and (1.38) are no longer necessarily equal to $\sqrt{\chi^2_{0.50}(p)}$. Therefore, more accurate robust distances are possibly obtained by applying the correction factor (1.19) to the robust distances computed with the one-step improvement of Rousseeuw and Leroy. Notice that the constant c of Rousseeuw and Van Zomeren (1991) becomes redundant once the robust distances are multiplied by our correction factor (1.19). In Section 1.3.3 we test both correction mechanisms. As will become clear below, it appears that the procedure in which correction factor (1.19) is applied to robust distances obtained from (1.37)

and (1.38) results in the most satisfactory approximations. We need to remark that (1.19) is a small sample correction factor since it will converge to 1 if the number of symmetrically distributed observations grows to infinity.

In short, the algorithm as we have used it in our simulations evolves according to the following steps:

Algorithm R (Resampling Algorithm). Robust distances for each case x_i are computed, using approximations of the MVE-estimates of location and scale. The latter are approximated respectively as the mean and corrected covariance of a subsample K^*. The covariance of this subsample has the lowest corresponding volume among Q different randomly drawn subsamples of $p+1$ data points.

R1. [Number of drawings Q] Compute

$$Q = \left[\frac{\ln(1-prob)}{\ln(1-(1-\varepsilon)^{p+1})} + 1 \right],$$

where ε is the maximum fraction of data pollution, and where *prob* is the probability that at least one of the subsamples is free of outliers. In our GAUSS routines we have set ε equal to 0.5, *prob* equal 0.99 and take 10 times as many subsamples as given by the above rule.

R2. [Drawing of a subsample] Draw at random a subsample K of $p+1$ different observations x_i.

R3. [Initial mean and covariance] Compute the mean (T_K) and the covariance (C_K) of the subsample K according to (1.29) and (1.30) respectively.

R4. [Objective value] Compute inflation factor (1.31) and accordingly the objective function's value (1.32).

R5. [Repetition] Repeat steps R2 through R4 for Q different subsamples. Keep the subsample, K^*, with the lowest value of the objective function, as computed under R4.

R6. [Scale estimate] Compute the covariance estimate (1.34) on subsample K^*.

R7. [Weights] Assign to each observation x_i a weight w_i according to (1.36), using the covariance estimate of step R6.

R8. [Rousseeuw and Leroy's one step improvement] Compute new estimations of location and scale according to (1.37) and (1.38) respectively.

R9. [Robust distances] Compute robust distances RD_i^* by inserting the location and scatter estimates of step R8 into (1.17).

R10. [Correction factor] Compute correction factor (1.19).

R11. [Final correction] Multiply each robust distance RD_i^* by the factor of step R10.

Table 1.1 Simulated 97.5th percentiles using Resampling and Projection Algorithms[a]

Algorithm[b]	$n=50$				
	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
I	0.936926	0.878970	0.854576	0.839810	0.829328
II	0.982440	0.957838	0.949937	0.946205	0.944832
III	0.968137	0.956340	0.953417	0.951283	0.950901
IV	0.968073	0.964506	0.962086	0.959474	0.958450
V	0.936398	0.865388	0.882428	0.886528	0.924196
VI	0.936398	0.978546	0.973650	0.967520	0.960898
	$n=250$				
I	0.963207	0.951092	0.942612	0.932714	0.922099
II	0.973080	0.964387	0.959334	0.952085	0.944556
III	0.961111	0.960795	0.961163	0.960228	0.959358
IV	0.972621	0.972188	0.971987	0.971478	0.971067
V	0.963284	0.937228	0.936116	0.939820	0.949292
VI	0.963284	0.986768	0.984296	0.980796	0.977084

[a] Each cell contains the mean of the estimated percentile over the 10,000 runs.

[b] I Resampling
II Resampling with correction factor (1.35)
III Resampling with Rousseeuw and Leroy's one step improvement *and* correction factor (1.35)
IV Resampling with Rousseeuw and Leroy's one step improvement *and* correction factor (1.19)
V Projection
VI Projection with correction factor (1.19)

1.3.3 Testing the MVE-Approximations

In order to assess the empirical performance of (1.19) we have tested different versions of both the resampling and projection algorithm to approximate (1.16). Table 1.1 reports our simulation results. Each experiment consisted of generating a standard normal random variable for which the MVE-distances were approximated, using various versions of the resampling algorithm (lines I-IV) and the projection algorithm (lines V-VI). Then, what fraction of the approximated squared MVE-distances exceeded the 97.5th percentile of the $\chi^2(p)$-distribution ($\hat{\alpha}$) was examined. This was repeated 10,000

times both for a small sample (n = 50) and for a large sample (n = 250), as well as for a number of explanatory variables. In Table 1.1 the mean outcomes of $1-\hat{\alpha}$ are reported.

The rows labelled (I) in Table 1.1 contain the results obtained when using the resampling algorithm without any correction. The impact of the correction factor (1.35) (due to Rousseeuw and Van Zomeren (1991)) on the resampling algorithm can be seen from the second row. Indeed, a significant improvement is observed although, apart from the one dimensional case, the MVE-distances are not adequately approximated. Observe also that the initial correction factor has much more impact in the small sample case (n =50) than in case of the large sample (n =250). This is due to the fact that the small sample correction factor is approaching 1 if n increases. The third row entries (III) involve the resampling algorithm with correction factor (1.35) and the one-step improvement as suggested by Rousseeuw and Leroy (1987, p.260). This leads to some improvement although in all cases considered the 97.5th percentile is underestimated. Both the one-step improvement of Rousseeuw and Leroy (1987) and correction factor (1.19) are used in the fourth algorithm. This version of the resampling algorithm has the best performance of all versions considered. Indeed, since the third algorithm underestimates the 97.5th percentile there is still room for additional refinement, which we achieve by applying (1.19) to this algorithm.

Except for the one-dimensional cases, the projection algorithm (V) almost never yields appropriate distance approximations. Applying correction factor (1.19) to this algorithm does considerably improve its performance. Note that the two versions of the projection algorithm are equivalent if the model contains only one regressor.

To conclude, in all cases considered the resampling algorithm with Rousseeuw and Leroy's one step improvement *and* correction factor (1.19) yields reliable approximations of the MVE distances. Further, applying correction factor (1.19) to the projection algorithm also results in appropriate distance approximations. When the sample is small the corrected projection algorithm outperforms the resampling algorithm whereas the corrected resampling algorithm is slightly more accurate than projection in the large sample experiment. In the first case we recommend using projection since this algorithm is much faster than the resampling algorithm. For the simulations concerning efficiency of the HBP GM estimator (reported below) and the other cases in this thesis where computation of MVE distances is required, however, we have used algorithm IV.

Section 1.4 Statistical Properties of the HBP GM Estimator

Our next aim is to bring to light the statistical properties of the HBP GM estimator. First we discuss several widely-used diagnostic regression statistics such as the coefficient of determination, scale estimate and t-values. Then we proceed with revealing the asymptotic behavior of the estimator by considering consistency, the breakdown point and the level of efficiency.

1.4.1 t-values, adjusted R^2 and Scale Estimate

In order to derive the covariance matrix of the HBP GM estimator, $cov(\hat{\beta}_{GM})$, we use a Taylor expansion around $\hat{\beta}_{GM} = \beta$ for the first order condition (1.5):

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \psi(r_i) x_i^T w_x(x_i) \\ &= \sum_{i=1}^{n} \psi(\varepsilon_i) x_i^T w_x(x_i) + \sum_{i=1}^{n} \frac{\partial \psi(r_i)}{\partial r_i} \frac{\partial r_i}{\partial \beta_{GM}} x_i^T w_x(x_i)(\hat{\beta}_{GM} - \beta). \end{aligned} \tag{1.39}$$

Evaluating (1.39) in $\hat{\beta}_{GM} = \beta$ and performing obvious manipulations gives

$$\sqrt{n}(\hat{\beta}_{GM} - \beta) \approx \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \psi(\varepsilon_i)}{\partial \varepsilon_i} x_i^T x_i w_x(x_i) \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi(\varepsilon_i) x_i^T w_x(x_i) \right\}. \tag{1.40}$$

Thus, in large samples, the covariance matrix of $\hat{\beta}_{GM}$ approximately equals

$$cov(\hat{\beta}_{GM}) = (X^T V(r) W_x(X) X)^{-1} X^T W_x(X) \Psi W_x(X) X (X^T V(r) W_x(X) X)^{-1}, \tag{1.41}$$

where Ψ is a diagonal matrix consisting of $(\psi(r_i))^2$, $i=1,...,n$, $V(r)$ is a diagonal matrix containing $\partial \psi(r_i)/\partial r_i$, $i=1,...,n$ and $W_x(X)$ is as defined as in equation (1.15) of Section 1.2.4. With (1.41) the usual t-values can be computed.[11]

Although any observation has bounded influence on the estimates of β, standard error breakdown may still occur (see Simpson, Ruppert and Carroll (1992), p. 441). Consequently, the t-statistics do not provide a robust test on the significance of the HBP GM estimates. A solution to robust testing of the hypothesis $\beta_2 = 0$, β_1 unspecified (where β_2 contains q components of the p-dimensional vector β, $\beta^T = (\beta_1^T \beta_2^T)$, $q<p$), is found by, among others, Markatou and Hettmansperger (1990). They robustify the traditional F-test for testing general linear hypotheses in the linear model to a so-called aligned generalized M-test. The M-test is constructed for the Generalized M-estimator and exhibits the appealing property of having a bounded influence function.[12] Furtermore, Markatou and Hettmansperger show that the robust M-test has asymptotically a chi-squared distribution with q degrees of freedom under the null hypothesis.

To examine the reliability of the usual t-values in clean and corrupted data sets we have performed a Monte Carlo study. Series of two independent standard normal one-dimensional

[11] Ali and Sharma (1996) investigate the robustness of F-tests when the null distribution is not necessarily normal. They find that the standard F-test is especially sensitive to the presence of leverage observations. Since the HBP GM estimator weighs these influential observations, Ali and Sharma's results do not apply here.

[12] See Hampel et al. (1986), p. 191, for a definition of the test influence function.

Table 1.2 Simulation Results Concerning Power Properties of the t-values and M-test[a]

Pollution	97.5th percentile/ 2.5th percentile, t-statistic	p-value, t-statistic	95th percentile, M-statistic	p-value, m-statistic
		N=200		
0%	2.002 -2.016	0.0547	3.503	0.0394
5%	2.187 -2.150	0.0719	3.239	0.0303
10%	2.285 -2.255	0.0818	2.781	0.0200
15%	2.408 -2.414	0.1005	2.353	0.0110
20%	2.647 -2.685	0.1227	2.105	0.0089
25%	2.963 -2.858	0.1504	1.784	0.0045
		n=1000		
0%	1.991 -2.035	0.0575	3.897	0.0514
5%	1.979 -2.046	0.0546	3.629	0.0433
10%	2.019 -2.013	0.0560	3.369	0.0368
15%	2.049 -2.112	0.0622	3.159	0.0297
20%	2.121 -2.145	0.0697	2.982	0.0241
25%	2.294 -2.219	0.0801	2.767	0.0187

[a] The number of replications is 10,000 for both series of length N=200 and N=1000. Under the null-hypothesis of $\beta_2 = 0$ the M-statistic has a chi-squared distribution with 1 degree of freedom, $\chi^2_{0.05}(1) = 3.84146$, $t_{0.025}(\infty) = 1.96$.

explanatory variables x_1 and x_2 were generated. The independent variable was constructed by $y = x_1 + \varepsilon$ where ε is a random variable also with a standard normal distribution. After generating the clean data, the t-value associated with the coefficient on x_2 and the Markatou-Hettmansperger M-

test, both based on application of our HBP GM estimator to a model which includes the regressors x_1 and x_2, are computed for a small and a large sample consisting of 200 and 1000 observations respectively. This experiment was repeated ten thousand times. The *2.5*-th and *97.5*-th percentile of the calculated values of the t-statistic and the *95*-th percentile of the M-test with corresponding p-values are reported in Table 1.2. Then we contaminated both the series of the independent and explanatory variables by replacing δ% of the observations by random values drawn from a normal distribution with zero mean and variance 100. The t-values and M-tests are computed for different data sets which are increasingly contaminated with steps of 5% up to the level of pollution of 25% (δ=0,5,10,15,20,25).

The simulation results reveal that the applied econometrician should be cautious when interpreting the t-values of the HBP GM estimator if the data set under investigation is relatively small and contaminated. In all cases considered, the t-statistic rejects too many times the true null-hypothesis if a significance level of 5% is employed. In the small sample case of maximum data pollution, the p-value reaches a level of approximately 15%. Thus, 15% of the computed absolute t-values exceeded the cutoff value of 1.96. On the contrary, the M-test tends to be first order stochastically dominated by the $\chi^2(1)$-distribution. The corresponding p-value decreases from almost 4% for the clean small sample data set to not even 0.5% when 25% of the data are polluted. Evidently, conlusions can not be drawn from Table 1.2 about the power properties of the M-test when the alternative hypothesis holds. On the other hand, the usual t-values are accurate for relatively large samples even if between 10 and 15 per cent of the data points do not follow the statistical pattern observed by the majority of the data. The computed p-value associated with the t-statistic is equal to 5.6% for the 10% pollution large sample case. This number is reasonably close to the significance level of 5% while extreme outliers corrupt the data. Note that for the large sample cases the p-values corresponding to the M-tests remain considerably higher.

A reasonable estimate of the goodness of fit (R^2) of the regression requires a correction factor. The final outcome of the HBP GM estimation procedure (see equation (1.15)) can be obtained by applying OLS to the weighted observations $X^* = X\sqrt{W_x(X)W_r(r)}$ and $y^* = y\sqrt{W_x(X)W_r(r)}$. Given the perception that the HBP GM technique boils down to weighted least squares, a logical choice for computation of the unadjusted coefficient of determination is:

$$A^2 = \frac{\sum_{i=1}^{n}(x_i^*\beta_{GM} - \bar{y}^*)^2}{\sum_{i=1}^{n}(y_i^* - \bar{y}^*)^2} \tag{1.42}$$

when the model contains an intercept and

$$A^2 = \frac{(X^*\beta_{GM})^T X^*\beta_{GM}}{y^{*T}y^*} \tag{1.43}$$

otherwise, where $\bar{y}^* = \frac{1}{n}\sum_{i=1}^{n} y_i^*$. A^2 is an inappropriate measure of fit since it is increasing in the number of explanatory variables. For that reason, the adjusted R^2

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-p}\right)(1-A^2) \tag{1.44}$$

is used to measure the explanatory power of the regression model (see Judge et al. (1988, pp. 212-213)).

Another statistic often reported in regression analysis is the estimator of the scale parameter σ^2. To obtain a robust estimate of the HBP GM regression standard deviation we adopt the following procedure. Consider a normally distributed random variable, u, with mean μ and variance σ^2. For a sample of n observations of u (denoted by $\boldsymbol{u}$) the sample median is a robust estimator of the location parameter μ. Consequently u - med(u) is symmetrically distributed around zero. If we take the median of the absolute value of u - med(u), we have a robust estimate of the 75th percentile of the initial distribution of u. Therefore it is straightforward to propose the Median Absolute Deviation (MAD), defined as

$$s = med(|u - med(u)|)/0.6745, \tag{1.45}$$

as a robust estimator for σ where the scaling factor 0.6745 is the 75th percentile of the standard normal distribution. We use as final estimation of the variance of the HBP GM regression the square of (1.45), where u_i is defined as

$$u_i = y_i - \boldsymbol{x}_i\hat{\beta}_{GM} \quad i = 1,...,n. \tag{1.46}$$

1.4.2 Breakdown Point

The Generalized M-estimator developed in this Chapter is B-robust; i.e. any observation has bounded influence on the estimate. B-robustness does not necessarily imply that a group of observations containing more than one outlier have bounded impact. Therefore, it is useful to consider also the notion of breakdown point. The breakdown point of an estimator is determined by the smallest amount of data contamination that can cause the estimator to take on arbitrary values. Suppose $\hat{\beta}$ is an estimator. Define $\eta(\delta;\hat{\beta}(\cdot),X)$ to be the supremum of $\|\hat{\beta}(X') - \hat{\beta}(X)\|$ for all X', where X' corresponds to the original data set with a fraction δ ($=m/n$) of the observations replaced by arbitrary values. The breakdown point of $\hat{\beta}(\cdot)$ on X is then formally defined as (see Donoho and Huber (1983))

$$b(T(\cdot),X) = \min\{\frac{m}{n}; \eta(m;T(\cdot),X) = \infty\}, \tag{1.47}$$

where n is the number of observations in X and m the number of original data points exchanged by any value.

Simpson, Ruppert and Carroll (1992) show that[13]

> Under reasonably general conditions the regression parameter estimates (one-step GM estimates) inherit the breakdown properties of the preliminary estimates of the regression parameters and the multivariate location and scale estimates of the design x's (1992, p.446).

One-step refers to one iteration of the Newton-Raphson type algorithm to solve Mallow's first order condition. In particular it is sufficient that the following conditions hold:

(B.1) $n-m \geq \frac{n}{2}+1 \geq p$

(B.2) $\psi(r_i)$ is odd and bounded,

(B.3) $w_r(r_i) = \frac{\psi(r_i)}{r_i} \geq d_0 > 0 \quad if \quad |r_i| \leq a,$

(B.4) $\frac{\partial \psi(r_i)}{\partial r_i} \geq d_1 > 0 \quad if \quad |r_i| \leq a$ for at least n-m good points,

(B.5) $\frac{\partial \psi(r_i)}{\partial r_i} \geq 0$ for all points,

(B.6) the set of 'good' points (i.e. $|r_i| < a$) must contain a linearly independent subset of size p,

(B.7) and a must strictly exceed the appropriate tuning constant κ (κ being the tuning constant 0.6745 in the formula of the LMS scale estimate, see Appendix A1 and Simpson, Ruppert and Carroll (1992), p.440).

Hereafter we will discuss conditions (B.1) to (B.7) with the intention of ascertaining the breakdown point of our GM-estimator. First, observe that the GM estimator is computed with a multi-step procedure (see formula (1.8)) instead of performing only one NR-iteration. If however the one-step GM estimator inherits the breakdown point of its preliminary estimators then this breakdown point is also carried over to the multi-step estimator by the argument of induction.

The first condition requires that at least half of the data are regular. This assumption together with (B.4) and (B.6) are needed to ensure that the Hessian (1.6) is positive definite and thus invertible. Condition (B.2) asserts that any vertical outlier has limited impact on the estimates. Evidently, the ψ-function associated with OLS does not fulfill (B.2). Therefore, a single outlying observation might have unbounded influence on the estimates and thus the breakdown point of the OLS estimator is nil. In contrast with the ψ-function of OLS, our GM estimator is based on the bounded Tukey's bi-square function. From (1.29) and (1.30) it is straightforward to recognize that (B.3) holds (see also Figure 1.2)

[13] Words within parentheses are added.

and to compute the number a which is enforced by condition (B.4). Recall that efficiency considerations led to the choice c=4.685 as appropriate specification of the bi-square function. Writing down the derivative of $\psi(\cdot)$

$$\begin{aligned} \frac{\partial \psi(r_i)}{\partial r_i} &= (1-(r_i/(c\sigma))^2)(1-5(r_i/(c\sigma))^2) \quad |r_i/\sigma|<c \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad |r_i/\sigma| \geq c. \end{aligned} \tag{1.48}$$

and evaluating when $\partial\psi(r_i)/\partial r_i = 0$ using (1.48) for $|r_i| < c$ brings about the constant a=2.095. Consequently, also condition (B.7) is verified. Only restriction (B.5) is left for verification. Here the ship runs ashore. Indeed, since the bi-square function is redescending, for the mild leverage points points (i.e. $a<|r_i|<c$) it is the case that $\partial\psi(r_i)/\partial r_i<0$, and (B.5) is violated. SRC comment on this situation by pointing out the following in their remark 2.1 on p. 441:

> If ψ is redescending then $\partial\psi(r_i)/\partial r_i$ can go negative. We conjecture that in this case it is possible to manipulate p data points so that the Newton-Raphson version of H_0 (the Hessian) equals 0.

In other words, the one-step GM estimates can diverge infinitely far from the starting preliminary estimates. Only p outliers might cause the breakdown of the one-step GM-estimator. By way of comparison, the Huber ψ-function (derived from (1.4)) satisfies all conditions (B.1)-(B.7). The GM-estimator constructed with (1.4) therefore can handle up to 50% pollution of the data since both the preliminary LMS and MVE estimators have breakdown points of 50%. In case of our GM estimator, however, convergence will not be achieved with a multi-step NR-algorithm if the Hessian becomes singular. The Hessian (1.6) has to be positive definite for all iterations of the NR-algorithm in order to find a local minimum of the objective function. We suggest adopting the following estimation procedure: Compute for each NR-iteration the smallest eigenvalue of the Hessian. If this eigenvalue is below some critical positive number or if the iterations outnumber a certain upperbound then our best course is to bring the NR-algorithm to a halt and report the preliminary LMS estimate as the final outcome. Otherwise, the converged GM-estimates are the final parameter estimates. This procedure inherits the high breakdown point of 50% of the preliminary estimators, while it is equivalent to the described GM-estimator based on Tukey's bi-square function in almost all of the regressions with either real or simulated data sources.

1.4.3 Consistency

Simpson, Ruppert and Carroll (1992) also provide sufficient conditions that guarantee the asymptotic normality and consistency of the GM estimator. The requirements placed on the asymptotic behavior of the (weighted) design are observed when the identification of leverage points is based on MVE

estimates. The other conditions are explicitly stated below:

(C.1) The errors $\varepsilon_1,...,\varepsilon_n$ are independent with distribution function F which has symmetric density around zero.

(C.2) The score function $\psi(r_i)$ is bounded and continuous,

(C.3) $E[\psi(\varepsilon v)] = 0$ and $E[\varepsilon v\frac{\partial\psi(\varepsilon v)}{\partial r_i}] = 0$ for any non negative scalar v,

(C.4) $\psi(r_i)$ has derivative $\frac{\partial\psi(r_i)}{\partial r_i}$ such that $\|\frac{\partial\psi(r_i)}{\partial r_i}\|_{sup} < \infty$ and $\|r_i\frac{\partial\psi(r_i)}{\partial r_i}\|_{sup} < \infty$, where $\|.\|_{sup}$ is the supremum norm,

(C.5) $\frac{\partial\psi(r_i)}{\partial r_i}$ has derivative $\frac{\partial^2\psi(r_i)}{\partial r_i^2}$ such that $\|\frac{\partial^2\psi(r_i)}{\partial r_i^2}\|_{sup} < \infty$, $\|r_i\frac{\partial^2\psi(r_i)}{\partial r_i^2}\|_{sup} < \infty$ and $\|r_i^2\frac{\partial^2\psi(r_i)}{\partial r_i^2}\|_{sup} < \infty$

and

(C.6) $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} var[\psi(\varepsilon_i)]w_x(x_i)^2x_i^Tx_i = A$ and $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E[\frac{\partial\psi(\varepsilon_i)}{\partial\varepsilon_i}]w_x(x_i)x_i^Tx_i = B$

for some symmetric positive definite matrices A and B.[14]

Conditions (C.2) and (C.3) are easily checked using (1.12) and (1.48), assuming that ε_i obeys (C.1) and remembering that all odd moments of a symmetric distribution are zero. The second derivative of the ψ-function can be written as:

$$\begin{aligned}\frac{\partial^2\psi(r_i)}{\partial r_i^2} &= (-2r_i/(c\sigma)^2)(6-10(r_i/(c\sigma))^2) \quad |r_i/\sigma|<c \\ &= 0 \qquad\qquad\qquad\qquad\qquad |r_i/\sigma|\geq c.\end{aligned} \tag{1.49}$$

Clearly, the first derivative (1.48) and second derivative (1.49) of the score function are bounded and become zero for large residuals. As a result (C.4) and (C.5) follow immediately. Finally, condition (C.6) is satisfied for an infinitely large sample of observations which are generated by a joint normal distribution. Requirement (C.6) asserts that the Hessian must converge to a symmetric positive definite

[14] As shown by SRC (1992), the HBP GM estimator produces root-n consistent estimates if there are heteroskedastic symmetric regression errors provided that conditions (C.1)-(C.6) apply and the NR-algorithm is used to solve Mallow's first order condition. For instance, in this case, the HBP GM estimates found by the scoring algorithm have the same rate of convergence $(n^{1/3})$ as the preliminary LMS estimates (see Rousseeuw and Leroy (1987), p.179).

Table 1.3 Simulation Results with Uncontaminated Data, $\beta_1=\beta_2=1$, n=200[a]

	OLS	LMS	GM
β_1	0.9989610	1.0023398	0.9991616
	0.0049485	0.0421095	0.0052413
	0.0049474	0.0421040	0.0052406
β_2	1.0012882	1.0009535	1.0011469
	0.0050393	0.0378368	0.0053645
	0.0050376	0.0378358	0.0053632

[a] Each cell contains the mean of the estimated parameter value, the mean squared error and the variance over the 10,000 runs.

matrix. In the previous sub-section we argued that, on top of the asymptotic result, the Hessian should be positive definite in any step of the NR-algorithm in order to secure the transmission of a high breakdown point. After having verified conditions (C.1)-(C.6) we conclude that the employed HBP GM estimator has asymptotically a normal distribution and produces consistent estimates.

1.4.4 Efficiency

To examine the efficiency of the HBP GM estimator compared to OLS and LMS under different types and fractions of data contamination we have performed a number of simulations. A single experiment began with generating a matrix of explanatory variables, x, of length 200, consisting of an intercept and one explanatory variable drawn from a standard normal distribution. Then a response variable, y, was created according to $y=x\beta+e$ with β_1 and β_2 set equal to unity and where e_i, $i=1,...,200$, are independent standard normally distributed. Given these data we re-estimated β using the OLS, LMS and HBP-GM estimators, the results of which are presented in Table 1.3.

Table 1.3 shows that OLS both yields the lowest variance and mean squared error. This is, of course, no surprise since under the data generating process outlined above OLS is the Uniformly Minimum Variance Unbiased Estimator. Note however that both the variance and the mean squared error of the HBP GM estimator are just a little higher. Recall that a M-estimator constructed with Tukey's bi-square function achieves an efficient level of 95% in comparison with OLS for large normally distributed data samples if the tuning constant c equals 4.685. M-estimators have unbounded influence functions since they only take care of vertical outliers. On the other hand, the HBP GM estimator downweighs leverage points independently of their vertical distance from the regression line. Therefore, we foresee an additional loss in efficiency. Division of the OLS and GM variances reveals that the HBP GM estimator obtains a relative efficiency level of 94.4% in the experiment we carried out. The alleged loss in efficiency thus is very modest. Clearly, efficiency considerations are not convincing enough to prefer OLS to GM, knowing that the former has a breakdown-point of 0%. On

Table 1.4 Simulation Results Concerning Efficiency of the HBP GM Estimator[a]

Leverage Points Added			Vertical Outliers Added			Leverage Points and Vertical Outliers Added		
OLS	LMS	GM	OLS	LMS	GM	OLS	LMS	GM
			5% Pollution					
0.200567	0.995782	0.998774	0.949351	0.998133	0.998666	0.196364	0.998468	0.998783
0.651876	0.041557	0.005486	0.032132	0.041833	0.005632	0.731822	0.041477	0.005538
0.012784	0.041539	0.005484	0.029566	0.041829	0.005630	0.085991	0.041474	0.005536
0.199080	0.994961	0.999903	0.953168	0.998214	1.000576	0.194158	0.999168	1.000575
0.654344	0.037870	0.005787	0.033310	0.038128	0.005697	0.739740	0.038582	0.005801
0.012872	0.037845	0.005787	0.031207	0.038125	0.005697	0.090358	0.038582	0.005800
			10% Pollution					
0.093830	0.991806	0.999576	0.898336	0.998771	0.998250	0.095341	0.996986	0.999447
0.823991	0.041783	0.005945	0.065004	0.041443	0.006232	0.866717	0.041690	0.005859
0.002847	0.041715	0.005945	0.054668	0.041441	0.006229	0.048310	0.041681	0.005859
0.093759	0.989552	0.999188	0.904547	0.997718	1.000113	0.095719	0.992524	1.000337
0.824204	0.038624	0.006159	0.065180	0.037663	0.006258	0.866904	0.038170	0.006165
0.002931	0.038515	0.006158	0.056068	0.037658	0.006258	0.049179	0.038114	0.006165
			15% Pollution					
0.057845	0.986834	0.999977	0.848411	0.999543	0.998582	0.057884	0.993765	1.000368
0.888980	0.041591	0.006472	0.107698	0.040655	0.007167	0.920049	0.040990	0.006512
0.001323	0.041417	0.006472	0.084718	0.040655	0.007165	0.032465	0.040951	0.006512
0.056900	0.984971	0.996509	0.847506	1.002215	0.997990	0.057706	0.991309	0.998740
0.890819	0.040720	0.006712	0.110691	0.038850	0.007157	0.920038	0.038975	0.006594
0.001383	0.040494	0.006700	0.087437	0.038845	0.007152	0.032120	0.038899	0.006592
			20% Pollution					
0.041390	0.980704	0.999659	0.804535	1.000447	0.998151	0.044029	0.989080	0.999559
0.919811	0.042421	0.006678	0.143066	0.041214	0.007736	0.938577	0.040697	0.006673
0.000877	0.042049	0.006678	0.104859	0.041213	0.007733	0.024696	0.040578	0.006673
0.040943	0.978557	0.995468	0.798557	0.996269	0.996417	0.041805	0.984524	0.997812
0.920667	0.040763	0.006809	0.149329	0.038140	0.007691	0.943166	0.038602	0.006821
0.000876	0.040303	0.006789	0.108750	0.038126	0.007678	0.025029	0.038362	0.006817
			25% Pollution					
0.030888	0.967603	0.998027	0.747312	0.998391	0.995502	0.031209	0.981777	0.998596
0.939851	0.048125	0.006839	0.190763	0.039194	0.008250	0.959140	0.042082	0.006932
0.000674	0.047076	0.006835	0.126912	0.039191	0.008230	0.020584	0.041750	0.006930
0.030397	0.963753	0.994769	0.749810	0.999797	0.997402	0.029549	0.982081	0.997499
0.940816	0.045688	0.007214	0.197492	0.037467	0.008487	0.961735	0.040121	0.007231
0.000687	0.044374	0.007186	0.134897	0.037467	0.008480	0.019959	0.039800	0.007225

[a] Each cell contains the mean, the mean squared error and the variance of the estimated parameter value over the 10,000 runs. Each first row concerns the intercept (β_1), each second row concerns the coefficient of the normally distributed explanatory variable (β_2) The length of the series equals 200 and $\beta_1=\beta_2=1$.

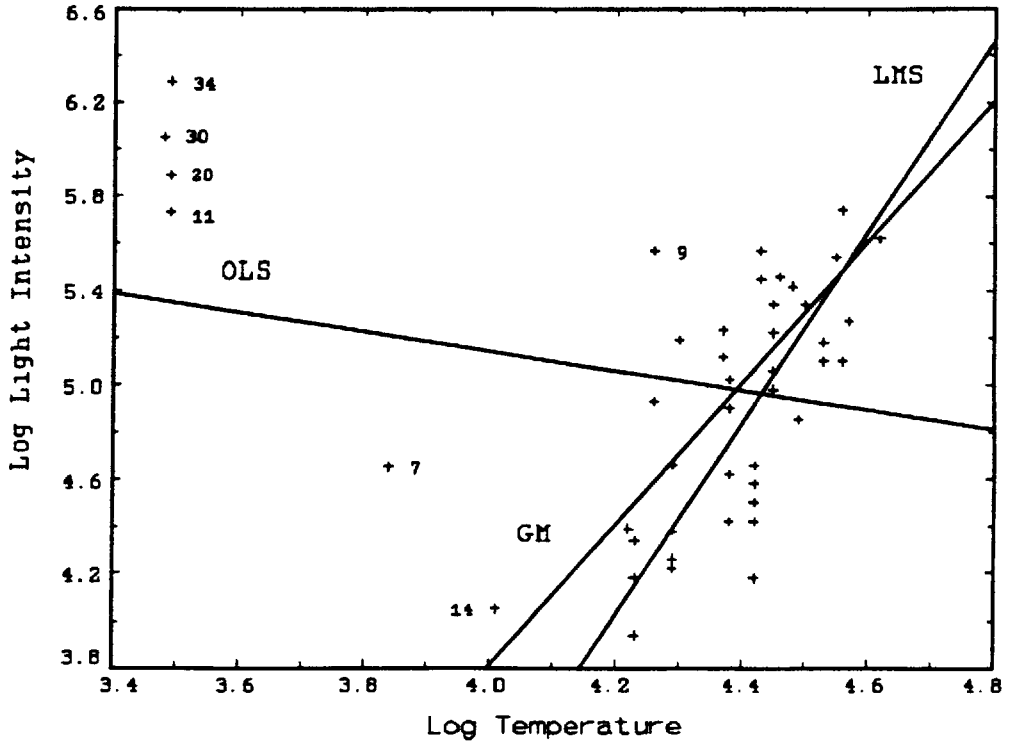
the other hand, Rousseeuw's LMS estimator performs less well in terms of efficiency. This is considered a setback, despite its appealing HBP. Note, however, that we need the LMS in order to compute the HBP GM estimator in the first place.

Next we successively corrupted the second column of the explanatory variables and the response variable by replacing (randomly drawn) δ% of the observations by random values drawn from a normal distribution with zero mean and variance 100. In terms of figure 1.1 we first added points like (d) and (b) in figure 1.1 seperately and then simultaneously. The last corruption may have involved adding observations like (c) in figure 1.1. With these polluted data we re-estimated β using all three estimators. This process was repeated 10,000 times for δ% equal to 5, 10, 15, 20 and 25, the results of which are summarized in Table 1.4. Indeed, we could go up to corrupt 50% of the data considering the breakdown point of both the LMS and HBP GM estimator. However, routine data are thought to contain 1%-10% gross errors (see Hampel et al. (1986), p.28). Clearly, Table 1.4 includes all interesting percentages of data contamination.

From Table 1.4 we can draw the following conclusions. A dramatic decay of OLS is noticed when data is contaminated. OLS is unable to produce reliable estimates in any of the cases considered, i.e. it breaks down completely. In terms of adequately estimating the unknown parameters, OLS is less sensitive to the construction of vertical outliers than to the construction of leverage points (this is also observed by Rousseeuw and Leroy (1987)), although applying this technique still results in considerable bias when vertical outliers corrupt the data. The explanation for this phenomenon lies in the way we corrupted the data. With respect to the model under which the original data are generated, constructing leverage points implies that these observations also become vertically outlying because the original slope parameter (β_2) differs from zero. The most influential observations in OLS regressions are characterized by a relatively high absolute value of the product of the concomitant explanatory variable and residual. Hence the stronger impact on the estimates of the constructed leverage points. On average LMS returns very reliable estimates, even if 25% of the data are corrupted in any way. In terms of precision however the HBP GM estimator is in most cases even superior to LMS.

Looking at the variance of the different estimates over the 10,000 runs we see that if only leverage points are constructed OLS is most efficient if 10% or more data contamination is involved (in case of only 5% data pollution by leverage points the HBP GM estimator is most efficient). This is however little consolation considering the bias in estimates this technique yields (even if only 5% of the data are replaced by leverage points OLS returns very unreliable estimates). In all other cases (i.e. when only vertical outliers or both leverage points and vertical outliers are constructed) the HBP GM estimator is most efficient. Moreover, under all types and fractions of data contamination the LMS technique is considerably less efficient than the HBP GM method. Finally, considering both efficiency and precision (i.e. comparing the mean squared errors) the HBP GM estimator is found to be superior to both OLS and LMS. Based on the statistical performance of the OLS, LMS and HBP GM estimators as revealed by Table 1.4, we conclude that the HBP GM estimation procedure is favoured.

Figure 1.3 Hertzsprung-Russel Diagram of the Star Cluster CYG OB1



Section 1.5 Stars, Light and Heat

To illustrate the importance of robust estimation when data contain outlying observations, we use an example from astronomy. In this field of science it is well known that the star cluster CYG OB1, which contains 47 stars in the direction of Cygnus, comprises four conspicuous stars, so called giants (see Rousseeuw and Leroy (1987)). The scatterplot of the logarithm of the effective temperature at the surface of a star (T_e) and the logarithm of its light intensity (L/L_0), the Hertzsprung-Russell diagram, for the star cluster CYG OB1, reveals that the celestial bodies can be divided into two groups: the majority of the stars, which are lying along a positively sloped band, and the four giant stars in the upper left corner (see Figure 1.3). Table 1.5 summarizes the data of star cluster CYG OB1 (observations 11, 20, 30 and 34 are giants).

OLS reveals a negative relation between the light intensity and the temperature of a star ($\hat{y}$ = 6.793 - 0.413x), although for 43 stars (which in astronomy are said to lie on the main sequence) this relation is positive. Clearly, the four giants trick OLS. The LMS estimate is insensitive to these huge objects and yields a line which fits the majority of observations properly ($\hat{y}$ = -12.964 + 4.046x

Table 1.5 Data for the Hertzsprung-Russell Diagram

Star Index	log T_e (x_i)	log(L/L_0) (y_i)	Star Index	log T_e (x_i)	log(L/L_0) (y_i)
1	4.37	5.23	25	4.38	5.02
2	4.56	5.74	26	4.42	4.66
3	4.26	4.93	27	4.29	4.66
4	4.56	5.74	28	4.38	4.90
5	4.30	5.19	29	4.22	4.39
6	4.46	5.46	30	3.48	6.05
7	3.84	4.65	31	4.38	4.42
8	4.57	5.27	32	4.56	5.10
9	4.26	5.57	33	4.45	5.22
10	4.37	5.12	34	3.49	6.29
11	3.49	5.73	35	4.23	4.34
12	4.43	5.45	36	4.62	5.62
13	4.48	5.42	37	4.53	5.10
14	4.01	4.05	38	4.45	5.22
15	4.29	4.26	39	4.53	5.18
16	4.42	4.58	40	4.43	5.57
17	4.23	3.94	41	4.38	4.62
18	4.42	4.18	42	4.45	5.06
19	4.23	4.18	43	4.50	5.34
20	3.49	5.89	44	4.45	5.34
21	4.29	4.38	45	4.55	5.54
22	4.29	4.22	46	4.45	4.98
23	4.42	4.42	47	4.42	4.50
24	4.49	4.85			

Source: Rousseeuw and Leroy (1987, p.27).

).[15] Also the HBP GM estimator ignores these observations. However, since the HBP GM estimator is more efficient than Rousseeuw's LMS, it uses more information contained in the data. In particular, observations 7 and 14 attract the HBP GM line ($\hat{y}$ = -7.132 + 2.741x). On the other hand, they cannot reverse the relation between light intensity and temperature, as exhibited by the majority of the stars.

[15] Rousseeuw and Leroy (1987) find a somewhat different LMS line ($\hat{y} = -12.298 + 3.898x$). This is due to the enormous increase in computer technology over the years which enables us to perform many more drawings to compute the LMS estimator. The difference in LMS estimates does not affect the main conclusions of this section.

Figure 1.4 Scatterplot of Standardized OLS Residuals Versus Mahalanobis Distances

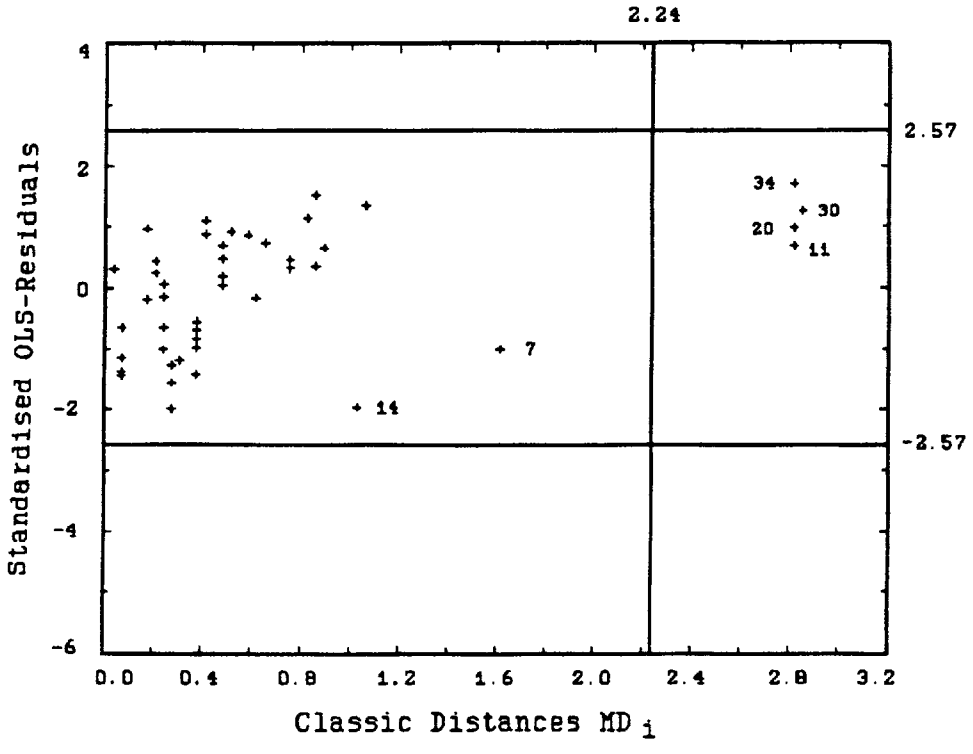


Table 1.6 reports the estimation results. It is striking to see that the OLS estimate is not only completely wrong (both coefficients have the wrong sign), but is also insignificant for the main explanatory variable. A well known phenomenon in astronomy is not revealed by this estimation technique. The LMS and GM estimators find the true, positive, relation, whereas the latter technique leaves no doubt as to the significance of the relation between the light intensity and the temperature of a star.

The scatterplot of standardized residuals versus x-distances is an aid in visualizing leverage points and vertical outliers (see Rousseeuw and Van Zomeren (1990)). For our regression results with OLS, LMS and HBP GM these plots are depicted in Figure 1.4, 1.5 and 1.6 respectively. Leverage points are defined as observations with a x-distance exceeding the square root of the 97.5th percentile of the $\chi^2(p)$-distribution.[16] If p is 1, this critical value is 2.24. In terms of Figures 1.4, 1.5 and 1.6, points to the right of the vertical line with x-distance 2.24 are called leverage points. The standardized residuals are evaluated with a standard normal distribution. For this probability density function, the

[16] See Section 1.2.2.

Table 1.6 Estimation Results, Hertzsprung-Russell Data[a]

Dependent Variable: Log Light Intensity			
	OLS	LMS	HBP GM
Constant	6.793	-12.964	-7.132
	(1.237)		(3.023)
Log Temperature	-0.413	4.046	2.741
	(-0.286)		(0.680)
$\bar{R}^2$	0.023		0.98
$\hat{\sigma}$	0.552	0.368[b]	0.440[c]

[a] Standard errors are within parentheses.
[b] See Appendix A1 for a description of this estimator.
[c] See Section 1.4.1 for a description of this estimator.

Table 1.7 Outliers in the CYG OB1 Data Identified by LMS and HBP GM

Estimation Technique	Leverage Points	Vertical Outlier	Leverage Point and Vertical Outlier	Weights based on x-distances $w_x(x_i)$	Weights based on residuals $w_r(r_i)$
			7	0.695	0.220
			11	0.430	0
HBP GM	14			0.991	0.976
			20	0.430	0
			30	0.430	0
			34	0.430	0
			7		
		9			
LMS			11		
	14				
			20		
			30		
			34		

Figure 1.5 Scatterplot of Standardized LMS Residuals Versus MVE Distances

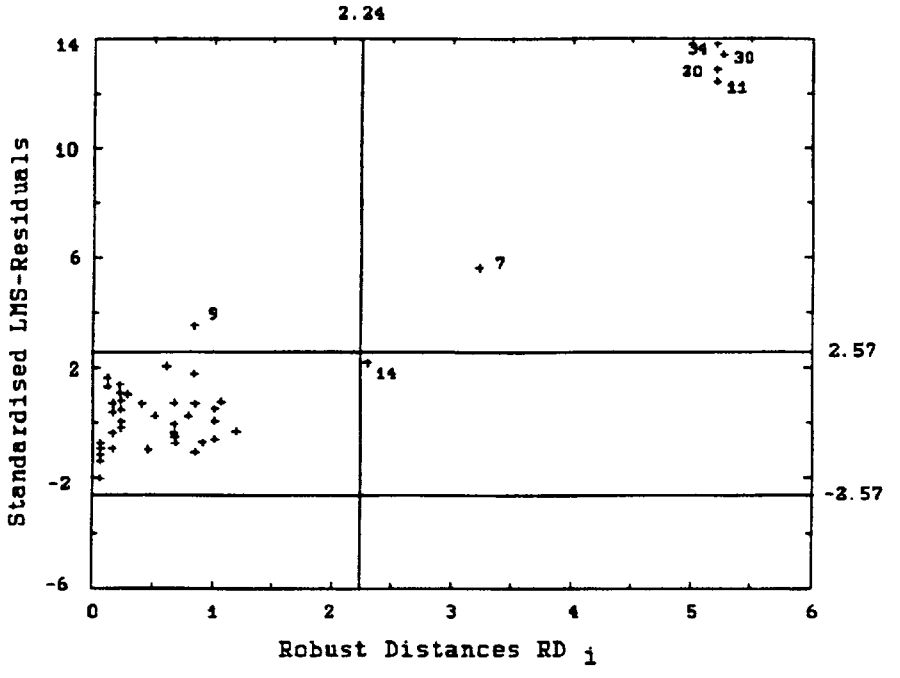
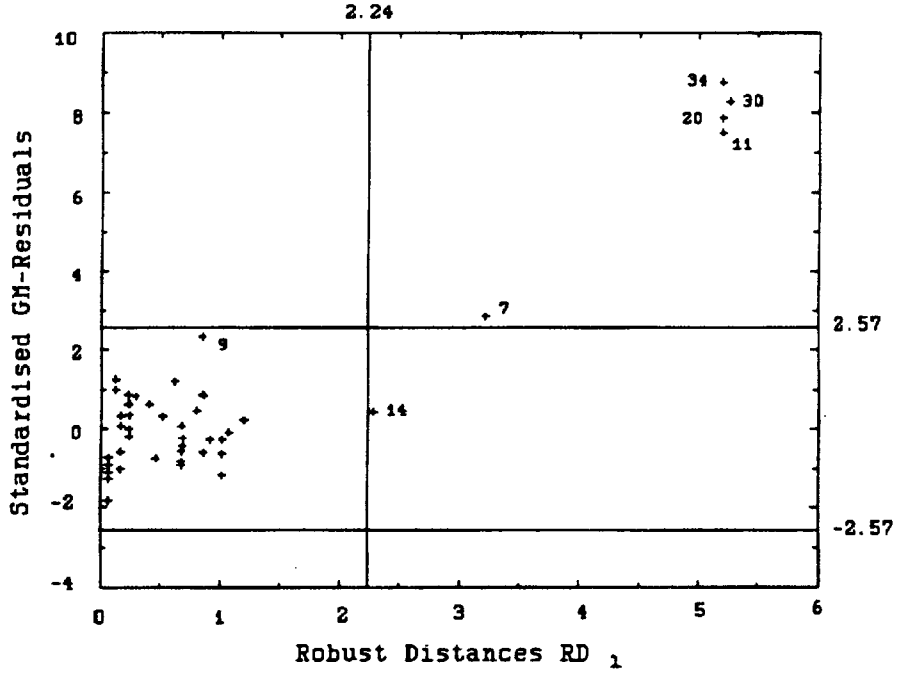


Figure 1.6 Scatterplot of Standardized HBP GM Residuals Versus MVE Distances



absolute value of 99% of the elements of its domain is less than 2.57. Points lying outside the strip indicated by this cutoff value are defined as vertically outlying. The link between Figures 1.4, 1.5 and 1.6 and Figure 1.1 is that points in the mid west areas of the former figures coincide with points (a) in the latter, points in the mid east are like (c), points in the north east area conform with (d) and points in the north west area correspond to (b).

According to the HBP GM and LMS estimates, five observations, including the four giants, are both a leverage point and a vertical outlier (see also Table 1.7). Star 14 is only a leverage point. LMS also identifies observation 9 as vertically outlying. On the other hand, OLS only recognizes the giant stars as leverage points but not as vertical outliers. Note that this implies that, according to the OLS result, the four giants are not characterized by a relation between light intensity and temperature which differs considerably from that associated with the majority of the data (compare Figure 1.1, in which leverage points (c) are in line with the regular observations (a)). Moreover, star 7 and 14 are not identified as horizontally outlying. The inability of the classical (Mahalanobis) x-distance to reveal these two observations as leverage points is due to the masking effect. The four giants corrupt the Mahalanobis measure such that mild leverage points remain undiscovered.

Section 1.6 Conclusions

The success of OLS and its generalizations in applied economic research is not justified by its performance on contaminated (real) data sets. If leverage points and/or vertical outliers are present, OLS not only produces unreliable parameter estimates but also incorrect t-values. Especially in economics where data are, in many cases, far from smooth, the use of OLS can lead to seriously wrong conclusions.

In this chapter we have described a reliable alternative, a Generalized M-estimator for the linear model, which has a high breakdown point, is consistent, has asymptotically a normal distribution and, as our simulation results indicate, is more efficient than Rousseeuw's LMS estimator. A Monte Carlo study reveals that the associated t-values are accurate in large samples but tend to reject the null-hypothesis (i.e. the unknown parameter is zero) in too many cases of the small sample experiment. An example from the astronomy illustrates the effects of outliers in a real data set and shows how these anomalous points can be highlighted with graphical devices.

Either the resampling or projection algorithm is used in the construction of the HBP GM estimator. We have proposed a correction factor which, according to our simulation results, improves both algorithms for estimation of the location and scatter of the design matrix.

In the subsequent two chapters we will present some robust specification tests and present solutions for the specific problems stemming from robust estimation of panel data.

Appendix A1 Least Median of Squares Regression

Rousseeuw (1984) introduced a simple yet elegant and robust estimator for the unknown parameter vector β in the classical linear model. The objective of the estimator is to minimize the median of the squared residuals:

$$\min_{\beta} med(y_i - x_i\beta)^2. \tag{A1.1}$$

The breakdown point of the resulting Least Median of Squares (LMS) estimator equals $([n/2]-p+2)/n$ for n observations and p explanatory variables which is as high as 50% when n goes to infinity (the notation [r] stands for the largest integer less than or equal to r).

To approximate (A1.1) we employ the resampling algorithm. This algorithm starts with the application of the OLS estimator on a randomly drawn sample of size p. Evidently, the resulting regression line perfectly fits the sub sample. Given this trial estimate, β_{LMS}, the objective value $med(y_i - x_i\beta_{LMS})^2$ is calculated with respect to the complete data set. A refinement for the intercept estimate is obtained by computing the midpoint of the shortest half (see sub-section 1.3.1 for a definition) of the numbers $y_i - x_i^*\beta^*_{LMS}$, $i=1,..,n$ where $x_i^* = x_{i,2}..x_{i,p}$ and $\beta^*_{LMS} = (\beta_{LMS_2},...,\beta_{LMS_p})^T$ (see Rousseeuw and Leroy (1987), p.201). We then proceed with the next drawing. The estimate with the lowest objective value is kept as the approximation of the LMS estimate.

If all distinct sub samples were to be enumerated, there were far too many estimates to be made for data sets of reasonable size. However Rousseeuw and Leroy (1987, p.198) show that the probability of getting a sub sample consisting of p non-outlying observations when the fraction of contamination equals δ is

$$1-(1-(1-\delta)^p)^m, \tag{A1.2}$$

where m is the required number of independent sub samples. Rewriting (A1.2) gives the number of drawings needed to get a *good* sub sample with probability Λ:

$$m = [\ln(1-\Lambda)/\ln(1-(1-\delta)^p)]. \tag{A1.3}$$

In our GAUSS program Λ is set equal to 0.99 and δ is set equal to 0.5. When approximating the LMS estimator we, however, used 10 times as many trials as prescribed by (A1.3).

The scale estimate associated with the LMS technique is the result of a two stage procedure (see Rousseeuw and Leroy (1987), p.202). An initial estimate is made according to

$$s^0 = (1+5/(n-p))\sqrt{med(r_i^2)}/0.6745, \tag{A1.4}$$

where $r_i = y_i - x_i\beta_{LMS}$. Given s^0, weights are determined according to

$$w_i = 1 \quad \textit{if } |r_i/s^0| \leq 2.5$$
$$= 0 \quad \textit{otherwise.} \tag{A1.5}$$

The second step yields the estimate of the standard deviation of the errors:

$$s_{LMS} = \sqrt{(\sum_{i=1}^{n} w_i r_i^2)/(\sum_{i=1}^{n} w_i - p)} \; . \tag{A1.6}$$

We use (A1.6) as scaling factor in Tukey's bi-square function (1.12).

Wagenvoort, Rien J.L.M. (1998), Robust estimation of panel data : with an application to investment equations
European University Institute DOI: 10.2870/75660

CHAPTER 2

B-Robust Estimation and Testing of the SUR Model With Fixed Effects and Heteroskedasticity

Section 2.1 Introduction

Panel data sets are constructed by the pooling of cross-sections and time series. These so-called longitudinal data sets enable researchers: (1) to draw conclusions from more reliable statistical inference as a result of an increase in the number of observations in comparison with a single cross-section or a single time series, (2) to incorporate individual/firm specific fixed or random effects in the regression model, (3) to assign each cross-sectional unit a different coefficient vector and (4) to model dynamics. The High Breakdown Point GM estimator as described in Chapter 1 is particulary suitable for the estimation of models using cross-sectional data (see Section 1.2.4). When estimating panel data models with our HBP GM estimator we may encounter several econometric issues including heteroskedastic regression errors due to heterogeneity among cross-sectional units, autocorrelated errors and also fixed effects which are correlated with the explanatory variables. In this chapter we present some robust diagnostic tests and B-robust regression methods for panel data to detect and to capture the aforementioned problems. These procedures are founded on the HBP GM estimation method but use also features of other well known estimators in order to gain efficiency.

Multiplicative heteroskedasticity and the occurence of outlying observations are not necessarily different names for the same phenomenon in a regression model. Smooth data, i.e. data which do not contain any outliers, might adequately fit a regression model with heteroskedastic errors. Neither is it the case that just a few (vertical) outlying observations in the data imply heteroskedasticity. On the other hand, large vertical outliers might coincide with large leverage points. In this case, we cannot distinguish between multiplicative heteroskedasticity and outlying observations. To test for heteroskedasticity we derive a robust version of the Pagan and Vella (1989) test associated with the HBP GM estimator. Monte Carlo results indeed show that heteroskedasticity is detected by the robust Pagan and Vella test when the generated homoskedastic regression model is polluted with leverage points. As shown by Simpson, Ruppert and Carroll (1992), the HBP GM estimator still produces root-n consistent estimates if there are heteroskedastic symmetric regression errors provided that the Newton-Raphson algorithm is used to solve the first

order condition (see also Section 1.4.3). We show by means of a simulation experiment that the HBP GM estimator extended with Harvey's (1976) two-step improvement, which corrects for multiplicative heteroskedasticity, is more efficient than the *standard* HBP GM estimator in a generated heteroskedastic regression model.

It is highly conceivable that regression errors are correlated over time in panel data models. For instance, it is not inconceivable that the errors belonging to a certain cross-sectional unit are all positive for every period whereas the opposite applies to another unit. Consider a panel data set with a small number of time periods but which is large in the number n of cross-sectional units. Here we might apply the Seemingly Unrelated Regression (SUR) model (see Zellner (1962)) with one equation for each time period. Peracchi (1991) develops an optimal bounded influence estimator for the SUR model. This estimator is a specific case of the general solution for B-robust estimation of panel data models proposed by Krishnakumar (1995). Peracchi (1991) derives the normal equations which define the B-robust estimator of the unknown components of the variance-covariance matrix (Σ) of the model error terms. After choosing the bounds on the influence functions of the estimators for respectively β and Σ, optimal solutions in the sense of estimators which have minimum asymptotic mean square error can be computed. We develop a similar B-robust estimator for the SUR model. However, instead of solving the first order condition associated with Σ and computing new accompanying weights, we use the regression residuals and corresponding weights of the first step HBP GM regression to determine a possible transformation of the observations in order to correct for autocorrelation. Then in the second step we proceed with the computation of the robust SUR (RSUR) estimates, which are considerably more efficient than the standard HBP GM estimates if autocorrelation occurs as is shown by a Monte Carlo study.

It is well known that the Generalized Least Squares (GLS) Estimator is inconsistent if the fixed effects are correlated with the independent variables and left aside as part of the error terms. Among a number of different solutions which have been proposed to eliminate these individual specific intercepts are: estimating in the *within* dimension of the data by calculating deviations around individual means or differencing the data over the time periods. For example, first differences are calculated by subtracting the last period value (first lag) from each current value of the cross-sectional observation. Although the GLS estimator is most efficient and consistent under the null hypothesis of uncorrelated fixed effects, the *difference* and *within* estimators generate consistent estimates even when the fixed effects are correlated. Griliches and Hausman (1986) in turn show that the reason for finding biased and significantly different *within* and *difference* estimates is possibly due to measurement errors in the explanatory variables. They propose an estimation strategy for panel data with errors in variables which is based on the Generalized Method of Moments estimator. In Chapter 3 we develop an alternative robust GMM estimator which continues to generate unbiased estimates when the panel data set with correlated fixed effects is contaminated (see Section 1.1 for different types of data pollution). In this case, the standard GMM estimator breaks down.

In short, Chapter 2 is organized as follows. After disclosing the panel data model with its underlying assumptions in section 2.2 we present a specification test on correlated fixed effects and a specification test on cross-sectional correlation in sections 2.3 and 2.4 respectively. In Section 2.5 we show how to extend the standard HBP GM estimator with Harvey's two-step improvement and derive the robust Pagan and Vella test. Both the robust testing and the estimation procedures concerning heteroskedasticity are evaluated by means of a Monte Carlo study in section 2.5.1 and 2.5.3 respectively. The subsequent section contains the generalized Durbin-Watson statistic of Bhargava, Franzini and Narendranathan (1982), the derivation of our B-robust SUR estimator and its variance-covariance matrix and the simulation results concerning autocorrelation. Finally, section 2.7 concludes Chapter 2.

Section 2.2 Model Assumptions

Consider the linear panel data model

$$y_{ti} = \eta_i + \boldsymbol{x}_{ti}\beta + \varepsilon_{ti}, \tag{2.1}$$

where ε_{ti} is the error term corresponding to cross-sectional unit i in period t and η_i, i=1,..,n are the fixed effects. y_{ti} is an observable dependent variable and $\boldsymbol{x}_{ti}$ is a row vector of length p of observable explanatory variables. There are T periods and n cross-sectional units, adding up to N observations. The fixed effects are eliminated from equation (2.1) by estimating either in *first differences* ($\hat{\beta}^{D}_{GM}$) or in the *within* ($\hat{\beta}^{W}_{GM}$) dimension of the data.

The model errors are assumed to exhibit the following properties. First, the associated distribution function F has a density symmetric around zero. Second, ε_{ti} is assumed to be orthogonal to $\boldsymbol{x}_{ti}$. In other words, the explanatory variables are treated as fixed or predetermined. Third, the n errors ε_{ti} t=s, i=1,..,n within one time period s (one equation) are assumed to be independent. Therefore, we use the Seemingly Unrelated Regression (SUR) model in a somewhat unusual way.

> The distinguishing features of Seemingly Unrelated Regressions as a method for pooling time-series and cross-sectional data are contemporaneous correlation in the disturbances and the assumption that each cross-sectional unit has a different coefficient vector (see Judge et al. (1988), p.444).

In contrast with this definition we assume that the parameters of the linear model are equal for each cross-sectional unit (model (2.1)) and there is zero cross-sectional correlation (see (*A*.3) below). However, we allow disturbances within each time-series to be correlated. Consequently, the following assumptions are made on the covariance structure of the error terms of the panel data

model (2.1):

(*A*.1) $E[\varepsilon_{it}]=0,\ i=1,..,n,\ t=1,..,T,$

(*A*.2) $E[\varepsilon_{it}x_{it}]=0,$

(*A*.3) $E[\varepsilon_{it}\varepsilon_{tj}]=0,\ if\ i\neq j,$

(*A*.4) $E[\varepsilon_{it}\varepsilon_{si}] = \sigma_{ts}.$

In the literature on panel data, (2.1) together with assumptions (*A*.1)-(*A*.4) is called the *covariance model* (see Mátyás and Sevestre (1992) p. 26).

Section 2.3 Testing For Correlated Fixed Effects

Rewriting the panel data model (2.1) as

$$y_{it} = x_{it}\beta + \vartheta_{it} \tag{2.2}$$

where $\vartheta_{it} = \varepsilon_{it} + \eta_i$, shows that Generalized Least Squares estimation of (2.2) generates inconsistent estimates if the fixed effects η_i are correlated with the explanatory variables. In this case the disturbances ϑ_{it} are not perpendicular to the independent variables. Since the *within* estimator does not exhibit this inconsistency, one can test for correlated fixed effects by constructing a Wu (1973) - Hausman (1978) test statistic of the form

$$m = (\beta_{GLS}-\beta^{W}_{LS})^{T}(cov(\beta_{GLS})-cov(\beta^{W}_{LS}))^{-1}(\beta_{GLS}-\beta^{W}_{LS}). \tag{2.3}$$

The Wu-Hausman test compares, under the null-hypothesis of correct model specification, a consistent and efficient estimator with an alternative consistent estimator which does not necessarily attain asymptotically the Cramer-Rao bound. In case of misspecification, the first estimator is inconsistent but the latter one has to be consistent. Under the null hypothesis of non-correlated fixed effects, m is asymptotically chi-squared distributed with p degrees of freedom.

Metcalf (1996) discusses powerful Wu-Hausman tests in panel data models with endogenous regressors. He shows that computing m from (2.3), where the GLS and within

estimators are replaced by their instrumental variable counterparts, is equivalent to calculating:

$$m=(\beta_W^{IV}-\beta_B^{IV})^T(cov(\beta_W^{IV})-cov(\beta_B^{IV}))^{-1}(\beta_W^{IV}-\beta_B^{IV}), \tag{2.4}$$

where β_B^{IV} are the instrumental variable estimates obtained with the *between* estimator. *Between* estimation boils down to computing time-averages for each cross-sectional unit and regressing the average value of the independent variable on the average value of the dependent variables. A remarkable result is that the more powerful test is not necessarily constructed with the more efficient GLS estimator. It can be advisable to use an incomplete set of instruments. See Metcalf (1996) for an explanation and details on the choice of the instruments.

In the application discussed in Chapter 5, we will test for the need to add fixed effects to the panel data model by comparing the HBP GM (RSUR) estimates of the model

$$y_{it} = c + x_{it}\beta + \varepsilon_{it} \tag{2.5}$$

with the HBP GM (RSUR) within or first difference counterparts using the Wu-Hausman statistic. c is an unknown intercept. If the null hypothesis $E[\eta_i|x_i]=c$ holds then the OLS (SUR) estimator (in levels) is consistent and asymptotically attains the Cramer-Rao bound. The HBP GM (RSUR) estimator however attains at the most 95% of the efficiency level of the OLS (SUR) estimator. We therefore foresee that the computed Wu-Hausman statistic only approximates the χ^2-distribution. Estimating (2.5) in levels returns biased coefficients under the alternative hypothesis that the fixed components η_i are not constant and correlated with the explanatory factors.

Section 2.4 Testing For Cross-sectional Independence

The Lagrange Multiplier test statistic due to Breusch and Pagan (1980)

$$C_{AVE}^2 = \binom{n}{2}^{-1}\sum_{i<j} c_{ij}^2 \tag{2.6}$$

is one measure to test the null hypothesis of zero cross-sectional correlation; c_{ij} is the Pearson correlation coefficient between the ith and jth cross-sectional units. The standardized version, $C_{STD} = n((T-1)C_{AVE}^2-1)/2$ has an asymptotic standard normal distribution. This desirable asympototic result can be shown to hold when the number of time periods T grows to infinity. Thus, the Breusch and Pagan test is especially useful in panels where T is relatively large with respect to n. Throughout this monograph, however, we consider panel data sets which are large in n but relatively small in T. It is a fact that C_{STD} is not distribution free if T remains fixed, even when the number of cross-sectional units grows to infinity. This is seen as a serious drawback.

Frees (1995) introduces a non-parametric version of C_{AVE}^2 which does not depend on the

parent distribution of ε_{ti} and has good power in finite samples[1],

$$R^2_{AVE} = \binom{n}{2}^{-1}\sum_{i<j} r^2_{ij}, \tag{2.7}$$

where r_{ij} is the Spearman rank correlation coefficient between the ith and jth cross-sectional units. The statistic (2.7) can be approximated with

$$R^{2\,*}_{AVE} = (n(n-1))^{-1}\sum_{t=1}^{T}\sum_{u=1}^{T}\left(\left(\sum_{i=1}^{n} Z_{i,t,u}\right)^2 - \sum_{i=1}^{n} Z^2_{i,t,u}\right) \tag{2.8}$$

Frees (1995) shows that $n(R^{2\,*}_{AVE}-(T-1)^{-1})$ converges in distribution to Q when $n\to\infty$, under the null hypothesis of *iid* errors with;[2]

$$Z_{i,t,u} = \frac{12\left(R_{it}-\frac{T+1}{2}\right)\left(R_{iu}-\frac{T+1}{2}\right)}{T^3-T}, \tag{2.9}$$

$$Q = a(T)(\chi^2(T-1)-(T-1)) + b(T)(\chi^2(T(T-3)/2)-(T(T-3)/2)), \tag{2.10}$$

$$a(T) = \frac{4(T+2)}{5(T-1)^2(T+1)},\quad b(T) = \frac{2(5T+6)}{5T(T-1)(T+1)} \tag{2.11}$$

where $(R_{i,1},...,R_{i,T})$ are the ranks of $(\varepsilon_{1i},...,\varepsilon_{Ti})$; $\chi^2(T-1)$ and $\chi^2(T(T-3)/2)$ are independent chi-squared random variables with $T-1$ and $T(T-3)/2$ degrees of freedom respectively. The lower bound of Q and an upper critical value of the Frees test can be established, given the number of time periods which must exceed three. We reject the null hypothesis of zero cross-sectional correlation if $R^{2\,*}_{AVE}$ is less than this lower bound or exceeds the upper 95th percentile of Q, the weighted sum of two χ^2-distributions.

The Frees (1995) test would be vulnerable to outliers if we compute $R^{2\,*}_{AVE}$ on the basis of the HBP GM residuals $r_{ti} = y_{ti}-x_{ti}\hat{\beta}_{GM}$. Instead a robust Frees test is achieved by employing the weighted residuals

$$rw_{ti} = r_{ti}\sqrt{w_x(x_{ti})w_r(r_{ti})}\ . \tag{2.12}$$

---

[1] If either negative or positive correlations occur, Frees (1995) recommends using a one-sided test. The statistic R^2_{AVE} is especially powerful when (1) *a priori* it is unclear whether positive or negative correlation prevail or (2) a mixture of negative and positive correlations occurs.

[2] It is assumed that the errors ε_{ti} have a continuous distribution.

Notice that these weighted residuals rw_{ti} are different from the ones which are obtained by multiplying the residuals by the weights associated with Tukey's bi-square function $\psi(r_{ti}) = r_{ti}w_r(r_{ti})$ (see section 1.2.3). The choice of (2.12) follows naturally once one realizes that computation of the HBP GM estimator is equivalent to applying Ordinary Least Squares to the weighted data (see formula (1.14) of section 1.2.4).

Section 2.5 Heteroskedasticity

2.5.1 The Pagan and Vella Diagnostic Test on Heteroskedasticity

To test the null hypothesis of homoskedastic errors in (2.1) a robust Pagan and Vella (1989) test statistic is derived. The Pagan-Vella test on heteroskedasticity evaluates the sample analogue of

$$E[v_{ti}(\varepsilon_{ti}^2 - \sigma^2)] = 0, \tag{2.13}$$

where the errors ε_{ti} have constant variance σ^2 that is unrelated to the variables v_{ti} under the null hypothesis. Below we will indicate the necessary ingredients to compute the test statistic which corresponds to the HBP GM regression. In this section autocorrelation as set down in assumption (A.4) and the fixed effects are neglected, i.e. $E[\varepsilon_{ti}\varepsilon_{si}]=\sigma_{ts}=0$, $t\neq s$ and $\eta_i=0$, $i=1,...,n$.

First, we start with scaling model (2.1) with the robust Median Absolute Deviation (MAD) estimate of the scale parameter σ (see section 1.4.1). Then, we choose v_{ti} equal to the square of the explanatory variables x_{ti}, $v_{ti}=x_{ti}^2$ and define[3]

$$m(q_{ti},\omega) = v_{ti}^T w_x(x_{ti})^2\Phi(\varepsilon_{ti}), \tag{2.14}$$

$$g(q_{ti},\omega) = x_{ti}^T w_x(x_{ti})\psi(\varepsilon_{ti}), \tag{2.15}$$

where $\omega=(\beta^T,\sigma^2)^T$, $\Phi(\varepsilon_{ti})=(\psi(\varepsilon_{ti}))^2-E[(\psi(\varepsilon_{ti}))^2]$, $\psi(r_{ti})=r_{ti}w_r(r_{ti})$, $q_{ti}=(y_{ti},x_{ti},v_{ti})$, $r_{ti}=y_{ti}-x_{ti}\hat{\beta}_{GM}$ and

$$\beta_{GM} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n} x_{ti}^T w_x(x_{ti}) w_r(r_{ti}) y_{ti}}{\sum_{t=1}^{T}\sum_{i=1}^{n} x_{ti}^T w_x(x_{ti}) w_r(r_{ti}) x_{ti}}. \tag{2.16}$$

The weight functions $w_x(x_{ti})$ and $w_r(r_{ti})$ are based on the identification of leverage points and

[3] We note that v_{ti} should not contain an intercept term.

vertical outliers respectively. These are discussed at length in section 1.2.2 and section 1.2.3. v_{ti}^T is the transpose of v_{ti}.

Using a Taylor approximation of $g(q_{ti},\omega)$ we can write

$$\hat{\omega} - \omega \approx -B^{-1}\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{n} x_{ti}^T w_x(x_{ti})\psi(r_{ti}), \tag{2.17}$$

where $\hat{\omega}$ is the robust estimate of ω, and

$$B = -\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{n} x_{ti}^T w_x(x_{ti})\frac{\partial\psi(r_{ti})}{\partial\beta}. \tag{2.18}$$

Now, Pagan and Vella (1989) show that $\sqrt{N}\hat{\imath}$ has an asymptotic normal distribution with zero mean under the null hypothesis, where

$$\hat{\imath} = \frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{n} m(q_{ti},\hat{\omega}). \tag{2.19}$$

Using (2.17) and a Taylor approximation of $\sqrt{N}\hat{\imath}$ we derive

$$\sqrt{N}\hat{\imath} \approx \sqrt{N}\left(\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{n} m(q_{ti},\omega) + A(\hat{\omega} - \omega)\right) = \frac{1}{\sqrt{N}}\left(I_p \mid -A(B^{-1}) \right)\begin{pmatrix}\sum_{t=1}^{T}\sum_{i=1}^{n} m(q_{ti},\omega) \\ \sum_{t=1}^{T}\sum_{i=1}^{n} g(q_{ti},\omega)\end{pmatrix} \tag{2.20}$$

where

$$A = \frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{n} v_{ti}^T w_x(x_{ti})^2\frac{\partial\Phi(r_{ti})}{\partial\beta} \tag{2.21}$$

and I_p is the identity matrix of dimension p. From (2.20) follows that the covariance matrix of $\sqrt{n}\,\hat{\imath}$ can be written as

$$V = \left(I_p \mid -A(B^{-1}) \right)\left(\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{n}\begin{pmatrix} m(q_{ti},\omega) \\ g(q_{ti},\omega)\end{pmatrix}\begin{pmatrix} m(q_{ti},\omega) & g(q_{ti},\omega)\end{pmatrix}\right)\left(I_p \mid -A(B^{-1}) \right)^T. \tag{2.22}$$

Under the null hypothesis, the test statistic

$$N\hat{\imath}^T\hat{V}^{-1}\hat{\imath} \tag{2.23}$$

is asymptotically $\chi^2(p)$ distributed where p is equal to the dimension of v_{ti} and the sample analogue of (2.22), $\hat{V}$, is a consistent estimate of V.

To examine the power properties of the robust Pagan-Vella test in clean and contaminated data sets we compute the test statistic (2.23) for the HBP GM regressions of y on x_1 in the course

Table 2.1 Simulation Results Concerning Power Properties of the Pagan-Vella Test on Heteroskedasticity[a]

	N=200		N=1000	
Pollution	95th Percentile	p-value	95th percentile	p-value
		Standard		
0%	3.520	0.0414	2.796	0.0244
		Robust		
0%	2.411	0.0158	2.313	0.0125
5%	2.437	0.0113	4.850	0.0980
10%	3.379	0.0333	9.027	0.4025
15%	4.204	0.0685	13.469	0.7059
20%	5.250	0.1303	18.043	0.8484
25%	5.950	0.1817	21.976	0.9043

[a] The number of replications is 10,000 for both series of length N=200 and N=1000. Under the null-hypothesis of homoskedasticity the test-statistic has a chi-squared distribution with 1 degree of freedom, $\chi^2_{0.05}(1) = 3.84146$.

of which the generated data sets of the Monte Carlo experiments of section 1.4.1 are used. Table 2.1 shows the p-values and the 95-th percentiles corresponding to the series of length ten thousand containing the Pagan-Vella statistics for both small and large samples consisting of 200 and 1000 observations respectively. A striking observation is the high p-values for the experiments with the polluted data sets. Although on average the regression residuals are not heteroskedastic, the null hypothesis of homoskedasticity is rejected many times. This result is explained by the way the outliers are constructed. That is, the contaminated data sets contain many leverage points which are also vertically outlying. These points cause the robust Pagan-Vella statistic to rightly reject the homoskedasticity hypothesis. Heteroskedasticity and the occurence of outliers are for some data sets two sides of the same coin. With this end in view it could be deduced from Table 2.1 that the robust Pagan-Vella statistic has less power in the small sample cases in comparison with the large sample cases. For example, if 15% of the data is polluted with possibly heteroskedastic observations then the p-value of 0.0685 associated with the small sample experiment just exceeds the 5% significance level whereas in case of the large sample experiment the null hypothesis is more than 7000 times (p-value equal to 0.7059) rejected.

The purpose of application of the Pagan-Vella test is to inquire whether efficiency gains can be obtained through correction for multiplicative heteroskedasticity using for instance Harvey's (1976) method. If the model errors are homoskedastic and the data sets are free of outlying

observations then the standard HBP GM estimator is more efficient than the HBP GM estimator extended with Harvey's two-step improvement (see the next sub-section for an exposition of this procedure). In this case, the robust Pagan-Vella test gives the right signal, namely, it will not reject homoskedasticity. Furthermore, if the majority of the errors are heteroskedastic then the Pagan-Vella statistic will successfully trigger correction for heteroskedasticity, irrespective of whether outliers are present. However, there is also the other possibility that the null hypothesis is rejected because outliers corrupt the data although the majority of the observations fit the homoskedastic model. Simulation results, not shown here, indicate that in this case the standard HBP GM estimator is superior to the adjusted HBP GM estimator in terms of efficiency. Consequently, the Pagan-Vella test may provide the wrong signal, even if some part of the data fit the heteroskedastic model. We conclude that the robust Pagan-Vella test is a reliable indicator of heteroskedasticity. The test decisively indicates the correction for heteroskedasticity if homoskedasticity of the disturbances is rejected but the data is clean or polluted with relatively few outiers. If a substantial portion of the data is contaminated with outliers and the test rejects the null hypothesis then further investigation of the data is required before adjusting for multiplicative heteroskedasticity.

2.5.2 Correction For Multiplicative Heteroskedasticity

Although the HBP GM estimator (2.16) returns heteroskedastic consistent estimates of the parameters of model (2.1) with heteroskedastic consistent standard errors, a more efficient estimator can be acquired by following Harvey (1976) if we are willing to assume that

$$E[\varepsilon_{ti}^2] = \sigma_{ti}^2 = \exp(v_{ti}\alpha), \tag{2.24}$$

where α is an unknown parameter vector and v_{ti} is a vector of observable explanatory variables of the variance equation. The first element of the k dimensional row vector v_{ti} is a constant and we choose the other components equal to the absolute value of the regressors x_{ti}, $v_{ti}=(1, |x_{ti}|)$, assuming that the explanatory factors X do not include a column of ones. The standard Harvey (1976) procedure computes weighted least squares estimates by means of a second step LS after a first step in which each cross-sectional observation is scaled by the concomitant estimated standard deviation.

The unknown parameter vector α is estimated with the help of a second HBP GM regression. The natural logarithm of the square of the first step HBP GM residuals r_{ti} is regressed on v_{ti}

$$LN(r_{ti}^2) = v_{ti}\alpha + \xi_{ti} \tag{2.25}$$

where ξ_{ti} is an error term. Notice that the size of the unweighted residual r_{ti}, which determines the

extent to which the observation in the original equation is vertically outlying, together with the explanatory variables v_{it} may be outlying in view of the plane determined by the variance equation. Therefore, new weights $w_v(v_{it})$ and $w_\xi(\xi)$ are determined. The HBP GM estimator $\hat{\alpha}_{GM}$ returns an initial inconsistent estimate of the constant term of the variance equation. The inconsistency of the first element of $\hat{\alpha}_{GM}$ follows from the fact that ξ_{it} converges in distribution to a variable ξ_{it}^* which is distributed as the logarithm of a χ^2 variate with one degree of freedom and mean, $E[\xi_{it}^*]=-1.2704$. Therefore, we add 1.2740 to $\hat{\alpha}_1$

$$\hat{\alpha}_{GM,1} = \hat{\alpha}_{GM,1} + 1.2704. \tag{2.26}$$

Then we proceed with the computation of the weights

$$w_s(v_{it};r_{it}) = \left(\sqrt{\exp(v_{it}\hat{\alpha}_{GM})}\right)^{-1} \tag{2.27}$$

based on the identification of multiplicative heteroskedasticity.

During the second step both sides of the original equation are multiplied by (2.27)

$$y_{it}w_s(v_{it};r_{it}) = x_{it}w_s(v_{it};r_{it})\beta + \varepsilon_{it}w_s(v_{it};r_{it}) \tag{2.28}$$

and the HBP GM method is applied once more to estimate β in (2.28). After this scaling, some observations which were traced as outliers in the first HBP GM regression may have disappeared as deviating points whereas regular observations in the first regression may show up as outliers in the final third robust regression. The formula for the HBP GM estimator extended with Harvey's two-step improvement reads:

$$\beta_{GM}^H = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n} x_{it}^T w_s(v_{it};r_{it}) w_x(x_{it}w_s(v_{it};r_{it})) w_r(r_{it}^H) w_s(v_{it};r_{it}) y_{it}}{\sum_{t=1}^{T}\sum_{i=1}^{n} x_{it}^T w_s(v_{it};r_{it}) w_x(x_{it}w_s(v_{it};r_{it})) w_r(r_{it}^H) w_s(v_{it};r_{it}) x_{it}}, \tag{2.29}$$

where $r_{it}^H = y_{it}w_s(v_{it};r_{it}) - x_{it}w_s(v_{it};r_{it})\beta_{GM}^H$. The covariance matrix of β_{GM}^H can be approximated in large samples by

$$cov(\beta_{GM}^H) = (D^{-1})X^T W_s(v;r) W_x(X W_s(v;r)) \Psi W_x(X W_s(v;r)) W_s(v;r) X (D^{-1}), \tag{2.30}$$

where

$$D = X^T W_s(v;r) V(r^H) W_x(X W_s(v;r)) W_s(v;r) X \tag{2.31}$$

Ψ and $W_s(v;r)$ are diagonal matrices consisting of $(\psi(r_{it}^H))^2$ and $w_s(v_{it};r_{it})$ respectively. $V(r^H)$

Table 2.2 Simulation Results Concerning Efficiency of the HBP GM Estimator When the Regression Errors Are Heteroskedastic, Experiment (I)[a]

β_{GM}	β^{H}_{GM}	β_{GM}	β^{H}_{GM}
0% Pollution		10% Pollution	
1.000149	1.000097	0.999869	1.000112
0.002212	0.000498	0.002689	0.000619
0.002212	0.000498	0.002689	0.000619
1.000239	1.000026	0.995826	0.998892
0.017134	0.007008	0.019211	0.007956
0.017134	0.007008	0.019193	0.007955

[a] Each cell contains the mean, the mean squared error and the variance of the estimated parameter value over the 10,000 runs. Each first row concerns the intercept (β_1), each second row concerns the coefficient of the normally distributed explanatory variable (β_2). The length of the series equals 200 and $\beta_1=\beta_2=1$.

and $W_x(X)$ are defined in section 1.4.1. Notice that the diagonal weighting matrix W_x is computed on the basis of the Minimum Volume Ellipsoid distances of $x_{ti}w_s(v_{ti};r_{ti})$. to the centre. Within the derivation of (2.30) one can show that $w_s(v_{ti};r_{ti})=1/\hat{\sigma}_{ti}$ is asymptotically independent from $\hat{\beta}^{H}_{GM}$.

2.5.3 Simulation Results

To investigate the efficiency of the HBP GM estimator extended with Harvey's two-step improvement, we carried out the following simulation experiment:

Experiment (I)

$$y_i = \beta_1 + x_i\beta_2 + \varepsilon_i,$$

$$\varepsilon_i = x_i\gamma_i$$

where γ_i and x_i are independent identical standard normally distributed (*iid* N(0,1)); $\beta_1=\beta_2=1$.

By construction, the generated random variables x_i^2 and ε_i^2 are correlated. This induces heteroskedasticity. The length of these series is equal to 200. Experiment (I) has been replicated ten thousand times both for *clean* and contaminated series. Outliers were created by replacing the values of both the response and the explanatory variable x_i for 10% of the observations (randomly drawn) by random values drawn from a normal distribution with zero mean and variance 100. Thus, both leverage points and vertical outliers were added.

We reestimated β of experiment (I) using the *standard* HBP GM estimator of Chapter 1 and the adjusted HBP GM estimator according to Harvey's two-step procedure. Table 2.2 contains the mean, the mean squared error and the variance of the estimated parameter value over the ten thousand runs.[4] Both estimators are consistent but the *adjusted* HBP GM estimator is considerably more efficient than the *standard* estimator. We observe a reduction of approximately 59% in the variance of the estimated coeffients β_{GM2}^{H} in comparison with the variance of $\hat{\beta}_{GM2}$. This gain in efficiency is similar in clean or polluted data sets. Evidently, these particular Monte Carlo results will not necessarily be obtained outside of our limited experiment. However, on the basis of our results, the HBP GM estimator extended with Harvey's two-step improvement is to be preferred to the *standard* estimator since its associated mean squared errors are smaller in all cases considered.

Section 2.6 Autocorrelation

2.6.1 The Generalized Durbin-Watson Test

To verify whether the first order autocorrelation coefficient (ρ) of the equation

$$\varepsilon_{ti} = \varepsilon_{t-1,i}\rho + \xi_{ti} \tag{2.32}$$

is significantly different from zero, we employ the generalized Durbin and Watson *bounds* test statistic

$$d_p = \frac{\sum_{t=2}^{T}\sum_{i=1}^{n}(e_{ti}-e_{t-1,i})^2}{\sum_{t=1}^{T}\sum_{i=1}^{n}e_{ti}^2}. \tag{2.33}$$

ξ_{ti} are errors and e_{ti} are the regression residuals resulting from the estimation of the panel data model (2.1). Bhargava, Franzini and Narendranathan (1982), hereafter BFN, derive the lower bound d_L^* and the upper bound d_U^* corresponding with d_p as a function of the number of time

[4] These results can be reproduced by choosing the seed for the random number generator of the *UNIX* machine equal to 1. The same seed number is used in experiment (III) of section 2.6.3.

periods T, the number of cross-sectional units n and the number of explanatory factors at a specified significance level. The null hypothesis of zero autocorrelation is rejected against the alternatives $\rho>0$ and $\rho<0$ if $d_p<d_L^*$ and $d_p>4-d_L^*$ respectively. Likewise, we accept the null hypothesis if $d>d_U^*$ and $d<4-d_U^*$ respectively. The generalized Durbin-Watson statistic converges to 2 for large samples in n if the data are free from outliers and $\rho=0$. BFN tabulate values of d_L^* and d_U^* for panel data sets of different size.

Two remarks are called for. First, application of the *standard* Durbin-Watson test requires specification of an intercept term in the linear regression model in order to have good performance in terms of power (see, among others, Judge et al. (1988), p. 399). The panel data model studied in this chapter contains individiual specific fixed effects instead of a constant. It turns out that the generalized Durbin-Watson statistic has the desirable power properties if the model is estimated in the within dimension of the data. However, the statistic d_p is not distribution invariant to the first difference transformation of the data, i.e. BFN's lower and upper bounds can not be applied when d_p is computed on the basis of the residuals from the first difference regression. In that case, BFN suggest consulting the Berenblut-Webb statistic (See BFN (1982), p.538). Second, it is assumed that the error terms ξ_{ti} of equation (2.32) are independent normally distributed with mean zero.

We compute the generalized Durbin-Watson statistic for the weighted residuals rw_{ti} which are given in formula (2.12). Using the unweighted residuals r_{ti} would result in a highly non-robust test. These unweighted residuals are normally distributed under normality of the model disturbances ε_{ti} and outlier free data. On the other hand, the weighted residuals rw_{ti} are not normally distributed under the central model distribution (i.e. ε_{ti} is normally distributed) even when outlying observations are not present in the data. As a consequence, the error terms ξ_{ti} of equation (2.32) are not necessarily drawn from a normal distribution under the null hypothesis $\rho=0$. Strictly speaking, the generalized Durbin-Watson test can not be applied to the weighted residuals. However, if relatively few outliers corrupt the data then the distribution of the rw_{ti} approximates to the normal distribution. We make the following suggestion as a practical solution. Start with the execution of a normality test on the weighted residuals. We use the Doornik-Hansen (1994) test for this purpose since it was shown to exhibit good finite (small) sample power properties. Then, if normality is not rejected, we proceed with the computation of the generalized Durbin-Watson statistic and decide not to apply this autocorrelation test otherwise. This ad hoc procedure is more reliable than application of the generalized Durbin-Watson statistic on the unweighted non-robust residuals, for there is the possibility that only a single observation causes the autocorrelation test to take on arbitrary values.

Consider the following experiment:

Experiment (II):

$$y_{ti} = \eta_i + x_{ti}\beta + \varepsilon_{ti},\ t=1,..,6,\ i=1,..,100.$$

Table 2.3 Simulation Results Concerning Power Properties of the Generalized Durbin-Watson Test, Experiment II[a]

Pollution, δ%	95th Percentile	p-value ($P[d_p]<d_L^*$)	p-value ($P[d_p]>4-d_L^*$)	Average of d_p
		OLS		
0%	2.133	0.040	0.050	2.000
	HBP GM, δ% of the Observations (Randomly Drawn) are Contaminated			
0%	2.129	0.066	0.046	1.992
5%	1.708	0.998	0	1.540
10%	1.511	1.000	0	1.339
15%	1.519	1.000	0	1.355
20%	1.598	1.000	0	1.432
25%	1.739	0.996	0	1.545
	HBP GM, The Very Same δ% Observations of Each Cross-section Are Contaminated For Each Time Period			
5%	2.130	0.070	0.046	1.988
10%	2.114	0.094	0.026	1.977
15%	2.115	0.092	0.032	1.976
20%	2.102	0.124	0.022	1.964
25%	2.095	0.218	0.018	1.940

[a] The generalized Durbin-Watson statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation and the absence of outliers. At the five per cent significance level, d_L^*=1.8660 and d_U^*=1.8731 when T=6, n=100 and the panel data model contains one explanatory variable (see Bhargava, Franzini and Narendranathan (1982), p. 537). The number of replications is equal to 500. Estimations are performed in the within dimension of the data.

The fixed effects η_i are drawn from the standard uniform distribution whereas x_{it} and ε_{it} are independent standard normally distributed. The coefficient β is set to one. There are T=6 periods and n=100 cross-sectional units. The total length of the series y_t is thus 600. Experiment (II) is repeated 500 times to evaluate the power properties of BFN's test in clean and contaminated data samples.

The simulated data is first transformed into the within dimension by calculating deviations around individual means. Next, the OLS estimator is applied to estimate β and to construct the OLS residuals. The second row of Table 2.3 contains the 95th percentile and the average value of the generalized Durbin-Watson statistics based on these ordinary residuals. Moreover, it reports the

empirical p-values associated with the probability that the null hypothesis is rejected against the alternative hypothesis of a significant positive autocorrelation coefficient $P[d_p<d_L^*=1.866]$ or the probability that significant negative autocorrelation $P[d_p>4-d_L^*]$ is detected. The subsequent row of Table 2.3 displays these four items attributed to the Durbin-Watson statistics based on the weighted HBP GM residuals. The two rows with the clean data results project abundantly clear similarities. The rounded average of the Durbin-Watson statistics of the series computed with both the OLS and the HBP GM weighted residuals is equal to two. The 95th percentiles and the p-values are also reasonably close. Thus, the generalized Durbin-Watson statistic constructed with the weighted residuals appears to be accurate for panel data sets which are free from outlying observations.

The results are dramatically different for the contaminated data sets where $\delta\%$ of the observations (randomly drawn) are replaced by random values which are drawn from the normal distribution with mean zero and variance 100. The percentage of corrupted data points is increased up to 25% with a step size equal to 5%. The autocorrelation test falsely claims the occurence of significant positive autocorrelation in almost all of the cases. Negative significant autocorrelation is never found. These results indicate that BFN's upper and lower bounds do not hold for the Durbin-Watson statistic if the weighted residuals are used to compute d_p and a substantial portion of the data is polluted. The data points which were replaced by possibly outlying numbers, were randomly chosen among the total set of observations. For many empirical panel data studies it happens that a certain cross-sectional unit when traced as an aberrant point is outlying for each time period. Therefore, we also consider the case where $\delta\%$ of the very same cross-sectional units are polluted in each time period. Consequently, $(1-\delta)\%$ of the cross-sectional observations is not polluted in any of the time periods. Still, $\delta\%$ of the total number of observations is replaced by random values of the normal distribution with mean zero and variance 100. The power properties of the generalized Durbin-Watson statistic are more satisfactory for the experiments with these polluted data sets. The average value of the test statistics remains close to two until the cases where more than 15% of the data is contaminated. The empirical probability $P[d_p>4-d_L^*]$ of reporting significant negative autocorrelation is about 3% if $\delta=15\%$ and so is close to the probability that d_p exceeds the critical value under the central model distribution. In this case, the p-value indicating the probability $P[d_p<d_L^*]$ is less than twice the significance level of 5% but is fairly high.

To summarize the simulation results, the generalized Durbin-Watson statistic based on the weighted residuals is a reliable indicator of autocorrelation if the data is free of outliers or when relatively few data points are deviating. The test might provide a wrong signal if a substantial number of the cross-sectional units is contaminated.

2.6.2 Robust Seemingly Unrelated Regression

In matrix representation, model (2.1) can be reformulated as

$$y = \eta \otimes l + X\beta + \varepsilon, \tag{2.34}$$

where l is a T-dimensional vector comprising ones, $\otimes$ is the symbol of the Kronecker product, $y=(y_1^T,...,y_T^T)^T$, $X=(X_1^T,...,X_T^T)^T$ and $y_t=(y_{t1},...,y_{tn})^T$, $X_t=(x_{t1}^T,...,x_{tn}^T)^T$. The cross-sectional units are compiled for each time period and these bundles of observations per period are stacked below each other. According to the assumptions (A.1)-(A.4) of Section 2.2 the variance-covariance matrix of ε can be written as $E[\varepsilon\varepsilon^T]=\Sigma\otimes I_n$, where the T-dimensional matrix Σ consist of the elements $E[\varepsilon_{ti}\varepsilon_{si}]=\sigma_{ts}$ and I_n is the n-dimensionial identitiy matrix. As usual, the fixed effect are eliminated by estimation in the within dimension of the data or in first differences. Below we will not introduce a new notation for these transformed dependent and independent variables but keep writing y and X.

Our robust SUR estimator is defined by the first order condition

$$X^T W_x(X)(\hat{\Omega}\otimes I_n)^{-1} W_r(r_{ti})(y - X\hat{\beta}_{RSUR}) = 0 \tag{2.35}$$

where $\hat{\Omega}$ contains the consistent estimates of the elements of the variance-covariance matrix of Tukey's bi-square errors $\psi(\varepsilon_{ti})=\varepsilon_{ti}w_r(\varepsilon_{ti})$ (see section 1.2.3). Notice that the probability limit of $\hat{\Omega}$ is different from the variance-covariance matrix of the original model errors. Adjustment for autocorrelation is achieved by the insertion of the inverse of the weighting matrix $(\hat{\Omega}\otimes I_n)$ in the usual HBP GM defining equation. The thus constructed robust SUR estimator is essentially different from Peracchi's (1991) bounded influence SUR estimator. Peracchi (1991) derives an additional GM score function in order to obtain consistent robust estimates of Σ which then are used to determine the autocorrelation adjustment in the normal equation for $\hat{\beta}_{RSUR}$.

The computation of the robust SUR estimator in (2.35) involves a two-step procedure. In the first round, the usual HBP GM estimates, either corrected for multiplicative heteroskedasticity or not, are computed to construct the T-dimensional square matrix $\hat{\Omega}$ with entries

$$\hat{\omega}_{ts} = \frac{1}{n}\sum_{i=1}^{n}\psi(r_{ti})\psi(r_{si}). \tag{2.36}$$

$r_{ti}=y_{ti}-x_{ti}\hat{\beta}_{GM}$ are the HBP GM regression residuals. The second round consists of solving (2.35) for the robust SUR estimator:

$$\hat{\beta}_{RSUR} = (X^T W_x(X)(\hat{\Omega}\otimes I_n)^{-1} W_r(r)X)^{-1}(X^T W_x(X)(\hat{\Omega}\otimes I_n)^{-1} W_r(r)y) \tag{2.37}$$

in case that no adjustment for heteroskedasticity is made and

$$\beta^{H}_{RSUR} = (X^{T}W_x(XW_s)W_s(\hat{\Omega}\otimes I_n)^{-1}W_sW_rX)^{-1}(X^{T}W_x(XW_s)W_s(\hat{\Omega}\otimes I_n)^{-1}W_sW_ry) \tag{2.38}$$

otherwise.[5] An iterative procedure such as the Newton-Raphson (NR) algorithm could be employed to solve the score function (2.35). For each NR step one could update the weight function w_r and the estimate of $\hat{\Omega}$. Instead, we perform only one iteration. The one-step RSUR estimator however exhibits the same asymptotic properties as the multi-step alternative.

Recall that the influence function measures the change in the estimates, returned by the estimator β, caused by an infinitesimal perturbation in the central model distribution F:

$$IF(z;\beta,F) = \lim_{s\downarrow 0}\frac{\beta((1-s)F+sG(z)) - \beta(F)}{s}. \tag{2.39}$$

The distribution function $G(z)$ puts mass 1 at the point $z = (y_i,x_i)$. Let us introduce the notation $F_s = (1-s)F + sG$ and define F_n as the empirical distribution function which assigns a probability of $1/N$ to each observation (y_{it},x_{it}). The distribution function F_s describes the data at hand of which $(1-s)\%$ is generated by F and $s\%$ by G, given β. β_{RSUR} provides a consistent estimate of these unknown parameters β. Thus the outliers are supposed to be drawn from the distribution function G. Then, the score function (2.35) can be written as

$$\int X_i^{T}W_x(X_i)\hat{\Omega}(F_n)^{-1}W_r(r_i)(y_i-X_i\beta_{RSUR}(F_n))dF_n=0, \tag{2.40}$$

where $y_i=(y_{1i},..,y_{Ti})^T$, $X_i=(x_{1i}^T,...x_{Ti}^T)^T$ and the diagonal matrices $W_x(X_i)$ and $W_r(r_i)$ contain the weights $(w_x(x_{1i}),..,w_x(x_{Ti}))$ and $(w_r(r_{1i}),..,w_r(r_{Ti}))$ respectively. By the same token, we may define $\beta_{SUR}(F_s)$ for any distribution function F_s:

$$\int X_i^{T}W_x(X_i)\hat{\Omega}(F_s)^{-1}W_r(r_i)(y_i-X_i\beta_{RSUR}(F_s))dF_s=0. \tag{2.41}$$

The normal equation (2.41) can be rewritten as:

$$(1-s)\int X_i^{T}W_x(X_i)\hat{\Omega}(F_s)^{-1}W_r(r_i)(y_i-X_i\beta_{RSUR}(F_s))dF + \tag{2.42}$$

$$s\int X_i^{T}W_x(X_i)\hat{\Omega}(F_s)^{-1}W_r(r_i)(y_i-X_i\beta_{RSUR}(F_s))dG=0.$$

---

[5] In the latter case, $\hat{\omega}_{ts}=\frac{1}{n}\sum_{i=1}^{n}\psi(r_{it}^{H})\psi(r_{is}^{H})$ where $r_{it}^{H}=y_{it}w_s(v_{it};r_{it})-x_{it}w_s(v_{it};r_{it})\beta^{H}_{GM}$

Given some regularity conditions (see, among others, Huber (1981) and Hampel et al. (1986)) which allow the interchange of integration and differentiation, we obtain the following identity by application of the chain rule when differentiating (2.42) with respect to s:

$$-\int X_i^T W_x \hat{\Omega}(F_s)^{-1} W_r(r_i)(y_i - X_i \beta_{RSUR}(F_s))dF + \int X_i^T W_x \hat{\Omega}(F_s)^{-1} W_r(r_i)(y_i - X_i \beta_{RSUR}(F_s))dG \tag{2.43}$$

$$- (1-s)\int X_i^T W_x \hat{\Omega}^{-1}(F_s) V(r_i) X_i \frac{\partial \beta_{RSUR}(F_s)}{\partial s} dF + (1-s)\int X_i^T W_x \frac{\partial \hat{\Omega}^{-1}(F_s)}{\partial s} W_r(r_i)(y_i - X_i \beta_{RSUR})dF$$

$$- s\int X_i^T W_x \hat{\Omega}(F_s)^{-1} V(r_i) X_i \frac{\partial \beta_{RSUR}(F_s)}{\partial s} dG + s\int X_i^T W_x \frac{\partial \hat{\Omega}^{-1}(F_s)}{\partial s} W_r(r_i)(y_i - X_i \beta_{RSUR})dG = 0$$

where $V(r_i)$ is a diagonal matrix consisting of $(\partial\psi(r_{1i})/\partial r_{1i},...,\partial\psi(r_{Ti})/\partial r_{Ti})$. The influence function $IF(z;\beta_{RSUR},F)=\partial\beta_{RSUR}(F_s)/\partial s$ is delivered by evaluating (2.43) in $s=0$, using the facts that $\beta_{RSUR}(F)=\beta$, $\hat{\Omega}(F)=\Omega$ and noticing that $E[r_{it}x_{it}]=0$ following assumption (A.2):

$$IF(z;\beta_{RSUR},F) = E_F[X_i^T W_x(X_i)\hat{\Omega}^{-1}V(r_i)X_i]^{-1}x_i^T W_x(x_i)\Omega^{-1}W_r(\varepsilon_i)(y_i - x_i\beta). \tag{2.44}$$

This influence function is only well defined if we assume that the expectations $E_F[\psi(\varepsilon)\psi(\varepsilon)^T]$ and $E_F[X_i^T W_x(X_i)\hat{\Omega}^{-1}V(r_i)X_i]$ exist and are matrices of full rank, i.e. nonsingular. The right hand side of equation (2.44) is clearly bounded. Leverage points x_i are downweighted by the function $W_x(x_i)$ whereas aberrant errors are bounded by the weights $W_r(\varepsilon_i)$. We conclude that our estimator β_{RSUR} is a reliable robust alternative to Zellner's (1962) SUR estimator since it has a bounded influence function.

The robust SUR two-step procedure developed in this section belongs to the category of estimators considered in Krishnakumar's (1995) *general methodology of robust estimation for panel data models*. So we may apply Lemma 1 (see Krishnakumar (1995), p.7) where it is demonstrated that the estimator has the appealing properties of being consistent and asymptotic normally distributed (see also Maronna and Yohai (1981)). Using a Taylor approximation of (2.35) one can show that the asymptotic variance-covariance matrix of β_{RSUR} can be consistently estimated by:

$$cov(\beta_{RSUR}) = (D^{-1})X^T W_x(X)(\hat{\Omega}^{-1}\otimes I_n)W_x(X)X(D^{-1})^T \tag{2.45}$$

where

$$D = XW_x(X)(\hat{\Omega}\otimes I_n)^{-1}V(r)X. \tag{2.46}$$

At this point, after taking care of heteroskedasticity and autocorrelation, it is worthwhile establishing whether anomalous observations have excessive impact on Zellner's (1962) SUR estimates. The robust Wu-Hausman type test statistic

$$h_R = (\beta^H_{RSUR}-\beta^H_{SUR})^T cov(\beta^H_{RSUR})^{-1}(\beta^H_{RSUR}-\beta^H_{SUR}) \tag{2.47}$$

is computed to test whether the adjusted RSUR estimates β^H_{RSUR} are significantly different from the GLS counterparts (see Hausman (1978) and Wu (1973)). The $\chi^2(p)$ distribution first order stochastically dominates h_R; p indicates the number of explanatory variables. Thus, under the null hypothesis that β^H_{SUR} is unbiased and efficient for smooth data sets h_R induces type II errors (acceptance of the null hypothesis when the null hypothesis is false) more frequently than the significance level associated with the chosen critical value (percentile of the $\chi^2(p)$ distribution) would predict. The common Wu-Hausman test statistic

$$h = (\beta^H_{RSUR}-\beta^H_{SUR})^T(cov(\beta^H_{RSUR})-cov(\beta^H_{SUR}))^{-1}(\beta^H_{RSUR}-\beta^H_{SUR}) \tag{2.48}$$

is well behaved under the null hypothesis but is inappropriate for (small) contaminated samples since outliers also corrupt the OLS covariance estimate. In particular, h may induce type II errors in polluted data sets.

2.6.3 A Monte Carlo Study

To examine the consistency and efficiency of the RSUR estimator in panel data sets with autocorrelated errors we performed experiment (III).

Experiment (III):

$$y_{ti} = \beta_1 + x_{ti}\beta_2 + \varepsilon_{ti},$$

$$\varepsilon_{ti} = \gamma_{ti} + v_i,\ i=1,..,100,\ t=1,..,4$$

where γ_{ti} are independent and identically normally distributed with variance 1/4 (*iid* N(0,1/4)), v_i, x_{ti} are *iid* N(0,1). The dimension of the vectors v is equal to n=100; The number of time periods equals T=4; Thus, the series y and x have length 400; $\beta_1 = \beta_2 = 1$.

Like experiment (I), experiment (III) was repeated 10,000 times for both *clean* and contaminated data sets. In the latter ones, ten percent of the observations have been replaced by a normal random variable with mean zero and variance 100. This creates anomalous observations in

Table 2.4 Simulation Results Concerning Efficiency of the HBP GM Estimator When the Regression Errors Are Autocorrelated, Experiment (III)[a]

β_{GM}	β_{RSUR}	β_{GM}	β_{RSUR}
0% Pollution		10% Pollution	
1.002323	1.002412	1.001934	1.002243
0.011120	0.011587	0.011533	0.012782
0.011114	0.011582	0.011529	0.012777
1.000364	1.000053	0.986842	0.986751
0.003490	0.001098	0.004920	0.003258
0.003490	0.001098	0.004747	0.003082

[a] Each cell contains the mean, the mean squared error and the variance of the estimated parameter value over the 10,000 runs. Each first row concerns the intercept (β_1), each second row concerns the coefficient of the normally distributed explanatory variable (β_2). The length of the series y equals 400 and $\beta_1=\beta_2=1$.

the form of both leverage points and vertical outliers. The disturbances $\varepsilon_{1i},...,\varepsilon_{4i}$ have innovation v_i in common. This causes correlation in the time-series of the panel data whereas the cross-sectional units are uncorrelated since the v_i ,$i=1,..,n$ and γ_{it}, $i=1,..,n$, $t=1,..,T$ are drawn independently. Correlated errors are constructed without omission of important explanatory variables, in the sense that there are no independent variables available to explain the correlation in the disturbances. In contrast, for instance Mizon (1995) gives an example of a data generating process where the lagged dependent variable is incorporated in the set of explanatory variables. Re-estimation of this model without including the lagged response variable, obviously also induces autocorrelated residuals. Here we need to remark that in view of model (2.1) the data of experiment (III) result after transformation of (2.1) in order to drop the fixed effects. These individual effects are supposed not to fully explain the autocorrelation in the original model disturbances. Evidently, within or first difference regression of experiment (III) would eliminate the autocorrelation.

The mean, the mean squared error and the variance of the parameter estimates for both the standard HBP GM method and the RSUR estimator over the 10,000 replications are reported in Table 2.4. The standard HBP GM estimator is consistent for experiment (III) where the disturbances are autocorrelated, even in the case of data pollution. Moreover, rather surprisingly, it produces the most accurate estimates of the intercept term β_1 as is shown by the corresponding mean squared errors of the HBP GM and RSUR estimates. The B-robust SUR estimator is considerably more efficient in generating consistent estimates of the parameter β_2. β_{RSUR} provides, among the two estimators considered here, the most accurate estimates of the coefficient of the

explanatory variable x_{i1}. However, the differences in the corresponding mean squared errors decline when the data is contaminated. Table 2.4 reveals that the variance associated with β_2 is about 68% smaller in case of RSUR estimation in comparison with standard HBP GM regression for smooth data sets. If 10% of the data is deviating from the central model distribution then this efficiency gain is diminished to approximately 35%. In section 2.5.3 we found that the extent to which the performance of the HBP GM technique extended with Harvey's two-step improvement is superior to standard HBP GM, is equal under the central heteroskedastic model distribution and the polluted version.

Mizon (1995) shows for his simulation experiment that autocorrelation correctors, for instance the Autoregressive Least Squares (ALS) estimator, might produce inconsistent parameter estimates. He concludes:

> Although it is important to test for serial correlation in the residuals of econometric models, it is rarely appropriate to 'autocorrelation correct' in response to rejecting the hypothesis of zero serial correlation (Mizon (1995), p. 285).

In the context of panel data models, we provided an intuitive widespread counterexample where the B-robust SUR estimator is to be preferred to the standard HBP GM method. Furthermore, Mizon (1995) continues by stating:

> The practice of 'autocorrelation correction' is an example of specific-to-general modelling, and so the example presented in section 2 is a particular illustration of the weakness of this modelling strategy (Mizon (1995), p. 285).

The nature of the serial correlation in experiment (III) is dissociated from the independent variables of the regression model, in the sense of that it is assumed that no explanatory factor is available.[6] In this respect, modelling strategy does not come into the play in our case. We did not verify what would have happened to the efficiency of the HBP GM estimator if the panel data regression model incorporated the lagged values of the response variable. For panels with a low number of time periods T, however, it might be unwise to lose one period consisting of n cross-sectional observations.

---

[6] Strictly speaking, one could introduce n firm dummies in the regression model to take account of the serial correlation. This results in $p+n$ explanatory variables in proportion to N observations. If n is large and T is small, as is often the case in empirical panel data sets, then this modelling strategy is not recommendable.

Section 2.7 Conclusions

Chapter 2 is concluded below by a short summary. In the context of a panel data model we have investigated how the standard High Breakdown Point GM estimator of Chapter 1 behaves when (1) specific individual fixed effects are correlated with the regressors and (2) regression errors are heteroskedastistic or (3) serially correlated. Robust specification tests have been introduced to test the assumptions underlying the standard linear model, including cross-sectional independence and uncorrelated fixed effects. It was shown, by means of limited Monte Carlo studies, that reliable application of the robust version of the Pagan-Vella statistic and the generalized Durbin-Watson test of Bhargava, Franzini and Narendranathan (1982) based on the weighted residuals is restricted. The first test decisively indicates non-adjustment for heteroskedasticity if the statistic cannot reject the null hypothesis of homoskedasticity but can be misleading if the homoskedastic regression model is contaminated with vertically outlying leverage points. The lower and upper bounds tabulated by BFN do only apply to the latter test if the normality of the weighted residuals cannot be rejected. Although our simulation experiments show that the standard HBP GM estimator is consistent under heteroskedasticity and autocorrelation, more efficient estimators can be obtained by extending the HBP GM method with Harvey's two-step improvement or by employing the Robust SUR estimator in case the errors of the regression model are respectively heteroskedastic or serially correlated. The influence function of the RSUR technique is derived and is bounded. The RSUR estimator is thus relatively insensitive to small departures from the central model distribution. Our simulation results reveal that robust autocorrelation correctors are more accurate in producing estimates of the covariance model than the classical solution of Zellner (1962). Since empirical data sets are thought to contain 1 to 10 percent pollution (see Hampel et al. (1986), p.28), the use of robust estimators is indispensable. The research topic of the next chapter covers B-robust estimation of simultaneous equations or models with erroneously measured explanatory factors.

CHAPTER 3

Two Stage GM and Robust Generalized Method of Moments

Section 3.1 Introduction

The orthogonality condition of the standard linear model is the next condition to be weakened after having relaxed the independence and homogeneity assumptions in the previous chapter. Weakening this condition is often of cucial importance when investigating the empirical relationship between economic variables. For instance, in simultaneous equations or when measurement errors occur in the explanatory factors one can not expect that the error terms of the regression model are perpendicular to the regressors. The occasion of measurement errors in real economic data sets arises many times as theoretical variables frequently do not coincide one-to-one with the observable variables. Likewise the classical estimation procedures such as OLS and GLS, the HBP GM technique and the B-robust SUR estimator are also biased in these circumstances. In response to this miscalculating of the unknown model parameters we introduce two stage generalized M-estimation following Krasker (1986)[1] in Section 3.2 and robustify the Generalized Method of Moments (see e.g. Hansen (1982)) in Section 3.3. Conditions are presented under which these B-robust alternatives to two stage least squares (2SLS) and GMM are consistent and asymptotically normal. Furthermore, we show that the influence function of the Robust GMM (RGMM) estimator is bounded. Section 3.4 contains a Monte Carlo study on the performance of various robust and non-robust estimators in a simulated panel data model with errors in variables which is extensively discussed by Griliches and Hausman (1986). Finally, conclusions and an outlook are presented in the last section.

Section 3.2 Two Stage Generalized M-Estimation

When applying the High Breakdown Point Generalized M-estimator to the standard linear model

$$y_i = x_i\beta + \varepsilon_i, \quad i=1,..,n \tag{3.1}$$

[1] See Krasker and Welsch (1985) for a closely related instrumental variable approach, the so-called Weighted Instrumental Variables (WIV) estimators.

to estimate the unknown parameter vector β it is assumed that the row vector x_i of length p of observable explanatory factors is orthogonal to the error term ε_i. As usual, y is an observable dependent variable and n is the number of observations. Further, the errors are supposed to be independent and identically distributed with bounded variance and zero expected value. Recall that, following the definitions of Rousseeuw and van Zomeren (1990), outliers are classified as leverage points (horizontal outliers) when observations exhibit outlying x_i whereas vertical outliers are observations which are positioned far from the plane through the majority of the data but are not necessarily outlying in the direction of x.

The aim of this chapter is to demonstrate how to obtain robust consistent estimates when the orthogonality condition that x_i and ε_i are independent, is not fulfilled. Suppose there are instrumental variables z_i, $i=1,...,n$ which are correlated with the explanatory factors x_i but independent of the error term ε_i; z_i is a row vector of dimension k, $k \geq p$. In imitation of the Two Stage Least Squares (2SLS) procedure introduced by Theil (1971) we propose to carry out Two Stage Generalized M-estimation (2SGM) as follows:

Stage 1. The explanatory variable x_{ij} (x_j is the jth column of X) is regressed on the instrumental variables z_i:

$$x_{ij} = z_i \gamma_j + \xi_{ij}, \tag{3.2}$$

where ξ_{ij} are error terms and γ_j is a column vector of lenght k. This returns the HBP GM estimate $\hat{\gamma}_{j,GM}$ with concomitant first step weights $w_z(\cdot)$ and $w_r(\cdot)$ based on the identification of leverage points in the instrumental variables and vertical outliers in relation to (3.2) respectively. The prediction of the jth column of X is then computed according to

$$\hat{x}_j = Z(Z^T W_z(Z) W_r(r_{1,j}) Z)^{-1} Z^T W_z(Z) W_r(r_{1,j}) x_j. \tag{3.3}$$

$r_{1,j}$ are the first stage HBP GM residuals associated with x_j. Notice that $W_r(r_{1,j})$ differs for every distinct column of X. Thus p separate HBP GM regressions are performed in stage 1.

Stage 2. Replacing the explanatory variables of the original equation by their robust projection on Z and applying the HBP GM technique once more provides the 2SGM estimates

$$\beta_{2SGM} = (\hat{X}^T W_x(\hat{X}) W_r(r_2) \hat{X})^{-1} \hat{X}^T W_x(\hat{X}) W_r(r_2) y. \tag{3.4}$$

$W_x(\hat{X})$ and $W_r(r_2)$ are diagonal matrices containing the second step HBP GM weights; r_2 are the second stage HBP GM residuals.

The derivation of a consistent and relatively efficient variance-covariance estimator for the 2SGM method is not as simple as determining an asymptotic reliable approximation of the

variance-covariance matrix of the HBP GM estimator (see Chapter 1, section 1.4.1). Consider a Taylor series expansion around $\hat{\beta}_{2SGM}=\beta$ for the first order condition associated with the second stage of the 2SGM procedure

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \hat{x}_i^T w_x(\hat{x}_i) w_r(r_{2i}) r_{2i} \\ &= \sum_{i=1}^{n} \hat{x}_i^T w_x(\hat{x}_i) w_r(\varepsilon_{2i}) \varepsilon_{2i} - \sum_{i=1}^{n} \frac{\partial \psi(\varepsilon_{2i})}{\partial \varepsilon_{2i}} \hat{x}_i^T \hat{x}_i w_x(\hat{x}_i)(\hat{\beta}_{2SGM}-\beta). \end{aligned} \tag{3.5}$$

Solving (3.5) for $\sqrt{n}(\hat{\beta}_{2SGM}-\beta)$ gives

$$\sqrt{n}(\hat{\beta}_{2SGM}-\beta) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \psi(\varepsilon_{2i})}{\partial \varepsilon_{2i}} \hat{x}_i^T \hat{x}_i w_x(\hat{x}_i)\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{x}_i^T w_x(\hat{x}_i)\psi(\varepsilon_{2i})\right). \tag{3.6}$$

In large samples, $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \psi(\varepsilon_{2i})}{\partial \varepsilon_{2i}} \hat{x}_i^T \hat{x}_i w_x(\hat{x}_i)$ can be approximated by $\frac{1}{n}\hat{D}$ where

$$\hat{D} = \hat{X}^T W_x(\hat{X}) V(r_2) \hat{X}. \tag{3.7}$$

Following the notation of Chapter 1 and Chapter 2, the diagonal matrix $V(r_2)$ contains the elements $\partial\psi(r_{2i})/\partial r_{2i}$, $i=1,..,n$; Ψ is a diagonal matrix consisting of $(\psi(r_{2i}))^2$, $i=1,..,n$. The matrix $\hat{D}$ is found by evaluating $V(\cdot)$ in r_2. The variance of the other term of the right hand side of (3.6), however, can not be adequately approximated by evaluating $\hat{X}^T W_x(\hat{X}) \Psi W_x(\hat{X}) \hat{X}$ in r_2. This result occurs because $\hat{X} W_x(\hat{X})$ is not independent of $\Psi(r_2)$.[2]

A solution to this problem is found by seperating the second stage HBP GM residuals into components that are perpendicular to the weighted explanatory factors. For notational convenience, we start with the derivation of the 2SGM variance-covariance matrix for the one-dimensional case $p=k=1$. From this starting point, the formula of the variance-covariance matrix in general dimensions can be easily deducted.

First, using (3.2) we derive

$$x_i-\hat{x}_i = (x_i-z_i\gamma) - z_i(z^T W_z(z) W_r(r_1) z)^{-1} z^T W_z(z) W_r(r_1)(x-z\gamma). \tag{3.8}$$

With (3.8), $\sum_{i=1}^{n} \hat{x}_i w_x(\hat{x}_i) w_r(r_{2i})(y_i-\hat{x}_i\hat{\beta}_{2SGM})$ can be seperated into

[2] Suppose A and B are two random variables, then $E[(AB)^2] = E[A]E[B^2]E[A]$ only if A and B are independent.

$$\sum_{i=1}^{n}\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})(y_i-x_i\beta_{2SGM}+(x_i-z_i\gamma)\beta_{2SGM}-z_i(z^TW_z(z)W_r(r_1)z)^{-1}z^TW_z(z)W_r(r_1)(x-z\gamma)\beta_{2SGM}). \tag{3.9}$$

Formula (3.9) can be rewritten as

$$\sum_{i=1}^{n}\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})(y_i-\hat{x}_i\beta_{2SGM}) = \sum_{i=1}^{n}\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})(y_i-x_i\beta_{2SGM} + (x_i-z_i\gamma)\beta_{2SGM}) \tag{3.10}$$

$$- \sum_{i=1}^{n}\hat{\gamma}z_i w_x(\hat{x}_i)w_r(r_{2i})z_i(z^TW_z(z)W_r(r_1)z)^{-1}z^TW_z(z)W_r(r_1)(x-z\gamma)\beta_{2SGM}.$$

After some manipulations we obtain

$$\sum_{i=1}^{n}\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})(y_i-\hat{x}_i\beta_{2SGM}) = \sum_{i=1}^{n}\Big(\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})(y_i-x_i\beta_{2SGM} + (x_i-z_i\gamma)\beta_{2SGM}) \tag{3.11}$$

$$- \hat{\gamma}(z^TW_x(\hat{x})W_r(r_2)z)(z^TW_z(z)W_r(r_1)z)^{-1}z_i w_z(z_i)w_r(r_{1i})(x_i-z_i\gamma)\beta_{2SGM} \Big).$$

In the case that p=k=1, formula (3.11) can be written as

$$\sum_{i=1}^{n}\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})(y_i-\hat{x}_i\beta_{2SGM}) = \tag{3.12}$$

$$\sum_{i=1}^{n}\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})(y_i-x_i\beta_{2SGM} + \left(1-\frac{(z^TW_x(\hat{x})W_r(r_2)z)w_z(z_i)w_r(r_{1i})}{(z^TW_z(z)W_r(r_1)z)w_x(\hat{x}_i)w_r(r_{2i})}\right)(x_i-z_i\gamma)\beta_{2SGM}).$$

Since the components $(y_i-x_i\beta_{2SGM})$ and $(x_i-z_i\gamma)$ are independent of $\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})$ by construction, we are now in the position to approximate the variance of $\sum_{i=1}^{n}\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})(y_i-\hat{x}_i\beta_{2SGM})$ in large samples by replacing $z_i\gamma$ by $\hat{x}_i$ in:

$$M = \sum_{i=1}^{n}\hat{x}_i w_x(\hat{x}_i)w_r(r_{2i})\Big(y_i-x_i\beta_{2SGM} + \tag{3.13}$$

$$\left(1-\frac{(z^TW_x(\hat{x})W_r(r_2)z)w_z(z_i)w_r(r_{1i})}{(z^TW_z(z)W_r(r_1)z)w_x(\hat{x}_i)w_r(r_{2i})}\right)(x_i-z_i\gamma)\beta_{2SGM}\Big)^2 w_r(r_{2i})w_x(\hat{x}_i)\hat{x}_i.$$

The variance-covariance matrix of the 2SGM technique is then computed as

$$vcov(\beta_{2SGM}) = \hat{D}^{-1}\hat{M}\hat{D}^{-1}. \tag{3.14}$$

For any dimensions of p and k, the matrix $\hat{M}$ can be calculated as $\hat{M} = QQ^T$ where

$$Q = \hat{X}^T W_x(\hat{X}) W_r(r_2)(R_2+P) - \hat{\Gamma}^T (Z^T W_z(\hat{X}) W_r(r_2) Z) BAC. \tag{3.15}$$

R_2 and P are diagonal matrices containing $(y_1-x_1\beta_{2SGM}),...,(y_n-x_n\beta_{2SGM})$ and $(x_1-\hat{x}_1)\beta_{2SGM},...,$ $(x_n-\hat{x}_n)\beta_{2SGM}$ respectively. The matrix B, consisting of k rows and kp columns, is constructed as follows:

$$B = \begin{bmatrix} \beta_1 & 0 & \cdot & 0 & \beta_2 & 0 & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \beta_p & 0 & \cdot & 0 \\ 0 & \beta_1 & \cdot & \cdot & 0 & \beta_2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \beta_p & \cdot & \cdot \\ \cdot & 0 & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \beta_1 & 0 & 0 & \cdot & \beta_2 & \cdot & \cdot & \cdot & \cdot & 0 & 0 & \cdot & \beta_p \end{bmatrix}, \tag{3.16}$$

where β_j is the jth element of β_{2SGM}. A is a block-diagonal matrix containing the matrices $(Z^T W_z(Z) W_r(r_{1,1}) Z)^{-1},...,(Z^T W_z(Z) W_r(r_{1,p}) Z)^{-1}$. Recall that $r_{1,j}$ are the first stage residuals corresponding to the HBP GM regression of x_j on Z. The dimensions of the matrix C are determined by the number of observations n (fixing the number of columns) and pk (fixing the number of rows). A typical column C_i is equal to

$$C_i = vec(z_i^T w_z(z_i)(x_i-\hat{x}_i) W_r(r_{1i,j})) = \begin{matrix} z_i^T w_z(z_i)(x_{i,1}-\hat{x}_{i,1}) w_r(r_{1i,1}) \\ \vdots \\ z_i^T w_z(z_i)(x_{i,p}-\hat{x}_{i,p}) w_r(r_{1i,p}) \end{matrix} \tag{3.17}$$

where $r_{1i,j}$ is the first stage HBP GM residual associated with the ith observation and the jth explanatory variable. Finally, notice that the p columns of $\hat{\Gamma}$ are built from the k-dimensional vectors $\hat{\gamma}_j$, $j=1,...,p$.

We will assume that the quality of the instruments and the distribution $F(q_i)$ of the observations $q_i = (y_i, x_i, z_i)$, $i=1,...,n$ induce the following conditions to hold:

(A.1) $\underset{n\to\infty}{plim}\ (n^{-1} Z^T \varepsilon) = 0,$

(A.2) $\underset{n\to\infty}{plim}\ \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \psi(r_{2i})}{\partial r_{2i}} \hat{x}_i^T \hat{x}_i w_x(\hat{x}_i) \right) = D(\beta,\Gamma),$

the probability limit $D(\beta,\Gamma)$ exists and is a finite positive definite matrix of dimension p, and

$$(A.3)\quad \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{x}_i^T w_x(\hat{x}_i)\psi(r_{2i})\right)_{\beta,\Gamma} \sim N(0,M).$$

Assumption (*A*.1) especially is crucial for the success of the 2SGM estimator and can be tested (see, *inter alia*, Davidson and Mackinnon (1993), p.241).

Proposition 1.

The 2SGM estimator as defined in (3.4) is consistent and asymptotic normally distributed if conditions (*A*.1)-(*A*.3) hold.

Proof:

Using the Taylor expansion (3.6) and conditions (*A*.1)-(*A*.2) it is straightforward to prove the consistency of the 2SGM estimator by employing Slutsky's theorem. Further, if a central limit theorem (see, for instance, Davidson and Mackinnon (1993), p.126) is applicable to $n^{-1/2}\hat{X}^T W_x(\hat{X})W_r(r_2)\varepsilon$ then condition (*A*.3) holds and so the asymptotic normality of the 2SGM estimator is guaranteed. Under conditions (*A*.1)-(*A*.3) the proof follows directly from Lemma 1 of Krishnakumar (1995). See also Krasker (1986) and Maronna and Yohai (1981).

Krasker (1986) derives restrictions on F and $\phi_i = \hat{x}_i^T w_x(\hat{x}_i)\psi(r_{2i})$ which induce conditions (*A*.2)-(*A*.3) to hold. Here we will not go into detail but only remark that the key requirement, ϕ_i is bounded and continuous, is fulfilled for our 2SGM estimator.

On two fronts the 2SGM procedure does not coincide with the 2SLS method. First, the way in which each method computes final residuals is strikingly different. These residuals can be used to construct the sum of squared residuals, the scale parameter or test statistics (for example, the Frees (1995) test on cross-sectional independence or the Durbin-Watson autocorrelation test). Two Stage Least Squares final residuals which are calculated according to $r_{2SLS,i} = y_i - x_i\beta_{2SLS}$ are more accurate predictions of the model error terms than the second step residuals $r_{2SLS,i} = y_i - \hat{x}_i\beta_{2SLS}$ if the data are free from outlying observations. Second step HBP GM weights are involved when computing the weighted 2SGM residuals. On the basis of these weighted predictions of the disturbances, more reliable (robust) statistics can be provided. As a consequence, we can not swap $\hat{x}_i$ for x_i in the equation

$$rw_{2SGM,i} = y_i\sqrt{w_x(\hat{x}_i)w_r(r_{2i})} - \hat{x}_i\sqrt{w_x(\hat{x}_i)w_r(r_{2i})}\beta_{2SGM}. \qquad (3.18)$$

without adjusting the weights w_x and w_r.[3] Second, the variance-covariance matrix (3.14) associated with the 2SGM method is consistent under heteroskedasticity whereas the covariance estimator of the standard 2SLS technique (see, inter alia, Davidson and Mackinnon, p.217) is not.

Section 3.3 Robust Generalized Method of Moments Estimation

Reconsider the Seemingly Unrelated Regression (SUR) model

$$y_{ti} = x_{ti}\beta + \varepsilon_{ti}, \tag{3.19}$$

as introduced in section 2.2 but whose fixed effects are omitted and whose orthogonality condition $E[\varepsilon_{ti}x_{ti}]=0$ is replaced by the instrumental variable counterpart $E[\varepsilon_{ti}z_{ti}]=0$. At this stage we will allow for heteroskedastic errors in model (3.19) but stick to the cross-sectional independence assumption. The covariance structure of the errors ($E[\varepsilon\varepsilon^T] = \Sigma$) is therefore described by:

(*B*.1) $E[\varepsilon_{ti}]=0$, $i=1,..,n$, $t=1,..,T$,

(*B*.2) $E[\varepsilon_{ti}\varepsilon_{tj}]=0$, *if* $i\neq j$,

(*B*.3) $E[\varepsilon_{ti}^2] = \sigma_{ti}^2$,

(*B*.4) $E[\varepsilon_{ti}\varepsilon_{si}] = \sigma_{ts}$.

The estimation method outlined below can be extended to the more general case of models with any finite error covariance matrix. For example, the Newey-West (1987) procedure can be employed to estimate the unknown components of Σ.

Although the 2SGM technique of Section 3.2 provides heteroskedasticity consistent estimates with concomitant heteroskedasticity consistent standard errors, a more efficient robust estimator can be developed by exploiting the relevant moment conditions. The Generalized Method of Moments (GMM) considers the *theoretical* unconditional moment condition

---

[3] Evidently, when new weights $w_x(x_i)$ and $w_r(y_i-x_i\beta_{2SGM})$ are determined then the relatively more accurate predictions $r_{2SGM,i} = y_i\sqrt{w_x(x_i)w_r(y_i-x_i\beta_{2SGM})} - x_i\sqrt{w_x(x_i)w_r(y_i-x_i\beta_{2SGM})}\,\beta_{2SGM}$ may serve to compute the various statistics.

$$E[w_{it}^T \varepsilon_{it}] = 0, \tag{3.20}$$

where w_{it} is any vector which belongs to any information set I_t at time t with $E[\varepsilon_{it}|I_t] = 0$. The traditional way to evaluate (3.20) is to equate its sample counterpart to zero[4]

$$\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{n} w_{it}^T(y_{it}-x_{it}\beta) = 0. \tag{3.21}$$

In the instrumental variable case, the *empirical* moment condition (3.21) can be written as

$$X^T Z(Z^T \Sigma Z)^{-1} Z^T(y-X\beta) = 0. \tag{3.22}$$

Replacing the diagonal elements of Σ by the squared 2SLS residuals and solving (3.22) for β gives the Two Stage Instrumental Variable Estimator (2SIV) first proposed by White (1982). Equation (3.22) is the first order condition associated with the objective function

$$\min_{\beta} (y-X\beta)^T Z(Z^T \Sigma Z)^{-1} Z^T(y-X\beta). \tag{3.23}$$

This criterium function expresses the square of the moment $Z^T(y-X\beta)$ weighted by its covariance matrix $Z^T \Sigma Z$.

The empirical moment condition (3.22) can be a very unreliable approximation of the theoretical condition (3.20) in finite samples if the data are contaminated. In order to robustify (3.22) we insert $W_z(\cdot)$ to downweigh leverage points in the instrumental variables. The entries of the N-dimensional diagonal matrix $W_z(Z)$,

$$w_z(z_{it}) = \min\left(1, \frac{\sqrt{\chi^2_{0.975}(k)}}{RD_{it}}\right), \tag{3.24}$$

are based on robust distances. These distances RD_i are computed with the Minimum Volume Ellipsoid (MVE) estimates of location and scale which are approximated through application of algorithm R of section 1.3.2. Furthermore, by making the connection between the error term $\varepsilon_{it} = y_{it}-x_{it}\beta$ and the second round 2SGM residuals, we propose multiplying both the response variable y_{it} and the explanatory factors x_{it} by the square root of the 2SGM weighting functions $w_x(x_{it})$ and $w_r(r_{it})$.[5] Note that we compute new weights for x_{it} using formula (3.24) since the

[4] Capital letter N defines the total number of observations.

[5] Recall that the computation of the HBP GM estimates is equivalent to applying OLS on weighted data (see formula (1.14) of section 1.2.4). The weighted data are constructed by multiplying both y_{it} and x_{it} by $\sqrt{w_x(x_{it})w_r(r_{it})}$.

second step 2SGM weights are based on $\hat{x}_{it}$.[6] The residuals are calculated as $r_{it} = y_{it} - x_{it}\hat{\beta}_{2SGM}$. The elements of the N-dimensional diagonal matrix $W_r(r)$ are founded on Tukey's bi-square function according to,

$$\begin{aligned} w_r(r_{it}) &= 0 && |r_{it}/\sigma_t| \geq c \\ &= (1-(r_{it}/(\sigma_t c))^2)^2 && |r_{it}/\sigma_t| < c. \end{aligned} \tag{3.25}$$

The standard deviation σ_t of the errors in period t is estimated by computation of the Median Absolute Deviation (MAD) of the n residuals r_{it}, $i=1,..,n$. Altogether, the robust normal equation becomes

$$X^T\sqrt{W_x(X)W_r(r)}\, W_z(Z)Z\, (Z^T W_z(Z)\Omega W_z(Z)Z)^{-1}\, Z^T W_z(Z)\sqrt{W_x(X)W_r(r)}\,(y-X\beta) = 0. \tag{3.26}$$

The appropriate covariance matrix Ω now is the one which consists of the variances and covariances of the weighted errors $\varepsilon_{it}\sqrt{w_r(\varepsilon_{it})w_x(x_{it})}$, i.e. $\Omega = E[\sqrt{W_r(\varepsilon)W_x(X)}\,\varepsilon\varepsilon^T\sqrt{W_x(X)W_r(\varepsilon)}\,]$.

Following White (1982), estimates for the diagonal components of the covariance matrix Ω are obtained by insertion of the squared weighted 2SGM residuals $r_{it}^2 w_x(x_{it})w_r(y_{it}-x_{it}\hat{\beta}_{2SGM})$ into the diagonal of $\hat{\Omega}$. In the circumstance of a large number of cross-sectional units n with respect to a relatively small number of time periods T, Zellner's (1962) solution,

$$\hat{\omega}_{ts} = \frac{1}{n}\sum_{i=1}^{n} r_{ti}\sqrt{w_r(r_{ti})w_x(x_{it})}\; r_{si}\sqrt{w_r(r_{si})w_x(x_{si})}\,, \tag{3.27}$$

is employed to obtain consistent estimates of the non-diagonal elements ω_{ts} of Ω which indicate the covariance between the weighted errors $\varepsilon_{it}\sqrt{w_r(\varepsilon_{it})w_x(x_{it})}$ of period t and the corresponding ones of period s. Non-diagonal elements of $\hat{\Omega}$ are zero for those which correspond to the covariances between different cross-sectional units. Solving the first order condition (3.26) with respect to β returns the Robust Generalized Method of Moments (RGMM) estimator

$$\begin{aligned} \beta_{RGMM} = &\left(X^T\sqrt{W_x(X)W_r(r)}\, W_z(Z)Z\, (Z^T W_z(Z)\hat{\Omega}W_z(Z)Z)^{-1}\, Z^T W_z(Z)\sqrt{W_x(X)W_r(r)}\, X\right)^{-1} \\ &\left(X^T\sqrt{W_x(X)W_r(r)}\, W_z(Z)Z\, (Z^T W_z(Z)\hat{\Omega}W_z(Z)Z)^{-1}\, Z^T W_z(Z)\sqrt{W_x(X)W_r(r)}\, y\right). \end{aligned} \tag{3.28}$$

Formula (3.28) is accompanied by two remarks. First, if the instrumental variable matrix Z has the same dimension as X, i.e. $p=k$, then the computation of the RGMM estimator reduces to

[6] Instead one could replace X by $\hat{X}$ in the RGMM defining equation. Then the second step 2SGM weights $w_x(\hat{x}_{it})$ can be employed. However, the derivation of the variance-covariance matrix of the resulting RGMM estimator is not at all straight forward (see Section 3.2 on the derivation of the variance-covariance matrix of the 2SGM estimator).

$$\beta_{RGMM}\Big|_{p=k} = \left(Z^T W_z(Z)\sqrt{W_x(X)W_r(r)}\,X\right)^{-1}\left(Z^T W_z(Z)\sqrt{W_x(X)W_r(r)}\,y\right). \tag{3.29}$$

From (3.29) we deduce that an adjustment for heteroskedasticity or autocorrelation is no longer made. Therefore, we propose constructing a block-diagonal instrument matrix Z with diagonal components Z_t, t=1,..,T in the case that p=k and the hypotheses of homogeneity and/or independence of the disturbances are rejected. Our second comment concerns the unknown constant c. Following section 1.2.3 this critical value is set to c=4.685, since in that way the HBP GM estimator reaches an efficiency level of 95% in comparison with the classical Maximum Likelihood procedure if the errors, and so the 2SGM residuals r_{ti}, are normally distributed. One could fix c to a value so as to induce the RGMM estimator to achieve 95% of the efficiency level of the classical GMM method in clean data samples. We however do not follow this route because c would not only depend on the distribution of the error terms but also be a function of, among other arguments, the unknown correlation between X and Z. For the simulation experiments we have carried out in the subsequent section, we find that for the choice of c=4.685 the efficiency of the RGMM estimator fairly approximates the 95% GMM level in the smooth data cases for different sets of instruments.

In order to derive the variance-covariance matrix of the RGMM estimator we approximate the RGMM normal equation with a Taylor expansion around β_{RGMM}=β:

$$\begin{aligned} 0 &= X^T\sqrt{W_x(X)W_r(r)}\,W_z(Z)Z\,(Z^T W_z(Z)\hat{\Omega}W_z(Z)Z)^{-1}\,Z^T W_z(Z)\sqrt{W_x(X)W_r(r)}\,(y-X\beta_{RGMM}) \\ &\approx X^T\sqrt{W_x(X)W_r(\varepsilon)}\,W_z(Z)Z\,(Z^T W_z(Z)\hat{\Omega}(\beta)W_z(Z)Z)^{-1}\,Z^T W_z(Z)\sqrt{W_x(X)W_r(\varepsilon)}\,\varepsilon - D(\beta_{RGMM}-\beta), \end{aligned} \tag{3.30}$$

where

$$D = \frac{\partial}{\partial\beta^T}\left(X^T\sqrt{W_x(X)W_r(r)}\,W_z(Z)Z\,(Z^T W_z(Z)\hat{\Omega}W_z(Z)Z)^{-1}\,Z^T W_z(Z)\sqrt{W_x(X)W_r(r)}\,r\right)_{\beta_{RGMM}=\beta}. \tag{3.31}$$

For large samples in n the matrix D can be consistently estimated by evaluating $D(\beta)$ in β_{2SGM}:

$$\begin{aligned} \hat{D} &= X^T\sqrt{W_x(X)}\,(\partial\sqrt{W_r(r)}\,/\partial r)PX \\ &+ X^T\sqrt{W_x(X)W_r(r)}\,W_z(Z)Z\,(Z^T W_z(Z)\hat{\Omega}W_z(Z)Z)^{-1}\,Z^T W_z(Z)\sqrt{W_x(X)}\,(\partial\sqrt{W_r(r)}\,/\partial r)RX \end{aligned} \tag{3.32}$$

$$+ X^{T}\sqrt{W_x(X)W_r(r)}\, W_z(Z)Z\,(Z^{T}W_z(Z)\hat{\Omega}W_z(Z)Z)^{-1}\, Z^{T}W_z(Z)\sqrt{W_x(X)W_r(r)}\, X.$$

P and R are N-dimensional diagonal matrices containing

$$z_{it}w_z(z_{it})(Z^{T}W_z(Z)\hat{\Omega}W_z(Z)Z)^{-1}Z^{T}W_z(Z)\sqrt{W_x(X)W_r(r)}\, r, \quad t=1,..,T,\ i=1,..,n \tag{3.33}$$

and $(y_{11}-x_{11}\beta_{2SGM}),...,(y_{Tn}-x_{Tn}\beta_{2SGM})$ respectively. $\partial\sqrt{W_r(r)}\,/\partial r$ is a diagonal matrix of dimension N with elements

$$\begin{aligned}\partial\sqrt{w_r(r_{it})}\,/\partial r_{it} &= 0 && |r_{it}/\hat{\sigma}_t|\geq c\\ &= -2r_{it}/(\hat{\sigma}_t c)^2 && |r_{it}/\hat{\sigma}_t|<c.\end{aligned} \tag{3.34}$$

On the basis of (3.30), we suggest computing the asymptotic heteroskedastic consistent variance-covariance matrix of the RGMM estimator as

$$vcov(\beta_{RGMM}) = \hat{D}^{-1}\hat{M}(\hat{D}^{-1})^{T}, \tag{3.35}$$

where $\hat{M} = X^{T}\sqrt{W_x(X)W_r(r)}\, W_z(Z)Z\,(Z^{T}W_z(Z)\hat{\Omega}W_z(Z)Z)^{-1}\, Z^{T}W_z(Z)\sqrt{W_x(X)W_r(r)}\, X.$

Analogously to the argument of Proposition 1 of Section 3.2 we impose the following conditions, in addition to (*A*.1)-(*A*.3), in order to secure the identification and consistency of the RGMM technique:

(*A*.4) $\underset{N\to\infty}{plim}\left(\frac{1}{N}\hat{D}\right)$

exists and converges to a finite positive definite matrix of dimension p,

(*A*.5) $\underset{N\to\infty}{plim}\left(\frac{1}{N}X^{T}\sqrt{W_x(X)W_r(r)}\, W_z(Z)Z\right) = A$

exists and has full rank p,

(*A*.6) $\underset{N\to\infty}{plim}\left(\frac{1}{N}Z^{T}W_z(Z)\hat{\Omega}W_z(Z)Z\right) = H$

exists and converges to a finite positive definite matrix of dimension k,

(*A*.7) $\underset{N\to\infty}{plim}\left(\frac{1}{N}\sum_{i=1}^{n}Z_i^{T}W_z(Z_i)\sqrt{W_x(X_i)}\,\frac{\partial(\sqrt{W_r(r_i)}\,R_i)}{\partial r_i}X_i\right) = B$

exists and has full rank p, where R_i is a T-dimensional diagonal matrix consisting of $r_{1i},..,r_{Ti}$.

Furthermore,

$$(A.8)\quad \left(\frac{1}{\sqrt{N}} Z^{T} W_{z}(Z)\sqrt{W_{x}(X)W_{r}(r)}\, r\right)_{\beta} \sim N(0,C),$$

where $C = E[Z_i^T W_z(Z_i)\hat{\Omega}_i W_z(Z_i)Z_i]$, $i=1,...,n$; $Z_i = (z_{1i}^T,...,z_{Ti}^T)^T$ and the T-dimensional symmetric matrix $\hat{\Omega}_i$ is defined similarly. Large sample properties of the classical GMM estimator are established in Hansen's seminal paper of 1982.

Proposition 2.

The RGMM estimator as defined in (3.28) is consistent and asymptotic normally distributed if conditions (*A*.1)-(*A*.8) hold.

Proof:

By solving the Taylor expansion (3.30) for $(\hat{\beta}_{RGMM}-\beta)$ and applying Slutsky's theorem when taking the probability limit of both sides of the resulting equation we obain:

$$\underset{N\to\infty}{plim}\left(\hat{\beta}_{RGMM}-\beta\right) = \underset{N\to\infty}{plim}\left(\frac{1}{N}D\right)^{-1} \underset{N\to\infty}{plim}\left(\frac{1}{N}X^{T}\sqrt{W_{x}(X)W_{r}(\varepsilon)}\, W_{z}(Z)Z\right) \tag{3.36}$$

$$\underset{N\to\infty}{plim}\left(\frac{1}{N}Z^{T}W_{z}(Z)\Omega W_{z}(Z)Z\right)^{-1} \underset{N\to\infty}{plim}\left(\frac{1}{N}Z^{T}W_{z}(Z)\sqrt{W_{x}(X)W_{r}(\varepsilon)}\,\varepsilon\right).$$

Under assumptions (*A*.4)-(*A*.6) equation (3.36) is well defined. The last term of the right hand side of (3.36) is equal to zero because of assumption (*A*.1). The consistency of the RGMM estimator is therefore shown, $plim(\hat{\beta}_{RGMM}) = \beta$. The asymptotic normality of $\hat{\beta}_{RGMM}$ is conjectured from equation (3.36) when it is valid to apply a central limit theorem (assumption A.8).

The estimator $\hat{\beta}_{RGMM}$ is B-robust in the sense that it produces reliable parameter estimates even if small perturbations in the central model distribution F occur. The traditional GMM method does not possess this appealing property. In the latter case, just a single observation $q_{it}=(y_{it},x_{it},z_{it})$ may cause the GMM estimator to return arbitrary estimates.

Proposition 3.

The influence function, $IF(q;\hat{\beta}_{RGMM},F)$, associated with the RGMM estimator (3.28) is bounded if conditions (*A*.1) and (*A*.5)-(*A*.7) hold.

Proof:

Following the same steps taken when obtaining the influence function associated with the robust SUR estimator in section 2.6.2, we define $F_s = (1-s)F + sG$ and evaluate the first order condition

(3.26) under F_s. As usual, F reflects the central model distribution whereas outliers are generated by the distribution function G. Hence we get the following equality:

$$0 = \left((1-s)E_F\left[X_i^T\sqrt{W_x(X_i)W_r(r_i)}\, W_z(Z_i)Z_i\right] + sE_G\left[X_i^T\sqrt{W_x(X_i)W_r(r_i)}\, W_z(Z_i)Z_i\right]\right)$$

$$\left((1-s)E_F\left[Z_i^T W_z(Z_i)\hat{\Omega}_i W_z(Z_i)Z_i\right] + sE_G\left[Z_i^T W_z(Z_i)\hat{\Omega}_i W_z(Z_i)Z_i\right]\right)^{-1} \tag{3.37}$$

$$\left((1-s)E_F\left[Z_i^T W_z(Z_i)\sqrt{W_x(X_i)W_r(r_i)}\, r_i\right] + sE_G\left[Z_i^T W_z(Z_i)\sqrt{W_x(X_i)W_r(r_i)}\, r_i\right]\right).$$

The influence function $IF(q;\hat{\beta}_{RGMM},F)$ is acquired by differentation of (3.37) with respect to s and evaluation of the resulting equation in s=0, using the assumptions $(A.1)$ and $(A.5)$-$(A.7)$:

$$IF(q;\hat{\beta}_{RGMM},F) = \left(E_F\left[X_i^T\sqrt{W_x(X_i)W_r(r_i)}\, W_z(Z_i)Z_i\right] E_F\left[Z_i^T W_z(Z_i)\hat{\Omega}_i W_z(Z_i)Z_i\right]^{-1}\right. \tag{3.38}$$

$$\left. E_F\left[Z_i^T W_z(Z_i)\sqrt{W_x(X_i)}\,\frac{\partial(\sqrt{W_r(r_i)}\, R_i)}{\partial r}X_i\right]\right)^{-1} \left(E_F\left[X_i^T\sqrt{W_x(X_i)W_r(r_i)}\, W_z(Z_i)Z_i\right]\right.$$

$$\left. E_F\left[Z_i^T W_z(Z_i)\hat{\Omega}_i W_z(Z_i)Z_i\right]^{-1} E_G\left[Z_i^T W_z(Z_i)\sqrt{W_x(X_i)W_r(r_i)}\, r_i\right]\right).$$

By choosing G equal to the distribution that puts mass 1 at the point $q=(y_{ti},x_{ti},z_{ti})$, we obtain

$$IF(q;\hat{\beta}_{RGMM},F) = (A(H^{-1})B)^{-1}A(H^{-1})z_{ti}^T w_z(z_{ti})\sqrt{w_x(x_{ti})w_r(\varepsilon_{ti})}\,(y_{ti}-x_{ti}\beta). \tag{3.39}$$

The distribution function G is supposed to generate the worst possible outlier at the maximum of the influence function. The error $\varepsilon_{ti} = (y_{ti}-x_{ti}\beta)$ is bounded by the weight $\sqrt{w_x(x_{ti})w_r(\varepsilon_{ti})}$ while $w_z(z_{ti})$ downweighs aberrant instruments. Thus the influence function (3.39) associated with $\hat{\beta}_{RGMM}$ is bounded.

Section 3.4 An Examination of the Performance of the Robust GMM estimator in Simulated Panel Data Sets With Measurement Errors in the Explanatory Variables

Griliches and Hausman (1986) use the traditional GMM technique in a panel data framework with measurement errors in the explanatory factors.[7] In this section, the performance of the Robust GMM estimator in such an environment is analysed. We closely follow the estimation strategy recommended by Griliches and Hausman (1986) but employ the robust methods of Section 3.2 and 3.3 in order to diminish the excessive impact of outlying observations.

Consider the linear panel data model

$$y_{ti} = \eta_i + x_{ti}\beta + \varepsilon_{ti}, \tag{3.40}$$

where ε_{ti} is the error term associated with cross-sectional unit i in period t and η_i is the ith individual specific fixed effect. There are again T periods and n individual units. Equation (3.40) represents the so-called covariance model or individual dummy variables model. It can be contrasted with the error components model and random coefficients model where the intercept is random (see Mátyás and Sevestre, pp. 26-28). Suppose, however, that the true equation reads

$$y_{ti} = \eta_i + s_{ti}\beta + \xi_{ti} \tag{3.41}$$

instead of the one presented in (3.40), where x_{ti} is the observable erroneous reflection of the unobservable variables s_{ti},

$$x_{ti} = s_{ti} + v_{ti}. \tag{3.42}$$

Then both first difference and within estimation of model (3.40) result in downward biased parameters because of the negative correlation between the composite disturbance term

$$\varepsilon_{ti} = \xi_{ti} - v_{ti}\beta \tag{3.43}$$

and the explanatory factors x_{ti}; ξ_{ti} are error terms and v_{ti} is a p-dimensional row vector of measurement errors. The Wu-Hausman statistic (see Wu (1973) and Hausman (1978))

$$h = (\beta_{GM}^{W} - \beta_{GM}^{D})^{T}(vcov(\beta_{GM}^{W}) - vcov(\beta_{GM}^{D}))^{-1}(\beta_{GM}^{W} - \beta_{GM}^{D}) \tag{3.44}$$

can be consulted in order to test the null hypothesis of the absence of measurement errors in the

[7] Arellano and Bond (1988) discuss a similar estimation methodology for panel data (computer package DPD). This method also lacks correction for outliers.

independent variables.[8] β_{GM}^{W} and β_{GM}^{D} are the HBP GM estimates in case of the within and the first difference regression respectively.

A standard econometric technique for regression models with endogenous or erroneously measured variables is instrumental variable estimation. When estimating panel data model (3.40), lagged and future values of the observable explanatory variables x_{it} may serve to construct instrumental variables z_{it},

$$Z = (P \otimes I_n)X. \tag{3.45}$$

I_n is the identity matrix of dimension n. The T-dimensional square matrix P must satisfy three sets of conditions:

$(C.1)$ $l^T P = 0$,

$(C.2)$ $E[Z^T X] = E[Z^T S]$, *i.e.* $E[X^T (P \otimes I_n)^T V] = 0$,

$(C.3)$ $E[X^T (P \otimes I_n)^T X] \neq 0$,

where l is a T-dimensional column vector with entries 1; $\otimes$ is the symbol of the Kronecker product. The first restriction is necessary to eliminate the fixed effects. The second requirement implies that the instrumental variables z_{it} should be uncorrelated with the measurement errors v_{it} and finally z_{it} should have a non-zero correlation with x_{it}. Note that by choosing the correct entries of the matrix P both the within and the first difference estimator can be written in the format $(Z^T X)^{-1} Z^T y$, of the 2SLS instrumental variable estimator.

An optimal set of instruments Z will contain all non-redundant instruments z_j, $j=1,\ldots,k$ which satisfy conditions $(C.1)$-$(C.3)$. An instrument z_j is called non-redundant if it provides information which is not delivered by the other instruments of the set. The number of non-redundant instruments equals the number of linearly independent P matrices which satisfy the three conditions stated above. Restriction $(C.1)$ requires that the sum of the elements of each column of P is equal to zero. This imposes T linear restrictions on the choice of P. In the case that the design matrix X is one-dimensional, condition $(C.2)$ can be rewritten as

$$\underset{n\to\infty}{plim} \frac{1}{n} \sum_{i=1}^{n} \sum_{r=1}^{T} \sum_{t=1}^{T} v_{it} v_{ti} P_{ti} = 0. \tag{3.46}$$

---

[8] h is asymptotically χ^2-distributed with p degrees of freedom (see Section 2.3).

Table 3.1 List of Instruments For Model (3.40) Under the Assumption of Identically Distributed Stationary Measurement Errors v_{it}, $T=4$ [a]

Difference Equations to be Estimated	Valid Instruments Under Uncorrelated Measurement Errors (β^{U}_{RGMM})	Valid Instruments Under Measurement Errors which are Correlated up to a MA(2) process (β^{C}_{RGMM})
$(y_{2i}-y_{1i}) = (x_{2i}-x_{1i})\beta$	$x_{1i}+x_{2i}$, x_{4i}	$x_{1i}+x_{2i}$
$(y_{3i}-y_{2i}) = (x_{3i}-x_{2i})\beta$	x_{1i}, $x_{2i}+x_{3i}$, x_{4i}	$x_{2i}+x_{3i}$
$(y_{4i}-y_{3i}) = (x_{4i}-x_{3i})\beta$	x_{1i}, $x_{3i}+x_{4i}$	$x_{3i}+x_{4i}$
$(y_{3i}-y_{1i}) = (x_{3i}-x_{1i})\beta$	x_{2i}	x_{2i}
$(y_{4i}-y_{2i}) = (x_{4i}-x_{2i})\beta$	x_{3i}	x_{3i}
$(y_{4i}-y_{1i}) = (x_{4i}-x_{1i})\beta$	x_{2i}, x_{3i}	
$\begin{pmatrix} y_{3i}-y_{2i} \\ y_{4i}-y_{1i} \end{pmatrix} = \begin{pmatrix} x_{3i}-x_{2i} \\ x_{4i}-x_{1i} \end{pmatrix}\beta$		$\begin{pmatrix} -x_{1i} \\ x_{2i} \end{pmatrix}, \begin{pmatrix} -x_{4i} \\ x_{3i} \end{pmatrix}$
$\begin{pmatrix} y_{2i}-y_{1i} \\ y_{4i}-y_{3i} \end{pmatrix} = \begin{pmatrix} x_{2i}-x_{1i} \\ x_{4i}-x_{3i} \end{pmatrix}\beta$		$\begin{pmatrix} x_{3i} \\ x_{2i} \end{pmatrix}, \begin{pmatrix} x_{4i} \\ x_{1i} \end{pmatrix}$
Total Number of Instruments	11	9

[a] Table 3.1 corresponds with Table 2 of Griliches and Hausman (1986) in condensed form.

Condition (3.46) reveals that the sum of the diagonal elements of P must equal zero if the measurement errors v_{it} are identically distributed and stationary.[9] For correlated errors following a Kth Moving Average (MA) process K additional restrictions have to be imposed on P. If we are willing to assume that condition ($C.3$) is not binding and that measurement errors associated with different columns of X are uncorrelated then the optimal number of instruments is equal to $T^2-(T+K+1)$ when the measurement errors associated with each column of X are identically distributed, stationary and correlated up to a MA(K) process. In the appendix A3 we clarify why optimal instrumental variable estimation of model (3.40), when the panel data comprise four periods, comes down to applying 2SLS (2SGM) on the difference equations of Table 3.1 (see also Table 2 of Griliches and Hausman (1986)) using the indicated instruments.

The difference equations of Table 3.1 are estimated with the help of the 2SGM method instead of the conventional instrumental variable technique 2SLS because of the argument that outliers may corrupt classical estimation methods. We employ the Robust Generalized Method of

[9] In case of non-stationary measurement errors, T extra conditions restrict the choice of P.

Moments Estimator of Section 3.3 in order to combine the 2SGM estimates of the differentdifference equations of Table 3.1 optimally. With this end in view, consider the *super* model of stacked difference equations

$$\tilde{y} = \tilde{X}\beta+\tilde{\varepsilon}, \tag{3.47}$$

where $\tilde{X}$ and $\tilde{y}$ are the stacked differenced independent and dependent variables respectively. It is assumed that the stacked errors $\tilde{\varepsilon}$ obey conditions $(B.1)$-$(B.4)$. The variance-covariance structure of these disturbances is therefore described by $E[\tilde{\varepsilon}\tilde{\varepsilon}^T] = \Sigma$. To summarize, by estimating model (3.47) with the B-robust estimator $\hat{\beta}_{RGMM}$ we make corrections for conditional heteroskedasticity, serial correlation in the regression errors, measurement errors in the independent variables and, last but not least, outlying observations.

Note that application of the compound 2SLS/GLS estimator, $(\hat{X}^T\hat{\Sigma}^{-1}\hat{X})^{-1}\hat{X}^T\hat{\Sigma}^{-1}y$, does not necessarily provide consistent estimates of the super model since the disturbances of one equation are not necessarily orthogonal to the explanatory variables or instruments of another equation.

We have conducted a limited Monte Carlo study in order to examine the consistency, efficiency and distribution of the various B-robust and classic estimators which have been mentioned so far. Consider the following data generating process:

Experiment (IV):

$$y_{ti} = \eta_i+s_{ti}\beta+\varepsilon_{ti},$$

$$s_{1i} = 3s^*_{1i}+s^*_{2i}+s^*_{3i}+s^*_{4i},$$

$$s_{2i} = s^*_{1i}+3s^*_{2i}+s^*_{3i}+s^*_{4i},$$

$$s_{3i} = s^*_{1i}+s^*_{2i}+3s^*_{3i}+s^*_{4i},$$

$$s_{4i} = s^*_{1i}+s^*_{2i}+s^*_{3i}+3s^*_{4i},$$

$$x_{ti} = s_{ti}+v_{ti},$$

where the columns of s^*_{1i}, s^*_{2i}, s^*_{3i}, s^*_{4i} are independent and identically normally distributed with variance 1/4 (*iid* N(0,1/4)); each column of v_{ti} and ε_{ti} is *iid* N(0,1); η_i are identically and uniformly distributed with variance 1; $S = (S_1^T,...,S_4^T)^T$; T=4, n=250 and p=2; The entries of the p-dimensional column vector β are equal to 1.

The explanatory factors in experiment (IV) are subject to stationary and uncorrelated measurement errors. To be in accordance with Table 3.1, the number of periods T equals four. The construction of the independent variables is such that lagged and future values of x_{ti} are valid

Table 3.2 Replication 1 of Experiment (IV), p=2[a]

	β^{W}_{GM}	β^{D}_{GM}	β^{U}_{RGMM}	β^{C}_{RGMM}	β^{U}_{GMM}	β^{C}_{GMM}
			0% Pollution			
X	0.942 (69.98)	0.856 (37.68)	0.999 (80.72)	0.998 (75.41)	0.996 (75.32)	0.999 (71.86)
	0.927 (60.98)	0.820 (31.06)	1.009 (75.23)	1.004 (64.49)	1.008 (72.94)	1.005 (66.76)
h	45.23	45.23				
			0.45	0.45		
					0.82	0.82
			0.64		0.64	
				0.0001		0.0001
			10% Pollution			
X	0.890 (44.44)	0.850 (31.17)	0.992 (65.25)	0.992 (55.56)	0.895 (20.46)	0.853 (15.00)
	0.835 (26.35)	0.738 (15.15)	1.003 (55.89)	0.987 (43.11)	1.054 (16.14)	1.057 (13.38)
h	17.59	17.59				
			1.40	1.40		
					1.39	1.39
			5.72		5.72	
				7.20		7.20

[a] t-values are within parentheses; $t(\infty)_{0.025}=1.96$, $\chi^2_{0.05}(2)=5.99$. β^{W}_{GM} and β^{D}_{GM} are the HBP GM estimates in case of *within* and *first difference* regression respectively. β^{U}_{RGMM} and β^{C}_{RGMM} are the Robust GMM estimates in case of *uncorrelated* stationary measurement errors and *MA(2)* stationary measurement errors respectively. Table 3.2 reports the standard Wu-Hausman (h) statistic for the cases($\beta^{W}_{GM}\leftrightarrow\beta^{D}_{GM}$), ($\beta^{U}_{RGMM}\leftrightarrow\beta^{C}_{RGMM}$), ($\beta^{U}_{GMM}\leftrightarrow\beta^{C}_{GMM}$), ($\beta^{U}_{RGMM}\leftrightarrow\beta^{U}_{GMM}$) and ($\beta^{C}_{RGMM}\leftrightarrow\beta^{C}_{GMM}$).

instruments (i.e. condition ($C.3$) is fulfilled). Since the calculation of the robust GMM estimates requires a substantial amount of computing time, we replicated experiment (IV) only 500 times for both *clean* and *polluted* data. Contaminated data sets are constructed by replacing ten percent (randomly drawn) of the observations (comprising the response and explanatory variables X) by drawings from three independent normal random variable with mean zero and variance 100. Consequently, vertical outliers in view of the relationships (3.2) and (3.40) may occur and leverage points are added to both the explanatory factors and the instruments.

The HBP GM estimator, either in the within or the first difference dimension of the data, produces estimates which are far away from the true parameter values as is shown in Table 3.2 for

Table 3.3 Simulation Results Concerning Efficiency, Consistency and Distribution of the RGMM Estimator When Measurement Errors in the Explanatory Variable Are Present, Experiment (IV), p=2[a]

β^{W}_{GM}	β^{D}_{GM}	β^{U}_{RGMM}	β^{C}_{RGMM}	β^{U}_{GMM}	β^{C}_{GMM}
		0% Pollution			
0.931682	0.848834	0.999240	1.000222	0.997243	0.997765
0.004937	0.023388	0.000301	0.000335	0.000304	0.000320
0.000270	0.000536	0.000301	0.000335	0.000296	0.000315
0.931101	0.846582	1.000179	1.000564	0.998163	0.998077
0.005020	0.024058	0.000306	0.000347	0.000291	0.000313
0.000273	0.000521	0.000306	0.000347	0.000288	0.000309
4.789	4.295	1.460	1.756	3.923	1.799
0.476	0.484	2.633	1.516	5.338	2.024
5.265	4.779	4.094	3.272	9.260	3.823
		10% Pollution			
0.861685	0.822615	0.995733	0.995489	0.963497	0.968178
0.020072	0.032732	0.000451	0.000497	0.005242	0.005436
0.000941	0.001267	0.000433	0.000476	0.003910	0.004423
0.862850	0.820582	0.997569	0.997033	0.964573	0.967978
0.019861	0.033708	0.000472	0.000554	0.005315	0.006162
0.001051	0.001517	0.000466	0.000545	0.004060	0.005136
2.190	2.553	1.522	0.116	18.139	21.030
7.594	1.154	0.715	0.898	17.692	49.610
9.784	3.707	2.237	1.014	35.831	70.640

[a] Each cell of the rows 3 and 6 contains the mean, the mean squared error and the variance of the estimated parameter values over the 500 runs. The length of the series y equals 1000 and $\beta_1=\beta_2=1$. β^{W}_{GM} and β^{D}_{GM} are the HBP GM estimates in case of *within* and *first difference* regression respectively. β^{U}_{RGMM} and β^{C}_{RGMM} are the Robust GMM estimates in case of *uncorrelated* stationary measurement errors and *MA(2)* stationary measurement errors respectively (β^{U}_{GMM} and β^{C}_{GMM} are the corresponding estimates which are based on GMM (*non-robust*), see Table 3.1). Each cell of the rows 4 and 7 contains the Doornik-Hansen (1994) statistics of symmetry, kurtosis and normality which are respectively distributed $\chi^2(2)$, $\chi^2(2)$ and $\chi^2(4)$. $\chi^2_{0.05}(2)=5.99$, $\chi^2_{0.05}(4)=9.49$.

the first replication of the contaminated and clean data sets. As predicted by Griliches and Hausman in 1986, our estimation results confirm that

> Errors of measurement will usually bias the first difference estimators downward (toward zero) by more than they will bias the within estimators (Griliches-Hausman (1986), p.95).

In agreement with this assertion the Wu-Hausman statistic rejects the null hypothesis of equivalence of the within and first difference estimates. In fact, the robust GMM estimators which correct for *uncorrelated* stationary measurement errors (β^{U}_{RGMM}) and *MA(2)* stationary measurement errors (β^{C}_{RGMM}) respectively, produce estimates close to one, that is to true β. Their classical counterparts, however, are not capable of returning estimates which are close to the parameters of the data generating process if anomalous observations are present. Evidently, outliers are responsible for these biased results. Notice that according to the Wu-Hausman test, the Robust GMM and the traditional GMM estimates for the first replication are not significantly different if the data are free from aberrant points. The estimates generated by β^{U}_{RGMM} and β^{C}_{RGMM} are in none of the cases presented in Table 3.2 significantly different as is again revealed by the Wu-Hausman test. One can interpret this result as a sign of the ability of the instrumental variables to solve the problem of non-orthogonality of the design matrix.

Table 3.3 reports the mean, the mean squared error and the variance of the HBP GM within and first difference estimates, the RGMM estimates in the case of uncorrelated and correlated measurement errors and the corresponding non-robust GMM versions over the 500 runs. The HBP GM estimators produce significantly biased estimates in all cases considered. It is of no surprise that, among the GMM estimators, the variance of the estimates is the lowest for the classical GMM method, using the set of instruments based on non-correlated and stationary measurement errors, since this technique provides the Uniformly Minimum Variance Unbiased estimator under the central model data generating process outlined above. Although the differences in efficiency between the B-robust and standard GMM estimators are small. The RGMM estimator β^{U}_{RGMM} obtains a relative efficiency level of 98.3% and 94.1% for the explanatory variables x_1 and x_2 respectively as is revealed by the division of the corresponding GMM and RGMM variances. More favorable efficiency characteristics in smooth data can obviously not compensate for the decay of the standard GMM estimator in polluted panel data sets. When outliers corrupt the data, β^{U}_{GMM} and β^{C}_{GMM} break down as can be seen from the relatively high mean squared errors and apparent inconsistency of the estimates. The breakdown point of the traditional GMM estimator is nil, i.e. only one observation may cause the estimator to produce any estimates. The robust alternatives β^{U}_{RGMM} and β^{C}_{RGMM} still provide accurate estimates for the 10% pollution case of Table 3.3. The robust GMM estimator which corrects for stationary and uncorrelated measurement errors is more efficient than the other which is founded on the set of instruments for stationary and correlated measurement errors. This result is inherent in the way experiment (IV) is built.

The Doornik-Hansen (1994) multivariate statistic is employed to test whether the parameter estimates are normally distributed. This finite sample test has a χ^2-distribution with four degrees of freedom. Our results presented in Table 3.3 indicate that the bounded influence estimators, HBP GM and RGMM, are normally distributed. If the data is contaminated with outliers then the GMM estimator no longer possesses the normality property. The Doornik-Hansen test values of 35.83 and 70.64 corresponding with β^{U}_{GMM} and β^{C}_{GMM} respectively, substantially exceed the critical value of 9.49 at the 5% significance level. The important implication of the rejection of the normality

hypothesis is that not only the classical GMM parameter estimates but also the concomitant standard errors and t-values are misleading.

To conclude, the B-robust Generalized Method of Moments technique developed in this chapter is to be preferred to classical estimation procedures when the explanatory factors are subject to measurement errors and data are contaminated.

Section 3.5 Summary and Outlook

The textbook approach towards econometric modelling is inadequate when economic theory does not coincide one-to-one with the statistical representation of the phenomenon under interest. Robust statistics provide a solution to reliable estimation and inference in this situation by considering neighborhoods of the standard parametric models. A favorable estimator within the class of robust estimators is the High Breakdown Point Generalized M-estimator which employs non-parametric tools to trace outlying observations and to diminish their impact on the estimates of the parametric model. The HBP GM estimator, however, only accounts for freaks in the data but leaves other important aspects of non-experimental data aside. If the fundamental orthogonality condition of the design is at issue, this estimator produces inconsistent estimates. The B-robust 2SGM and RGMM estimators which are presented in this chapter do not display this deficiency. The latter is more efficient than the first if the model errors exhibit heteroskedasticity or autocorrelation. Our simulation results stress the superiority of the RGMM estimator in panel data sets with erronously measured explanatory factors in comparison to the classical generalized method of moments.

One critique often put forward by econometricians who are sceptical about robust methods is that robust estimators diminish the influence of outliers without taking into consideration the reasons for their peculiarity. Outlying observations stemming from copying or computation errors can be appropriately corrected and included in the data. If the theoretical model only applies to a sub set of the data one could choose, for example, a switching regime model (e.g. see Hansen's (1996) threshold regression method) or statistical models with mixtures of distributions (see, inter alia, Titterington, Smith and Makov (1985), Robert (1994)). The first argument is implicitly founded on the assumption that it is possible to distinguish outlying observations from regular points in the data at hand. The use of robust estimators, however, is indispensable in discovering these outliers because of the masking effects which were already mentioned in the introduction. The second argument deserves more attention. In the presence of outliers, the threshold regression method may not only produce inadequate parameter estimates but can also make the wrong division of the sample into regimes. It seems fruitful to robustify the threshold regression technique since the switching regime model suits many economic analyses. Robert (1996) noted that, in the light of mixtures of classical distributions,

> Their appeal goes beyond the mere modelling of heterogeneous populations with homogeneous subgroups as in discrimination and outlier detection, since this modelling covers a wide range reaching towards nonparametric statistics (Robert (1996), front page).

Comparison and integration of robust methods and models with mixtures of distributions is another interesting topic for future research.

Appendix A3 The Choice of Instruments

Griliches, Hausman and Meyer (see the appendix of Griliches and Hausman (1986)) show for p=1 that the optimal number of instruments in case of stationary, identically distributed and uncorrelated measurement errors is equal to $T^2-(T+1)$ when estimating model (3.40) of Section 3.4. There are eleven transformation matrices P which fulfill the requirements $C1$-$C3$ if T=4. In this appendix we give examples of two such P matrices to clarify Table 3.1.

For T=4, we have $y = (y_1^T,...,y_4^T)^T$, $X=(x_1^T,...,x_4^T)^T$ where $y_1,...,y_4$ are column vectors of dimension n. The P matrices corresponding to the first equation of Table 3.1, under the assumption of stationary, identically distributed and uncorrelated measurement errors (second column), are equal to

$$P_1 = \begin{matrix} -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \qquad \text{(A3.1)}$$

and

$$P_2 = \begin{matrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}. \qquad \text{(A3.2)}$$

Now we will demonstrate that estimation of

$$y = \eta\otimes l + X\beta + \varepsilon \qquad \text{(A3.3)}$$

with instrumental variables $Z = (P_1\otimes I_n)X$ is equivalent to applying 2SLS (2SGM) to the difference equation

$$y_2 - y_1 = (X_2 - X_1)\beta + \xi_1 \qquad \text{(A3.4)}$$

using the instrument X_1+X_2. I_n is the n-dimensional identity matrix and l is a T-dimensional vector with entries 1; The vector ξ_1 comprises error terms. The Two Stage Least Squares estimator simplifies to

$$\beta_{2SLS} = (Z^T X)^{-1} Z^T y \qquad \text{(A3.5)}$$

if X and Z have the same dimension. Inserting $Z = (P_1\otimes I_n)X$ into (A3.5) gives

$$\beta_{2SLS} = \frac{((P_1\otimes I_n)X)^T y}{((P_1\otimes I_n)X)^T X} = \frac{(X_1+X_2)^T(y_2-y_1)}{(X_1+X_2)^T(X_2-X_1)}. \qquad \text{(A3.6)}$$

Wagenvoort, Rien J.L.M. (1998), Robust estimation of panel data : with an application to investment equations
European University Institute DOI: 10.2870/75660

which clarifies the second row of Table 3.1. Note that the 2SLS estimates of difference equation (A3.4) are calculated as

$$\beta_{2SLS} = ((X_2-X_1)^T Z(Z^T Z)^{-1} Z^T (X_2-X_1))^{-1}(X_2-X_1)^T Z(Z^T Z)^{-1} Z^T (y_2-y_1), \tag{A3.7}$$

where the instrumental variables are equal to $Z = (X_1+X_2, X_4)$ under the assumption of stationary, identically distributed and uncorrelated measurement errors. This formula can not be expressed in terms of P. The introduction of the P matrices, however, facilitate the discovery of the appropriate instruments for equation (A3.3).

PART II
INVESTMENT
AND FINANCIAL POLICY

PART II

INVESTMENT AND TECHNOLOGY

CHAPTER 4

A Literature Review of the Relevance of Financial Policy

Section 4.1 Modigliani and Miller's Proposition I

The purpose of a risk management program is to increase the value of a firm by altering its financial structure. For instance, *off-balance sheet hedging* instruments such as futures, forwards, options and swaps can be used to change the variability in future profits or cash flows. To the extent that trading in these financial assets reduces company's risk, hedging might be desirable although it creates another financial claim on the company. *On-balance sheet hedging* by means of improving the debt/equity ratio might prevent a situation of financial distress and as a consequence allows a firm to increase the amount of outside money when cash flow is temporarily insufficient to finance new investment projects. Before turning to the question, raised in Chapter 5, of whether Dutch manufacturing companies which are quoted on the Amsterdam stock exchange could benefit from pursuing risk management programs, we first give a brief literature review of the rationale for an active financial policy.

In the world of Modigliani and Miller (1958) there are no transaction costs, taxes or borrowing constraints and capital markets are perfect in the sense that any two assets which are perfect substitutes for each other must sell, in equilibrium, at the same price. Given this setting, Modigliani and Miller prove the irrelevance of financial policy in their famous Proposition I.

> *Proposition I (Modigliani and Miller (1958), p.268):*
> The market value of any firm is independent of its capital structure and is given by capitalizing its expected return at the rate ρ_k appropriate to its class.

Proposition I asserts that the investment and financing decisions can be taken separately from each other since the value of a firm is only determined by its real assets (the left-hand side of the balance sheet) and does not depend on the proportions of different financial claims on the company (the right-hand side of the balance sheet). The so-called principle of *value additivity* holds true; the value of the firm's securities combined is equal to the sum of their values considered separately.

The intuition behind Proposition I can be easily grasped by considering two companies which have identical real assets but differ in their financial structure. One company uses both stocks and debt to finance its investment projects whereas the other is unleveraged. An investor who follows strategy A buys 10% of the stock of the leveraged company whereas an investor of strategy B buys 10% of

the stock of the unleveraged company and borrows on his own account 10% of the value of the leveraged company's debt. Noting that both investment strategies have the same distribution of returns (0.1(profits-interest)), in perfect capital markets, portfolio A ($0.1E_L = 0.1(V_L-D_L)$) has the same price as portfolio B ($-0.1D_L+0.1V_U = 0.1(V_U-D_L)$), where the symbols E_L,D_L,V_L and V_U denote equity, debt, total value of the leveraged firm and the value of the unleveraged firm respectively. Consequently, the market values of the leveraged and unleveraged company are equal in equilibrium. Clearly, if the assumptions of Modigliani and Miller (1958) hold true, a firm can raise its value by dismissing the expensive risk managers since they only mess up the company's financial structure without contributing to the firm's profits. In fact, investors who care about their portfolio risk can always buy and sell hedging instruments themselves.

Since the pioneering work of Modigliani and Miller (1958), a substantial number of articles appeared where the irrelevance theorem is called into question. Nowadays, according to the traditional view, the optimal capital structure is found at the point where the weighted-average cost of capital is minimized. In the remainder of this chapter, we relax the assumptions of Modigliani and Miller one by one to gain insight into the failure of Proposition I. In Section 4.2 we follow the work of Detemple, Gottardi and Polemarchakis (1995), who claim that financial policy is relevant if the asset market is incomplete. In the subsequent section, game theoretical arguments are put forward to explain credit rationing in equilibrium. The empirical relevance of credit rationing is established and the implications of borrowing constraints for risk management are briefly analyzed. One of the more important reasons, why Proposition I often can not be applied, are tax shields. If future interest expenses on debt can be deducted from future taxes then the financial manager of a company should use an adjusted cost of capital, which takes into account the present value of the tax shield, when discounting the cash flows of an investment project (see e.g. Brealey and Myers (1988), Chapter 19). In this case, the acceptance of a new investment plan might depend on the way it can be financed. Mauer and Triantis (1994) find that the impact of debt financing on the firm's investment and operating decisions is economically insignificant. In Section 4.4 we will shed light on their arguments. Finally, Section 4.5 contains some concluding remarks.

Section 4.2 Incomplete Capital Markets

According to Detemple, Gottardi and Polemarchakis (1995) an asset market is *complete* if all contingent contracts are priced and all re-allocations of revenue are attainable.[1] This means that investors can actually take all conceivable positions in the contingent contracts at the prevailing prices. They show that, when the markets are incomplete, the range of attainable re-allocations of the firm's

[1] As noted by Detemple et al. (1995), the asset market may be incomplete as a result of differences in access to information sources between different players (the bank, stockholders, the management of the firm etc.). However, information asymmetry is not a necessary condition for incompleteness of the asset market.

revenue is dependent on the firm's financial policy if, for instance, there are secondary assets which have a payoff structure that is non-linear with the payoff to equity. To clarify this statement, consider an economy with a single firm which collects revenue $\alpha(2,s)$ depending on the state of nature (s=1,..,S) in period 2. Three different types of assets are sold on the financial market in the first period: equity, debt and a call option on the firm's equity. The debt of the firm is assumed to be riskless whereas the return on equity and the call option depend on the amount of revenue brought in by the company. The total dividend payed out to the stockholders of the company in period 2 is equal to

$$d(2,s) = \alpha(2,s)-D, \quad s=1,..,S. \tag{4.1}$$

where the gross return on the debt, D, is equal to the face value of the loan since we assume a zero risk-free rate of interest. For simplicity, there are only two periods in the model of Detemple et al. (1995), therefore, the payoff to the call option with exercise price k in the second period is equal to

$$(d(2,s)-k)^{+} = \max(d(2,s)-k,0) \tag{4.2}$$

since the value of the equity coincides with the dividend payoff $d(2,s)$. In summary, the matrix of asset payoffs in this simple two-period economy is

$$R(D) = \begin{matrix} \alpha(2,1)-D & (\alpha(2,1)-D-k)^{+} & D \\ \vdots & \vdots & \vdots \\ \alpha(2,s)-D & (\alpha(2,s)-D-k)^{+} & D \\ \vdots & \vdots & \vdots \\ \alpha(2,S)-D & (\alpha(2,S)-D-k)^{+} & D \end{matrix} \tag{4.3}$$

The columns of (4.3) represent the return on stock, options and debt respectively. The range of attainable redistributions of revenue is determined by the column span of asset payoffs, $[R(D)]$. Let us assume that the following assumptions hold:

Assumption 1. There are more than three different states of nature, i.e. the asset market composed of three assets is incomplete, S>3.

Assumption 2. Equity and debt are not perfect substitutes,

$$\alpha(2,S)>\alpha(2,S-1)\geq...>k>...\geq\alpha(2,2)>\alpha(2,1). \tag{4.4}$$

At this point it is easy to show that different financial policies generate matrices of asset payoffs with different column spans, ($[R(D)] \neq [R(\hat{D})]$ *if* $D \neq \hat{D}$). We conclude that the matrix consisting of the possible asset payoffs to the investor is dependent on the debt policy of the firm if the asset market is incomplete and a call option is written on the firm's stock.

Detemple et al. (1995) then proceed by introducing H consumers into the economy and define a financial policy to be *generically weakly relevant* if and only if distinct financial policies result in distinct consumption allocations almost everywhere and, if it does, it is robust to perturbations of the

economy. Financial policy is called *strongly relevant* if it also affects the market value of the firm. They show in their *proposition 1* that competitive equilibrium consumption allocations are distinct whenever the asset market is incomplete, if the number of agents exceeds three (H>3) and the agents are sufficiently diverse in endowments and/or preferences. Three examples of different economies are presented to highlight proposition 1. In example 1, where the consumers have identical preferences and endowments, the neutrality property of financial policy holds. In example 2 (variety in endowments), financial policy is weakly relevant since it affects the consumption allocations but does not influence the firm value whereas in example 3 (heterogenous preferences and different endowments) the strong result is obtained, i.e. the market value of the company also changes with the firm's financial strategy. Example 3 refutes Modigliani and Miller's Proposition I of Section 4.1 in the sense that, when comparing the values of an arbitrary firm across equilibria, changes in financial policy, in general, have an effect. As argued by Detemple et al. (1995), at a given equilibrium, the values of two identical companies which only differ in the capital structure must be the same if capital markets are perfect.

Financial policy is generically relevant whenever the asset market is incomplete. This may happen even when secondary assets such as options do not exist. For instance, when equity holders have limited liability, the payoff to equity depends non-linearly on the financial policy. For this case, the matrix of asset returns can be written as

$$R(D)^{*} = \begin{matrix} 0 & D \\ \vdots & \vdots \\ 0 & D \\ \alpha(2,s^{*})-D & D \\ \vdots & \vdots \\ \alpha(2,S)-D & D \end{matrix} \qquad (4.5)$$

where $\alpha(2,s^{*})\geq D$. Evidently, the column span of $R(D)^{*}$ is not independent of the financial policy D. Secondary assets and limited liability are observed in many financial markets. Therefore, it is highly unlikely that in practice the manager of the firm can separate his investment and financial decisions as predicted by Modigliani and Miller in 1958.

Section 4.3 Credit Constraints

4.3.1 A Theoretical Explanation For Credit Rationing Equilibria

One of the assumptions underlying Modigliani and Miller's Proposition I is free access to capital markets for every borrower. If investors are liquidity constrained or cannot borrow against the same terms as enterprises do then they might be willing to pay a premium for the stock of the leveraged firm in comparison with the unleveraged one. The question arises why lenders put up barriers other

than increases in interest rates to restrict the amount of loanable funds. After all, lenders could raise the price of credit in case of excess demand for loans. Before turning to this question we provide some arguments for why corporations have an incentive to enter debt markets.

A major reason for the existence of debt in a corporate firm can be found in the argument that internal funds and the issuing of new equity may raise insufficient funds to undertake the firm's projects. Myers and Majluf (1984) show that if managers with inside information about investment projects look after the interests of the existing stockholders then firms may refuse to issue new shares, even if it means giving up a project with a positive net present value.[2] In addition, there are several strategic reasons for the fact that many firms borrow funds on the credit market. As observed by Jensen and Meckling (1976), the management of a firm may undertake actions which are in conflict with the interests of the owners of the firm when the stockholders can neither perfectly nor costlessly monitor the management. One can think of managers who spend too much money on offices, cars and the secretarial staff. The stockholders can limit these aberrant actions of the managers by financing a part of the activities of the firm with debt. The monitoring activities of the bank might keep the management of the firm under control. Moreover, the managers fear bankruptcy even if the bank monitors poorly, as they lose their job if the firm goes bankrupt. Another reason may be found in the tax subsidy on interest payments (see Section 4.4).

Recent literature emphasizes the prevalence of imperfect information in financial markets. In particular, debt markets are generally characterized by asymmetric information since the borrower is better informed than the lender about the value of the project that will be undertaken. The lender may restrict the borrower's use of debt (e.g. see Jaffee and Russell (1976) and Stiglitz and Weiss (1981)) because of the problems of moral hazard and adverse selection. *Equilibrium* credit rationing was understood as a status quo of the credit market where the lender is not willing to change interest rates and to supply less or more funds to the borrowers even though credit is in excess demand. Generally speaking, if demand exceeds supply, then prices will rise until demand and supply are equated at the new Walrasian equilibrium price. However, the price of credit may have sorting and incentive effects. The expected return on loans is increasing initially but then decreasing in the interest rate charged since a higher price of capital causes a rise in the riskiness of the pool of loans.[3] Thus, given that the lender maximizes his expected profits, imperfect information in financial markets can lead to restrictions on the amount of debt.

Many authors writing about credit rationing use their own definition of a credit rationing equilibrium. Essentially, from these definitions it follows that a credit rationing equilibrium is

[2] Furthermore, the asymmetric information problems in the credit market (see below), which cause restricted debt markets, are believed to distort also the equity market (see Stiglitz (1988)).

[3] As is shown by de Meza and Webb (1987), if low-risk projects have a higher mean expected return than high-risk projects then banks experience favourable selection instead of adverse selection when raising the interest rate. Companies with high-risk projects do not apply for a loan when the interest rate is high. In this situation, a credit rationing equilibrium is not sustainable.

characterized by a group of borrowers who cannot obtain the amount of credit they wish. As formulated by Blanchard and Fischer (1989):

> *Type 1 (proportional)* credit rationing occurs when an individual cannot borrow as much as he or she wants at the going interest rate. *Type 2 (random)* credit rationing occurs when, among identical borrowers, some who wish to borrow are able to do so, while others cannot.

The moral hazard and adverse selection problems mentioned above follow from the fact that the management of the firm has private information about its investment project and actions, which is not available to the lender. A more recent idea which was also applied to the credit market is that of costly state verification. *Ex ante*, i.e. before uncertainty is resolved, the borrower knows as much as the lender, but there exists *ex post* asymmetry of information because the borrower is able to observe the return of his project costlessly, while the lender can not. Among others, Gale and Hellwig (1985) and Williamson (1986) show that equilibrium credit rationing may occur because of *ex post* asymmetric information.

It is important to notice that if the bank can sort the borrowers according to specific characteristics, the adverse selection problem is reduced. The interest rate is not the only term of the contract between the lender and the borrower, the amount of the loan and the amount of collateral or equity the bank demands of loan applicants will also affect both the behavior of borrowers and the distribution of borrowers. If low-risk borrowers are willing to pledge more collateral than high-risk borrowers then collateral may serve as a screening device (see Bester and Hellwig (1987)). A higher amount of outside collateral generates higher losses to the entrepeneur in the case that the entrepeneur cannot meet the payments to the creditor which are specified in the debt contract. This might give an incentive to the borrower to choose less risky projects if the amount of collateral demanded by the lender is high. Bester (1994) argues that a high degree of collateralization makes debt renegotiation more likely to occur when there is a dead-weight loss associated with the transfer of the borrower's assets to the lender. It is also shown that especially high-risk entrepreneurs are willing to offer collateral if there is a chance of debt renegotiation because collateral agreements not only punish the entrepreneur when the investment project fails but also makes default less attractive to the entrepreneur when the project is successful. A company defaults if it dishonestly reports a low return to the creditor while the actual realized return on the project is high. High risk entrepreneurs are shown to have a higher equilibrium likelihood of dishonesty. Therefore, outside collateral can be used especially to lower the expected cost of bankruptcy of risky firms. This reasoning is clearly in contrast with the screening explanation. A further remark has to be made about the possibility of banks influencing the actions of the firms. If we think of a multi-period setting, then the interest rate can be reduced for non-defaulters when time passes on. This would give borrowers the incentive to invest in relatively less risky projects if they have the possibility to choose between projects and therefore the moral hazard problem could be reduced.

In the model of Stiglitz and Weiss (1981), it is assumed that each project costs a fixed amount.

Bester and Hellwig (1987) consider the case where firms may vary the level of investment. They show that a type 2 credit rationing equilibrium is no longer an outcome of the model if the production functions of the firms exhibit decreasing returns to scale. Suppose the economy is in a state of credit rationing. Then, a bank could gain by giving credit to the rationed firms and lowering the amount of debt of the others. The expected return per unit of investment has increased for each firm because of a diminishing marginal product as a function of increasing investment. The expected return per unit of loan of the bank has also increased because the bank can ask for a higher interest rate while the available funds are spread more evenly over the population of loan applicants. Thus, a type 2 credit rationing equilibrium is not sustainable. However, it can be shown that a type 1 credit rationing equilibrium may still occur.

4.3.2 Implications of Credit Rationing

If firms with good investment opportunities are excluded from the credit market then there exists scope for monetary policy by a central government which can have real effects on the economy. As noted by Hillier and Ibrahimo (1993):

> ...if the initial equilibrium is characterized by credit rationing it is quite possible that an increase in the supply of loanable funds (e.g. by the depositors of the bank) may lead to an increase in the volume of loans without a restriction in the interest rate (see p.285, words in brackets are added).

A larger number of investment projects with positive net present value are undertaken and contribute to the total real output of an economy. From a social welfare point of view, however, it might also be possible that social efficiency requires credit to be even more tightly rationed compared to the initial credit rationing equilibrium. Hillier and Worrall (1992) explain this result by pointing out that a rise in the interest rate charged increases the amount of loans supplied by the banks but also raises the expected monitoring costs as more borrowers are supposed to default. These higher expected average monitoring costs are passed on to all of the borrowers of the bank whereas social efficiency requires that each borrower should bear the marginal cost of monitoring him self.

> Monitoring costs act as a negative externality which is not fully internalised (Hillier and Worrall (1992), p. 11).

Since the effect of monetary policy on social welfare might be ambiguous, other ways of reducing the risk of ending up in a credit rationing equilibrium should be considered. Companies could lower the chance of having insufficient funds to finance their future investment plans by adopting hedging policies or signing contracts with credit commitments. A debt contract with a credit commitment guarantees the commitment holder that he can borrow according to need up to a certain

limit. Morgan (1994) shows that bank credit commitment reduces bankruptcy risk and therefore allows borrowers to obtain higher loan limits if bankruptcy is costly.

In the absence of credit constraints, a risk neutral firm has no incentive to enter financial markets for secondary assets such as futures and forwards. In contrast to perfect capital markets, let us assume that this company is capital constrained. The management of the firm has to choose the bundle of inputs before the production process is started. One of the productive factors has a stochastic input price whereas the other factor prices are fixed. Consequently, the firm faces uncertainty about total costs of production when making operating decisions. The profit maximizing firm is credit constrained in the sense that it would increase production if it had more funds at its disposal. Vercammen (1994) shows that this company may choose to hedge the financial risk although it is risk neutral. The optimal factor hedge ratio is chosen such that the uncertain factor price has about the same correlation with total factor expenditures as exists in case of unconstrained profit maximization. This is achieved by taking either a short or a long position in the input with uncertain factor price, which is traded on a forward or future market. Liquidity constrained corporations conducted by risk neutral managers may benefit from active risk mangement programs.

4.3.3 Empirical Evidence Regarding Credit Rationing

It is interesting to establish the empirical significance of credit rationing because of the important implications it has for the economy. Several approaches have been adopted. Berger and Udell (1992) test for credit rationing by explaining the stickiness of commercial loan rates. They suggest that equilibrium credit rationing is not a significant macroeconomic phenomenon. Interest rates on 1,103,933 loans of banks in the USA between 1977 and 1988 are sticky with respect to open-market rates. But nearly half of the observed loan rate stickiness prevails on debt contracts with credit commitment. According to Berger and Udell, borrowers who signed these contracts are by definition free of credit constraints. One could expect that a larger proportion of new loans are issued under commitment when the credit market is tight, i.e. open-market rates are high. Berger and Udell however do not find such behaviour in the credit market and therefore question the empirical relevance of equilibrium credit rationing as a result of informational asymmetries.

Another empirical test is performed by modelling the optimal rate of investment and testing whether financial variables like the debt to assets ratio play a role as explanatory factors. Among others, Bond and Meghir (1992, 651 UK firms, sample period 1974-1986), Whited (1992, 325 USA firms, 1972-1986) and Rondi, Sembenelli and Zanetti (1994, 44 Italian firms, 1964-1988) estimate the Euler equation of a dynamic model of borrowing and investment with liquidity constraints.[4] Financial distress seems to have major consequences for the investment behavior of firms and thus credit rationing is claimed.

[4] See Chapter 5 for the results on a panel data set comprising Dutch manufacturing companies.

An approach closely related to the latter one is the estimation of Tobin's Q-model of investment. Given the underlying assumptions of the Q-model (i.e. a credit market with perfect information), the only determinant of the rate of investment is "Tobin's Q", the market value of the firm relative to its replacement costs (see Tobin (1969) and Chapter 5). Fazzari, Hubbard and Petersen (1988a, 1988b, 422 USA firms, 1970-1984), Hoshi, Kashyap and Scharfstein (1991, 145 Japanese firms, 1977-1982), Blundell, Bond, Devereux and Schiantarelli (1992, 532 UK firms, 1975-1986), Schaller (1993, 212 Canadian firms, 1973-1986), Becht and Ramirez (1993, 29 German firms, 1907-1912), and Alonso and Bentolila (1994, 68 Spanish firms, 1985-1987) found cash flow as a significant explanatory variable of investment in addition to Tobin's Q and conclude that this is evidence of credit rationing.

The reported coefficients on Q, in the studies which were mentioned above, are in many cases surprisingly low and sometimes even significantly negative. Blundell et al. attribute the unsatisfactory explanatory power of empirical Q-models to the endogeneity of Q and correlated firm-specific fixed effects. A similar explanation for these underestimated parameters is the presence of measurement errors in Tobin's Q since the value of equity as determined on the stock market is used in the construction of Q.[5] Equity prices, in general, are thought to display more volatile behaviour than the underlying real values would suggest (see Marcus (1989)). We argue that aside from measurement errors in Q there is another reason for possible misleading conclusions on the effect of Q on the rate of investment. Namely, firm level data often contain severe outlying observations which have too much influence on the estimates of the model under investigation if standard econometric techniques are used, as is advocated in Part I of the thesis. Re-estimation of the study of Schaller (1993) with our robust HBP GM estimator which downweighs outliers (see Chapter 1) results in considerably higher and more significant coefficients on Q.

As is suggested by Schaller (1993), a significant coefficient on cash flow is not a proof of the existence of a credit rationing equilibrium as defined by Blanchard and Fischer (1989) (see Section 4.3.1). For it is not inconceivable that companies with a high cash flow have investment opportunities which are not captured by Q or firms may prefer internal funds to outside finance (the "pecking order" (financing hierarchy) argument, see Myers (1984)). In these cases, a positive relationship between cash flow and investment may evidently come up without any company facing external liquidity constraints. Firms are free to borrow at some interest rate which is higher than the return received on deposits. In order to determine whether the difference between the costs of internal and external financing is due to asymmetric information, Schaller (1993) defines different classes within his sample of Canadian firms based on the access they have to the credit market. Mature, non-manufacturing or concentrated ownership companies are thought to have better outside financing opportunities than young,

[5] See Appendix A5.1, Chapter 5.

manufacturing or dispersed ownership companies.[6] As an example, cash flow has a stronger effect on the investment expenditures of dispersed ownership corporations than concentrated ownership corporations whereas there is little difference in investment opportunities as measured by Q. Therefore it is claimed that imperfect financial markets cause underinvestment by Canadian firms. It is noteworthy that with our robust HBP GM estimation method these differences in sensitivity of investment to cash flow between different classes of firms almost disappear.

Although, after reviewing the empirical literature, it remains unclear whether asymmetric information distorts the credit market there seems consensus about the widely held belief that investors and firms can only borrow funds up to a certain limit. Evidently, credit constraints do undermine the reasoning behind Modigliani and Miller's Proposition I.

Section 4.4 Tax Shields

Another assumption made by Modigliani and Miller (1958), which conflicts with real life observations, is the absence of taxes. Generally speaking, corporate taxes favour debt financing whereas personal taxes favour equity financing. In the extreme case of zero taxes on equity returns and a debt investors' personal tax rate identical to the one corporations have to pay, the net present value of the debt financing effect (tax shield) is zero given that debt holders are fully compensated by means of a higher pretax return. In this situation the manager of a company can make his decisions on investment and capital structure independently of each other. Usually, the management takes into account the side effects of the financing possibilities when evaluating an investment opportunity. Mauer and Triantis (1994) study the complex interaction effects between investment, operating and financing decisions in a dynamic model with operating adjustment costs and recapitalization costs. They find the remarkable result that, in spite of the presence of tax shields with positive present value, the impact of debt financing on the firm's investment and operating decisions is economically insignificant. Thus, firms can actually split the investment, operating and financing decisions. However, Mauer and Triantis (1994) do reject Proposition I, i.e. the value of a firm changes with its financial policy because of tax benefits. To gain insight into these results we present a brief summary of the dynamic model studied by Mauer and Triantis.[7]

Consider a company that produces a single commodity which is sold in a perfectly competitive market at the per unit price P. P is stochastic and its future values are unknown to the decision makers of the company at the beginning of the time period. The company is flexible in production in the sense

[6] Rondi, Sembenelli and Zanetti (1994) conclude from panel data on Italian companies that the investment expenditures of state-owned relative to privately-owned or quoted relative to unquoted firms are more (!) sensitive to cash flow.

[7] Mauer and Triantis (1994) rely on numerical solutions since the model is too complicated to derive analytical results.

that it can shut down or resume operations at the expense of fixed exit and entry costs respectively. There are no personal taxes and interest expenses can be deducted from the company's taxable profits. The tax system therefore favors the issuance of debt. The firm pays both fixed and variable recapitalization costs when altering its financial structure. The recapitalization costs are increasing in the amount of new debt. Mauer and Triantis recognize that the costs of debt financing are not only incured by companies with high probability of bankruptcy or in financial distress but also to firms which execute *on-balance sheet* hedging (recapitalization) plans, in terms of improving the debt-equity ratio, to prevent financial distress.

As the operating adjustment costs decrease the operating profits increase and the volatility of the firm's value falls. This drop in the variance of the company's value is put down to the fact that the managers shut down the factory when the commodity's price P is temporarily low and resume production when it is high. The value of the interest tax shield increases since interest expenses can be deducted from the more stable and higher income.[8] Thus, Mauer and Triantis (1994) predict a positive relationship between firm leverage and production flexibility. However, production flexibility (lower operating adjustment costs) and financial flexibility (lower recapitalization costs) are to a certain extent substitutes regarding the effect on the present value of the tax shield. After all, the company will respond with changes in the optimal capital structure along the path of output price realizations if the recapitalization costs are small. Hence, the effect of production flexibility on the value of the interest tax shield is lower if financial flexibility is higher.

Following the reasoning above, one is tempted to argue that financial policy will have a strong influence on investment and operating decisions. Especially if recapitalization costs are small then a levered company might accept a new investment plan (or resume the old production plan) at a lower output price than the equivalent unlevered firm since the interest tax shield is also contributing to the levered firm's profits. Nevertheless, Mauer and Triantis (1994) do not find significant differences in the operating and investment strategies of the two companies:

> Indeed, our analysis shows that if a levered firm uses the investment and operating policies of an equivalent unlevered firm, there is a negligible loss in firm value. These results are in sharp contrast to traditional static capital budgetting analyses that conclude that tax-advantaged debt financing can make the difference between project acceptance or rejection. In a static setting, however, the value of interest tax shields over the entire life of an investment project bear on the investment decision because the firm does not have the option to delay investment. By contrast, in a dynamic setting the firm foregoes earning interest tax shields only over the period of time that it chooses to delay the investment. Our analysis shows that the loss in tax shield value from waiting for additional uncertainty resolution is not large enough to encourage the firm to significantly deviate from the investment and operating policies of an equivalent

[8] A firm which makes losses on its business activities does not pay any taxes so cannot profit from tax deductions.

unlevered firm (Mauer and Triantis (1994), pp. 1272-1273).

It is important to notice that although the levered and unlevered companies follow almost the same production and investment strategies, they differ substantially in total firm value. The value of the levered company exceeds the value of the equivalent fully equity financed company since the levered company pays less taxes due to the interest tax shield. Once again, we reject Modigliani and Miller's Proposition I.

Section 4.5 Concluding Remarks

In contrast to Modigliani and Miller's separability theorem, firms may have a "pecking order" (financing hierarchy) when choosing between sources of finance. Internal sources may be less costly than debt financing or issuing new equity due to transaction costs, agency problems, taxes and bankruptcy costs. In most cases, firms appear to prefer using internal funds to external finance and, if internal funds are insufficient, debt instead of equity as a source of incremental funding of investment projects (e.g. see Fazzari et al. (1988a, 1988b) and Myers and Majluf (1984)).

Financial policy is relevant when the financial markets are incomplete and credit rationing distorts the optimal allocation of funds. Secondary assets, limited liability, and liquidity constraints do not appear out of the blue, therefore, hedging strategies determine the range of attainable reallocations of the firm's revenue and consequently firm value. For instance, we have seen that it is optimal for a capital constrained risk neutral company which faces input price uncertainty to trade in forwards or futures in order to reduce company risk. In order to explain credit rationing equilibria we discussed Stiglitz and Weiss's (1981) argument that the lender might restrict the borrower's use of debt as a consequence of asymmetric information in financial markets. Two contradicting strands of literature which examine empirical evidence regarding credit rationing were mentioned. On one hand, Berger and Udell (1992) argue that the observed interest rate stickeness in the USA between 1977 and 1988 is not the result of restricted credit markets. On the other hand, financial variables such as the cash flow to capital ratio or the debt to assets ratio are important explanatory factors in empirical investment models. This finding is held as evidence of liquidity constraints. Finally, we conclude that the introduction of valuable interest tax shields has consequences for the firm's optimal financial policy. Although the capital structure matters for the value of the company for tax reasons, Mauer and Triantis (1994) predict no major impact on the company's investment and operating decisions.

This survey on the relevance of financial policy is far from exhaustive and just mentions briefly some important issues in the theory of finance. For further reading on the capital structure puzzle we refer to Myers (1984) and Harris and Raviv (1991). Hillier and Ibrahimo (1992) can be consulted for an excellent review of the literature on credit rationing and, among others, Froot, Scharfstein and Stein (1993) deserve credit for their analytical examination of the implications of capital market imperfections for risk management.

CHAPTER 5

Empirical Evidence Regarding the Sensitivity of Dutch Corporate Investment to the Financial Environment

Section 5.1 Introduction

In the previous chapter we looked at some of the theoretical arguments which are put forward to explain the failure of Modigliani and Miller's well-known Proposition I. Most financial experts are in agreement about the non-separability of corporate investment and financing decisions and the existence of a wedge between the cost of internal and outside finance. In general, the financial markets are not thought to be perfect.

Using a new balance sheet panel data set of 117 Dutch manufacturing companies we investigate in this chapter whether credit rationing distorts the allocation of funds over investment projects in the Netherlands. Do Dutch manufacturing companies face binding liquidity constraints when choosing their investment expenditures? Would, on average, a corporation in the Netherlands be willing to increase its level of investment if it could borrow more funds on the credit market? Do these firms distinguish between internally generated funds and external finance when setting their optimal investment objectives?

To address these questions we empirically analyse two versions of a neoclassical investment model with convex capital adjustment costs: Tobin's Q-model (Tobin (1969)) and the Euler equation approach which was originally used by Jorgenson (1963). Bond and Meghir (1992) solve the first order condition of an optimization problem under uncertainty where the firm's choice between debt financing and retained earnings is taken simultaneously with the investment decision. Within their framework we derive the Q-model of investment and discuss the empirical implications for companies which, *a priori*, belong to different financial regimes. Using the robust estimation techniques which are developed in part I of the thesis, both the Euler equation and the Q-model are estimated to test whether Modigliani and Miller's proposition I can be applied to Dutch manufacturing industries which are quoted on the Amsterdam stock exchange.

The outline of Chapter 5 is as follows: In the subsequent section we derive the Q-model in line with Bond and Meghir's (1992) paper. The data is described in Section 5.3 while in Section 5.4 we handle the practical problem of choosing the sorting key for the division of the sample into different financial regimes. In Section 5.5 we explain the adopted econometric methodology.

Estimation results of the Q-model of investment and the Euler equation are presented in Section 5.6. Section 5.7 concludes. Appendix A5.1 contains the construction and definitions of the variables of both investment models. The tables reporting the estimation results are placed in Appendix A5.2.

Section 5.2 Neoclassical Investment Theory

In this Section we formally derive the Q-model of investment with endogenous financial policy by closely following the dynamic investment and financing setting considered by Bond and Meghir (1992). Financial markets are perfect in Hayashi's (1982) seminal paper where he introduces an empirically testable model of the Q-theory by proposing conditions for which the marginal Q is equal to the average Q (see Section 5.2.2). Thus, the optimal choice of finance does not enter the story. In contrast, here we explicitly consider internal funds, debt and equity as financing sources to which different price tags dangle. First we will deal with the crucial ingredients and steps necessary to solve the Euler equation of the dynamic simultaneous optimization problem analyzed by Bond and Meghir (1992). The interested reader is referred to this article for a more complete exposition of the model.[1] We then exploit their framework to derive the Q-model rather simply but neatly.

An important aspect of the model under consideration is capital adjustment costs, which may result from explicit installation, renovation and dismantling costs or losses in production during the transition period. Firms have an incentive to spread a proposed substantial change in the capital stock over time since these adjustment costs are assumed to be convex. To introduce notation and to define the companie's optimization problem we begin with listing the underlying assumptions of the investment model.

Assumptions

(A.1) Mission of the company

The Managers of the firm are assumed to act in the interests of their current stockholders, i.e. they maximize the sum of discounted dividends (firm value, V_t). Stockholders are risk neutral. Thus, the appropriate discount factor for the one period ahead dividend payment is equal to $\beta'_{t,1} = 1/(1+i_t)$ where i_t is the risk-free rate of interest. The stockholder pays personal taxes m_t for each unit of received dividend D_t and pays the effective capital gains tax rate z_t on a unit of capital gains made between period t and t+1. Tax parameters for period t+1 are supposed to be

[1] Blundell, Bond and Meghir delivered an excellent survey on econometric models of company investment, see chapter 17 of *The Econometrics of Panel Data, Handbook of Theory and Applications*, Mátyás and Sevestre (1992).

known to the investors at time t.

(A.2) Rational Expectations

The managers are thought to make rational decisions and to have rational expectations. Expected values of future variables are evaluated at realised values. This introduces measurement errors and therefore instrumental variable estimators are needed to obtain consistent parameter estimates.

(A.3) Formation of the capital stock

The capital stock K_{it} of firm i in period t is determined by the level of investment I_{it} and the previous capital stock:

$$K_{it} = I_{it} + (1-\delta_i)K_{t-1,i} \tag{5.1}$$

where δ_i is the constant rate of economic depreciation of firm i. Investment goods are immediately productive subject to convex adjustment costs which are linearly homogenous in investment and capital. The symmetric convex adjustment cost function is written as

$$G(I_t,K_t) = \frac{1}{2}bK_t\left(\frac{I_t}{K_t}-c\right)^2. \tag{5.2}$$

The constant c reflects the natural rate of investment at which adjustment costs are zero.

(A.4) Tax treatment

The firm is allowed to deduct its interest payments made on the outstanding debt B_t from the pre-tax firm's profits which are taxed at the corporate tax rate τ_t.

(A.5) Production and output market conditions

The production function $F(K_t,L_t)$ exhibits constant returns to scale; Net output $Y_t = F(K_t,L_t) - G(I_t,K_t)$ is linearly homogenous in capital and the variable input factors L_t. The market where the output of the firm is sold for a price p_t per unit can be characterized by imperfect competition. The price elasticity of demand $d>1$ is assumed constant and defines the number $\alpha = 1-\frac{1}{d} > 0$. w_t denotes the vector of prices of the variable factor inputs. The optimal choice for L_t is such that the marginal product of variable input factors is equal to the ratio $\frac{\partial F}{\partial L} = \frac{w}{\alpha p}$. This last statement allows us to compute the derivatives of the the revenue function

$$\Pi(K_t,L_t,I_t) = p_tF(K_t,L_t) - p_tG(I_t,K_t) - w_tL_t - p_t^II_t \tag{5.3}$$

with respect to K_t and I_t without specifying the production function. p_t^I reflects the price of capital goods.

(A.6) Financial Markets

To finance new investment projects the firm can choose between either using retained earnings, issuing new equity N_t or borrowing on the debt market. There are no arbitrage possibilities on the financial markets and according to our null-hypothesis the markets for credit are perfect. Furthermore, we impose the following transversality condition

$$\lim_{T\to\infty} \left(\prod_{j=t}^{T} \beta_j \right) B_T = 0, \quad \forall t \tag{5.4}$$

which prevents the firm from borrowing an infinite amount to pay out as dividends.[2]

Bond and Meghir (1992) explicitly consider bankruptcy, taxes on the lender's income, and deadweight costs of issuing equity and paying dividend. Omission of these elements of the investment model simplifies the mathematics but does not change the final Euler equation which is to be estimated. We therefore do not take into accounct explicitly the effects of the probability of bankruptcy etc. on the investment and financing choices.[3]

5.2.1 Derivation of The Euler Equation

The capital market arbitrage condition dictates that the cum-dividend value of the firm's shares V_t minus dividend plus new equity, is expected to grow at the rate of return on default-free bonds

$$(1+(1-m_{t+1})i_t)(V_t-(1-m_t)D_t+N_t) = E_t[V_{t+1}] - \xi_{t+1}(E_t[V_{t+1}]-V_t-N_t) \tag{5.5}$$

to the after-tax value of the company's equity in period t+1. $\xi_{t+1} = z_t(1+(1-m_{t+1})i_t)$ is the value of taxes paid by the marginal stockholder on one unit of capital gains in period t+1. By stretching equation (5.5) we obtain the firm's value at time t

$$V_t = E_t \left(\sum_{j=0}^{\infty} \beta'_{t+j}(\gamma_{t+j}D_{t+j}-N_{t+j}) \right) \tag{5.6}$$

---

[2] In a multi-period economy, we may observe a "rational ponzi game" or bubble. A rational ponzi game is the policy of a debtor to roll over all principal repayments and interest forever, i.e. to finance repayments by issuing new debt. O'Connell and Zeldes (1988) and Hammond (1987) show conditions under which rational ponzi games may exist. The basic result established by O'Connell and Zeldes is that ponzi games require the participation of an infinity of agents. There can be a finite number of individual decision makers in the credit market in each period. However, the number of agents must grow to infinity over time. To rule out rational ponzi games, we assume a finite number of agents over time and impose the tranversality condition, $\lim(T\to\infty)\ \Gamma(T)B_T = 0$, where $\Gamma(T)$ is the inverse of the discount factor of period T. The transversality condition implies that an individual chooses in the end not to be on the lending side of a ponzi game and every individual faces a budget constraint which restricts his possible consumption schemes.

[3] Following Bond and Meghir's notation we would have $m_t^B=0$, $f_t=0$, $q'_{t-1}=0$, $X_t=0$ *and* $\theta_t=1$.

which is maximized with respect to the size of the capital stock, the financial structure and the input factors of the production process, under the restriction that dividends

$$D_t = \Pi_t + N_t + B_t - (1+(1-\tau_t)i_{t-1})B_{t-1} \tag{5.7}$$

are positive. The relative tax advantage between dividend income and capital gains is assimilated by the tax parameter

$$\gamma_t = \frac{1-m_t}{1-z_t}. \tag{5.8}$$

Bond and Meghir (1992) show in their appendix that the equity value can be rewritten as:

$$\begin{aligned} V_t &= E_t\left(\sum_{j=0}^{\infty}\beta^j_{t+j}(\gamma_{t+j}\Pi_{t+j}+(\gamma_{t+j}-1)N_{t+j})\right) - \gamma_t(1+(1-\tau_t)i_{t-1})B_{t-1} \\ &+ E_t\left(\sum_{j=1}^{\infty}\beta^j_{t+j}\gamma_{t+j}\tau_{t+j}i_{t+j-1}B_{t+j-1}\right) + E_t\left(\sum_{j=1}^{\infty}(\beta^j_{t+j-1}\gamma_{t+j-1}-\beta^j_{t+j}\gamma_{t+j}(1+i_{t+j-1}))B_{t+j-1}\right) \end{aligned} \tag{5.9}$$

Using (5.9) we form the Hamiltonian

$$\begin{aligned} H_t &= V_t + \lambda^D_t(B_t-(1+(1-\tau_t)i_{t-1})B_{t-1} + \Pi_t + N_t) \\ &+ \beta^1_{t+1}\lambda^D_{t+1}(B_{t+1}-(1+(1-\tau_{t+1})i_t)B_t + \Pi_{t+1} + N_{t+1}) \\ &+ \textit{restrictions of period } t+2 \textit{ and higher,} \end{aligned} \tag{5.10}$$

where λ^D_t denotes the Lagrange multiplier of the restriction on dividends. At this point we relax the perfect capital market condition. Firms with high leverage ratio's are supposed to find themselves in a situation of financial distress or bankruptcy more frequently than low levered firms. Therefore, banks require a higher interest rate on debt supplied to the high levered firms. Let us assume that the interest rate on the company's debt $i(B_t/p_t{}^tK_t)$,, depends only on the ratio $(B_t/p_t{}^tK_t)$ and is increasing in its argument. This specification is chosen to show that the debt to capital ratio is entering the Euler equation and Q-model when the financial markets are imperfect. Unfortunately, as we will discover in Section 5.3, in this case both the Euler equation and the Q-model version of the investment model under consideration, do not display linear one-to-one relationships between the investment capital ratio and its explanatory factors.

Observing that

$$V_t = \gamma_t\Pi_t + (\gamma_t - 1)N_t - \gamma_t(1+(1-\tau_t)i_{t-1})B_{t-1} +$$
$$\gamma_t B_t + E_t\left(\beta'_{t+1}\gamma_{t+1}(i_t(\tau_{t+1}-1)-1)B_t\right) + \textit{terms of period } t+1 \textit{ and higher} \tag{5.11}$$

and differentiating (5.10) with respect to B_t gives

$$-v_t\frac{B_t}{p_t{}^I K_t} = -(\gamma_t+\lambda_t^D) + E_t\left(\beta'_{t+1}(\gamma_{t+1}+\lambda_{t+1}^D)(1+i_t(1-\tau_{t+1}))\right) \tag{5.12}$$

where $\frac{\partial i_t(x)}{\partial x} = i' \geq 0$ and[4]

$$v_t = E_t\left(\beta'_{t+1}(\lambda_{t+1}^D+\gamma_{t+1})i'(1-\tau_{t+1})\right). \tag{5.13}$$

Differentiating the Hamiltonian again to find the first-order condition associated with I_t results in the equation

$$\frac{\partial\Pi_t}{\partial K_t} = \frac{-\partial\Pi_t}{\partial I_t} - E_t\left(\beta'_{t+1}\frac{(\gamma_{t+1}+\lambda_{t+1}^D)i'_t(1-\tau_{t+1})B_t^2}{(\gamma_t+\lambda_t^D)p_t{}^I K_t^2}\right) - E_t\left(\beta'_{t+1}\frac{\lambda_{t+1}}{\gamma_t+\lambda_t^D}\right) \tag{5.14}$$

and applying the envelope theorem when differentiating H_t with respect to K_{t-1} returns

$$\lambda_t = \frac{\partial H_t}{\partial K_{t-1}} = (\gamma_t+\lambda_t^D)(1-\delta)\frac{\partial\Pi_t}{\partial K_t} + E_t\left(\beta'_{t+1}\lambda_{t+1}(1-\delta)\right) +$$
$$(\gamma_t+\lambda_t^D)(1-\tau_t)i'_{t-1}\frac{B_{t-1}^2}{p_t{}^I K_{t-1}^2} + E_t\left(\beta'_{t+1}(\gamma_{t+1}+\lambda_{t+1}^D)(1-\delta)(1-\tau_{t+1})i'_t\frac{B_t^2}{p_t{}^I K_t^2}\right), \tag{5.15}$$

where λ_t measures the shadow value of one marginal unit of the capital stock in period t-1. Using (5.14) in (5.15) gives

$$\lambda_t = -(\gamma_t+\lambda_t^D)(1-\delta)\frac{\partial\Pi_t}{\partial I_t} + (\gamma_t+\lambda_t^D)(1-\tau_t)i'_{t-1}\frac{B_{t-1}^2}{p_t{}^I K_{t-1}^2}. \tag{5.16}$$

Inserting (5.16) twice in (5.15) and rearranging terms solves the Euler equation

[4] v_t differs from the one reported on page 35, formula (A1), by Bond and Meghir (1992). Our formula assures that v_t is positive whereas (A1), although it is claimed that it can be shown to be positive, can be negative. We presume the occurence of a typing error.

$$E_t\left(-\beta'_{t+1}(\gamma_{t+1}+\lambda^D_{t+1})(1-\delta)\frac{\partial \Pi_{t+1}}{\partial I_{t+1}}\right) = -(\gamma_t+\lambda^D_t)\frac{\partial \Pi_t}{\partial I_t} - (\gamma_t+\lambda^D_t)\frac{\partial \Pi_t}{\partial K_t} - 2v_t\frac{B_t^2}{p_t^I K_t^2}. \tag{5.17}$$

From the revenue function (5.3), the capital adjustment cost function and the production and output market conditions we obtain the equations

$$\frac{\partial \Pi_t}{\partial I_t} = -b\alpha p_t\frac{I_t}{K_t} + bc\alpha p_t - p_t^I \tag{5.18}$$

and

$$\frac{\partial \Pi_t}{\partial K_t} = \alpha p_t\frac{Y_t}{K_t} - w_t\frac{L_t}{K_t} + b\alpha p_t\left(\frac{I_t}{K_t}\right)^2 - bc\alpha p_t\frac{I_t}{K_t}. \tag{5.19}$$

Under the null-hypothesis that Modigliani and Miller's Proposition I holds true ($\gamma_t = \gamma_{t+1} = 1$, and $\lambda^D_t = \lambda^D_{t+1} = 0$) we use (5.18) and (5.19) in the Euler equation (5.17) to derive the following empirically verifiable linear relationship which explains the future investment capital ratio:

$$\frac{I_{t+1}}{K_{t+1}} = c(1-\omega_{t+1}) + (1+c)\omega_{t+1}\frac{I_t}{K_t} - \omega_{t+1}\left(\frac{I_t}{K_t}\right)^2 - \frac{\omega_{t+1}}{b\alpha}\left(\frac{C}{K}\right)_t + \frac{\omega_{t+1}}{b\alpha}J_t + \frac{\omega_{t+1}}{b(d-1)}\frac{Y_t}{K_t} - 2\frac{(1+i_t)v_t}{b\alpha(1-\delta)}\left(\frac{B}{K}\right)_t + \varepsilon_{t+1}, \tag{5.20}$$

where we have defined the adjusted discount factor $\omega_t = p/(\beta_{t+1}(1-\delta)p_{t+1})$, the cash flow capital ratio $(C/K)_t = (p_t^I/p_t)(p_tY_t - w_tL_t)/p_t^IK_t$, the debt to capital ratio squared $(B/K)_t^2 = (p_t^I/p_{t+1})(B_t/(p_t^IK_t))^2$, the user cost of capital

$$J_t = \frac{p_t^I}{p_t}\left(1-\frac{p_{t+1}^I(1-\delta)}{(1+i_{t+1})p_t^I}\right) \tag{5.21}$$

and ε_{t+1} as the forecast error.

To actually arrive in the position to estimate the Euler equation we have to assume that the coefficients of equation (5.20) are constant over time and across firms. If capital markets are perfect then the coefficient on the debt to capital ratio squared is equal to zero since for this case $v_t = 0$. The coefficient on the output capital ratio reflects the degree of output market imperfection. The user cost of capital will not be explicitly measured but its contribution to the investment capital ratio is assumed to be absorbed by the time dummies (ϕ_t) and firm-specific fixed effects (η_i) which are added to (5.20):

$$\frac{I_{it}}{K_{it}} = \beta_1(I/K)_{t-1,i} - \beta_2(I/K)^2_{t-1,i} - \beta_3(C/K)_{t-1,i} + \beta_4(Y/K)_{t-1,i} - \beta_5(B/K)^2_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}. \tag{5.22}$$

The subscript i (i=1,..,n) is the company indicator and the beta's are the unknown parameters. The results of estimating (5.22) for different groups of Dutch manufacturing firms are reported in Section 5.6.1.

5.2.2 Derivation of the Q-model of investment

Another strand of the literature on empirical research on the optimal rate of investment, estimates Tobin's Q-model. A company will invest in an additional unit of capital if the marginal value of the investment exceeds the marginal costs, i.e. if marginal q is greater than marginal investment costs. Marginal q is difficult to measure but Hayashi (1982) shows that under the assumptions of perfect competition, perfect capital markets and first-degree homogeneity of the production function and the capital adjustment cost function that marginal q is equal to average Q. Once again, adjustment costs associated with new investment in fixed capital are also a crucial feature of the Q-theory of investment.

Given the assumption that V_t is homogenous of degree 1 in K_{t-1}, within the framework of the previous sub-section the Q-model specification is found by using $\partial V/\partial K_{t-1} = V/K_{t-1}$, rewriting (5.16) to

$$\frac{\partial \Pi_t}{\partial I_t} = \frac{-\lambda_t}{(\gamma_t+\lambda_t^D)(1-\delta)} + \frac{(1-\tau_t)i'_{t-1}}{(1-\delta)}\frac{B^2_{t-1}}{p_t{}'K^2_{t-1}} \tag{5.23}$$

and then inserting this formula into equation (5.18):

$$\frac{I_t}{K_t} = c + \frac{1}{b\alpha}\left(\frac{V_t}{(\gamma_t+\lambda_t^D)(1-\delta)p_t{}'K_{t-1}}-1\right)\frac{p_t{}'}{p_t} - \frac{(1-\tau_t)i'_{t-1}}{b\alpha(1-\delta)}\left(\frac{B_{t-1}}{p_t{}'K_{t-1}}\right)^2\frac{p_t{}'}{p_t}. \tag{5.24}$$

Under the conditions of Modigliani and Miller's debt irrelevance theorem (γ_t = 1 and λ_t^D = 0), the variable associated with debt drops out of the Q-equation since i' = 0 and Tobin's Q is computed as

$$Q_t = \left(\frac{B_t+E_t-F_t}{(1-\delta)p_t{}'K_{t-1}}-1\right)\frac{p_t{}'}{p_t}. \tag{5.25}$$

E_t is the value of equity and F_t are the firm's financial assets (see Appendix A5.1). Schiantarelli

and Georgoutsos (1990) show that current output should enter equation (5.24) with a negative sign when the firm is a monopolistic competitor.[5] Following the empirical studies so far presented in the literature on the Q-theory (see Chapter 4, Section 4.3.3) we include the cash flow capital ratio to allow for liquidity constraints resulting from imperfections in the financial markets under the alternative hypothesis that Proposition I is rejected. Altogether with time-dummies and firm specific fixed effects, we obtain the following econometric specification

$$\frac{I_{it}}{K_{it}} = \beta_1 Q_{it} - \beta_2 (B/K)^2_{t-1,i} - \beta_3 (Y/K)_{it} + \beta_4 (C/K)_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it} \qquad (5.26)$$

where $(B/K)^2_{t-1} = (B_{t-1}/p_t{}'K_{t-1})^2(p_t{}'/p_t)$. Note that the beta's of equation (5.26) are so-called *deep* parameters which only depend on the firm's technology and output market conditions so that the Lucas (1976) critique does not apply here.

Many comments have to be made about the Q-model based test of credit rationing. First, if the cost of internal funds substantially differs from the cost of external finance, it is possible that a firm is willing to increase its level of investment when internally generated funds are available but is not when it has to borrow funds on the debt market. Thus, a credit rationing equilibrium in the sense of a group of borrowers who cannot obtain the amount of credit they wish (see Section 4.3.1), cannot be claimed only on the basis of a coefficient on cash which is significantly different from zero. When we refer in this chapter to liquidity constraints, these might arise both from internal and external restrictions. Secondly, as argued by Blinder (see Comments and Discussion Fazzari et al. (1988a)) it is hard to believe that firms will underinvest when cash flow is temporarily low but, at the same time, the stock of liquidity is sufficiently high. Stock variables rendering the availability of internal funds for new investment should be added to the Q-equation. Unfortunately, our data set does not provide numbers on short-term deposits or other stocks of that sort. Thirdly, as noticed by Poterba (see Comments and Discussion Fazzari et al. (1988a)) a high level of current cash flow may signal future profit opportunities which are not captured by Q when the efficient market hypothesis is violated. Equity prices, which are used in the construction of Q, display erratic behaviour and are not universally believed to be a true mirror of the real underlying value of the firm. In this case, it is evident that a positive relationship between investment and cash flow is to be expected without necessarily having debt ceilings.

A more dramatic problem with the verification of the Q-theory of investment is the recognition of the fact, implied by (5.24), that Tobin's Q as measured in (5.25) is uninformative or misleading with regard to the optimal level of the investment capital ratio if $\gamma_t + \lambda_\tau^D \neq 1$. The non-existence of a one-to-one relationship between the investment of a liquidity constrained company

[5] Furthermore, Schiantarelli and Georgoutsos (1990) find that in this case marginal q can not be equated to average Q and propose a dynamically richer structure of the model by incorporating future values of the investment capital ratio and Tobin's Q. We, however, did not attempt to measure this distortion in our Q-variable.

and Q was recognized before by Hayashi (1985) and Chirinko (1987), who explicitly consider the corporate financing side of the Q-theory. The same line of reasoning can be applied to the Euler equation (5.20) of the previous sub-section. Bond and Meghir (1992) suggest distinguishing three different financial regimes. Firms of regime 1 pay out positive dividends and do not issue new equity. These firms are expected to have generated sufficient earnings to carry out all their valuable investment plans. Thus, the shadow value of internal finance as measured by the Lagrange multiplier λ_t^D can be set equal to zero. The companies of the other two financial regimes do not remit dividends. Firms which are assigned to regime 3 have sufficient attractive investment opportunities to finance their investment projects partly or fully by issuing new equity whereas the regime 2 firms are reluctant to do so. If debt markets are restrictive then the investment expenditures of the corporations of the latter category clearly depend on the availability of internal finance. By making the assumption of a constant tax discrimination parameter γ_t it is argued that the Euler equation holds for firms which are either in regime 1 or 3 for two subsequent periods. Bond and Meghir (1992) incorporate extra coefficients for all explanatory variables of the econometric model to allow the parameters of the Euler equation to be different for regime 2 firms with respect to the others. However, we need to remark that these additional parameters are not necessarily constant across firms and over time which might invalidate the econometric results. Unless the test on the null-hypothesis of perfect capital markets indicates that the type-2 company's shadow value of internal finance is also equal to zero. To conclude, equation (5.26) is only well specified for corporations which are not liquidity constrained.

The derivation of an Euler equation or Q-model which is observed by funds restricted companies is, to the best of the author's knowledge, an unsolved issue in the corporate finance literature. Our testing procedure to examine the validity of Proposition I consists of the following steps: First, we decide on the significance of the parameters of debt and cash in the Q-specification using the full sample of companies. Significant coefficients on these financial variables reject the null-hypothesis of perfect capital markets. Then, we repeat this first step for two different groups of firms. We distinguish a class of firms with free access to capital markets (group 1). In this case, cash and debt should have minor influence on the investment decisions and Tobin's Q is well defined. For the other group of liquidity constrained companies (group 2), debt and cash can be of great importantance for the investment choices but, on the other hand, Tobin's Q will have less explanatory power.

Tabel 5.1 Partition of the 117 Dutch Manufacturing Companies in Ten Sectors

Industry Sector	SBI-code	Number of Companies
Metallurgical Industry	3300-3999	21
Building and Wood Industry	3200-3299 5100-5299	6
Textile, Clothing, Paper and Printing Industry	2200-2799 4000-4099	21
Petrolium and Chemical Industry	1200-1999 2800-3199	10
Agriculture, Fishery and Food Industry	100-399 2000-2199	11
Wholesale Trade	6100-6499	23
Retail Trade	6500-6699	5
Railway, Road Transport, Shipping and Aviation	7100-7799	7
Service Industry and Machine Renting	8400-8599 9000-9999	7
Other		6

Section 5.3 The Data Sources

Panel data of Dutch corporations is delivered by *Financieel Economisch Lexicon* (REACH) and DATASTREAM.[6] The balanced data set contains 117 Dutch manufacturing companies from ten industry sectors with SBI-codes between 100 and 9999 and runs from 1986 to 1990 (see Table 5.1).[7] Firm specific annual data on: the capital stock, sales, investment, short debt, long debt, cash flow and depreciation have been made available by REACH. The definition and the construction of the variables are given in Appendix 1. Annual data on: the price deflator for capital goods, the producer price index and the market value of stock are taken from DATASTREAM. Unfortunately, 22 firms drop out when estimating the Q-model of investment since DATASTREAM only provides equity prices for 95 companies of our panel data set.

[6] Financieel Economisch Lexicon is published by Delwel, The Hague, The Netherlands and is available on CD-ROM under the name REACH (Review and Analysis of Companies in Holland).

[7] SBI-codes are used by the *Centraal Bureau voor de Statistiek* (CBS) to subdivide manufacturing companies into different categories.

Section 5.4 Financial Regimes

We employ an *a priori* empirical sorting key to classify our data set which is different from the theoretical one described in Section 5.2. The division rule should partition the sample according to exogenous characteristics of the firms, on *a priori* grounds, into a group of firms which are expected to have liquidity constraints and a group of firms which are free to borrow on the capital markets. Dividend policy is endogenous. At time t, the bank can only guess the degree of credit worthiness of the firm in subsequent periods. The creditor might decide to tighten the provision of new funds to the debtor at present and in the future even if the company remits positive dividends in the current period. Furthermore, dividend policy can serve as a signaling device for shareholders or banks. Firms might be unwilling to change their dividend policy even when this means giving up valuable investment opportunities. A practical reason for not taking up dividend policy as the seperating rule is the empirical observation that, for the sample period 1987-1990, only 83 out of 468 observations would belong to the group 2 companies.[8] These observations cannot be ascribed to an invariable group, i.e. most companies distribute positive dividends but some firms do not pay dividends only for a brief span of time. Since our estimation method requires a balanced data set, we would be left with a very small group of liquidity constrained firms.

The full sample is divided into two classes of firms based on the mean of the debt to capital ratio squared over the period 1986-1989, where this distributive code is relatively low for 48 group 1 borrowers in comparison with 47 group 2 borrowers. The debt to capital ratio is believed to be a measure of financial distress according to the investment model expounded in Section 5.2. Recall that the econometric specifications claim a negative relationship between investment and this regressor. Although we realize that the debt to capital ratio does not necessarily coincide with the usual criterion for fianancial distress as given by the relative amount of debt in the financial structure. Especially for firms of the service sector, the debt to capital ratio might be quite misleading about the degree of indebtedness. Note that our data and investment model apply to production companies for which the size of the capital stock matters for generating profits. If the creditors attribute to the highly levered companies a higher probability of bankruptcy, then they might be inclined to restrict future increments in the level of debt of group 2 companies.[9] Or from another point of view, if the capital stock can be sold on a second-hand market then it can be contributed as collateral to the debt contract up to a certain limit of the debt to capital ratio. Our sorting key also has its drawbacks. As in the case with dividend policy, the seperating rule is endogenous since it is based on future unknown variables. An exogenous measure would be delivered by considering the past level of the debt to capital ratio. However, we have chosen the average of this ratio over the sample period of the explanatory variables because in this way we do

[8] In case of the Q-model only 46 out of 380 items would be classified as group 2 observations.

[9] Throughout the rest of Chapter 5 we refer to highly levered companies to indicate the ones which have relatively high debt to capital ratio's.

not have to assume that a firm is necessarily punished with a liquidity constraint when it just temporarily faces an unfavourable capital structure. Evidently, firms can swap from one group to the other once their debt to capital ratio has improved or worsened permanently. The division rule employed does not take into account this possible change in the company's status. We note that, given the short sample period, our global measure might effectively order the data.

Section 5.5 The Estimation Procedure

The procedure for the estimation of the Euler equation and the Q-model is as follows:

First, we employ the GM estimator as developed in Chapter 1 when estimating equations (5.22) and (5.26) in levels, first differences and in the "within" dimension of the data by subtracting firm averages over time. The firm-specific fixed effects are omitted but a constant is included in the level regressions. The Wu (1973) - Hausman (1978) statistic is used to test whether incorporating fixed effects in the regression model is necessary, by comparing the estimates from the regression in levels with the ones obtained by estimating in first differences or within. The robust version of the Pagan-Vella (1989) test statistic is computed to test for homogeneity of the error terms whereas the generalized Durbin-Watson (see Bhargava et al. (1982)) statistic is consulted to examine first-order autocorrelation in the weighted residuals. We test for cross-sectional correlations using the Frees (1995) test based on the weighted residuals.
Second, the GM estimates of the first step of the estimation procedure are corrected for heteroskedasticity and autocorrelation by application of the methods which are described in Chapter 2. Then, we proceed with testing whether the first difference and within estimates are significantly different using again the Wu-Hausman statistic to discover possible endogeneity of the explanatory variables or measurement errors.

Third, we use the Robust GMM estimator which assures consistent estimates in the case of simultaneity as is explained in Chapter 3. The two-stage GM estimator which is used in the first round of the RGMM method is applied to a stacked model where we have put the explanatory variables below each other in the matrix. Correction for autocorrelation using Zellner's solution (see Chapter 3) is made in the first stage of the 2SGM procedure (orthogonal projection) and in the second round of the Robust GMM technique (the final formula of the RGMM estimator). In the first case, not only the correlation between error terms of the same firm over different time-periods but also the correlation of the first stage disturbances is removed for the same firm, in the same time period but for different explanatory variables (Q etc.) when regressed on the instruments. We decide not to correct explicitly for heteroskedasticity neither in the first 2SGM round nor in the second round of the RGMM estimator but note that the covariance matrix associated with the robust GMM method is heteroskedasticity consistent. To determine the validness of the instruments we apply Hansen's (1982) test on overidentifying restrictions. Since the sample period is small, we prefer including another year in the sample period, as correction for autocorrelation becomes

feasible, to obtaining one more lag of the instrumental variables. Furthermore, Doornik-Hansen's (1994) normality test is applied on the weighted RGMM residuals.[10] Although the Robust GMM estimator is asymptotically normal under the conditions of a central limit theorem, even if this test rejects normality of the residuals, computation of the Doornik-Hansen statistic might be useful for indicating misspecification of the econometric model. Finally, the Wu-Hausman statistic, which this time is based on the difference between the classic and robust GMM estimates, is computed to evaluate the importance of robust estimation techniques.

Section 5.6 Empiricism

Table 5.2 contains the median values of the investment to capital ratio and its explanatory variables (of both the Q-model and the Euler equation approach) for each year over the period 1986-1990. After a period of economic stagnation at the beginning of the decade, 1983 was a turning point towards an episode of prosperity (see *Centraal Economisch Plan* 1986-1990).[11] Labour costs remained behind in the increase of labour productivity. The growth in world trade picked up considerably which boosted foreign demand. The Dutch econonomy is to a large extent dependent on the expansion of the economy of their trading partners given that in 1989 almost 70% of production is exported! Dutch industrial firms acquired high-quality capital goods and became more efficiently organized. These positive effects together were the cradle for a rise in firm profits, consumption and employment levels. Only in 1986-1987 was a slow-down in industrial activity experienced. Not least because of the weak dollar against the guilder which harmed the competitiveness of Dutch products. During the period 1983-1990 the investment climate was excellent for most of the Dutch companies. Our data meshes with this picture. The median of Tobin's Q over the full set of firms climbed from 0.500 in 1987 to 0.697 in 1990. The investment capital ratio, the cash flow capital ratio and output capital ratio, on average, show a moderate increase. Whereas, the debt to capital ratio squared in the case of the Q-model sample of firms, on average, decreased in 1989, after an increase in 1987, almost to the value of this ratio in 1986.

It is remarkable that low leveraged companies have considerably lower investment expenditures, cash flow and output levels than the high leveraged companies (compare column 4 and 5 of Table 5.2). The firms which were labelled as the financially healthy companies are also the ones with, on average, less thriving investment opportunities as measured by Tobin's Q. These primary results are in accordance with the predictions of the neoclassical Q-theory of investment. Apparently, companies with sound investment projects are willing to increase their debt to capital

[10] We note that, strictly speaking, the weighted RGMM residuals are not normally distributed under the central model distribution. However, the distribution of these weighted residuals approximates to the normal distribution.

[11] *Centraal Economisch Plan* is published yearly by the *Centraal Plan Bureau* (CPB).

Table 5.2 Median Values of the Investment Variables of both the Euler Equation Approach and Tobin's Q-model

	Euler Equation 117 companies	Q-model 95 companies	Q-model Group 1 48 companies	Q-model Group 2 47 companies
Investment Capital Ratio (I/K)				
1990	0.183	0.181	0.160	0.201
1989	0.172	0.173	0.150	0.206
1988	0.195	0.187	0.190	0.180
1987	0.177	0.171	0.166	0.193
1986	0.178			
Investment Capital Ratio Squared $((I/K)^2)$				
1989	0.031			
1988	0.038			
1987	0.036			
1986	0.035			
Tobin's Q				
1990		0.697	-0.042	1.923
1989		0.580	-0.148	1.727
1988		0.477	0.004	1.677
1987		0.500	-0.074	1.543
Cashflow Capital Ratio (C/K)				
1989	0.334	0.320	0.276	0.401
1988	0.323	0.308	0.239	0.377
1987	0.305	0.289	0.258	0.372
1986	0.312	0.305	0.250	0.354
Output Capital Ratio (Y/K)				
1990		4.751	2.274	9.193
1989	4.542	4.053	1.995	8.252
1988	4.360	4.040	2.098	7.150
1987	4.331	4.073	2.154	7.387
1986	4.369			
Debt to Capital Ratio squared $(B/K)^2$				
1989	2.019	1.683	0.797	5.125
1988	1.816	1.624	0.647	4.952
1987	2.014	1.753	0.681	4.080
1986	1.850	1.684	0.801	4.149

ratio in order to finance their promising projects and have access to the financial markets to do so. By interpreting only Table 5.2 one cannot distinguish between the investment opportunity argument (high Tobin's Q causes high I/K) and the argument that Modigliani-Miller's Proposition I does not hold, to explain the perceivable positive relationship between cash flow and investment. To test the null-hypothesis of perfect capital markets (Proposition I), we need to estimate the Q-model of investment and decide whether cash flow has significant explanatory power in addition to Tobin's Q.

5.6.1 Estimation Results, Euler Equation

Table 5.3 (see Appendix 2) reports the results concerning the first and second step of the estimation procedure outlined in Section 5.5. At first sight, the empirical performance of the Euler equation is poor. Only the output capital ratio enters the Euler equation with the correct sign on a significant coefficient. The coefficient on cash flow is positive and significantly different from zero contrary to the prediction of equation (5.22). This result might be interpreted as evidence for the failure of Modigliani and Miller's Proposition I. Evidently, if capital markets are not perfect, then cash flow can have a positive influence on the choice of the level of investment. The robust Pagan-Vella tests do not reject the assumption of homoskedastic error terms. In fact, correcting the GM estimates for heteroskedasticity does not substantially alter the outcomes (compare column 1 with column 2). The Frees test informs us about the independency of the observations by examining the cross-sectional correlation which appears statistically insignificant. The generalized Durbin-Watson statistic, though, indicates that the residuals are first-order autocorrelated.[12] We remark, in view of the discussion of section 2.6.1, that the upper and lower bounds as reported by Bhargava, Franzini and Narendranathan (1982) apply since in our case relatively few outliers are present in the data. Column 3 contains the heteroskedasticity and autocorrelation corrected GM estimates, which are not very different from the unadjusted counterparts. Adding the fixed effects to the econometric model is crucial for the parameter estimates as is revealed by the Wu-Hausman test. The coefficients from the level regressions are significantly different from the first difference or within parameter estimates. This result might reflect variations in the user cost of capital among corporations. Finally, the within estimates for their part differ significantly from the first difference estimates, thereby, calling for instrumental variable estimation techniques.

In order to gain efficiency, we only instrument those explanatory variables which are conceivably simultaneously determined with the investment capital ratio of time t. Considering the information structure and the decision process faced by the manager at time t-1, we include the lagged debt to capital ratio squared, the lagged investment capital ratio and its square in this

[12] The generalized Durbin-Watson statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation only in case of the "within" regressions. In the case of first difference regression we do not test for autocorrelation.

category. *A priori*, it is less clear whether the lagged cash flow to capital and output to capital ratio's are endogenous. Our estimation results of Table 5.4 indicate that these latter ratio's can be taken as exogenous variables. Hansen's test on overidentifying restrictions rejects the appropriateness of the first-lag instruments. Instruments with time labels t-2 and t-3, though, fullfill the necessary conditions to be considered as valid instruments (see Chapter 3). Outside instruments do not seem to be necessary. Note that adding the t-3 instruments restricts the sample period to one year. Thus, correction for autocorrelation in the second round of the B-robust GMM method is no longer feasible. With this end in view, we realize that the coefficients of column 3 might be more precisely estimated by leaving out the t-3 instruments but including the year 1989 in the sample period. The estimation results of this exercise are presented in Table 5.5. The performance of the Euler equation does not really improve.[13] Having sufficient internal funds and being a monopolistic competitor on the output market seem to be of great importance for the investment behaviour of Dutch manufacturing firms. We compute the F-statistic

$$F = (R\hat{\beta}_{RGMM})^{T}(Rvcov(\hat{\beta}_{RGMM})R^{T})^{-1}(R\hat{\beta}_{RGMM})/J \tag{5.27}$$

to test the null hypothesis that the coefficients on cash flow and debt are equal to zero, $R\hat{\beta}_{RGMM} = 0$, and conclude that Proposition I is rejected. As usual, the matrix R defines the J restrictions on the parameters β. Neither the standard nor the robust Wu-Hausman test rejects the equivalence of the classic and robust GMM estimates which is in accordance with the observation that the parameter estimates of columns 2 and 3 of Table 5.5 are very similar. However, the Doornik-Hansen test rejects normality in the classic case whereas it does not indicate possible misspecification of the econometric model in the robust regression case. Just a few outliers might be held responsible for the excessive value of the classical Doornik-Hansen statistic.

The classification of the firms by the debt to capital ratio appears to work well. As can be verified in Table 5.6, the F-test rejects Proposition I for the liquidity constrained (group 2) companies but does not reject the separability of the financial policy and the investment choices of the group 1 firms. This time, at least for the case of the group 2 regression, we reap the harvest of having developed the robust estimation techniques in the first part of the thesis, since the robust GMM estimator produces significantly different estimates from the classic one. Not a single coefficent is significantly different from zero for the regression which uses the observations on the unconstrained group of firms. We remark that for this case the normality of the weighted residuals is rejected which once again stresses the relative success of the empirical Euler equation.

[13] Adding the dividend capital ratio to the econometric specification or instrumenting all the explanatory variables does not lead to basically different conclusions.

5.6.2 Estimation Results, Q-model

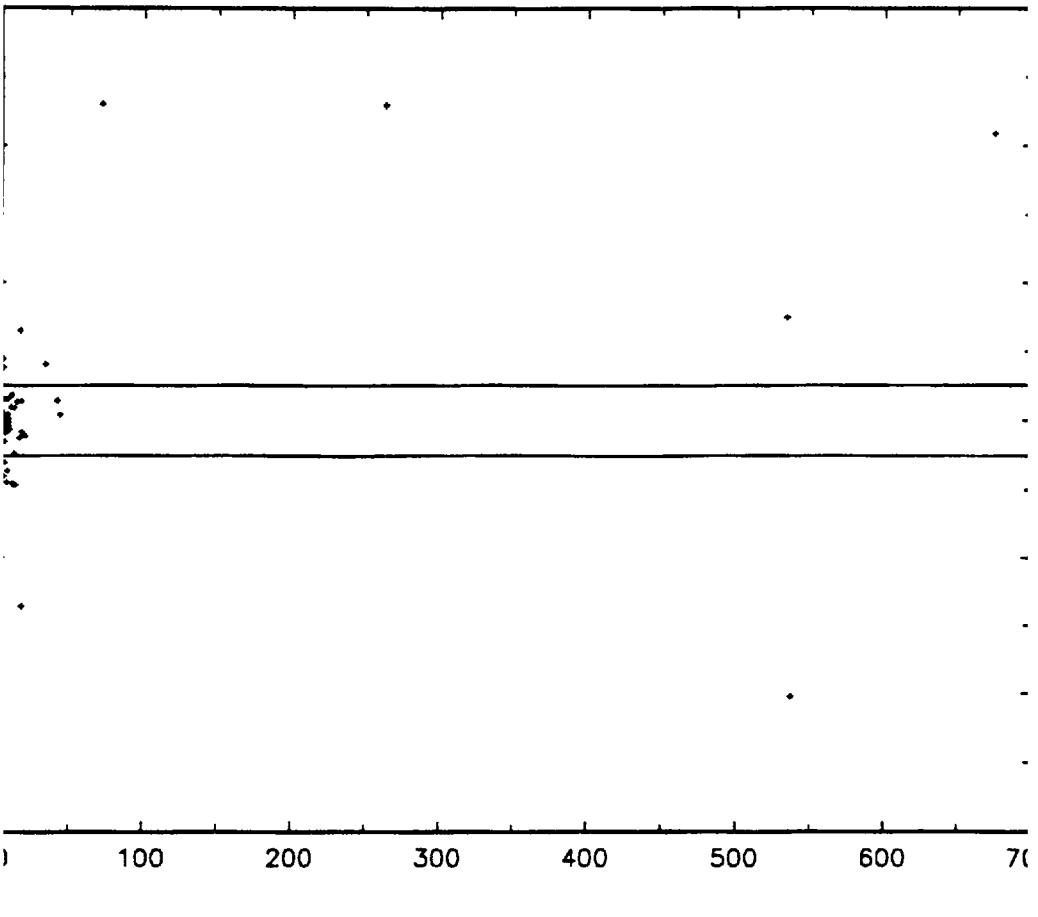
In contrast with the Euler equation, the Q-model fits the Dutch investment data extremely well. The GM estimates are shown in Table 5.7. All coefficients enter significantly different from zero. Only the debt to capital ratio is the black sheep of the explanatory variables with a coefficient of the wrong sign. As follows from Tables 5.8-5.11 this odd result might be explained by the endogeneity problem. Indeed, the Wu-Hausman test rejects the equality of the first difference and the within estimates in case of the heteroskedasticity and autocorrelation adjusted GM regressions. If we instrument the debt to capital ratio then its influence on the investment capital ratio becomes negligible. By anology with Table 5.3, the specification tests report similar econometric issues which need to be handled: the generalized Durbin-Watson statistics indicate that the residuals are not free of autocorrelation and firm specific effects can not be left out of the econometric model. This time the robust Pagan-Vella tests reject the null hypothesis of homoskedastic errors. From the t-values associated with the GM estimates of the second column one notices a small gain in efficiency on account of the correction for multiplicative heteroskedasticity. The Frees test legitimatizes the fact that we consider the cross-sectional units to be independent. Compared with previous empirical studies, the ability of Tobin's Q to predict the investment capital ratio is remarkably good.

Initially we make an orthogonal projection of Tobin's Q, the lagged debt to capital ratio squared and the current output capital ratio on their instruments. Instruments of period t-1 are again found to be invalid (see Table 5.8). Using t-2 and t-3 instruments only ensures that the robust GMM method produces consistent estimates if the lagged cash flow capital ratio is also included in the set of endogenous regressors. This result can be checked by comparing the robust Hansen (1982) statistic of the third with the fourth column.

The efficiency gain from leaving out the t-3 instruments but adding 1989 to the sample period is considerable which follows from a comparison of the last column of Table 5.8 with the second column of Table 5.9. For the full sample of firms, the conditions of the standard Q-model without financial regimes can not be rejected, if we consider all explanatory factors of equation (5.26). However, after leaving out the debt to capital ratio, the coefficient on the output capital ratio and cash flow capital ratio turn out to be significant. Again, liquidity constraints and the degree of competition in the output market are important determinants of the investment decisions of at least some manufacturing firms of the sample.[14] The weighted robust regression residuals are strongly autocorrelated but follow a normal distribution as is revealed by the generalized Durbin-Watson and the Doornik-Hansen tests respectively. The classic GMM procedure completely breaks down. Its dramatic decay is demonstrated by a substantial loss in the precision of the estimates, and hence also the t-values, and the relatively high values of the standard and robust

[14] Including the current cash flow capital ratio to the Q-model does not substantially change the reported estimation results and the corresponding added coefficient appears not significantly different from zero.

re 5.1 Robust Distances Versus Scaled 2SGM-Residuals*



cut-off values determining leverage points and vertical outliers are $\sqrt{\chi^2_{0.025}(5)}$ = 3.58 and 2.5
ively. The robust distances of the predicted explanatory variables and the scaled 2SGM-resid
horizontal and vertical axis respectively.

ausman statistics. These vulnerable GMM estimates result from just a few very large
data set as is shown in Figure 5.1 where the robust distances of the predicted expl
es are plotted against the scaled 2SGM residuals (see Chapter 1 for a detailed expos
ns of this sort). The leverage points especially corrupt the classic GMM results si
tionally large horizontal outliers also appear to be vertically outlying. The company
onding to the five smallest and largest residuals and the five largest robust distan
in Table 5.10. Combining Table 5.10 with Figure 5.1 leads us to conclude that o
nies put a stick in the wheel of the GMM estimator. It cannot be emphasized enou
ant it is to consider outlying observations in micro data. According to the classic
es, Tobin's Q does not explain much variation in the investment capital ratio whe
ust GMM regression case this is the most important explanatory variable.

Table 5.10. Vertical Outliers and Leverage points, detected by 2SGM

Total Number of Outliers = 66 Out of 190 Observations	Company	Year
Largest Positive Residual. Total Number of Vertical Outliers = 23		
1	VIBA	1989
2	A'dam Rubber	1990
3	A'dam Rubber	1989
4	Emba	1990
5	VRG	1989
Smallest Negative Residuals. Total Number of Vertical Outliers = 23		
1	Emba	1989
2	VIBA	1990
3	Econosto	1990
4	ACF	1989
5	VRG	1990
Largest Robust Distances. Total Number of Leverage points = 59		
1	A'dam Rubber	1989
2	Emba	1989
3	Emba	1990
4	A'dam Rubber	1990
5	VIBA	1989

Table 5.11 was drawn up from robust estimation of the Q-model seperately for two different financial regimes. The non-significance of the coefficient on the cash flow to capital ratio is conspicuous for the group of borrowers with relatively low debt levels. However, when we leave out the cash flow from the regression model then the explanatory power of Tobin's Q diminishes considerably. This result might indicate that the group 1 firms do not want to borrow on the capital markets, although they have free access, until they finish their internal available funds. Notice that for this class of firms, the coefficient on Tobin's Q is substantially higher than the ones reported for the liquidity constrained companies. Naturally, for the full set of firms (see Table 5.9) the parameter on Q approximates the average value. Tobin's Q-model is less accurate for the group 2 borrowers which is in accordance with our comment of Section 5.2.2. If a corporation is liquidity constrained then the measure of Q as dictated by equation (5.25) can be uninformative about the firm's real investment opportunities. Thus, although there seems to be no direct reason to reject Proposition I for the group 1 firms (the F statistic is far below the critical value 3), the Lagrange

multiplier of the financing constraint may still not equal zero since Tobin's Q may not be not a perfect image of the marginal value of extending the capital stock with one extra unit. But the empirical observation is that average Q remains the most important driving force behind the capital goods expenditures of the group 1 borrowers.

These results are reversed for the liquidity constrained companies. In this case, cash flow has more explanatory power than Tobin's Q and the F-test rejects the debt irrelevance theorem. Although we still do not find a significant relationship between the relative indebtedness of the company with respect to the capital stock and the investment capital ratio. One could argue that there is a lack of clarity about the existence of external financing barriers but the fact that internal financing constraints draw a bill on the investment choice is eminently clear. Even within the group of highly leveraged companies, firms, on average, do not suffer substantially from under-investment in their production processes because a high debt to capital ratio position putting them in a state of financial distress.

The output capital ratio is especially relevant for the group 1 regressions. Why monopolistic competitors would choose lower debt to capital ratio's is not at all immediately obvious. We leave this empirical finding unexplained.

For all regressions presented in Table 5.11, the Doornik-Hansen test does not reject the hypothesis that the weighted residuals are normally distributed. These results render the successful performance of the Q-specification of the neoclassical investment model under consideration. Once again, the generalized Durbin-Watson tests report severe first-order autocorrelation in the regression residuals. Recall that we deal with this problem by making a correction for autocorrelation with Zellner's solution (see Chapter 3).

Section 5.7 Conclusion

The explanatory power of empirical Q-models recently presented in the literature is often low. We argue that, among other reasons, outlying observations in the employed panel data sets can be blamed for corrupting standard econometric techniques resulting in misleading conclusions on the impact of Tobin's Q. Estimating the Q-model of investment with a Robust GMM estimator returns plausible results for a sample of 95 Dutch manufacturing companies. The RGMM estimator did convincingly solve the problem of securing consistent relatively efficient estimates in a model with conceivable simultaneity and autocorrelated errors.

We conclude by giving a reply to the questions which were raised in the introduction to this chapter. At least some of the firms of our Dutch panel data set are (internally) liquidity constrained when setting their investment objectives since cash flow plays an important role as an explanatory variable of the level of investment expenditures of group 2 borrowers in addition to Tobin's Q. These companies may raise their expected value by adopting an active hedging policy to lower the risk of having insufficient funds, when cash flow is temporarily low, to finance new

investment opportunities. In many cases, off-balance sheet hedging instruments such as options, futures and other financial derivatives can be exploited to diminish the capricious nature of the cash flow. Though it remains questionable whether the relatively highly leveraged companies are barred from the capital markets. Financial distress, as measured by the debt to capital ratio, does not have major consequences for the investment decisions of the majority of the Dutch corporations. The reason for the positive impact of cash flow on the investment choice may be imputed to the motive that managers of Dutch manufacturing companies prefer internally generated funds above external funds as a source of financing investment projects (pecking order argument). The estimation results associated with the Q-model reject Modigliani and Miller's Proposition I which stems from the difference between the external and internal costs of finance while at the same time firms have free access to the capital markets.

The empirical performance of the Euler equation is unsatisfactory for our panel data set which comprises only 117 Dutch firms.

We recall that the Euler equation and the Q-model are well specified only for companies with a particular financial regime. It is for future work to take up the challenge to develop an investment model which characterizes the optimal level of investment of the financially restricted firms in terms of direct observable explanatory factors.

Appendix A5.1 Data and Variable Construction

C (cash)	= net profits plus depreciation
Y (output)	= sales
I	= net investment in plant, property and equipment
K	= the capital stock
NK	= market value of the net capital stock
Q	= Tobin's Q
E	= market value of equity
B	= book value of total debt
p^I	= deflator for investment
p	= producer price index
D	= dividend
DEPR	= depreciation of the capital stock
F	= financial assets, this includes fixed interest securities, shares, investment in associates, partnerships and so on.

Table A5.1 contains the codes corresponding to the variables on which data is supplied by REACH. The original data set contains 142 companies over the ten-year period 1981-1990 and represents well the Dutch manufacturing industries which are quoted on the Amsterdam stock exchange. Note that international corporations such as Royal Dutch Ptl. (Shell), Philips, Unilever and Akzo Nobel are excluded. Maximizing the number of observations under the restriction that the resulting data set is balanced, leaves 117 companies in the case of the Euler equation approach and 95 companies in the case of the Q-model for the five-year period 1986-1990.

In order to convert the book value of the net capital stock into its market value, we use the perpetual inventory method (see Ward (1976) and Salinger and Summers (1983)):

$$NK_{it} = \frac{p_t^I}{p_{t-1}^I}\left(NK_{t-1,i} - DEPR_{t-1,i}\right) + I_{it} \qquad \text{(A5.1)}$$

The market value of the net capital stock for the first year is set equal to the book value $NK_{1986,i}$. The data set does not break down the capital stock into its components such as buidings, machinery, utility plants in service and transport equipment which prevents the application of different depreciation rates to different types of capital goods. Moreover, we are not able to construct the market value of the gross capital stock since we lack the figures on cumulative (over time) depreciaton. Therefore, we anticipate that the replacement value of the capital stock will be substantially higher than the one provided by A5.1. As a result, Tobin's Q, (I/K) etc. will generally

Table A5.1. Codes in Financieel Economisch Lexicon (FEL)

Variable	FEL Code
Capital Stock	M
Sales	AA
Investment	WIB
Short Debt	U
Long Debt	T
Cash Flow	WH
Dividend	KA
Financial Assets	N
Depreciation	WG

be overstated (see Table 5.2). Since the conceivable undervaluation of the capital stock affects monotonically both sides of the investment equations it does not obscure the estimation of the deep parameters of the investment models.

Output is approximated by sales which do not reflect the changes in the stock of finished goods. Our data on inventories include raw materials etc. and for this reason can not be used to construct a closer approximation for output.

The total amount of debt (B) of a firm is equal to the sum of the firm's short and long debt. Short-term debts are debts with maturities of less than one year. The book value of debt may differ from the market value because interest rates fluctuate. This difference in value seems especially relevant for long-term debt. Since additional information on the maturities of the long-term debts is not available, we cannot compute the market value of total debt.[15] However for the firms under investigation, on average, only approximately 20% of the level of debt can be ascribed to long-term debt. Schaller (1993) notes that his estimation results do not differ substantially if Tobin's Q is calculated with the market value of debt instead of the book value. The market value of equity (E) is calculated as the end of year share price multiplied by the number of ordinary shares in issue (datatype (MV) in DATASTREAM). Using the book value of financial assets (F) we calculate Q as

$$Q_{ti} = \left(\frac{B_{ti}+E_{ti}-F_{ti}}{NK_{ti}-I_{ti}}-1\right)\frac{p_t'}{p_t}. \qquad \text{(A5.2)}$$

[15] Whited (1992) uses the method of Brainard, Shoven and Weiss (1980) to construct the maturity distribution of book debt. Our data set does not cover enough years to apply this method successfully.

The median marginal value of one unit of capital ($\partial V_t/\partial K_{t-1}$) (see section 5.2.2) associated with measure (A5.2) ranges approximately from 0.97 in 1988 to 1.18 in 1990. These numbers are in conformity with the neoclassical theory of investment which predicts that $\partial V_t/\partial K_{t-1}$ will oscillate around one. Q is a highly simplified measure of the investment opportunities of a firm since, among other things, individual tax parameters are not taken into account. See Summers (1981) and Hayashi (1982) for a derivation of a tax-adjusted Q.[16] Perfect and Wiles (1994) compare different estimators of Tobin's Q to determine whether empirical analyses are sensitive to alternative constructions of Q and conclude that the simple Q-ratio, defined by the sum of the firm's common stock, the estimated market value of preferred stock and the book value of total debt divided by the bookvalue of total assets, produces empirical results that differ from four alternative relatively more complex estimators.

The *Economics Codes* in DATASTREAM of the producer price index (p_t) and the price deflator of gross fixed capital formation (p_t^I) for The Netherlands are NLI63...F and NLIPDINV respectively.

Variable Definitions:

Euler Equation

$$(I/K)_{it} = \frac{I_{it}}{NK_{it}},$$

$$(C/K)_{it} = \frac{p_t^I}{p_t}\left(\frac{C_{it}}{NK_{it}}\right)$$

$$(Y/K)_{it} = \frac{p_t^I}{p_t}\left(\frac{Y_{it}}{NK_{it}}\right)$$

$$(B/K)^2_{it} = \frac{p_t^I}{p_{t+1}}\left(\frac{B_{it}}{NK_{it}}\right)^2,$$

Q-model

$$(I/K)_{it} = \frac{I_{it}}{NK_{it}},$$

$$(C/K)_{it} = \frac{C_{it}}{NK_{it}},$$

$$(Y/K)_{it} = \frac{Y_{it}}{NK_{it}},$$

$$(B/K)^2_{t-1,i} = \frac{p_t^I}{p_t}\left(\frac{B_{t-1,i}}{(p_t^I/p_{t-1}^I)NK_{t-1,i}}\right)^2.$$

[16] Since data on the present value of the tax allowances associated with investments made before 1986 is not available we do not compute a tax-adjusted Q.

Appendix A5.2 Tables Containing the Estimation Results

Table 5.3 Generalized M-Estimates of the Euler Equation*

$$\frac{I_{it}}{K_{it}} = \beta_1(I/K)_{t-1,i} - \beta_2(I/K)^2_{t-1,i} - \beta_3(C/K)_{t-1,i} + \beta_4(Y/K)_{t-1,i} - \beta_5(B/K)^2_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}$$

	β_{GM}	β^H_{GM}	β^H_{RSUR}
"Within", Sample Period 1987-1990,		468 observations	
$(I/K)_{t-1}$	0.028 (0.25)	0.012 (0.12)	0.132 (0.98)
$(I/K)^2_{t-1}$	-0.014 (-0.005)	-0.002 (-0.01)	-0.202 (-0.74)
$(C/K)_{t-1}$	0.232 (5.19)	0.265 (5.80)	0.267 (4.99)
$(Y/K)_{t-1}$	0.024 (4.39)	0.026 (4.74)	0.022 (3.88)
$(B/K)^2_{t-1}$	-0.001 (-2.59)	-0.001 (-1.55)	-0.001 (-1.80)
Adjusted R^2	0.26	0.22	0.22
Heteroskedasticity, Robust Pagan-Vella (1989)	11.47		
Cross-sectional Correlation, Frees (1995)	-0.23	-0.001	-0.31
Generalized Durbin-Watson, Bhargava et al. (1982)	1.69	1.76	1.77
Doornik-Hansen (1994), Weighted Residuals	2.25	4.61	25.55
Wu (1973) - Hausman (1978) (Within ↔ Constant)	43634.29	1989.67	57.39
First Difference, Sample Period 1988-1990		351 Observations	
$(I/K)_{t-1}$	-0.345 (-3.05)	-0.310 (-2.68)	-0.270 (-2.51)
$(I/K)^2_{t-1}$	0.267 (1.22)	0.197 (0.89)	0.173 (0.86)
$(C/K)_{t-1}$	0.238 (5.44)	0.230 (4.90)	0.225 (4.55)
$(Y/K)_{t-1}$	0.021 (2.95)	0.021 (3.15)	0.019 (3.04)
$(B/K)^2_{t-1}$	0.001 (1.91)	0.001 (1.35)	0.001 (1.12)
Adjusted R^2	0.28	0.22	0.20
Heteroskedasticity, Robust Pagan-vella (1989)	10.80		
Doornik-Hansen (1994), Weighted Residuals	0.04	3.62	22.47
Wu (1973)-Hausman (1978) (First Dif. ↔ Constant)	904.72	261.14	378.33
Wu (1973)-Hausman (1978) (Within ↔ First Dif.)	143.35	44.60	49.71

*t-values are within parentheses, $t(\infty)_{0.025} = 1.96$. The robust Pagan-Vella statistic and the Wu-Hausman test statistics are asymptotically $\chi^2(8)$ distributed. The Wu-Hausman (within/first difference ↔ constant) statistic compares the within/first difference estimates with the estimates of the original model (5.22) where the fixed effects are left out but a constant is included. The lower bound and upper 95th percentile of the Frees statistic are respectively -0.67 and 1.21. The Doornik-Hansen test has a χ^2-distribution with two degrees of freedom, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(8) = 15.51$. The Bhargava test statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation. The time-dummies ϕ_t are not reported. The estimator β^H_{GM} corrects for heteroskedasticity whereas the estimator β^H_{RSUR} corrects for both heteroskedasticity and autocorrelation according to the procedures which are outlined in Chapter 2.

Table 5.4 Robust GMM Estimates of the Euler Equation; Determining the Optimal Set of Instruments*

$$\frac{I_{it}}{K_{it}} = \beta_1(I/K)_{t-1,i} - \beta_2(I/K)^2_{t-1,i} - \beta_3(C/K)_{t-1,i} + \beta_4(Y/K)_{t-1,i} - \beta_5(B/K)^2_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}$$

Sample Period	1989-1990	1990
Instruments	t-1, t-2	t-2, t-3
	$(I/K)_{t-1}$, $(I/K)^2_{t-1}$, $(B/K)^2_{t-1}$	$(I/K)_{t-1}$, $(I/K)^2_{t-1}$, $(B/K)^2_{t-1}$
Number of Observations	234	117
"Within"		
$(I/K)_{t-1}$	0.426 (1.46)	0.080 (0.13)
$(I/K)^2_{t-1}$	-0.317 (-0.31)	-0.066 (-0.04)
$(C/K)_{t-1}$	0.269 (1.52)	0.178 (2.50)
$(Y/K)_{t-1}$	0.043 (1.60)	0.033 (1.68)
$(B/K)^2_{t-1}$	-0.011 (-0.53)	-0.012 (-0.62)
Generalized Durbin-Watson, Bhargava et al. (1982)	1.24	
Doornik-Hansen (1994), Weighted Residuals	0.79	1.74
Test on Overidentifying Restrictions, Robust Hansen (1982)	18.56	2.38
Wu (1973) - Hausman (1978) (RGMM ↔ GMM, Standard/Robust)	238.36, 36.84	0.02, 6.79

*t-values are within parentheses, $t(\infty)_{0.025} = 1.96$. The robust Hansen test on overidentifying restrictions is asymptotically $\chi^2(3)$ distributed. The standard Wu-Hausman statistic has asymptotically a $\chi^2(6)$ distribution for column 2 (1 time dummy) and a $\chi^2(5)$ distribution for column 3 (no time dummy). The Doornik-Hansen test has a χ^2-distribution with two degrees of freedom, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(3) = 7.81$, $\chi^2_{0.05}(5) = 11.07$, $\chi^2_{0.05}(6) = 12.59$. The Bhargava test statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation. The time-dummies ϕ_t are not reported.

Table 5.5 Classic and Robust GMM Estimates of the Euler Equation*

$$\frac{I_{it}}{K_{it}} = \beta_1(I/K)_{t-1,i} - \beta_2(I/K)^2_{t-1,i} - \beta_3(C/K)_{t-1,i} + \beta_4(Y/K)_{t-1,i} - \beta_5(B/K)^2_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}$$

Sample Period 1989-1990	Robust GMM	Classic GMM
Instruments	t-2 $(I/K)_{t-1}$, $(I/K)^2_{t-1}$ $(B/K)^2_{t-1}$	t-2 $(I/K)_{t-1}$, $(I/K)^2_{t-1}$ $(B/K)^2_{t-1}$
Number of Observations	234	
"Within"		
$(I/K)_{t-1}$	-0.072 (-0.19)	0.152 (0.80)
$(I/K)^2_{t-1}$	0.547 (0.77)	0.286 (1.06)
$(C/K)_{t-1}$	0.200 (2.77)	0.221 (7.68)
$(Y/K)_{t-1}$	0.033 (4.39)	0.012 (2.22)
$(B/K)^2_{t-1}$	-0.005 (-1.82)	-0.0003 (-0.72)
Generalized Durbin-Watson, Bhargava et al. (1982)	1.15	1.13
Doornik-Hansen (1994), weighted residuals	1.14	105.75
F-test ($\beta_3=\beta_5=0$)	4.41	29.93
Wu (1973) - Hausman (1978) (RGMM $\leftrightarrow$ GMM, Standard/Robust)	9.65, 11.88	9.65, 11.88

*t-values are within parentheses, $t(\infty)_{0.025} = 1.96$. The standard Wu-Hausman statistic has asymptotically a $\chi^2(6)$ distribution. The Doornik-Hansen test has a χ^2-distribution with two degrees of freedom, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(6) = 12.59$. The Bhargava test statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation. The time-dummies ϕ_t are not reported. $F_{0.05}(2,\infty) = 3.00$.

Table 5.6 Robust GMM Estimates of the Euler Equation, Group 1 and Group 2*

$$\frac{I_{it}}{K_{it}} = \beta_1(I/K)_{t-1,i} - \beta_2(I/K)^2_{t-1,i} - \beta_3(C/K)_{t-1,i} + \beta_4(Y/K)_{t-1,i} - \beta_5(B/K)^2_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}$$

Sample Period 1989-1990	Group 1 Low Debt	Group 2 High Debt
Instruments	t-2 $(I/K)_{t-1}$, $(I/K)^2_{t-1}$ $(B/K)^2_{t-1}$	t-2 $(I/K)_{t-1}$, $(I/K)^2_{t-1}$ $(B/K)^2_{t-1}$
Number of Observations	118	116
"Within"		
$(I/K)_{t-1}$	-0.650 (-1.43)	0.296 (1.34)
$(I/K)^2_{t-1}$	1.457 (1.73)	-0.112 (-0.35)
$(C/K)_{t-1}$	0.281 (1.41)	0.150 (1.26)
$(Y/K)_{t-1}$	0.052 (0.87)	0.023 (2.54)
$(B/K)^2_{t-1}$	0.005 (0.03)	-0.002 (-2.46)
Generalized Durbin-Watson, Bhargava et al. (1982)	0.96	1.10
Doornik-Hansen (1994), Weighted Residuals	10.47	0.16
F-test ($\beta_3=\beta_5=0$)	1.01	10.13
Wu (1973) - Hausman (1978) (RGMM ↔ GMM, Standard/Robust)	1.13, 2.90	62.49, 17.09

*t-values are within parentheses, $t(\infty)_{0.025} = 1.96$. The standard Wu-Hausman statistic has asymptotically a $\chi^2(6)$ distribution. The Doornik-Hansen test has a χ^2-distribution with two degrees of freedom, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(6) = 12.59$. The Bhargava test statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation. The time-dummies ϕ_t are not reported. $F_{0.05}(2,\infty) = 3.00$.

Table 5.7 Generalized M-Estimates of the Q-model of Investment*

$$\frac{I_{it}}{K_{it}} = \beta_1 Q_{it} - \beta_2 (B/K)^2_{t-1,i} - \beta_3 (Y/K)_{it} + \beta_4 (C/K)_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}$$

	β_{GM}	β^H_{GM}	β^H_{RSUR}
"Within", Sample Period 1987-1990		380 observations	
Q_t	0.100 (3.74)	0.110 (8.16)	0.102 (7.84)
$(B/K)^2_{t-1}$	0.009 (2.18)	0.004 (2.94)	0.004 (3.30)
$(Y/K)_t$	-0.039 (-4.88)	-0.038 (-5.53)	-0.038 (-5.64)
$(C/K)_{t-1}$	0.248 (5.06)	0.297 (5.78)	0.295 (5.41)
Adjusted R^2	0.42	0.40	0.43
Heteroskedasticity, Robust Pagan-Vella (1989)	19.40		
Cross-sectional Correlation, Frees (1995)	0.16	0.63	0.15
Generalized Durbin-Watson, Bhargava et al. (1982)	1.54	1.56	1.61
Doornik-Hansen (1994), Weighted Residuals	2.03	5.45	21.83
Wu (1973) - Hausman (1978) (Within ↔ Constant)	42.99	32.10	50.59
First Difference, Sample Period 1988-1990		285 observations	
Q_t	0.133 (6.13)	0.138 (8.67)	0.134 (8.90)
$(B/K)^2_{t-1}$	0.009 (2.04)	0.011 (2.64)	0.009 (2.32)
$(Y/K)_t$	-0.046 (-6.96)	-0.050 (-7.93)	-0.048 (-8.16)
$(C/K)_{t-1}$	0.245 (4.43)	0.260 (5.00)	0.284 (5.54)
Adjusted R^2	0.51	0.46	0.54
Heteroskedasticity, Robust Pagan-vella (1989)	17.03		
Doornik-Hansen (1994), Weighted Residuals	0.42	5.21	21.09
Wu (1973)-Hausman (1978) (First Dif. ↔ Constant)	68.94	89.12	88.31
Wu (1973)-Hausman (1978) (Within ↔ First Dif.)	24.77	2.61	23.81

*t-values are within parentheses, $t(\infty)_{0.025} = 1.96$. The robust Pagan-Vella statistic and the Wu-Hausman test statistics are asymptotically $\chi^2(7)$ distributed. The Wu-Hausman (within/first difference ↔ constant) statistic compares the within/first difference estimates with the estimates of the original model (5.26) where the fixed effects are left out but a constant is included. The lower bound and upper 95th percentile of the Frees statistic are respectively -0.67 and 1.21. The Doornik-Hansen test has a χ^2-distribution with two degrees of freedom, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(7) = 14.07$. The Bhargava test statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation. The time-dummies ϕ_t are not reported. The estimator β^H_{GM} corrects for heteroskedasticity whereas the estimator β^H_{RSUR} corrects for both heteroskedasticity and autocorrelation according to the procedures which are outlined in Chapter 2.

Table 5.8 Robust GMM Estimates of the Q-model of Investment; Determining the Optimal Set of Instruments*

$$\frac{I_{it}}{K_{it}} = \beta_1 Q_{it} - \beta_2 (B/K)^2_{t-1,i} - \beta_3 (Y/K)_{it} + \beta_4 (C/K)_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}$$

Sample Period	1989-1990	1990	1990
Instruments	t-1, t-2	t-2, t-3	t-2, t-3
	Q_t, $(B/K)^2_{t-1}$, $(Y/K)_t$	Q_t, $(B/K)^2_{t-1}$, $(Y/K)_t$	Q_t, $(B/K)^2_{t-1}$, $(Y/K)_t$, $(C/K)_{t-1}$
Number of Observations	190	95	95
"Within"			
Q_t	0.131 (4.49)	0.074 (0.02)	0.110 (1.83)
$(B/K)^2_{t-1}$	0.012 (0.49)	-0.005 (-0.001)	-0.001 (-0.04)
$(Y/K)_t$	-0.054 (-1.67)	-0.021 (-0.02)	-0.029 (-1.71)
$(C/K)_{t-1}$	0.393 (1.42)	0.184 (0.14)	0.112 (1.07)
Generalized Durbin-Watson, Bhargava et al. (1982)	0.90		
Doornik-Hansen (1994), Weighted Residuals	2.13	1.15	0.80
Test on Overidentifying Restrictions, Robust Hansen (1982)	16.47	10.45	8.38
F-test (β_2=β_4=0)	4.06	1.35	0.87
Wu (1973) - Hausman (1978) (RGMM ↔ GMM, Standard/Robust)	36.44, 50.93	4.67, 4.18	12.08, 4.75

*t-values are within parentheses, $t(\infty)_{0.025} = 1.96$. The robust Hansen test on overidentifying restrictions is asymptotically $\chi^2(3)$ and $\chi^2(4)$ distributed in case of column 2, 3 and 4 respectively. The standard Wu-Hausman statistic has asymptotically a $\chi^2(5)$ distribution for column 2 (1 time dummy) and a $\chi^2(4)$ distribution for columns 3 and 4 (no time dummy). The Doornik-Hansen test has a χ^2-distribution with two degrees of freedom, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(3) = 7.81$, $\chi^2_{0.05}(4) = 9.49$, $\chi^2_{0.05}(5) = 11.07$. The Bhargava test statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation. The time-dummies ϕ_t are not reported. $F_{0.05}(2,\infty) = 3.00$.

Table 5.9 Classic and Robust GMM Estimates of the Q-model of Investment*

$$\frac{I_{it}}{K_{it}} = \beta_1 Q_{it} - \beta_2 (B/K)^2_{t-1,i} - \beta_3 (Y/K)_{it} + \beta_4 (C/K)_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}$$

Sample Period 1989-1990	Robust GMM	Robust GMM	Classic GMM
Instruments	t-2	t-2	t-2
	Q_t, $(B/K)^2_{t-1}$, $(Y/K)_t$, $(C/K)_{t-1}$	Q_t, $(Y/K)_t$, $(C/K)_{t-1}$	Q_t, $(B/K)^2_{t-1}$, $(Y/K)_t$, $(C/K)_{t-1}$
Number of Observations		190	
"Within"			
Q_t	0.143 (4.94)	0.109 (3.89)	0.008 (0.32)
$(B/K)^2_{t-1}$	0.002 (0.38)		-0.0003 (-0.93)
$(Y/K)_t$	-0.029 (-1.56)	-0.041 (-2.59)	-0.007 (-0.73)
$(C/K)_{t-1}$	0.195 (1.48)	0.324 (2.00)	0.305 (2.32)
Generalized Durbin-Watson, Bhargava et al. (1982)	0.96	1.01	0.91
Doornik-Hansen (1994), Weighted Residuals	1.69	2.46	105.74
F-test (β_2=β_4=0)	1.56		3.12
Wu (1973) - Hausman (1978) (RGMM ↔ GMM, Standard/Robust)	59.78, 97.40	12.56, 30.99	59.78, 97.40

*t-values are within parentheses, $t(\infty)_{0.025} = 1.96$. The standard Wu-Hausman statistic has asymptotically a $\chi^2(4)$ distribution for column 3 and a $\chi^2(5)$ distribution for columns 2 and 4. The Doornik-Hansen test has a χ^2-distribution with two degrees of freedom, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(4) = 9.49$, $\chi^2_{0.05}(5) = 11.07$. The Bhargava test statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation. The time-dummies ϕ_t are not reported. $F_{0.05}(2,\infty) = 3.00$.

Table 5.11 Robust GMM Estimates of the Q-model of Investment, Group 1 and Group 2*

$$\frac{I_{it}}{K_{it}} = \beta_1 Q_{it} - \beta_2 (B/K)^2_{t-1,i} - \beta_3 (Y/K)_{it} + \beta_4 (C/K)_{t-1,i} + \phi_t + \eta_i + \varepsilon_{it}$$

Sample Period 1989-1990	Group 1, Low Debt		
Instruments	t-2 Q_t, $(B/K)^2_{t-1}$, $(Y/K)_t$, $(C/K)_{t-1}$	t-2 Q_t, $(B/K)^2_{t-1}$, $(Y/K)_t$	t-2 Q_t, $(Y/K)_t$, $(C/K)_{t-1}$
Number of Observations		96	
"Within"			
Q_t	0.340 (4.51)	0.311 (0.92)	0.297 (4.04)
$(B/K)^2_{t-1}$	-0.021 (-0.25)	0.004 (0.02)	
$(Y/K)_t$	-0.136 (-3.04)	-0.119 (-1.49)	-0.121 (-2.27)
$(C/K)_{t-1}$	0.306 (1.39)		0.241 (1.19)
Bhargava et al. (1982)	1.04	1.02	0.97
Doornik-Hansen (1994)	0.85	1.49	2.26
F-test (β_2=β_4=0)	0.98		
Wu (1973) - Hausman (1978) (RGMM ↔ GMM, Standard/Robust)	-1.36, 1.56	-5.18, 2.50	-3.28, 16.40
Sample Period 1989-1990	Group 2, High Debt		
Instruments	t-2 Q_t, $(B/K)^2_{t-1}$, $(Y/K)_t$, $(C/K)_{t-1}$	t-2 Q_t, $(Y/K)_t$, $(C/K)_{t-1}$	t-2 Q_t, $(B/K)^2_{t-1}$, $(C/K)_{t-1}$
Number of Observations		94	
"Within"			
Q_t	0.054 (0.56)	0.036 (0.91)	0.045 (2.28)
$(B/K)^2_{t-1}$	0.004 (0.16)		-0.002 (-0.29)
$(Y/K)_t$	-0.013 (-0.26)	-0.012 (-0.66)	
$(C/K)_{t-1}$	0.330 (1.35)	0.338 (3.08)	0.358 (2.01)
Bhargava et al. (1982)	0.94	1.09	1.13
Doornik-Hansen (1994)	0.70	2.55	2.97
F-test (β_2=β_4=0)	2.30		4.18
Wu (1973) - Hausman (1978) (RGMM ↔ GMM, Standard/Robust)	-3.13, 1.78	1.29, 2.01	27.71, 27.94

*t-values are within parentheses, $t(\infty)_{0.025}$ = 1.96. The standard Wu-Hausman statistic has asymptotically a $\chi^2(4)$ distribution for column 3 and 4, and a $\chi^2(5)$ distribution for column 2. The Doornik-Hansen test has a χ^2-distribution with two degrees of freedom, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(4) = 9.49$, $\chi^2_{0.05}(5) = 11.07$. The Bhargava test statistic tends to 2 for large samples (in n) under the null-hypothesis of zero autocorrelation. The time-dummies ϕ_t are not reported. $F_{0.05}(2,\infty) = 3.00$.

REFERENCES

Ali, M.M. and Sharma, S.C., 1996, "Robustness to Nonnormality of Regression *F*-tests", *Journal of Econometrics*, vol. 71, 175-205.

Alonso, C. and Bentolila, S., 1994, "Investment and Q in Spanish Industrial Firms", *Oxford Bulletin of Economics and Statistics*, vol. 56, no. 1, 49-65.

Arellano, M. and Bond, S., 1988, "Dynamic Panel Data Estimation Using DPD - A Guide For Users", IFS Working Paper, no. 88/15.

Beaton, A.E. and Tukey, J.W., 1974, "The fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectro-Scopic Data", *Technometrics*, Vol.16, 147-85.

Becht, M. and Ramirez, C., 1993, "Financial Capitalism in Pre-World War I Germany: The Role of the Universal Banks in the Financing of German Mining Companies 1906-1912", *EUI Working Paper*, ECO no.93/26.

Berger, A.N. and Udell, G.F., 1992, "Some Evidence on the Empirical Significance of Credit Rationing." *Journal of Political Economy*, vol. 100, no.5, 1047-1077.

Bester, H., 1994, "The Role of Collateral in a Model of Debt Renegotiation", *The Journal of Money, Credit and Banking*, Vol. 26, No.1, 72-86.

Bester, H. and Hellwig, M., 1987, "Capital Markets and Moral Hazard" in Bamberg G. and Spreman K., ed.: *Agency Theory, Information, and Incentives*, Heidelberg.

Bhargava, A., Franzini, L. and Narendranathan, 1982, "Serial Correlation and the Fixed Effects Model", *Review of Economic Studies*, 533-549.

Blanchard, O.J. and Fischer, S., 1989, "Lectures on Macroeconomics", *M.I.T. Press*, Cambridge.

Blundell, R., Bond, S., Devereux, M. and Schiantarelli, F., 1992, "Investment and Tobin's Q: Evidence from Company Panel Data." *Journal of Econometrics*, vol.51, 233-257.

Bond, S. and Meghir, C., 1992, "Dynamic Investment Models And The Firm's Financial Policy", *UCL Discussion Paper*, no. 92-14.

Brainard, W.C., Shoven J.B. and L. Weiss (1980) "The Financial Valuation of the Return to Capital." *Brookings Papers on Economic Activity* 2, 453-512.

Brealey, R.A. and Myers, S.C., 1988, "Principles of Corporate Finance", third edition, *McGraw-Hill*, New York.

Breusch, T.S. and Pagan, A.R., 1980, "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics", *Review of Economic Studies*, vol. 47, 239-253.

Centraal Plan Bureau, 1986-1990, *Centraal Economisch Plan*, The Hague, The Netherlands.

Chen, C. and Liu, L.M., 1993, "Joint Estimation of Model Parameters and Outlier Effects in Time Series", *Journal of the American Statistical Association*, Vol.88, 284-97.

Chirinko, R.S., 1987, "Tobin's Q and Financial Policy", *Journal of Monetary Economics*, 19, 69-87.

Davidson, R. and Mackinnon, J.G., 1993, *Estimation and Inference in Econometrics*, Oxford University Press.

De Meza, D. and Webb, D.C., 1987, "Too Much Investment: A Problem of Asymmetric Information",

Quarterly Journal of Economics, vol. 102, pp. 281-292.

Detemple, J., Gottardi, P. and Polemarchakis, H.M., 1995, "The Relevance of Financial Policy", *European Economic Review*, vol. 39, 1133-1154.

Donoho, D.L. and Huber, P.J., 1983, "The notion of breakdown point", in *A Festschrift for Erich Lehmann*, edited by P. Bickel, K. Doksum and J.L. Hodges, Jr., Wadsworth. Belmont, CA.

Doornik, J.A. and Hansen, H.,1994, "A Practical Test for Univariate and Multivariate Normality", mimeo, Nuffield College, Oxford.

Fazzari, S., Hubbard, R.G. and Petersen, B., 1988a, "Finance Constraints and Corporate Investment", *Brookings Papers on Economic Activity*, 1, 141-195.

Fazzari, S., Hubbard, R.G. and Petersen, B, 1988b, "Investment, Financing Decisions, and Tax Policy", *The American Economic Review*, 78, 200-205.

Frees, E.A., 1995, "Assessing cross-sectional correlation in panel data", *Journal of Econometrics*, vol. 69, 393-414.

Froot, K.A., Scharfstein, D.S. and Stein, J.C., 1993, "Risk Management: Coordinating Corporate Investment and Financing Policies", *The Journal of Finance*, vol. XLVIII, no. 5, 1629-1658.

Gale, D. and Hellwig, M., 1985, "Incentive-Compatible Debt Contracts: The One-Period Problem", *Review of Economic Studies*, vol. 52, 647-663.

Gasko, M. and Donoho, D., 1982, "Influential Observation in Data Analysis", *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 104-9.

Gómez, V. and Maravall, A., 1994, "Program Tramo: Time Series Regression with Arima Noise, Missing Observations, and Outliers; Instructions for the User", Working Paper ECO 94/31, *European University Institute*.

Griliches, Z. and Hausman, J.A., 1986, "Errors in Variables in Panel Data", *Journal of Econometrics*, vol. 31, 93-118.

Hammond, P.J., 1987, "On the possibility of Perfect Capital Markets", mimeo, Stanford University, California.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., 1986, "Robust Statistics. The Approach Based on Influence Functions", *Wiley*, New York.

Hansen, B.E., 1996, "Threshold Effects in Panel Data: Estimation, Testing, and Inference, mimeo, Boston College.

Hansen, L.P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators",*Econometrica*, vol. 50, no. 4, 1029-1054.

Harris, M. and Raviv, A., 1991, "The Theory of Capital Structure", *The Journal of Finance*, vol. XLVI, no. 1, 297-355.

Harvey, A.C., 1976, "Estimating Regression Models With Multiplicative Heteroskedasticity", *Econometrica*, vol. 44, no. 3, 461-465.

Hausman, J.A., 1978, "Specification Tests in Econometrics", *Econometrica*, vol. 46, 1251-1271.

Hayashi, F., 1982, "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica*, vol. 50, no. 1, 213-224.

Hayashi, F., 1985, "Corporate Finance Side of the Q Theory of Investment", *Journal of Public Economics*, vol. 27, 261-280.

Hendry, D.F., (1987), "Econometric Methodology: a Personal Perspective", in Bewley, T.F. (ed.), *Advances in Econometrics: Fifth World Congress*, vol.2, Ch. 10, Cambridge University Press.

Hillier, B. and Ibrahimo, M.V., 1993, "Asymmetric Information and Models of Credit Rationing", *Bulletin of Economic Research*, vol. 45, no.4, 271-304.

Hillier, B. and Worrall, T., 1992, "The Welfare Implications of Costly Monitoring in the Credit Market", mimeo, *University of Liverpool*, Liverpool.

Hinloopen, J., 1997, *Research and Development, Product Differentiation and Robust Estimation*, Ph.D. thesis, European University Institute, Florence.

Hinloopen, J. and Wagenvoort, J.L.M., 1995, "Robust Estimation: An Example", Working Paper ECO no. 95/2, *European University Institute*, Florence.

Hinloopen, J. and Wagenvoort, J.L.M., 1997, "On the Computation and Efficiency of a HBP-GM Estimator: Some Simulation Results", *Computational Statistics and Data Analysis*, vol.25, no. 1, 1-15.

Hoshi, T., Kashyap, A. and Scharfstein, D., 1991, "Corporate Structure, Liquidity, and Investment: Evidence from Japanese Industrial Groups", *Quarterly Journal of Economics*, vol. 106, 33-60.

Huber, P.J., 1964, "Robust Estimation of a Location Parameter", Annual of Mathematical Statistics, vol.35, 73-101.

Huber, P.J., 1973, "Robust Regression: Asymptotics, Conjectures and Monte Carlo, *The Annals of Statistics*, 1, 799-821.

Huber, P.J., 1981, "Robust Statistics", *Wiley*, New York.

Jaffee, D.M. and Russell, T., 1976, "Imperfect Information, Uncertainty, and Credit Rationing", *Quarterly Journal of Economics*, 651-666.

Jensen, M.C. and Meckling, W.H., 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", *Journal of Financial Economics*, 3, 305-360.

Jorgenson, D.W. (1963) "Capital Theory and Investment Behaviour, *American Economic Review* 53.

Judge, G.G., Hill R.C., Griffiths, W.E., Lütkepohl, H. and Lee, T.C., 1988, *Introduction to the Theory and Practice of Econometrics*. Wiley, New York.

Kloek, T., 1996, Correspondence Address: Erasmus University Rotterdam, Department of Econometrics and Operations Research, PO Box 1738, 3000 DR Rotterdam, The Netherlands.

Krasker, W.S., 1986, "Two-Stage Bounded-Influence Estimators for Simultaneous-Equations Models", *Journal of Business & Economic Statistics*, vol. 4, no. 4, 437-444.

Krasker, W.S. and Welsch, R.E., 1982, "Efficient Bounded-Influence Regression Estimation", *Journal of the American Statistical Association*, vol. 77, no. 379, 595-604.

Krasker, W.S. and Welsch, R.E., 1985, "Resistant Estimation For Simultaneous-Equations Models Using Weighted Instrumental Variables", *Econometrica*, vol. 53, no. 6, 1475-1488.

Krishnakumar, J., 1995, "A General Methodology of Robust Estimation for Panel Data Models", *Cahier 95.02*, University of Geneva, Department of Econometrics.

Lucas, R.E., 1976, "Econometric Policy Evaluation: A Critique", in Brunner, K. and Meltzer, A. (eds.), "The Phillips Curve and Labour Markets", *Carnegie-Rochester Conference Series*, vol 1.

Mallows, C.L., 1975, "On Some Topics in Robustness", Technical Memorandom, *Bell Telephone Laboratories*, Murray Hill, NJ.

Marcus, A.J., 1989, "An Equilibrium Theory of Excess Volatility and Mean Reversion in Stock Market Prices", *NBER Working Paper*, no. 3106.

Markatou, M. and Hettmansperger, T.P., 1990, "Robust Bounded-Influence Tests in Linear Models", *Journal of the American Statistical Association*, vol. 85, no. 409, 187-190.

Maronna, R.A., Bustos, O.H. and Yohai, V.J., 1979, "Bias- and Efficiency-Robustness of General M-estimators for Regression With Random Carriers", in Gasser, T. and Rosenblatt, M. (eds.), *Smoothing Techniques for Curve Estimation*, New York, Springer-Verlag, 91-116.

Maronna, R.A. and Yohai, V.J., 1981, "Asymptotic Behavior of General M-estimates for Regression and Scale with Random Carries", *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 58, 7-20.

Mátyás, L. and Sevestre, P., 1992, *The Econometrics of Panel Data, Handbook of Theory and Applications*, Kluwer, Dordrecht.

Mauer, D.C. and Triantis, A., 1994, "Interactions of Corporate Financing and Investment Decisions: A Dynamic Framework", *The Journal of Finance*, vol. XLIX, no. 4, 1253-1277.

Metcalf, G.E., 1996, "Specification testing in panel data with instrumental variables", *Journal of Econometrics*, 71, 291-307.

Mizon, G.E., 1995, "A Simple Message for Autocorrelation Correctors: Don't", *Journal of Econometrics*, vol. 69, p.267-288.

Modigliani, F. and Miller, M.H., 1958, "The Cost of Capital, Corporation Finance And the Theory of Investment", *The American Economic Review*, vol. XLVIII, no. 3, 261-297.

Morgan, D.P., 1994, "Bank Credit Commitments, Credit Rationing, and Monetary Policy", *Journal of Money, Credit, and Banking*, vol.26, no.1, 87-101.

Myers, S.C., 1984, "The Capital Structure Puzzle", *The Journal of Finance*, vol. XXXIX, no. 3, 575-592.

Myers, S.C. and Majluf, N.S., 1984, "Corporate Financing and Investment Decisions when Firms have Information that Investors do not have", *Journal of Financial Economics* 13, 187-221.

Newey, W.K. and West, K.D., 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, vol. 55, 703-708.

O'Connell, S.A. and Zeldes, S.P., 1988, "Rational Ponzi Games", *International Economic Review*, vol. 29, no. 3, 431-450.

Pagan, A. and Vella, F., 1989, "Diagnostic Tests for models Based on Individual Data: A Survey", *Journal of Applied Econometrics*, vol. 4, S29-S59.

Peracchi, F., 1991, "Bounded-Influence Estimators for the SURE Model", *Journal of Econometrics*, vol. 48, 119-134.

Perfect, S.B. and K.W. Wiles, 1994, "Alternative constructions of Tobin's q: An empirical comparison." *Journal of Empirical Finance*, vol. 1, 313-341.

Robert, C.P., 1994, "Mixtures of Distributions: Inference and Estimation", *Série des Documents de Travail du CREST*, no. 9413, INSEE.

Rondi, L., Sembenelli, A. and Zanetti, G., 1994, "Is Excess Sensitivity of Investment to Financial Factors Constant Across Firms? Evidence From Panel Data on Italian Companies", *Journal of Empirical Finance*, vol. 1, 365-383.

Rousseeuw, P.J., 1984, "Least Median of Squares Regression", *Journal of the American Statistical Association*, Vol. 79, No.388, December, 871-80.

Rousseeuw, P.J. and Yohai, V., 1984, "Robust regression by means of S-estimators", in *Robust and Nonlinear Time Series Analysis*, edited by J. Franke, W. Härdle and R.D. Martin, Lecture Notes in Statistics No. 26, Springer Verlag, New York, 256-272.

Rousseeuw, P.J., 1985, "Multivariate Estimation with High Breakdown Point", in *Mathematical Statistics and Applications*, Grossman, W., Pflug, G., Vincze, I. and Wertz, W. (eds.), *Riedel Publishing*, Dordrecht, Vol.B, 283-97.

Rousseeuw, P.J. and Leroy, A.M., 1987, "Robust Regression and Outlier Detection", *Wiley*, New York.

Rousseeuw, P.J. and van Zomeren, B.C., 1990a, "Unmasking Multivariate Outliers and Leverage Points", *Journal of American Statistical Association*, Vol.85, No.411, September, 633-39.

Rousseeuw, P.J. and van Zomeren, B.C., 1991, "Robust Distances: Simulations and Cutoff Values", in Stahel, W. and Weisberg, S. (eds.), "Directions in Robust Statistics and Diagnostics, Part II", *Springer Verlag*, New York (NY), p.195-203.

Schaller, H., 1993, "Asymmetric Information, Liquidity Constraints, and Canadian Investment", *Canadian Journal of Economics*, vol. XXVI, 552-574.

Schiantarelli, F. and D. Georgoutsos (1990) "Monopolistic Competition and the Q Theory of Investment.", *European Economic Review*, 34, 1061-1078.

Simpson, D.G., Ruppert, D. and Carroll, R.J., 1992, "On One-Step GM Estimates and Stability of Inferences in Linear Regression", *Journal of the American Statistical Association*, Vol.87, No.418, June, 439-50.

Spanos, A., 1986, *Statistical Foundations of Econometric Modelling*, Cambridge University Press.

Spanos, A., 1993, "On Theory Testing in Econometrics: The Case of the Efficient Market Hypothesis", mimeo, University of Cyprus, Cyprus.

Stiglitz, J.E., 1988, "Money, Credit, and Business Fluctuations", *Economic Record*, vol. 64, pp. 307-322.

Stiglitz, J.E. and Weiss, A., 1981, "Credit Rationing in Markets with Imperfect Information", *The American Economic Review*, vol. 71, no. 3, 393-410.

Salinger, M.A. and L.H. Summers (1983) "Tax Reform and Corporate Investment: A Microeconomic Simulation Study", in Martin Feldstein, ed.: *Behavioral Simulation Methods in Tax Policy Analysis*, University of Chicago Press, Chicago.

Summers, L.H. (1981) "Taxation and Corporate Investment: A q-theory Approach, *Brookings Papers*

on Economic Activity, vol. 1, 67-127.

Theil, H., 1971, *Principles of Econometrics*, Wiley, New York.

Titterington, D.M., Smith, A.F.M. and Makov, U.E., 1985, *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.

Tobin, J., 1969, "A General Equilibrium Approach to Monetary Theory", *Journal of Money, Credit and Banking*, 1, 15-29.

Vercammen, J., 1994, "Capital Constraints and Optimal Hedging in Factor Markets", *Economic Letters*, 45, 433-439.

Ward, M., 1976, *The Measurement of Capital, The Methodology of Capital Stock Estimates in OECD Countries*, OECD, Paris.

White, H., 1982, "Instrumental Variables Regression with Independent Observations", *Econometrica*, vol.50, no.2, 483-499.

Whited, T.M., 1992, "Debt, Liquidity Constraints, and Corporate Investment: Evidence from Panel Data", *The Journal of Finance*, vol. XLVII, no. 4, 1425-1460.

Williamson, S.D., 1986, "Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing", *Journal of Monetary Economics*, 18, 159-179.

Wu, D., 1973, "Alternative Tests of Independence Between Stochastic Regressors and Disturbances." *Econometrica*, vol. 41, 733-750.

Zellner, A., 1962, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias", *Journal of the American Statistical Association*, 57, 348-368.