SFB 649 Discussion Paper 2006-050

Robust Econometrics

Pavel Čížek*
Wolfgang Härdle**

* Department of Econometrics and Operations Research,
Universiteit van Tilburg, The Netherlands
** Institute for Statistics and Econometrics and
C.A.S.E. – Center for Applied Statistics and Economics,
School of Business and Economics,
Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Robust Econometrics

P. Čížek and W. Härdle*

Econometrics often deals with data under, from the statistical point of view, non-standard conditions such as heteroscedasticity or measurement errors and the estimation methods need thus be either adopted to such conditions or be at least insensitive to them. The methods insensitive to violation of certain assumptions, for example insensitive to the presence of heteroscedasticity, are in a broad sense referred to as robust (e.g., to heteroscedasticity). On the other hand, there is also a more specific meaning of the word 'robust', which stems from the field of robust statistics. This latter notion defines robustness rigorously in terms of behavior of an estimator both at the assumed (parametric) model and in its neighborhood in the space of probability distributions. Even though the methods of robust statistics have been used only in the simplest setting such as estimation of location, scale, or linear regression for a long time, they motivated a range of new econometric methods recently, which we focus on in this chapter.

*This work was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

The concepts and measures of robustness are introduced first (Section 1), followed by a most common types of estimation methods and their properties (Section 2). Various econometric methods based on these common estimators are discussed later in Section 3, covering tasks from time series regression over GMM estimation to simulation-based methods.

1 Measures of robustness

Robustness properties can be formulated within two frameworks: qualitative and quantitative robustness. *Qualitative robustness* is concerned with the situation in which the shape of the underlying (true) data distribution deviates slightly from the assumed model. It focuses on questions like stability and performance loss over a family of such slightly deviating distributions. *Quantitative robustness* considers the situation in which the sensitivity of estimators to a proportion of aberrant observations is studied.

A simple example can make this clear. Suppose one has collected a sample on an individual's income (after say 10 years of schooling) and one is interested in estimating the mean income. If $\{x_i\}_{i=1}^n$ denotes the logarithm of this data and we suppose that they have a cumulative distribution function (cdf) F, assumed to be $N(\mu, \sigma^2)$, the maximum likelihood estimator (MLE) is $\bar{x} = \int u dF_n(u) = T(F_n)$, where $F_n(u) = n^{-1} \sum_{i=1}^n I(x_i \leq u)$, and $\mu = \int u dF(u) = T(F)$. Qualitative robustness asks here the question: how well will μ be estimated if the true distribution is in some neighborhood of

F? Quantitative robustness would concentrate on: will $T(F_n)$ be bounded if some observations $x_i \to \infty$? In fact, the last question is easy to answer: if $x_i \to \infty$ for some i, $T(F_n) = \bar{x} \to \infty$ as well. So we can say here in a loose sense that $\bar{x}$ is not quantitatively robust.

Formalities

In the following we present a mathematical setup that allows us to formalize the robustness thoughts.

The notion of the sensitivity of an estimator T is put into theory by considering a model characterized by a cdf F and its neighborhood $\mathcal{F}_{\varepsilon,G}$: distributions $(1-\varepsilon)F + \varepsilon G$, where $\varepsilon \in (0, 1/2)$ and G is an arbitrary probability distribution, which represents data contamination. Hence, not all data necessarily follow the pre-specified distribution, but the ε-part of data can come from a different distribution G. If $H \in \mathcal{F}_{\varepsilon,G}$, the estimation method T is then judged by how sensitive or robust are the estimates $T(H)$ to the size of $\mathcal{F}_{\varepsilon,G}$, or alternatively, to the distance from the assumed cdf F. Two main concepts for robust measures analyze the sensitivity of an estimator to infinitesimal deviations, $\varepsilon \to 0$, and to finite (large) deviations, $\varepsilon > 0$, respectively. Despite generality of the concept, easy interpretation and technical difficulties often limit our choice to point-mass distributions (Dirac measures) $G = \delta_x, x \in \mathbb{R}$, which simply represents an (erroneous) observation at point $x \in \mathbb{R}$. This simplification is also used in the following text.

The influence of infinitesimal contamination on an estimator is characterized by the *influence function*, which measures the relative change in estimates caused by an infinitesimally small amount ε of contamination at x (Hampel et al., 1986). More formally,

$$IF(x; T, F) = \lim_{\varepsilon \to 0} \frac{T\{(1-\varepsilon)F + \varepsilon\delta_x\} - T(F)}{\varepsilon}. \quad (1)$$

For each point x, the influence function reveals the rate at which the estimator T changes if a wrong observation appears at x. In the case of sample mean $\bar{x} = T(F_n)$ for $\{x_i\}_{i=1}^n$, we obtain

$$\begin{aligned} IF(x; T, F_n) &= \lim_{\varepsilon \to 0} \left[(1-\varepsilon) \int u dF_n(u) + \varepsilon \int u d\delta_x(u) - \int u dF_n(u) \right] / \varepsilon \\ &= \lim_{\varepsilon \to 0} \left[- \int u dF_n(u) + \int u d\delta_x(u) \right] = x - \bar{x}. \end{aligned}$$

The influence function allows us to define various desirable properties of an estimation method. First, the largest influence of contamination on estimates can be formalized by the *gross-error sensitivity*,

$$\gamma(T, F) = \sup_{x \in \mathbb{R}} IF(x; T, F), \quad (2)$$

which under robustness considerations be finite and small. Even though such a measure can depend on F in general, the qualitative results (e.g., $\gamma(T, F)$ being bounded) are typically independent of F. Second, the sensitivity to small changes in data, for example moving an observation from x to $y \in \mathbb{R}$, can be measured by the *local-shift sensitivity*

$$\lambda(T, F) = \sup_{x \neq y} \frac{\|IF(x; T, F) - IF(y; T, F)\|}{\|x - y\|}. \quad (3)$$

Also this quantity should be relatively small since we generally do not expect that small changes in data cause extreme changes in values or sensitivity of estimates. Third, as an unlikely large or distant observations may represent data errors, their influence on estimates should become zero. Such a property is characterized by the *rejection point*,

$$\rho(T, F) = \inf_{r>0}\{r : IF(x; T, F) = 0, \|x\| \geq r\}, \tag{4}$$

which indicates the non-influence of large observations.

Alternatively, behavior of the estimator T can be studied for any finite amount ε of contamination. The most common property looked at in this context is the estimator's bias $b(T, H) = E_H\{T(H)\} - E_F\{T(F)\}$, which measures a distance between the estimates for clean data, $T(F)$, and contaminated data, $T(H), H \in \mathcal{F}_{\varepsilon,G}$. The corresponding *maximum-bias curve* measures the maximum bias of T on $\mathcal{F}_{\varepsilon,G}$ at any ε:

$$B(\varepsilon, T) = \sup_{x\in\mathbb{R}} b\{T, (1-\varepsilon)F + \varepsilon\delta_x\}. \tag{5}$$

Although the computation of this curve is rather complex, Berrendero and Zamar (2001) provide general methodology for its computation in the context of linear regression.

The maximum-bias curve is not only useful on its own, but allows us to define further scalar measures of robustness. The most prominent is the *breakdown point* (Hampel, 1971), which is defined as the smallest amount ε

of contamination that can cause an infinite bias:

$$\varepsilon^*(T) = \inf_{\varepsilon \geq 0} \{\varepsilon : B(\varepsilon, T) = \infty\}. \tag{6}$$

The intuitive aim of this definition specifies the breakdown point $\varepsilon^*(T)$ as the smallest amount of contamination that makes the estimator T useless. Note that in most cases $\varepsilon^*(T) \leq 0.5$ (He and Simpson, 1993). This definition and the upper bound however apply only in simple cases, such as location or linear regression estimation (Davies and Gather, 2005). The most general definition of breakdown point formalizes the idea of "useless" estimates in the following way: an estimator is said to break down if, under contamination, it is not random anymore, or more precisely, it can achieve only a finite set of values (Genton and Lucas, 2003). This definition is based on the fact that estimates are functions of observed random samples and are thus random quantities themselves unless they fail. Although the latter definition includes the first one, the latter one may generally depend on the underlying model F, for example in time-series context.

2 Estimation approaches

Denote by F_n an empirical distribution function (edf) corresponding to a sample $\{x_i\}_{i=1}^n \in \mathbb{R}$ drawn from a model based on probability distribution F. Most estimation methods can be defined as an extremum problem, minimizing a contrast $\int h(z, \theta) dF(z)$ over θ in a parameter space, or as a solution of an equation, $\int g(z,\theta)dF(z) = 0$ in θ. The estimation for a given sample utilizes finite-sample equivalents of these integrals, $\int h(z,\theta)dF_n(z)$ and $\int g(z,\theta)dF_n(z)$, respectively.

Consider the pure location model $X_i = \mu + \sigma\varepsilon_i, i = 1, \ldots, n$, with a known scale σ and $\varepsilon \sim F$. The cdf of X is then $F\{(x-\mu)/\sigma\}$. With a quadratic contrast function $h(x,\theta) = (x-\theta)^2$, the estimation problem is to minimize $\int (x-\theta)^2 dF\{(x-\mu)/\sigma\}$ with respect to θ. For known F, this leads to $\theta = \mu$ and one sees that, without loss of generality, one can assume $\mu = 0$ and $\sigma = 1$. For the sample $\{x_i\}_{i=1}^n$ characterized by edf F_n, the location parameter μ is estimated by

$$\hat{\mu} = \arg\min_{\theta} \int (x-\theta)^2 dF_n(x) = n^{-1}\sum_{i=1}^{n} x_i = \bar{x}.$$

Note that for $g(x,\theta) = x - \theta$, the parameter μ is the solution to $\int g(x,\theta)dF(x) = 0$. The estimator may therefore be alternatively defined through $\mu = T(F) = \int u dF(u)$.

As indicated in the introduction, this standard estimator of location performs unfortunately rather poorly under the sketched contamination model. Estimating a population mean by the least squares (LS) or sample mean

$\bar{x} = T(F_n)$ has the following properties. First, the influence function (1)

$$
\begin{aligned}
IF(x;T,F) &= \lim_{\varepsilon\to 0} \frac{T\{(1-\varepsilon)F + \varepsilon\delta_x\} - T(F)}{\varepsilon} \\
&= \lim_{\varepsilon\to 0} \frac{\{(1-\varepsilon)\int udF(u) + \varepsilon x\} - \int udF(u)}{\varepsilon} \\
&= \lim_{\varepsilon\to 0} \varepsilon^{-1}\{-\varepsilon \int udF(u) + \varepsilon x\} \\
&= x - \int udF(u) = x - T(F).
\end{aligned}
$$

Hence, the gross-error sensitivity (2) $\gamma(T,F) = \infty$, the local-shift sensitivity (3) $\lambda(T,F) = 0$, and the rejection point (4) $\rho(T,F) = \infty$. Second, the maximum-bias (5) is infinite for any $\varepsilon > 0$ since

$$\sup_{x\in\mathbb{R}} \|T\{(1-\varepsilon)F + \varepsilon\delta_x\} - T(F)\| = \sup_{x\in\mathbb{R}} \| - \varepsilon T(F) + \varepsilon x\| = \infty.$$

Consequently, the breakdown point (6) of the sample mean $\bar{x} = T(F_n)$ is zero, $\varepsilon^*(T) = 0$.

Thus, none of robustness measures characterizing the change of T under contamination of data (even infinitesimally small) is finite. This behavior, typical for LS-based methods, motivated alternative estimators that have the desirable robust properties. In this section, the M-estimators, S-estimators, and τ-estimators are discussed as well as some extensions and combination of these approaches. Even though there is a much wider range of robust estimation principles, we focus on those already studied and adopted in various areas of econometrics.

2.1 M-estimators

To achieve more flexibility in accommodating requirements on robustness, Huber (1964) proposed the *M-estimator* by considering a general extremum estimator based on $\int \rho(z,\theta)dF(z)$, thus minimizing $\int \rho(z,\theta)dF_n(z)$ in finite samples. Providing that the first derivative $\psi(z,\theta) = \partial\rho(z,\theta)/\partial\theta$ exists, an M-estimator can be also defined by an implicit equation $\int \psi(z,\theta)dF_n(z) = 0$.

This extremely general definition is usually adopted to a specific estimation problem such as location, scale, or regression estimation. In a univariate location model, $F(z)$ can be parametrized as $F(z-\theta)$ and hence one limits $\rho(z,\theta)$ and $\psi(z,\theta)$ to $\rho(z-\theta)$ and $\psi(z-\theta)$. In the case of scale estimation, $F(z) = F(z/\theta)$ and consequently $\rho(z,\theta) = \rho(z/\theta)$ and $\psi(z,\theta) = \psi(z/\theta)$. In linear regression, $z = (x,y)$ and a zero-mean error term $\varepsilon = y - x^\top\theta$. Analogously to the location case, one can then consider $\rho(z,\theta) = \rho(y - x^\top\theta)$ and $\psi(z,\theta) = \psi(y - x^\top\theta)x$, or more generally, $\rho(z,\theta) = \rho(y - x^\top\theta, x)$ and $\psi(z,\theta) = \psi(y - x^\top\theta, x)$ (GM-estimators). Generally, we can express $\rho(z,\theta)$ as $\rho\{\eta(z,\theta)\}$, $\psi\{\eta(z,\theta)\}$, where $\eta(z,\theta) \sim F$.

Some well-known choices of univariate objective functions ρ and ψ are given in Table 1; functions $\rho(t)$ are usually assumed to be non-constant, non-negative, even, and continuously increasing in $|t|$. This documents flexibility of the concept of M-estimators, which include LS and quantile regression as special cases.

On the other hand, many of the ρ and ψ functions in Table 1 depend

Table 1: Examples of ρ and ψ functions used with M-estimators.

		$\rho(t)$	$\psi(t)$
Least squares		t^2	$2t$
Least absolute deviation		$\|t\|$	$\text{sign}(t)$
Quantile estimation		$\{\tau - I(x < 0)\}x$	$\tau - I(x < 0)$
Huber:	for $\|t\| \leq c$	t^2	$2t$
	for $c < \|t\|$	$c\|t\|$	$c\,\text{sign}(t)$
Hampel:	for $\|t\| \leq a$	t^2	$2t$
	for $a < \|t\| \leq b$	$a\|t\|$	$a\,\text{sign}(t)$
	for $b < \|t\| \leq c$	$\frac{ac}{c-b}t - \frac{a}{c-b}t^2\,\text{sign}(t)$	$a(c - \|t\|)/(c - b)$
	for $c < \|t\|$	$a\|t\|$	0
Biweight (Tukey)		$-(c^2 - t^2)^3 I(\|t\| \leq c)/6$	$t(c^2 - t^2)^2 I(\|t\| \leq c)$
Sine (Andrews)		$-c\cos(x/c)I(\|t\| \leq \pi c)$	$\sin(x/c)I(\|t\| \leq \pi c)$

on one or more constants $a, b, c \in \mathbb{R}$. If an estimator T is to be invariant to the scale of data, one can apply the estimator to rescaled data, that is, to minimize $\int \rho\{(z-\theta)/s\}dF_n(z)$ or to solve $\int \psi\{(z-\theta)/s\}dF_n(z) = 0$ for a scale estimate s like the median absolute deviation (MAD). Alternatively, one may also estimate parameters θ and scale s simultaneously by considering $\rho(z, \{\theta, s\}) = \rho\{(z-\theta)/s\}$ or

$$\psi(z, \{\theta, s\}) = \{\psi_l(z, \{\theta, s\}), \psi_s(z, \{\theta, s\})\}.$$

Let us now turn to the question how the choice of functions ρ and ψ determines the robust properties of M-estimators. First, the influence function of an M-estimator can generally depend on several quantities such as its asymptotic variance or the position of explanatory variables in the regression case, but the influence function is always proportional to function $\psi(z, b)$. Thus, the finite gross-error sensitivity, $\gamma(T, F) < \infty$, requires bounded $\psi(t)$ (which is not the case of LS). Similarly, the finite rejection point, $\rho(T, F) < \infty$, leads to $\psi(t)$ being zero for all sufficiently large t (the M-estimators defined by such a ψ-function are called redescending). Hampel et al. (1986) shows how, for a given bound on $\gamma(T, F)$, one can determine the most efficient choice of ψ function (e.g., the skipped median, $\psi(t) = \text{sign}(t)I(|t| < K), K > 0$, in the location case).

More formally, the optimality of M-estimators in the context of qualitative robustness can be studied by the *asymptotic relative efficiency* (ARE) of an estimator $\hat{\theta}^1$ relative to another estimator $\hat{\theta}^2$:

$$ARE(\hat{\theta}^1, \hat{\theta}^2) = \frac{\text{as. var}(\hat{\theta}^1)}{\text{as. var}(\hat{\theta}^2)}. \tag{7}$$

For example, at the normal distribution with $\hat{\theta}^1$ and $\hat{\theta}^2$ being the least absolute deviation (LAD) and LS estimators, ARE equals $2/\pi \approx 0.64$. Under the Student cdf t_5, the ARE of the two estimators climbs up to ≈ 0.96. For Huber's M-estimator, we see that its limit cases are the median for $c \to 0$ and the mean for $c \to \infty$. At the normal distribution and for $c = 1.345$, we have ARE of about 0.95. This means that this M-estimator is almost as efficient as MLE, but does not lose so drastically in performance as the standard mean under contamination because of the bounded influence function.

Whereas the influence function of M-estimators is closely related to the choice of its objective function, the global robustness of M-estimators is in a certain sense independent of this choice. Maronna et al. (1979) showed in linear regression that the breakdown point of M-estimators is bounded by $1/p$, where p is the number of estimated parameters. As a remedy, several authors proposed *one-step M-estimators* that are defined, for example, as the first step of the iterative Newton-Raphson procedure, used to minimize $\int \rho(z, \theta) dF(z)$, started from initial robust estimators $\hat{\theta}^0$ of parameters and $\hat{s}^0$ of scale (see Welsh and Ronchetti, 2002, for an overview). Possible initial estimators can be those discussed in Sections 2.2 and 2.3. For example for an M-estimator of location $\hat{\theta}$ defined by a function $\psi(x, \theta) = \psi(x - \theta)$, its one-step counterpart can be defined at sample $\{x_i\}_{i=1}^n$ by

$$\hat{\theta} = \hat{\theta}^0 + \hat{s}^0 \sum_{i=1}^n \psi\left(\frac{x_i - \hat{\theta}^0}{\hat{s}^0}\right) \Big/ \sum_{i=1}^n \psi'\left(\frac{x_i - \hat{\theta}^0}{\hat{s}^0}\right),$$

where $\hat{\theta}^0$ and $\hat{s}^0$ represent initial robust estimators of location and scale like the median and MAD, respectively. Such one-step estimators, under certain conditions on the initial estimators, preserve the breakdown point of the initial estimators, and at the same time, have the same first-order asymptotic distribution as the original M-estimator (Simpson et al., 1992, and Welsh and Ronchetti, 2002). Further development of such ideas include an adaptive choice of parameters of function ψ in the iterative step (Gervini and Yohai, 2002).

2.2 S-estimators

An alternative approach to M-estimators achieving high breakdown point (HBP) was proposed by Rousseeuw and Yohai (1984). The S-*estimators* are defined by minimization of a scale statistics $s^2(z,b) = s\{\eta(z,b)\}$ defined as the M-estimate of scale,

$$\int \rho[\eta(z,b)/s\{\eta(z,b)\}]dF_n(z) = K = \int \rho(t)dF(t),$$

at the model distribution F; the functions ρ and η are those defining M-estimators in Section 2.1. More generally, one can define S-estimators by means of any scale-equivariant statistics s^2, that is, $s\{c\eta(z,b)\} = |c|s\{\eta(z,b)\}$. Under this more general definition, S-estimators include as special cases LS and LAD estimators. Further, they encompass several well-known robust methods including least median of squares (LMS) and least trimmed squares (LTS): whereas the first defines the scale statistics $s^2\{\eta(z,b)\}$ as the median of squared residuals $\eta(z,b)$, the latter used the scale defined by the sum of h smallest residuals $\eta(z,b)$. In order to appreciate the difference to M-estimators, it is worth pausing for a moment and to present LMS, the most prominent representative of S-estimators, in the location case:

$$\arg\min_{\theta} \operatorname{med}\{(x_1-\theta)^2,\ldots,(x_n-\theta)^2\}.$$

Due to its definition, the S-estimators have the same influence function as the M-estimator constructed from the same function ρ. Contrary to M-estimators, they can achieve the highest possible breakdown point $\varepsilon^* = 0.5$. For example, this is the case of LMS and LTS. For Gaussian data, the most efficient (in the sense of ARE (7) among the S-estimators with $\varepsilon^* = 0.5$ is however the one corresponding to $K = 1.548$ and ρ being the Tukey biweight function, see Table 1. Given the HBP of S-estimators, their maximum-bias behavior is of interest too. Although it depends on the function ρ and constant K (Berrendero and Zamar, 2001), Yohai and Zamar (1993) proved that LMS minimizes maximum bias among a large class of (residual admissible) estimators, which includes most robust methods.

An important shortcoming of HBP S-estimation is however its low ARE: under Gaussian data, efficiency relative to LS varies from 0% to 27%. Thus, S-estimators are often used as initial estimators for other, more efficient methods. Nevertheless, if an S-estimator is not applied directly to sample observations, but rather to the set of all pairwise differences of sample observations, the resulting generalized S-estimator exhibits higher relative efficiency for Gaussian data, while preserving its robust properties (Croux et al., 1994; Stromberg et al., 2000).

2.3 τ-estimators

The S-estimators improve upon M-estimators in terms of their breakdown-point properties, but at the cost of low Gaussian efficiency. Although one-step M-estimators based on an initial S-estimate can remedy this deficiency to a large extent, their exact breakdown properties are not known. One of alternative approaches, proposed by Yohai and Zamar (1988), extend the principle of S-estimation in the following way. Assuming that ρ_1 and ρ_2 are non-negative, even, and continuous functions, the M-estimate $s^2(z,\theta) = s^2\{\eta(z,\theta)\}$ of scale can be defined as in the case of S-estimation,

$$\int \rho_1[\eta(z,\theta)/s\{\eta(z,\theta)\}]dF_n(z) = K = \int \rho_1(t)dF(t).$$

Next, the τ-estimate of scale is defined by

$$\tau^2(z,\theta) = s^2\{\eta(z,\theta)\} \int \rho_2[\eta(z,\theta)/s\{\eta(z,\theta)\}]dF_n(z)$$

and the corresponding *τ-estimator* of parameters θ is then defined by minimizing the τ-estimate of scale, $\tau^2(z,\theta)$.

As a generalization of S-estimation, the τ-estimators include S-estimators as a special case for $\rho_1 = \rho_2$ because then $\tau^2(z, \theta) = \theta s^2(z, \theta)$. On the other hand, if $\rho_2(t) = t^2$, $\tau^2(z, \theta) = \int \eta^2(z, \theta) dF_n(z)$ is just the standard deviation of model residuals. Compared to S-estimators, the class of τ estimators can improve in terms relative Gaussian efficiency because its breakdown depends only on function ρ_1, whereas its asymptotic variance is function of both ρ_1 and ρ_2. Thus, ρ_1 can be defined to achieve the breakdown point equal to 0.5 and ρ_2 consequently adjusted to reach a pre-specified relative efficiency for Gaussian data (e.g., 95%).

3 Methods of robust econometrics

The concepts and methods of robust estimation discussed in Sections 1 and 2 are typically proposed in the context of a simple location or linear regression models, assuming independent, continuous, and identically distributed random variables. This however rarely corresponds to assumptions typical for most econometric models. In this section, we therefore present an overview of developments and extensions of robust methods to various econometric models. As the M-estimators are closest to the commonly used LS and MLE, most of the extensions employ M-estimation. The HBP techniques are not that frequently found in the economics literature (Zaman et al., 2001; Sapra, 2003) and are mostly applied only as a diagnostic tool.

In the rest of this section, robust estimation is first discussed in the context of models with discrete explanatory variables, models with time-dependent observations, and models involving multiple equations. Later, robust alternatives to general estimation principles, such as MLE and generalized method of moments (GMM), are discussed. Before doing so, let us mention that dangers of data contamination are not only studied only from the theoretical point of view. There is a number of studies that check the presence of outliers in real data and their influence on estimation methods. For example, there is evidence of data contamination and its adverse effects on LS and MLE in the case of macro economic time series (Balke and Fomby, 1994; Atkinson, Koopman, and Shephard, 1997), in financial time series (Sakata and White, 1998; Franses, van Dijk, and Lucas, 2004), marketing data (Franses, Kloek, and Lucas, 1999), and many other areas. These adverse effects include biased estimates, masking of structural changes, and creating seemingly nonlinear structures, for instance.

3.1 Discrete variables

To achieve a HBP, many robust methods such as LMS often eliminate a large portion of observations from the calculation of their objective function. This can cause non-identification of parameters associated with categorical variables. For example, having data on income $\{y_i\}_{i=1}^n$ of men and women, where gender is indicated by $\{d_i\}_{i=1}^n \in \{0, 1\}$, one can estimate the mean income of men and women by a simple regression model $y_i = a + bd_i$. If a HBP method such as LMS or LTS is used to estimate the model and it eliminates a large portion of observations from the calculation (e.g., one half of them), the remaining data could easily contain only income of men or only income of women, and consequently, the mean income of one of the groups could not be then identified. Even though this seems unlikely in our simple example, it becomes more pronounced as the number of discrete variables grows, see Hubert and Rouseeuw (1997) for an example.

A common strategy employs a robust estimator with a HBP for a model with only continuous variables, and using this initial estimate, the model with all variables is estimated by an M-estimator. Such a combined procedure preserves the breakdown point of the HBP estimator: even though a misclassified values of categorical explanatory variables can bias the estimates, this bias will be bounded in common models as the categorical variables are bounded as well. See Hubert and Rousseeuw (1997) and Maronna and Yohai (2000), who combine an initial S-estimator with an M-estimator.

3.2 Time series

In time series, there are several issues not addressed by the standard theory of robust estimation because of time-dependency of observations. First, the asymptotic behavior of various robust methods has to be established; see Koenker and Machado (1999), Koenker and Xiao (2002) for L_1 regression, Künsch (1984) and Bai (1997) for M-estimators and Sakata and White

(2001), Zinde-Walsh (2002), and Čížek (2005) for various S-type estimators. In these cases, the results are usually established for general nonlinear models.

Second, the effects of data contamination are more complex and wide spread due to time-dependency: an error in one observation is transferred, by means of a model, to other ones close in time. The possible effects of outliers in time series are elaborated by Chen and Liu (1993) and Tsay et al. (2000), for instance. The first work also offers a sequential identification of outliers (an alternative procedure based on τ-estimators is offered by Bianco et al., 2001). Consequently, the robust properties in time series differ from those experienced in cross-sectional data. For example, the breakdown point is asymptotically zero in the case of M-estimators (Sakata and White, 1995) and can be much below 0.5 for various S-estimators (Genton and Lucas, 2003).

A further issue specific to time series is testing for stationarity of a series. Effects of outliers are in this respect similar to those of neglected structural changes. To differentiate between random outliers and real structural changes, robust tests for change-point detection were proposed by Gagliardini et al. (2005), Fiteni (2002), and Fiteni (2004); the last paper uses τ-estimation. The asymptotics of M-estimators under unit-root assumption and the corresponding tests were established, for example, by Lucas (1995), Koenker and Xiao (2004), and Haldrup et al. (2005). An early reference is

Franke et al. (1984).

3.3 Multivariate regression

An important application of robust methods in economics concerns the multivariate regression case. This is relatively straightforward with exogenous explanatory variables only, see Koenker and Portnoy (1990), Bilodeau and Duchesne (2000), and Lopuhaä (1992) for the M-, S-, and τ-estimation, respectively. Estimating general simultaneous equations models has to mimick either three-stage LS or full information MLE (Marrona and Yohai, 1997). Whereas Koenker and Portnoy (1990) follow with the weighted LAD the first approach, Krishnakumar and Ronchetti (1997) use M-estimation together with the second strategy.

3.4 General estimation principles

There are naturally many more model classes, for which one can construct robust estimation procedures. Since most econometric models can be estimated by means of MLE or GMM, it is however easier to concentrate on robust counterparts of these two estimation principles. There are other estimation concepts, such as nonparametric smoothing, that can employ robust estimation (Härdle, 1982), but they go beyong the scope of this chapter.

First, recent contributions to robust MLE can be split to two groups. One simply defines a weighted maximum likelihood, where weights are computed from an initial robust fit (Windham, 1995; Markartou et al., 1997). Alternatively, some erroneous observations can be excluded completely from the likelihood function (Clarke, 2000; Marazzi and Yohai, 2004). This approach requires existence of an initial robust estimate, and thus, it is not useful for models, for which there are no robust methods available. The second approach is motivated by the S-estimation, namely LTS, and defines the maximum trimmed likelihood as an estimator maximizing the product of the h largest likelihood contribution; that is, those corresponding only to h most likely observations (Hadi and Luceno, 1997). This estimator was studied mainly in the context of generalized linear models (Müller and Neykov, 2003), but its consistency is established in a much wider class of models (Čížek, 2004).

Second, more widely used GMM also attracted attention from its robustness point of view. A special case, instrumental variable estimation, was studied, for example, by Wagenvoor and Waldman (2002) and Kim and Muller (2006). See also Chernozhukov and Hansen (2006) for instrumental variable quantile regression. More generally, Ronchetti and Trojani (2001) proposed an M-estimation-based generalization of GMM, studied its robust properties, and design corresponding tests. This work became a starting point for others, who extended the methodology of Ronchetti and Trojani (2001) to robustify simulation-based methods of moments (Genton and Ronchetti, 2003; Ortelli and Trojani, 2005).

References

[1] Atkinson, A. C., Koopman, S. J., and Shephard, N. (1997). Detecting shocks: outliers and breaks in time series. *Journal of Econometrics* 80, 387–422.

[2] Balke, N. S., and Fomby, T. B. (1994). Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics* 9, 181–200.

[3] Berrendero, J. R, and Zamar, R. H. (2001). Maximum bias curves for robust regression with non-elliptical regressors. *The Annals of Statistics* 29(1), 224–251.

[4] Bianco, A., Ben, M. G., Martnez, E., and Yohai, V. J. (2001). Regression models with ARIMA errors. *Journal of Forecasting* 20, 565–579.

[5] Bilodeau , M., and Duchesne, P. (2000). Robust estimation of the SUR model. *Canadian Journal of Statistics* 28, 277–288.

[6] Chen, C., and Liu, L.-M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association* 88, 284–297.

[7] Chernozhukov, V., and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, in press.

[8] Čížek, P. (2004). General trimmed estimation: robust approach to nonlinear and limited dependent variable models. CentER Discussion paper 2004/130, Tilburg University. Submitted to the Econometric Theory.

[9] Čížek, P. (2005). Least trimmed squares in nonlinear regression under dependence. *Journal of Statistical Planning and Inference*, in press.

[10] Clarke, B. R. (2000). An adaptive method of estimation and outlier detection in regression applicable for small to moderate sample sizes. *Probability and Statistics* 20, 25–50.

[11] Croux, C., Rousseeuw, P. J., Hossjer, O. (1993). Generalized S-estimators. *Journal of the American Statistical Association* 89, 1271–1281.

[12] Davies, L., and Gather, U. (2005). Breakdown and groups. *Annals of Statistics* 33, 988–993.

[13] Fiteni, I. (2002). Robust estimation of structural break points. *Econometric Theory* 18, 349–386.

[14] Fiteni, I. (2004). τ-estimators of regression models with structural change of unknown location. *Journal of Econometrics* 119, 19–44.

[15] Franses, P. H., Kloek, T., and Lucas, A. (1999). Outlier robust analysis of longrun marketing effects for weekly scanning data. *Journal of Econometrics* 89, 293–315.

[16] Franses, P. H., van Dijk, D., and Lucas, A. (2004). Short patches of outliers, ARCH and volatility modelling. *Applied Financial Economics* 14, 221–231.

[17] Franke, J. Härdle, W., and Martin, R. D. (1984). *Robust and Nonlinear Time Series Analysis*, Springer, Berlin.

[18] Gagliardini, P., Trojani, F., and Urga, G. (2005). Robust GMM tests for structural breaks. *Journal of Econometrics* 129, 139–182.

[19] Genton, M. G., and Lucas, A. (2003). Comprehensive definitions of breakdown-points for independent and dependent observations. *Journal of the Royal Statistical Society, Series B* 65, 81–94.

[20] Genton, M. G., and Ronchetti, E. (2003). Robust indirect inference. *Journal of the American Statistical Association* 98, 67–76.

[21] Gervini, D., and Yohai, V. J. (2002). A class of robust and fully efficient regression estimators. *The Annals of Statistics* 30(2), 583–616.

[22] Hadi, A. S., and Luceño, A. (1997). Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics & Data Analysis* 25(3), 251–272.

[23] Haldrup, N., Montans, A., and Sanso, A. (2005). Measurement errors and outliers in seasonal unit root testing. *Journal of Econometrics* 127, 103–128.

[24] Hampel, F. R. (1971). A general qualitative definition of robustness, *Annals of Mathematical Statistics* 42, 1887–1896.

[25] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust statistics: The approach based on influence function.* Wiley, New York.

[26] Härdle, W. (1982). Robust regression function estimation. *Journal of Multivariate Analysis* 14, 169–180.

[27] He, X., and Simpson, D. G. (1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation. *The Annals of Statistics* 21(1), 314–337.

[28] Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73–101.

[29] Hubert, M., and Rousseeuw, P. J. (1997). Robust regression with both continuous and binary regressors. *Journal of Statistical Planning and Inference* 57, 153–163.

[30] Kim, T.-H., and Muller, C. (2006). Two-stage Huber estimation. *Journal of Statistical Planning and Inference*, in press.

[31] Koenker, R., and Machado, J. A. F. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association* 94, 1296–1310.

[32] Koenker, R., and Portnoy, S. (1990). M-estimation of multivariate regressions. *Journal of the American Statistical Association* 85, 1060–1068.

[33] Koenker, R., and Xiao, Z. (2002). Inference on the Quantile Regression Process. *Econometrica* 70(4), 1583–1612.

[34] Koenker, R., and Xiao, Z. (2004). Unit root quantile autoregression inference. *Journal of the American Statistical Association* 99, 775–787.

[35] Krishnakumar, J., and Ronchetti, E. (1997). Robust estimators for simultaneous equations models. *Journal of Econometrics* 78, 295–314.

[36] Kunsch, H. (1984). Infinitesimal Robustness for Autoregressive Processes. *The Annals of Statistics* 12(3), 843–863.

[37] Lopuhaä, H. (1992). Multivariate τ-estimators. *Canadian Journal of Statistics* 19, 307–321.

[38] Lucas, A. (1995). An outlier robust unit root test with an application to the extended Nelson-Plosser data. *Journal of Econometrics* 66, 153–173.

[39] Marazzi, A., and Yohai, V. J. (2004). Adaptively truncated maximum likelihood regression with asymmetric errors. *Journal of Statistical Planning and Inference* 122, 271–291.

[40] Markatou, M., Basu, A., and Lindsay, B. (1997). Weighted likelihood estimating equations: the discrete case with applications to logistic regression. *Journal of Statistical Planning and Inference* 57(2), 215–232.

[41] 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 T. Gasser and M. Rossenblatt (eds.) *Smoothing Techniques for Curve Estimation.* Springer, Berlin, 91–116.

[42] Maronna, R. A., and Yohai, V. J. (1997). Robust estimation in simultaneous equations models. *Journal of Statistical Planning and Inference* 57(2), 233–244.

[43] Maronna, R. A., and Yohai, V. J. (2000). Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference* 89, 197–214.

[44] Müller, C. H., and Neykov, N. (2003). Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *Journal of Statistical Planning and Inference* 116(2), 503–519.

[45] Ortelli, C., and Trojani, F. (2005). Robust efficient method of moments. *Journal of Econometrics* 128, 69–97.

[46] Ronchetti, E., and Trojani, F. (2001). Robust inference with GMM estimators. *Journal of Econometrics* 101, 37–69.

[47] Rousseeuw, P. J., and Leroy, A. M. (1984). Robust regression by means of S-estimators. In J. Franke, W. Härdle, and R. D. Martin (eds.) *Robust and Nonlinear Time Series Analysis*, Springer, 256–272.

[48] Sakata, S., and White, H. (1995). An alternative definition of finite-sample breakdown point with application to regression model estimators. *Journal of the American Statistical Association* 90, 1099–1106.

[49] Sakata, S., and White, H. (1998). High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. *Econometrica* 66, 529–567.

[50] Sakata, S., and White, H. (2001). S-estimation of nonlinear regression models with dependent and heterogeneous observations. *Journal of Econometrics* 103, 5–72.

[51] Sapra, S. K. (2003). High-breakdown point estimation of some regression models. *Applied Economics Letters* 10(14), 875–878.

[52] 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* 87, 439–450.

[53] Stromberg, A. J., Hossjer, O., and Hawkins, D. M. (2000). The least trimmed differences regression estimator and alternatives. *Journal of the American Statistical Association* 95, 853–864.

[54] Tsay, R. S., Pena, D., and Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika* 87(4), 789–804.

[55] Wagenvoor, R., and Waldman, R. (2002). On B-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics* 106, 297–324.

[56] Welsh, A. H., and Ronchetti, E. (2002). A journey in single steps: robust one-step M-estimation in linear regression. *Journal of Statistical Planning and Inference* 103, 287–310.

[57] Windham, M. P. (1995). Robustifying model fitting. *Journal of the Royal Statistical Society, Series B* 57(3), 599–609.

[58] Yohai, V. J., and Zamar, R. H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association* 83, 406–413.

[59] Yohai, V. J., and Zamar, R. H. (1993). A minimax-bias property of the least a-quantile estimates. *The Annals of Statistics* 21(4), 1824–1842.

[60] Zaman, A., Rousseeuw, P. J., and Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economics Letters* 71(1), Pages 1–8.

SFB 649 Discussion Paper Series 2006

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Calibration Risk for Exotic Options" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
002 "Calibration Design of Implied Volatility Surfaces" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
003 "On the Appropriateness of Inappropriate VaR Models" by Wolfgang Härdle, Zdeněk Hlávka and Gerhard Stahl, January 2006.
004 "Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification" by Harald Uhlig, January/February 2006.
005 "British Interest Rate Convergence between the US and Europe: A Recursive Cointegration Analysis" by Enzo Weber, January 2006.
006 "A Combined Approach for Segment-Specific Analysis of Market Basket Data" by Yasemin Boztuğ and Thomas Reutterer, January 2006.
007 "Robust utility maximization in a stochastic factor model" by Daniel Hernández–Hernández and Alexander Schied, January 2006.
008 "Economic Growth of Agglomerations and Geographic Concentration of Industries - Evidence for Germany" by Kurt Geppert, Martin Gornig and Axel Werwatz, January 2006.
009 "Institutions, Bargaining Power and Labor Shares" by Benjamin Bental and Dominique Demougin, January 2006.
010 "Common Functional Principal Components" by Michal Benko, Wolfgang Härdle and Alois Kneip, Jauary 2006.
011 "VAR Modeling for Dynamic Semiparametric Factors of Volatility Strings" by Ralf Brüggemann, Wolfgang Härdle, Julius Mungo and Carsten Trenkler, February 2006.
012 "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms" by Carsten Trenkler, February 2006.
013 "Penalties and Optimality in Financial Contracts: Taking Stock" by Michel A. Robe, Eva-Maria Steiger and Pierre-Armand Michel, February 2006.
014 "Core Labour Standards and FDI: Friends or Foes? The Case of Child Labour" by Sebastian Braun, February 2006.
015 "Graphical Data Representation in Bankruptcy Analysis" by Wolfgang Härdle, Rouslan Moro and Dorothea Schäfer, February 2006.
016 "Fiscal Policy Effects in the European Union" by Andreas Thams, February 2006.
017 "Estimation with the Nested Logit Model: Specifications and Software Particularities" by Nadja Silberhorn, Yasemin Boztuğ and Lutz Hildebrandt, March 2006.
018 "The Bologna Process: How student mobility affects multi-cultural skills and educational quality" by Lydia Mechtenberg and Roland Strausz, March 2006.
019 "Cheap Talk in the Classroom" by Lydia Mechtenberg, March 2006.
020 "Time Dependent Relative Risk Aversion" by Enzo Giacomini, Michael Handel and Wolfgang Härdle, March 2006.
021 "Finite Sample Properties of Impulse Response Intervals in SVECMs with Long-Run Identifying Restrictions" by Ralf Brüggemann, March 2006.
022 "Barrier Option Hedging under Constraints: A Viscosity Approach" by Imen Bentahar and Bruno Bouchard, March 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

023 "How Far Are We From The Slippery Slope? The Laffer Curve Revisited" by Mathias Trabandt and Harald Uhlig, April 2006.
024 "e-Learning Statistics – A Selective Review" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, April 2006.
025 "Macroeconomic Regime Switches and Speculative Attacks" by Bartosz Maćkowiak, April 2006.
026 "External Shocks, U.S. Monetary Policy and Macroeconomic Fluctuations in Emerging Markets" by Bartosz Maćkowiak, April 2006.
027 "Institutional Competition, Political Process and Holdup" by Bruno Deffains and Dominique Demougin, April 2006.
028 "Technological Choice under Organizational Diseconomies of Scale" by Dominique Demougin and Anja Schöttner, April 2006.
029 "Tail Conditional Expectation for vector-valued Risks" by Imen Bentahar, April 2006.
030 "Approximate Solutions to Dynamic Models – Linear Methods" by Harald Uhlig, April 2006.
031 "Exploratory Graphics of a Financial Dataset" by Antony Unwin, Martin Theus and Wolfgang Härdle, April 2006.
032 "When did the 2001 recession *really* start?" by Jörg Polzehl, Vladimir Spokoiny and Cătălin Stărică, April 2006.
033 "Varying coefficient GARCH versus local constant volatility modeling. Comparison of the predictive power" by Jörg Polzehl and Vladimir Spokoiny, April 2006.
034 "Spectral calibration of exponential Lévy Models [1]" by Denis Belomestny and Markus Reiß, April 2006.
035 "Spectral calibration of exponential Lévy Models [2]" by Denis Belomestny and Markus Reiß, April 2006.
036 "Spatial aggregation of local likelihood estimates with applications to classification" by Denis Belomestny and Vladimir Spokoiny, April 2006.
037 "A jump-diffusion Libor model and its robust calibration" by Denis Belomestny and John Schoenmakers, April 2006.
038 "Adaptive Simulation Algorithms for Pricing American and Bermudan Options by Local Analysis of Financial Market" by Denis Belomestny and Grigori N. Milstein, April 2006.
039 "Macroeconomic Integration in Asia Pacific: Common Stochastic Trends and Business Cycle Coherence" by Enzo Weber, May 2006.
040 "In Search of Non-Gaussian Components of a High-Dimensional Distribution" by Gilles Blanchard, Motoaki Kawanabe, Masashi Sugiyama, Vladimir Spokoiny and Klaus-Robert Müller, May 2006.
041 "Forward and reverse representations for Markov chains" by Grigori N. Milstein, John G. M. Schoenmakers and Vladimir Spokoiny, May 2006.
042 "Discussion of 'The Source of Historical Economic Fluctuations: An Analysis using Long-Run Restrictions' by Neville Francis and Valerie A. Ramey" by Harald Uhlig, May 2006.
043 "An Iteration Procedure for Solving Integral Equations Related to Optimal Stopping Problems" by Denis Belomestny and Pavel V. Gapeev, May 2006.
044 "East Germany's Wage Gap: A non-parametric decomposition based on establishment characteristics" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.
045 "Firm Specific Wage Spread in Germany - Decomposition of regional differences in inter firm wage dispersion" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

046 "Produktdiversifizierung: Haben die ostdeutschen Unternehmen den Anschluss an den Westen geschafft? – Eine vergleichende Analyse mit Mikrodaten der amtlichen Statistik" by Bernd Görzig, Martin Gornig and Axel Werwatz, May 2006.
047 "The Division of Ownership in New Ventures" by Dominique Demougin and Oliver Fabel, June 2006.
048 "The Anglo-German Industrial Productivity Paradox, 1895-1938: A Restatement and a Possible Resolution" by Albrecht Ritschl, May 2006.
049 "The Influence of Information Costs on the Integration of Financial Markets: Northern Europe, 1350-1560" by Oliver Volckart, May 2006.
050 "Robust Econometrics" by Pavel Čížek and Wolfgang Härdle, June 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".