FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

FACHBEREICH WIRTSCHAFTS-
WISSENSCHAFTEN

Lehrstuhl für Statistik und Ökonometrie

Diskussionspapier

88 / 2012

Robustness Properties of Quasi-Linear Means with Application to the Laspeyres and Paasche Indices

Ingo Klein

Vlad Ardelean

Lange Gasse 20 · D-90403 Nürnberg

Robustness Properties of Quasi-Linear Means with Application to the Laspeyres and Paasche Indices

Ingo Klein[1], Vlad Ardelean

Department of Statistics and Econometrics
University of Erlangen-Nürnberg, Germany

Abstract

Li, Fang & Tian (1994) assert that special quasi-linear means should be preferred to the simple arithmetic mean for robustness properties. The strategy that is used to show robustness is completely detached from the concepts well-known from the theory of robust statistics. Robustness of estimators can be verified with tools from robust statistics, e.g. the influence function or the breakdown point. On the other hand it seems that robust statistics is not interested in quasi-linear means. Therefore, we compute influence functions and breakdown points for quasi-linear means and show that these means are not robust in the sense of robust statistics if the generator is unbounded. As special cases we consider the Laspeyres, the Paasche and the Fisher indices.

Keywords: Quasi-linear mean, robustness, influence function, breakdown point, Laspeyres index, Paasche index, Fisher index

[1]Correspondence Author: Ingo Klein, Department of Statistics and Econometrics, University of Erlangen- Nürnberg, D-90403 Nuremberg, Lange Gasse 20, E-Mail: ingo.klein@wiso.uni-erlangen.de

1 Introduction

A lot of attention in robust statistics in devoted to the modification of the arithmetic mean. Important price indices, i.e. Laspeyres, Paasche and Fisher are quasi-linear means. Only the article published by Li, Fang & Tian (1994) discusses robustness properties of the QLM, but their notion of robustness is detached form well-known concepts used in robust statistics, such as the influence function or the breakdown point.

For particular measures, such as the kurtosis, the influence function is given in Ruppert (1987) or Groeneveld (1991). The influence function of three particular broad classes of measures is discussed by Klein (1998). Recently, Groeneveld (2011) investigates the influence function of the coefficient of variation. Essama-Nssah und Lambert (2011) discuss the influence function of particular measures that are important in the field of income distribution. All these works exclude the QLM.

A quasi-linear mean is defined as follows (see for example Jarczyk (2007, p. 3)):

Definition 1 *Let $X_1, \ldots, X_n$ be random variables, $g_1, \ldots, g_n \geq 0$ and $u : [a, b] \to \mathbb{R}$ be a strictly monotone and continuous function with inverse u^{-1}. Then M_u*

$$M_u = u^{-1}\left(\sum_{i=1}^{n} u(X_i)\frac{g_i}{\sum_{j=1}^{n} g_j}\right), \quad g_i \geq 0$$

is called a quasi-linear mean (QLM). Furthermore, u is the generator of the quasi-linear mean.

The special case $g_i = 1$ for $i = 1, 2, \ldots, n$ is called quasi-arithmetic mean (QAM) (see f.e. Aczél (1966, p. 276), Bullen et al. (1988, p. 215)).

In a first, step Li, Fang & Tian (1994) discuss the estimation of a location parameter from observations that are generated form the following model:

$$x_i = \mu + \nu_i, \quad i = 1, 2, \ldots, n$$

where $\nu_1, \ldots, \nu_n$ are stochastically independent, identically distributed random variables with mean 0.

Let $\rho(x,\mu) \geq 0$. In order to obtain the estimated location parameter $\hat{\mu}$ we have to minimize

$$\sum_{i=1}^{n} \rho(x_i, \mu),$$

with regard to μ.

The estimate $\hat{\mu}$ corresponds to a QLM when the function ρ is set to

$$\rho(x_i, \mu) = (u(x_i) - u(\mu))^2 \frac{g_i}{\sum_{j=1}^{n} g_j}, \quad i = 1, 2, \ldots, n.$$

Li, Fang & Tian (1994) present known results regarding QLM and their corresponding objective function. These results are summarized in Table 1.

QLM		objective function ρ
Arithmetic mean	$1/n \sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n}(x_i - \mu)^2$
Median	$\text{med}(x_1, \ldots, x_n)$	$\sum_{i=1}^{n} \lvert x_i - \mu \rvert$
Harmonic mean	$n / \sum_{i=1}^{n} 1/x_i$	$\sum_{i=1}^{n} x_i(\mu/x_i - 1)^2$
Geometric mean	$\left(\prod_{i=1}^{n} x_i\right)^{1/n}$	$\sum_{i=1}^{n}(\ln(x_i/\mu))^2$
Gen. harm. mean (type 1)	$\sum_{i=1}^{n} x_i^{p-1} / \sum_{i=1}^{n} x_i^{p-2}$	$\sum_{i=1}^{n} x_i^{p}(\mu/x_i - 1)^2$
Gen. harm.mean (type 2)	$\left(n / \sum_{i=1}^{n} x_i^{q}\right)^{-1/q}$	$\sum_{i=1}^{n}(x_i^{q} - \mu^{q})^2$
Gen. geom. mean	$\left(\prod_{i=1}^{n} x_i^{x_i^{2r}}\right)^{1/\sum_{i=1}^{n} x_i^{2r}}$	$\sum_{i=1}^{n}\left(x_i^{r} \ln(x_i/\mu)\right)^2$

Table 1: Quasi-linear mean (QLM) and corresponding objective function

The generalized harmonic mean of type 2 is the unweighted power mean

$$\left(\frac{1}{n} \sum_{i=1}^{n} x_i^{q}\right)^{1/q}.$$

Gini (1938) discusses a weighted version of the power mean by setting the weights:

$$g_i = x_i^{r}, \quad i = 1, 2, \ldots, n$$

Li, Fang & Tian (1994) argue that the generalized means they propose are more robust than the arithmetic mean. In their notion, a mean is more robust if a test for location based on this particular mean holds its significance level if one observations is an outlier. This statement is based on a simulation study.

The robustness considerations of Li, Fang & Tian (1994) are surprising since robust statistics also uses objective functions of the form $\sum_{i=1}^{n} \rho(x_i, \mu)$ (see Huber (1981), P. 43). These type of estimators are called M-estimators and the specification of ρ determine the robustness properties of the estimator.

In this article we want to validate the informal considerations of Li, Fang & Tian (1994) with tools from robust statistics. In section 2, QLM are introduced as functionals and we calculate the influence function. The breakdown point is discussed in the third section. Laspeyres and Paasche indices are weighted arithmetic resp. harmonic means of relative prices. We present the influence function and the breakdown points for these indices and the Fisher index, too.

2 Influence function for quasi-linear means

The influence function has two important uses. First, it measures the relative contribution of an observation to the value of an estimate. Second the influence function is related to the variance of the estimator. This relationship is shown i.e. in Huber (1981, p.14). If the influence function is bounded, the impact an outlier has on the estimate and the asymptotic variance are bounded as well.

2.1 General weight functions

Since the influence function is the Gâteaux derivative of the functional, we will first derive the representation of the QML as a functional.

Let (X, Y) be a pair of random variables with distribution function F. The following results is also valid, if Y is a random vector. Furthermore, let u be the generator of the QML and let h be a measurable non-negative function.

The most general representation of a QML as a statistical functional is given by:

$$\begin{aligned} T(F) &= u^{-1}\left(E(u(X)g(X,Y))/E(g(X,Y))\right) \\ &= u^{-1}\left(\int u(x)\frac{g(x,y)}{E(g(X,Y))}dF(x,y)\right), \end{aligned}$$

given that $E(u(X))$ and $E(g(X,Y))$ exist. For the remainder of this paper we assume that both expectations exist. This representation includes all QLM that are summarized in Table 1. This representation is equivalent to (1) with the weight function g depending on X as well as other random variables.

The influence function of a statistical functional $T(F)$ is defined as the Gâteaux derivative of the functional T in direction of the Dirac measure $\delta_{x,y}$, $(x,y) \in \mathcal{X}$, where $\mathcal{X}$ denotes the carrier of the distribution F, see Hampel (1968, 1974).

Definition 2 *The influence function $T(F)$ is given by*

$$IF(x,y;T,F) = \lim_{\varepsilon \to 0_+} \frac{T(F_\varepsilon(\delta_{x,y})) - T(F)}{\varepsilon},$$

where $F_\varepsilon(\delta_{x,y}) = (1-\varepsilon)F + \varepsilon\delta_{x,y}$ and $x,y \in \mathcal{X}$.

The following theorem states the influence function for a quasi-linear mean.

Theorem 1 *Let T be the statistical functional of a QLM. Then the influence function of T for the special distribution F with $u'(T(F)) \neq 0$ is given by:*

$$IF(x,y;T,F) = \frac{u(x)g(x,y)E_F(g(X,Y)) - g(x,y)E_F(u(X)g(X,Y))}{E_F(g(X,Y)^2)} \frac{1}{u'(T(F))}.$$

Proof: The functional T evaluated at F_ε is:

$$T(F_\varepsilon) = u^{-1}\left((1-\varepsilon)\frac{E_F(u(X)g(X,Y))}{E_F(g(X,Y))} + \varepsilon\frac{u(x)g(x,y)}{g(x,y)}\right)$$

Differentiating with respect to ε and taking the limit $\varepsilon \to 0$ from above leads to the postulated result after some algebra. □

Both the generator u and the weight function g determine the robustness properties of the quasi-linear mean. If both functions are bounded the same holds for the influence function. The inverse function u^{-1} does not play any role w.r.t to robustness properties.

2.1.1 Special case: constant weight

If the weights are constant with $P(g(X,Y) = 1) = 1$, the influence function simplifies to

$$IF(x;T,F) = \frac{u(x) - E(u(X))}{u'(T(F))}.$$

The boundedness of the influence function is determined exclusively by the function u.

Example 1

1. *Let* $u(x) = x^q$ *for* $x > 0$ *and* $q \neq 0$.

   (a) *For* $q > 0$ *obviously* $u(x) \to \infty$ *for* $x \to \infty$ *and* $u(x) \to 0$ *for* $x \to 0$, *so the influence function is only bounded from below.*

   (b) *For* $q < 0$ *similar results hold,* $u(x)$ *is only bounded from above.*

2. *The function* $u(x) = \ln x$ *is unbounded for* $x \in \mathbb{R}^+$.

3. *The choice of* $u(x) = \arctan(x)$ *for* $x \in \mathbb{R}$ *results in a bounded influence function. The quasi-arithmetic mean that corresponds to this generator is used by Premaratne & Bera (2005) to measure skewness.*

2.1.2 Special case: weight function $g(x)$

In the special case of g depending on x only, the influence function simplifies to

$$IF(x; T, F) = \frac{u(x)g(x)E_F(g(X)) - g(x)E(u(X)g(X))}{E_F(g(X)^2)} \frac{1}{u'(T(F))}.$$

The product $u(x)g(x)$ determines if the influence function is bounded or not.

Li, Fang & Tian (1994) discuss the weight function

$$g(x) = x^r, \quad x > 0, \ r \in \mathbb{R}.$$

We combine the two unbound generators from Example 1 with the proposed weight function.

1. Let $u(x)g(x) = x^q x^r \to 0$ for $x \to \infty$, if $r + q < 0$. (I.e. $q < -r$) .

2. Let $u(x)g(x) = (\ln x)x^r \to 0$ for $x \to \infty$, if $r < 0$.

This indicates that with appropriate weight functions the influence function can at least be half-bounded.

3 Breakdown point for the quasi-linear mean

The concept of a breakdown point was introduced by Hampel in his Ph.D. dissertation 1968 (see also Hampel (1971)) and was further developed by Huber (1981) and Donoho & Huber (1983). Roughly speaking, the breakdown point is the maximal proportion of 'bad' observations an estimator can tolerate before the estimate takes an arbitrary large value. Recently, Davies & Gather (2005) studied formal aspects of the breakdown point.

Starting point for the finite-sample breakdown point is the realization of a random sample

$$\boldsymbol{x}^0 = (x_1, \ldots, x_n)$$

and the functional T evaluated at the empirical function associated with $\boldsymbol{x}^0$. From the original random sample $\boldsymbol{x}^0$ m of the components are replaced with arbitrary values, infinity is also allowed.

Denote the modified random sample with $\boldsymbol{x}^{(m)}$ and by $T_n(\boldsymbol{x}^m)$ the value of the estimator for this sample.

Definition 3 *The finite sample breakdown point of the estimator* T_n *for the sample* $\boldsymbol{x}^0$ *is given by*

$$\varepsilon_n^*(T_n, \boldsymbol{x}^{(0)}) = \frac{m^*(\boldsymbol{x}^{(0)})}{n},$$

where $m^*(\boldsymbol{x}^{(0)})$ *denotes the smallest non-negative integer for which the following holds true*

$$\sup_{\boldsymbol{x}^{(m)}} ||T_n(\boldsymbol{x}^{(m)})|| = \infty.$$

The fraction m^*/n denotes the smallest percentage of observations that can be replaced with arbitrary values before the estimator breaks down. If for an estimator the number m^* is independent of $\boldsymbol{x}^0$ the breakdown point is defined as follows

$$\varepsilon^* = \lim_{n\to\infty} \varepsilon_n^*.$$

This has the advantage that the breakdown point does not depend on the sample size n.

The breakdown point is not altered if, instead of replacing m values, the sample is augmented with m arbitrary values.

Example 2 (Mean) *Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, the breakdown point of the mean is*

$$\varepsilon_n^*(\bar{X}_n, \boldsymbol{x}^{(0)}) = \frac{1}{n}.$$

The breakdown point of the mean is independent of the start sample $\boldsymbol{x}^{(0)}$

$$\lim_{n\to\infty} \varepsilon_n^*(\bar{X}_n, \boldsymbol{x}^{(0)}) = 0$$

This definition of the breakdown point is only valid if the range of the estimator is $\mathbb{R}$. If the range of the estimator is a compact subset of $\mathbb{R}$ say $[a, b]$ we discuss the behaviour of the estimator when one or more observations tend to a (breakdown in a) or to b (breakdown in b)

The breakdown point of the quasi-linear mean depends on the function u and its range resp. its domain. As the breakdown point is influenced by the choice of the weight function we will discuss only the breakdown points of the quasi-arithmetic mean.

$$M_u = u^{-1}\left(\sum_{i=1}^n u(X_i)\right).$$

We discuss the breakdown point for different choices of u.

1. Let $u : \mathbb{R} \to \mathbb{R}$ be an absolutely monotonic increasing and surjective function.

   In this case M_u has the same breakdown point as the arithmetic mean.
   For $x_j \to \pm\infty$ the same holds for $u(x_j) \to \pm\infty$ for any $j \in \{1, 2, \ldots, n\}$.
   This implies $\sum_{i=1}^n u(x_i) \to \pm\infty$ and $M_u \to \pm\infty$ since $u^{-1} : \mathbb{R} \to \mathbb{R}$ absolutely monotonic increasing and surjective. An example for an absolutely monotonic increasing and surjective generator is

   $$u(x) = \mathrm{sign(x)|x|^r}, \quad \mathrm{r} > 1.$$

2. Let $u : \mathbb{R}^+ \to \mathbb{R}$ be absolutely monotonic increasing and surjective (i.e. $u(0) = -\infty$ and $u(\infty) = \infty$, $u^{-1}(-\infty) = 0$, $u^{-1}(\infty) = \infty$). Due to similar consideration as in case 1 M_u breaks down if at least one observation $x_j \to +\infty$. The breakdown point is also $1/n$.

   Furthermore, since the range of u is bounded we can check if M_u breaks down at 0. If $x_j \to 0$, then $u(x_j) \to -\infty$, so $\sum_{i=1}^n u(x_i) \to -\infty$ and $M_u \to 0$.

   An example for this function is $u(x) = \ln x$. This generator corresponds to the geometric mean.

3. Let $u : \mathbb{R}^+ \to \mathbb{R}^+$ absolutely monotonic increasing and surjective (i.e. $u(0) = 0$ and $u(\infty) = \infty$, $u^{-1}(0) = 0$, $u^{-1}(\infty) = \infty$).

   As before, the breakdown point is $1/n$, since for $x_j \to \infty$, $u(x_j) \to \infty$.

   The breakdown at 0 occurs only if all n observations are 0, the breakdown point is 1.

   Setting $u(x) = x^r$, $r \in \mathbb{Q}$ gives an example.

4. Let $u : \mathbb{R}^+ \setminus \{0\} \to \mathbb{R}^+ \setminus \{0\}$ absolutely monotonic decreasing and surjective.

   The estimator M_u breaks down at 0. If $x_j \to 0$, then $u(x_j) \to \infty$ and $\sum_{i=1}^n x_i \to \infty$, so $M_u \to 0$. The breakdown point is $1/n$.

   The generator of the harmonic mean ($u(x) = 1/x$ is monotone decreasing and surjective, so it can only break down at 0.

5. Let $u : \mathbb{R} \to (a, b)$ absolutely monotonic increasing and surjective with $-\infty < a < b < \infty$, $\lim_{x \to -\infty} u(x) = a$ and $\lim_{x \to b} u(x) = b$. Then $a \leq M_u \leq b$. $M_u = a$ resp. $M_u = b$, if $x_i \to -\infty$ for all $i = 1, 2, \ldots, n$ resp. $x_i \to \infty$ for all $i = 1, 2, \ldots, n$. The breakdown point is 1.

   Examples for this type of generator u are absolutely monotonic increasing distribution functions $u(x) = F(x)$ F. In this case $a = 0$ and $b = 1$.

4 Robustness of special measures

4.1 Price indices

Price indices measuring the cost of living are based on the price of of different goods and services needed for everyday live. The prices are collected at a regular base (say monthly). Commonly used price indices are the weighted mean of the change in price relatively to a base period. We assume that the price and the quantity of the goods and services are fixed and contain no outlying observations. Outliers can only occur in the current period. For the german consumer price index it is rather realistic to assume that outliers in the current prices can occur. Every month about 600 price collectors record 300000 single prices for about 750 goods and services.

For different aspects of the theory of index numbers we refer the reader to Diewert und Nakamura (1993) and Selvanathan & Rao (1994).

4.1.1 Laspeyres index

The Laspeyres price index is the weighted arithmetic mean of the relative prices weighted with the quantities in the base period. These weights are known and non stochastic. The price for a good or service for the current period is the arithmetic mean of the repeatedly collected prices.

The Laspeyres price index has the following representation as a functional

$$T_L(F) = \frac{\sum_{i=1}^{k} E[p_{it}]q_{i0}}{\sum_{i=1}^{k} p_{i0}q_{i0}},$$

where F is the distribution of all the prices in the current period $(p_{1t}, \ldots, p_{kt})$. The expected value replaces the arithmetic mean of the collected prices for goods or services.

Theorem 2 *Let T_L be the functional of the Laspeyres price index with deterministic prices p_{i0} and quantities q_{i0} for $i = 1, 2, \ldots, k$. Let F be the k-dimensional distribution of the price vector $(p_{1t}, \ldots, p_{kt})$. Then the influence function of T_L at $(p_{1t}^*, \ldots, p_{kt}^*)$ is given by:*

$$IF(p_{1t}^*, \ldots, p_{kt}^*, T_L, F) = \frac{\sum_{i=1}^{k}(p_{it}^* - E[p_{it}])q_{i0}}{\sum_{i=1}^{k} p_{i0}q_{i0}}.$$

Proof: Let F_ε be the distribution that is contaminated at $(p_{1t}^*, \ldots, p_{kt}^*)$, then

$$T_L(F_\varepsilon) = \frac{\sum_{i=1}^{k} \left(E[p_{it}] + \varepsilon(p_{it}^* - E[p_{it}])\right) q_{i0}}{\sum_{i=1}^{k} p_{i0}q_{i0}}.$$

Differentiating with respect to ε results in

$$\frac{dT_L(F_\varepsilon)}{d\varepsilon} = \frac{\sum_{i=1}^{k}(p_{it}^* - E[p_{it}])q_{i0}}{\sum_{i=1}^{k} p_{i0}q_{i0}}.$$

This derivative is independent of ε and is therefore the influence function. □

The influence function of the Laspeyres price index is unbounded. This implies that the Laspeyres price index is sensitive towards outliers in the gathered prices for the current period. The breakdown point is $1/k$ if k prices are collected.

4.1.2 Paasche index

In contrast to the Laspeyres price index, the weight of the Paasche price index is the quantity of the goods and services in the current period. This implies that the collected prices as well as the collected quantities can possibly be contaminated with outliers.

The functional corresponding to the Paasche price index, if only the prices in the base period are deterministic, is given by

$$T_P(F) = \frac{\sum_{i=1}^{k} E[p_{it}q_{it}]}{\sum_{i=1}^{k} p_{i0}E[q_{it}]},$$

where F is the distribution of all prices and quantities$(p_{1t}, \dots, p_{kt}, q_{1t}, \dots, q_{kt})$.

Theorem 3 *Let T_P be the functional of the Paasche price index with deterministic prices p_{i0} $i = 1, 2, \dots, k$. Let F be the k dimensional distribution of the combined vector of prices and quantities $(p_{1t}, \dots, p_{kt}, q_{1t}, \dots, q_{kt})$. Then the influence function of T_P at $(p^*_{1t}, \dots, p^*_{kt}, q^*_{1t}, \dots, q^*_{kt})$ is:*

$$IF(p^*_{1t}, \dots, p^*_{kt}, q^*_{1t}, \dots, q^*_{kt}; T_P, F) = \frac{\sum_{i=1}^{k}(p^*_{it} - T(F)p_{i0})q^*_{it}}{\sum_{i=1}^{k} p_{i0}E[q_{it}]}.$$

Proof: Let again be F_ε the distribution that is contaminated at $(p^*_{1t}, \dots, p^*_{kt}, q^*_{1t}, \dots, q^*_{kt})$, then

$$T_P(F_\varepsilon) = \frac{\sum_{i=1}^{k} (E[p_{it}q_{it}] + \varepsilon(p^*_{it}q^*_{it} - E[p_{it}q_{it}]))}{\sum_{i=1}^{k} p_{i0} (E[q_{it}] + \varepsilon(q^*_{it} - E[q_{it}]))}.$$

Differentiating with respect to ε results in

$$\frac{dT_P(F_\varepsilon)}{d\varepsilon} = \frac{N}{D}$$

with

$$N = \sum_{i=1}^{k} (p^*_{it}q^*_{it} - E[p_i t q_{it}]) \sum_{i=1}^{k} (E[q_{it}] + \varepsilon(p^*_{it} - E[p_{it}])) p_{i0}$$
$$- \sum_{i=1}^{k} (E[p_{it}q_{it}] + \varepsilon(p^*_{it}q^*_{it} - E[p_{it}q_{it}])) \sum_{i=1}^{k}(q^*_{it} - E[q_{it}])p_{i0}$$

and

$$D = \left(\sum_{i=1}^{k} (E[p_{it}] + \varepsilon(q^*_{it} - E[q_{it}])) p_{i0} \right)^2.$$

For $\varepsilon \to 0^+$ the numerator N converges to

$$\sum_{i=1}^{k} p_{it}^* q_{it}^* \sum_{i=1}^{k} E[q_{it}] p_{i0} - \sum_{i=1}^{k} E[p_{it} q_{it}] \sum_{i=1}^{k} q_{it}^* p_{i0}$$

and the denominator converges to

$$\left(\sum_{i=1}^{k} p_{i0} E[q_{it}] \right)^2 ,$$

the influence function is

$$IF(p_{1t}^*, \ldots, p_{kt}^*, q_{1t}^*, \ldots, q_{kt}^*; T_P, F) = \frac{\sum_{i=1}^{k} (p_{it}^* - T(F) p_{i0}) q_{it}^*}{\sum_{i=1}^{k} p_{i0} E[q_{it}]}. \quad \square$$

Similar to the influence function of the Laspeyres price index, the influence function of the Paasche price index is not bounded and the breakdown point is $1/k$, where k is the number of considered goods and services.

4.1.3 Fisher index

Functions of different indices inherit the robust properties of the individual indices. The Fisher price index is the geometric mean of the Laspeyres price index and the Paasche price index. The influence function for the Fisher price index T_F is given by

$$\begin{aligned} IF(p_{1t}^*, \ldots, p_{kt}^*, q_{1t}^*, \ldots, q_{kt}^*; T_F, F) = &\frac{1}{2}((T_L(F) T_P(F))^{-1/2} (IF(q_{1t}^*, \ldots, q_{kt}^*; T_L, F) T_P(F) \\ &+ IF(p_{1t}^*, \ldots, p_{kt}^*, q_{1t}^*, \ldots, q_{kt}^*; T_P, F) T_L(F))). \end{aligned}$$

The influence function is unbounded and the functional has a breakdown point of $1/k$.

5 Summary

Li, Fang & Tian (1994) assert that generalized means should be preferred over the arithmetic mean due to its robust properties. The way they determine the robustness of different means are does not comply with well known and established tools from robust statistics. In this article we have derived the influence function and the breakdown point for quasi-linear means in order to investigate their robustness. It has been shown that they are not robust in the sense of an unbound influence function and a sufficiently large breakdown point. Price indices, i.e. Laspeyres, Paasche and Fisher, can be written as quasi-linear mean and are therefore not robust.

References

1. Aczél, J. (1966). *Lectures on functional equations and their applications.* Academic Press, New York.

2. Bullen, P.S., Mitrinović, D.S., Vasić, P.M. (1988). *Means and their inequalities.* D. Reidel Publ. Co., Dordrecht.

3. Davies, P. L. & Gather, U. (2005). Breakdown and groups. *Annals of Statistics.* 977-1035.

4. Diewert, W.E. & Nakamura A.O. (1993) *Essays in index number theory.* Vol I. Elsevier Science Publishers B.V, North-Holland.

5. Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. In: *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

6. Essama-Nssah, B. & Lambert, P.J. (2011). Influence functions for distributional statistics. ECINEQ Working papers No. 236.

7. Gini, C. (1938). Di una formula comprensiva delle medie. *Metron* **13**, 3-22.

8. Groeneveld, R.A. (1991). An influence function approach to describing the skewness of a distribution. *American Statistician* **45**, 97-102.

9. Groeneveld, R.A. (2011). Influence functions for the coefficient of variation, its inverse, and CV comparisons. *Communications in Statistics - Theory and Methods* **40**, 4139-4150

10. Hampel, F.R. (1968). *Contributions to the theory of robust estimation.* Ph.D.–Thesis, Berkeley.

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

12. Hampel, F.R. (1974). The influence curve and its role in robust statistics. *Journal of the American Statistical Association* **69**, 383-393.

13. Huber, P.J. (1981). *Robust statistics.* Wiley, New York

14. Jarczyk, J. (2007), When lagrangian and quasi-arithmetic mean coincide. *Journal of Inequalities in Pure and Applied Mathematics* **8**, 1-11.

15. Klein, I. (1998). Einflussfunktionen höherer Verteilungsmaßzahlen. Diskussionspapier der Lehrstühle für Statistik der Universität Erlangen-Nürnberg. Nr. 21.

16. Li, X., Fang, W. & Tian, Q. (1994). Error criteria analysis and robust data fusion. *1994 IEEE International Conference on Acoustics, Speech and Signal Processing*, 37-40.

17. Premaratne, G. & Bera, A.K. (2005). A Test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics* **3**, 169-187.

18. Ruppert, D. (1987). What is Kurtosis? An influence function approach. *American Statistician* **41**, 1-5.

19. Selvanathan, E.A. & D.S. Prasada Rao (1994). *Index numbers: a stochastic approach.* Macmillan, London.