DIU Journal of Business and Entrepreneurship Volume 02, No. 01, 2007
https://doi.org/10.36481/diujbe.v02i1.11v00t11

Application of Robust Regression and Bootstrapping in Purchasing Power Parity Analysis

A. A. M. Nurunnabi*
Mohammed Nasser **

Abstract: This article is an attempt to show how robust regression, a computer based statistical technique introduced by P.J.Huber in 1973 and later developed by Rousseeuw (1984), Rousseeuw and Yohai (1984), and many others, can helps us in cases where OLS totally fails due to outliers, leverage points and non-normality of error distribution. To infer from the estimators obtained from robust regression we generally need, especially for small samples, bootstrapping (resampling) technique that is also a computer intensive statistical technique introduced by Efron (1979), and later developed in many directions. This talk illustrates the whole thing by an example using data extracted from the Big Mac. Index with a purchasing power parity analysis.

1. Introduction

The purpose of regression analysis is to fit a model to quantify relations among observed variables .The classical linear model assumes a relation of the type

$$y_i = x_i^T \beta + \varepsilon_i \qquad (1.1)$$

where, $i = 1,2,.....,n$, y_i's are the values of a response variable Y, x_i are values of explanatory variables, β is an unknown vector of parameters of length p, ε_i is a sequence of random errors. To fit a model we have to estimate and test the value of β.

The most well known method is the Least squares (LS) method. The LS stands on the assumptions that (i) x_i is a vector of deterministic variable; (ii) ε_i are normally distributed; (iii) ε_i's are identically and independently distributed random variables with mean 0 and variance σ^2. LS estimation (LSE) of β is the value of β that minimizes the residual sum of squares, r_i^2,

$$\sum r^2 = \sum (y_i - x_i^T \beta)^2 \qquad (1.2)$$

* Assistant Professor, Department of Business Administration, Uttara University, Dhaka.
** Professor, Department of Statistics, University of Rajshahi, Rajshahi.

From the time of its invention (in the beginning of the nineteenth century) it was a cornerstone of statistics. In spite of its mathematical beauty and computational simplicity LS method is now being criticized in case of long tailed distribution, for the inefficiency because of the heavy sensitivity to outliers. Outliers inflate the standard errors, error variance, confidence interval becomes stretched, and the estimators cannot become asymptotically consistent. To solve these problems, many research have proposed on the robust estimators, which are insensitive to outliers. LMS, LTS, robust MM regression are most commonly used robust regression techniques. Construction of a robust regression model is an iterative process. Necessary calculation with fruitful computational speed and sophisticated graphical representation are impossible without the help of modern computer technology. But to infer from the estimators obtained from robust regression we generally need, especially for small samples, bootstrapping technique that is also a computer intensive resampling technique introduced by Efron (1979) and later developed in many directions.

This article is organized as follows: section 2 holds the study methodology and data source, section 3 contains different aspects of literature review, notion of outliers, leverage points, finite breakdown point and influence observations, why and how LSE is hopeless for estimating parameters and how we can overcome the problems of LS are discussed in section 3.1. In section 3.2 we have a brief discussion on how diagnostics and robust regression reach the same goal in different ways. Section 3.3 upholds different types of robust regression. In section 3.4 we describe some limitations of robust regression where as in section 3.5 we briefly narrate the gist of bootstrap technique and pinpoint how it helps us in dealings with those limitations. Section 4 illustrates the whole thing with purchasing power parity analysis (PPP). Conclusion is given in section 5,

2. Study Methodologies and Data Source

We consider the data set given by The Economist (Gujarati, 1995), published the Big. Mac. Index of 31 countries as a crude and hilarious, measure of whether international currencies at their current exchange rate, judged by the theory of purchasing power parity (PPP). We study and test how outliers or unexpected observations occur in a real data set and OLS method fails to estimate and predicts the regression model in presence of outliers and with the violations of normality in error distribution. We also use bootstrap as a simulation technique to infer from the robust regression estimates. In these purposes we use robust regression as well as regression diagnostics. We use Minitab version 12 (1998), SPSS version 11 and12, and S-Plus 2000, and PROGRESS (Program for Robust Regression) for our necessary computations.

3. Literature Review

3.1 Problems of OLS method and Classifications of Outliers in Regression

It is well known that the least squares estimation (LSE) of the regression parameter β has the smallest variance among all the unbiased estimates when the errors are normally distributed. But the LSE is extremely sensitive to atypical data (an observation that is apart from the bulk of the data is treated as outlier (Staudte and, Sheather, 1990). Outliers are the observations that do not follow the pattern of the majority of the data, no matter how big the sample size is. Outlier can deviate in three ways: i) the change in the direction of response variable generally measured by absolute magnitude of standardized/studentized residual of the observation, ii) the deviation in the space of explanatory variable(s), deviated points in x-direction called leverage points, usually measured by magnitude of diagonal elements of 'Hat' matrix, and iii) the other is in both directions (X and Y). An observation (y_i, x_i) is called influential if it has drastic effect on LSE, measured by different types of diagnostic measures such as Cook's distance (CD) (Cook, 1977), DFFITS, DFBETAS (Belsley, Kuh, and Welsch, 1980), etc. An outlier may or may not be influential. However, sometimes even a single outlier can totally make LSE unfaithful. LSE has finite breakdown point 1/n (finite breakdown point can be intuitively defined as the ratio of minimum number of contaminated observations that can blow up the estimator (Hampel *et al.*, 1986)). Lack of stability of LSE is a serious problem for estimating the parameters, and for the lack of normality assumption on error terms, we cannot test the reliability of the estimated parameters by using the common test procedures like t, F or chi-square. We can depend on robust regression that possesses some stability in variance and bias under derivations from the regression model. In short, robust regression techniques provide answers similar to the least-squares regression when the data are linear with normally distributed errors, but differ significantly from the result of least-squares method when the errors don't satisfy the normality conditions or/and when the data contain significant outliers.

3.2. Diagnostics Versus Robust Regression

"Robustness and diagnostics are complementary approaches to the analysis of data, any one of the two alone is not good enough". (Rousseeuw, 1984). 'In robust statistics, one seeks new inferential methods that are rather insensitive to, or robust against, certain types of failures in the parametric model, so good answers are obtained even in some assumptions are only approximately true. Diagnostics have taken traditionally a somewhat different view. Rather than modifying the fitting method, diagnostics condition on the fit using standard methods to attempt to diagnose incorrect assumptions, allowing the analyst to modify them and refit under the new set of assumptions.' (Stahel and

Weisberg, 1991). Field diagnostics is a combination of graphical and numerical tools. It is design to detect and delete the outliers first and then to fit the 'good' data by least squares. On the contrary, a robust regression first wants to fit a regression to the majority of the data and then to discover the outliers as those points that possess large residuals from the robust output. Hence the goal of robustness is to safeguard against deviation from the assumptions and diagnostics is used to find and identify deviations from the assumptions. Packages Minitab and SPSS have facilities for modern tools of diagnostics.

3.3. Different Types of Robust Regression

There are different types of robust regression techniques are available in literature. The most commonly used methods are as follow:

3.3.1 Robust M (GM) Estimator

In regression Huber P.J (1973) introduced M estimator that he had developed in 1964 to estimate location parameter robustly. The name "M-estimator" (Huber, 1964) comes from "generalized maximum likelihood". Robust M-estimators attempt to limit the influence of outliers and based on the idea of replacing the squared residuals $r_i^{\ 2}$ used in LS estimation with less rapidly increasing loss-function of the data value and parameter estimate, yielding minimize $\sum_{i=1}^{n} \rho(r_i)$; where is a symmetric, positive-definite function generally with a unique minimum at zero. Differentiating this expression with respect to the regression coefficients yields

$$\sum_{i=1}^{n} \psi(r_i)x_i = 0 \qquad (3.3.1.1)$$

where, ψ, the derivative of ρ, whose normalized is called the influence function that measures the influence of an observation on the value of the parameter estimate. If now we define a weight function $w(z) = \frac{\psi(z)}{z}$ then equation (3.3.1.1) becomes

$$\sum w(r_i)r_i \frac{\delta r_i}{\delta \beta_j} = 0; \qquad j = 1,2,\ldots\ldots, p. \qquad (3.3.1.2)$$

This is exactly the system of equations that we obtain.

M-estimators are statistically more efficient (at a model with Gaussian errors) than L_1 regression while at the same time they are still robust with respect to outlying y_i. It has two drawbacks: i) estimators are not equivariant (in many cases the nature of the observations is such that changing the units of measurement should have the conclusion unaltered, Davies 1993.) and ii) finite breakdown point is 1/n due to outlying x_i. To remove these demerits Mallows (1975) suggested,

$$\sum_{i=1}^{n} w(x_i)\psi(r_i / \hat{s})x_i = 0 \qquad (3.3.1.3)$$

where, σ is estimated simultaneously by s^ and later some proposed some variants of this form, but nobody was able to attain maximum possible breakdown point.

3.3.2 Least Median of Squares Regression (LMS)

LMS was proposed by Hampel (1975) and further developed by Rousseeuw (1984). Instead of minimizing the sum of squared residuals, Rousseeuw proposed minimizing their median as follows:

$$\underset{\hat{\beta}}{\text{minimize}}\ \underset{i}{med}\ r_i^2 \ . \qquad (3.3.2.1)$$

This estimator effectively trims almost the (n/2) observations having the largest residuals, and uses the maximal residual value in the remaining set as the criterion to be minimized. Its breakdown point is $\left(\frac{[n/2] - p + 2}{n}\right)$ for p-dimensional data set i.e., it attains maximum possible breakdown point $\frac{1}{2}$ at usual models but unfortunately it possesses poor asymptotic efficiency. So later robust statisticians developed LTS and MM estimators. In spite of that the family of LMS has excellent global robustness.

3.3.3 Least Trimmed Squares (LTS) Method

LTS was introduced by Rousseeuw (1984) and is given by

$$\min_{\hat{\beta}} \sum_{i=1}^{h} r_{i:n}^{2} \qquad (3.3.3.1)$$

where, $r_{1:n}^2 \le \ldots \le r_{n:n}^2$ denote the ordered squared residuals and $h$is to be chosen between $\frac{n}{2}$ and n. The LTS estimators search for the optimal subset of size h who's least squares fit has the smallest sum of squares residuals. Hence, the LTS estimate of β is then the least square estimate of that subset of size h. For the data comes from continuous distribution breakdown points of LTS equals $\min(n-h+1, h-p+1)/n$, we have $h = \frac{n+p+1}{2}$ yields the maximum breakdown point, is asymptotically 50%, whereas h=n gives the ordinary least squares with breakdown point=1/n. LTS has the properties such as affine equivariance and asymptotic normality. Its influence function is bounded for both (directions; response and explanatory) the vertical outliers and bad leverage points. Moreover, LTS regression has several advantages over LMS. Its objective function is smoother, making LTS less 'jumpy' (i.e., sensitive to local effects) than LMS. LTS has better statistical efficiency than LMS, because of its asymptotically normal property (Hossjer, 1994), where as LMS has a lower convergence rate (Rousseeuw 1984). This also makes the LTS more suitable than the LMS as a starting point for two-step estimators such as the MM – estimators (Yohai, 1987) and generalized M-estimators (Simpson, Ruppert and Carrol, 1992, Cookley and Hettmansperger, 1993). It also fails to fit a correct model when large number of clustered outliers exits and with more than 50% outliers in the data. The performance of this method has recently been improved by the FAST-LTS (Rousseeuw and Van Driessen, 1999) and Fast and robust bootstrap for LTS (Willems, G. and Stefan V. A, 2004).

3.3.4 S -Estimates and MM -Estimates of Regression

S-PLUS computes a robust M-estimate $\hat{\beta}$, which minimizes the objective function

$$\sum_{i=1}^{n} \rho\left(\frac{y_i - x_i^T\beta}{\hat{s}}\right) \tag{3.3.4.1}$$

Where $\hat{s}$ is a robust scale estimate for the residuals and ρ is a particular optimal symmetric bounded loss function, described below.

Alternatively $\hat{\beta}$ is a solution of the estimating equation

$$\sum_{i=1}^{n} x_i \psi\left(\frac{y_i - x_i^T\beta}{\hat{s}}\right) = 0 \tag{3.3.4.2}$$

$\psi = \rho'$ is a redescending (nonmonotonic) function. A key issue is that since ρ .is bounded, it is nonconvex, and the minimization above can have many local minima. Correspondingly, the estimating equation above can have multiple solutions. S-PLUS deals with this by computing highly robust initial estimates of $\hat{\beta}$ and $\hat{s}$ with breakdown point 0.5, using the S-estimate approach described below and computes the final estimate as the local minimum of the M-estimate objective function nearest to the initial estimate. We refer to an M-estimate of this type and computed in this special way as an MM-estimate, a term introduced by Yohai (1987).

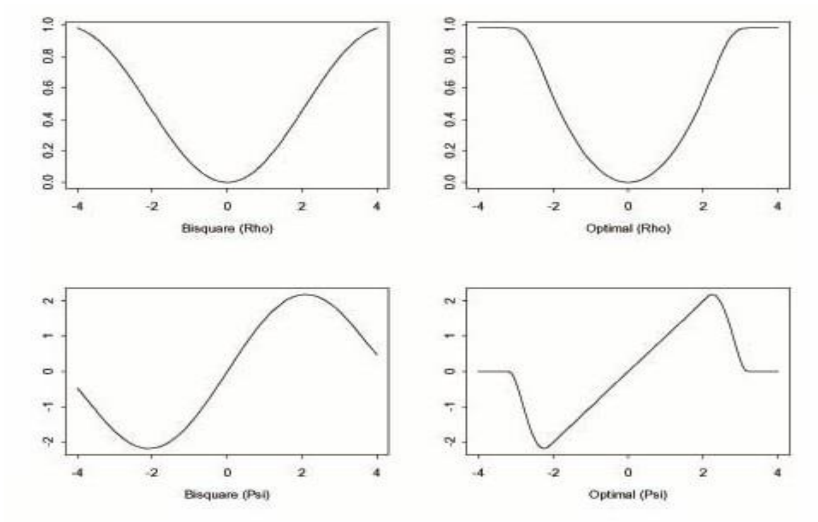

Fig 3.3.4.1 Tukey's bisquare loss function and optimal loss function (Yohai and Zamar, 1998) in S-Plus.

The key to obtaining a good local minimum of the M-estimation objective function, when using a bounded, non-convex loss function is to compute a highly robust initial estimate < ^{0}. S-PLUS does this by using the S-estimate method introduced by Rousseeuw and Yohai (1984), The S-estimate approach has as its foundation an M-estimate of an unknown scale parameter for observations y_1, y_2,............,y_n , assumed to be robustly centered (that is, by subtracting a robust location estimate). The M-estimate s^ is obtained by solving the equation

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i}{s}\right)=.5 \qquad (3.3.4.3)$$

where, . is a symmetric, bounded function. It is known that such a scale estimate has a breakdown point of one-half (Huber, 1981), and that one can find min-max bias robust M-estimates of scale.

Following this M- estimate Rousseeuw and Yohai (1984) introduced regression S-estimate method. Let us consider the linear regression model modification.

$$\frac{1}{n-p}\sum \rho\left(\frac{y_i - x_i^T \beta}{s(\beta)}\right) = .5 \qquad (3.3.4.4)$$

For each value of β, we have a corresponding robust scale estimate $\hat{s}(\beta)$. The regression S-estimate (which stands for "minimizing a robust scale estimate") is the value $\hat{\beta}^0$ that minimizes $\hat{s}(\beta)$:

$\hat{\beta}^0 = \arg\min_\beta \hat{s}(\beta)$ and corresponding minimized functional value is *S*-estimate of σ. This presents another nonlinear optimization, one for which the solution is traditionally found by a random resampling algorithm, followed by a local search (Yohai, Stahel, and Zamar, 1991). Once the initial S-estimate of $\hat{\beta}^0$ is computed, the final M-estimate is obtained as the nearest local minimum of the M-estimate objective function.

3. 4. Demerits of Robust Regression

Though robust regression technique gives us better result as an alternative of OLS, it has some deficiencies – (i) The diversity of estimator types and the necessary choices of tuning constants, combined with a lack of guidance for these decisions (ii) The lack of simple procedures for inference or reluctance to use the straight forward inference based on asymptotic. Here we have to take some help from the re-sampling technique like

'bootstrapping' especially for small samples (iii) unfamiliarity with interpretation of results from a robust analysis.

3. 5. The Bootstrap Procedure

3. 5.1 Bootstrap Standard Error, Bias and Confidence Interval (CI)

Bootstrap technique proposed by Efron (1979) is such a procedure, which creates a huge number of sub-samples from a pre-observed data set by a simple random sampling with replacement that could be later used to investigate the nature of the population without having any assumption about them. This computer-based technique is used for estimating standard error, bias, confidence interval and other statistical measures. Bootstrap is a very useful technique indeed but caution must be taken while considering this technique.

Le us wish to estimate a parameter of interest $\theta = t(F)$ on the basis of X, we calculate an estimate $\hat{\theta} = S(X)$ from X. Efron (1979) proposed the bootstrap technique where $X^* = (X_1^*, X_2^*, ..., X_n^*)$ is called the 'bootstrap sample'. Practically we draw a large number (say B) of bootstrap samples, which may be represented as $X^{*1}, X^{*2}, ..., X^{*B}$. Here the bootstrap replication corresponding to each bootstrap sample is

$$\hat{\theta}^{*b} = S(X^{*b}), b = 1,2,..,B \qquad (3.5.1.1)$$

The estimated standard error becomes

$$\hat{se}_B = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} \left[\hat{\theta}^*(b) - \hat{\theta}^*(.) \right]^2 \right\}^{\frac{1}{2}} \qquad (3.5.1.2)$$

where, $\hat{\theta}^*(.) = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*(b)$. The main advantage of using (3.5.1.2) as an estimate of standard error is that $\hat{se}_B$ approaches $se_{\hat{F}_n}$ as B goes to infinity (Efron and Tibshirani, 1993). In this process it does not require any parametric assumption and this estimator can reduce the bias significantly.

In parametric statistics we need generally large sample normal theory to construct confidence intervals. Through the use of bootstrap we can obtain more accurate intervals without large sample normal theory assumptions. There are several ways of constructing intervals. At first we introduce an interval that is known as bootstrap-*t* interval. After generating B bootstrap samples we compute

$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}}{\hat{se}^*(b)} \qquad (3.5.1.3)$$

where, $\hat{\theta}^*(b) = s(X^{*b})$ is the value of $\hat{\theta}$ for the bootstrap sample X^{*b}.$\hat{\theta}^*(b)$ and $\hat{se}^*(b)$ is the estimated standard error of $\hat{\theta}^*$ for the bootstrap sample X^{*b}. The -th percentile of $Z^*(b)$ is estimated by the value $\hat{t}^{(\alpha)}$ such that $\alpha = \frac{\{Z^*(b) \le \hat{t}^{(\alpha)}\}}{B}$

Finally the bootstrap–t interval is

$$\left[\hat{\theta} - \hat{t}^{(1-\alpha)}\hat{se}_B ,\ \hat{\theta} - \hat{t}^{(\alpha)}\hat{se}_B\right]$$

Here we introduce another approach to CI based on percentile of the distribution of statistic. It is generally known as bootstrap percentile method**.** For bootstrap replications, let us define

$\hat{\theta}_{lo} = \hat{\theta}^{*(\alpha)}$ =100   -th percentile of B bootstrap replications

$\hat{\theta}_{up} = \hat{\theta}^{*(1-\alpha)}$ = 100(1-   )-th percentile of B bootstrap replications

Then the bootstrap percentile interval of intended coverage (1-2   ) is given by

$$\left(\hat{\theta}_{lo}, \hat{\theta}_{up}\right) = \left(\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}\right)$$

It is well known that standard normal interval is neither second order accurate nor transformation respecting where as bootstrap-t interval is second order accurate, but not transformation respecting, and percentile interval is transformation respecting but not first order accurate. So another type of interval, the bias corrected and accelerated (BC_a) interval that satisfies both conditions is considered in the bootstrap literature. The BC_a interval of intended coverage (1-2   ) is defined as

$$\left(\hat{\theta}_{lo}, \hat{\theta}_{up}\right) = \left(\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}\right)$$

where, $\alpha_1 = \phi\left[\hat{Z}_0 + \dfrac{\hat{Z}_0 + Z^{(\alpha)}}{1-\hat{a}\left(\hat{Z}_0 + Z^{(\alpha)}\right)}\right]$ and $\alpha_2 = \phi\left[\hat{Z}_0 + \dfrac{\hat{Z}_0 + Z^{(1-\alpha)}}{1-\hat{a}\left(\hat{Z}_0 + Z^{(1-\alpha)}\right)}\right]$.

Here ϕ is the standard normal cumulative distribution function: $Z^{(\alpha)}$, the $100\alpha\, th$ percentile point of standard normal distribution; $\hat{z}_0$, the bias correction factor and $\hat{a}$, the acceleration factor. to consider the dependence of standard error on the parameter itself. "At the present level of development, the BC_a intervals are recommended for general use, especially for nonparametric problems" (Efron and Tibshirani, 1993) and for small sample in inexact parametric case.

3.5.2 Bootstrapping Regression

The least square estimate is the solution to the so-called normal equations

$$X^T X\beta = X^t Y$$

And is given by the formula

$$\hat{\beta} = (X^T X)^{-1} X^t y \qquad (3.5.2.1)$$

There are two general ways to bootstrap a regression like this; we can treat the predictors as random, potentially changing from sample or as fixed. Random-*X* resampling is also called case/pairs resampling, and fixed –*X* resampling is also called model-based resampling. Pairs bootstrapping involves choosing random samples in pairs from the original data set. Stratified bootstrapping involves resampling the data in groups defined by the magnitude of *X*. Residuals bootstrapping involves calculating the residuals, $\hat{\varepsilon}_i = y_i - x_i\hat{\beta}$, resampling these, and computing a new y_i^*.

Fixed -*X* Resamping (Bootstrap 1)

How can we generate bootstrap replications when the model matrix *X* is fixed? One way to produce is to treat the fitted values $\hat{y}_i$ from the model as giving the expectation of the response for the bootstrap samples. Attaching a random error to each $\hat{y}_i$ produces a fixed –*X* bootstrap samples, $y_r^* = \{y_{ri}^*\}$.

The errors could be generated parametrically from a normal distribution with mean 0 and variance σ^2, if we are willing to assume that the errors are normally distributed or non-parametrically, by resampling residuals from the original regression. We would then regress the bootstrapped values y_r^* on the fixed *X* matrix to obtain bootstrap replications of regression coefficients.

Let $\varepsilon_r^* = (\varepsilon_{r1}^*, \varepsilon_{r2}^*, \ldots \varepsilon_{rn}^*)$ be the *r*th bootstrap errors, Where *r* = 1,2,...B, and *B* is the bootstrap replications. Then the bootstrap responses $y_r^* = \{y_{ri}^*\}$ are generated by

$$y_{ri}^* = x_i\hat{\beta} + \varepsilon_{ri}^* \qquad (3.5.2.2)$$

which is called model based resampling of linear regression model, bootstrapping the residuals of linear regression model or Bootstrap 1 method of linear regression model.

The bootstrap least –squares estimate $\hat{\beta}^*$ is the minimizer of the residual squared error for the bootstrap data

$$\sum_{i=1}^{n}\left(y^{*}_{ri} - x\hat{\beta}^{*}\right)^{2} = \min_{r}\sum_{i=1}^{n}\left(y^{*}_{ri} - xb\right)^{2} \qquad (3.5.2.3)$$

The normal equation (3.5.2.1), applied to the bootstrap data, gives

$$\hat{\beta}^{*}_{r} = \left(X^{T}X\right)^{-1}X^{T}y^{*}_{r}. \qquad (3.5.2.4)$$

In this case we do not need Monte Carlo simulations to figure out bootstrap standard errors for the components of $\hat{\beta}^{*}$. An easy calculation gives a closed form expression for the $\text{var}_{\hat{F}}\left(\hat{\beta}^{*}_{i}\right) = \text{var}_{\infty}\left(\hat{\beta}_{j}\right)$, the ideal bootstrap variance is

$$\text{var}\left(\hat{\beta}^{*}_{r}\right) = \left(X^{T}X\right)^{-1}X^{T}\,\text{var}\left(y^{*}_{r}\right)X\left(X^{T}X\right)^{-1}$$

$$= \hat{\sigma}^{2}_{F}\left(X^{T}X\right)^{-1}$$

Since $\text{var}\left(y^{*}_{r}\right) = \hat{\sigma}^{2}_{F} I$ where I is the identity matrix. and $\hat{\sigma}^{2}_{F} = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}^{2}$

Therefore, the algorithm to generate simulation datasets and corresponding parameter estimates is as follows:

Algorithm (Model-based resampling in linear regression)

For $r = 1,2,\ldots,B$,

Step 1

(a) Data matrix X is fixed

(b) Random sample $\varepsilon^{*}_{r} = \{\varepsilon^{*}_{ri}\}$,

(c) The responses $y^{*}_{r} = \{y^{*}_{ri}\}$ is generated

$$y^{*}_{ri} = x_{i}\hat{\beta} + \varepsilon^{*}_{ri} \qquad i = 1,2,\ldots,n$$

Step 2

Least squares regression is fit to (X, y^{*}_{r}) giving estimates β^{*}_{rj}, $\text{var}\left(\beta^{*}_{rj}\right)$, $j = 1,2,\ldots,p$

Random - *X* Resampling (Bootstrap 2)

It is a completely different approach using the data as a sample of pairs, z_i=(y_i, xi)= ($y_i, x_{i1}, x_{i2}, \ldots \quad x_{ip}$), (i = 1,2,..., n). In random –X resampling, we simply select B samples of pairs $\left\{z^*_{bi_j} : i_j \in \{1,2....,n\right\}$ with replacement, b=1,2,---B; fit the model for each bootstrap sample and each time estimate the coefficients . Using B no. of values of estimators we estimate bootstrap standard error, bootstrap CI etc.

(*In this work we follow Bootstrap2.*)

4. Analyses and Discussion

We now show how robust regression techniques give us more accurate results comparing to LS with an analysis of purchasing power parity analysis. The Big Mac. Index of 31 countries as a crude and hilarious, measure of whether international currencies are at their current exchange rate, as judged by the theory of purchasing power parity (The PPP holds that a unit of currency should be able to buy the same bundle of goods in all countries). Consider the regression equation as: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where the response variable y_i is the Actual Exchange Rate (AER), the explanatory variable is PPP and u_i is the error term. Since the proponents of PPP argue that, in the long run, currencies tend to move toward their PPP, we have to estimate and test the null hypothesis H_0: β_0=0, β_1=1 and the alternatives H_1: $\beta_0 \neq 0$, $\beta_1 \neq 1$, especially on β_1. At first, we perform a scatter plot using 'S-Plus' (Fig: 4.1, appendix) and we see that one observation is apart from the bulk of the data in both directions, so we suspect this point as an outlier (leverage point). Now we entry the data under the 'Program for Robust Regression' (PROGRESS), we get the fitted values from ordinary Least Squares (LS) method and a robust regression technique "Least Median Squares (LMS) regression"; looking at the results from both the methods we find weakness of LS method, as for example we get –118.76 and .92 successively from the LS and LMS against actual exchange rate 1.00 (Argentina), table 4.3. But if we consider the case for Poland we get LS predicts more accurate than LMS. This means that LS is extremely affected by outlier. From the Fig: 4.2 we see that LS fit the line wonderfully with the outlier in the factor space while Fig 4.4 (in appendix) indicates that it is also too influential. We get the value for coefficient of determination, R^2 (Adjusted) =. 993, table 4.1,hence the leverage point is treated itself as good leverage, it improves the precision of the regression coefficient, F = 4341.62 (p = .0000), but how vague the result is! We know that inference of LS depends on the assumption that errors are normally distributed. We see from normal probability plot Fig: 4.3 that normality has seriously broken. To judge it in probabilistic term we calculate RM value (Imon 2003)= 370.5067 (whose p-value =. 0000). We can now safely infer that assumption of normality is strongly unreliable. There fore, we cannot use the LS method. We then use the methods of robust regression LMS and MM, both the methods give us slightly different fitted values comparing with LS, see Fig: 4.5. Next, we perform the two above methods without the leverage point. We see all the techniques give more accurate result without leverage point. A drastically changes occur in LS after the elimination of the case 22. Fig: 4.6 show it clearly; the fitted lines are approximately same in all the methods. We get calculated value of t (25.95) is greater than tabulated t (2.045) with 29 degrees of freedom, i.e. we may reject the H_0: β_1=1; there is no strong evidence that according to the proponent of PPP currencies tend to move toward their PPP. WLS gives reverse direction about β_1. When we consider the LMS and MM robust techniques for the estimates, we depend on 'bootstrapping'. Traditionally researchers have relied on different versions of central limit theorem and normal approximation to obtain confidence intervals. These techniques are valid only if the statistic or some known transformation on it is asymptotically normally distributed and sample size is large. If the normality assumption does not hold, then the traditional method cannot be used to obtain confidence intervals. We have already mentioned in the section 3.4 that with the availability of modern computing power, researchers need no longer rely on asymptotic theory to estimate the distribution of a statistic. They may use re- sampling methods, for either normal or non normal distributions. (Chen. C.), to estimate the confidence intervals (CI). Using the pair wise bootstrap we have got the 95% CI for β_1= (.5799,1.451) in the case of LMS, This make decision in favor of β_1=1.

Table: 4.1 Model summary of LS method.

Model Parameters					R^2	Adj. R^2	Std. Err.	DW	F-Value	p-Value
Coef	Esitim	Std. Err.	t-Value	p-Value						
β_0	-121.3	61.87	-1.96	0.06	.993	.993	332.73	2.013	4341.62	0.000
β_1	1.651	0.025	25.95	0.000						

Table 4.2 Basic findings of AER and PPP regression.

Findings **Criteria**	**Alarming values**	**Cut-off Values**	**No.** **Obs.**	**Comments:**
Standardized Residuals	-4.62 4.94 -2.2 a	Absolute value < 3	18 22 26	Observation 22 is an outlier as well as a leverage point and an influential observation of extreme type. But observation 18 is an outlier but not a leverage point. It shows significant influence by DFFITs, not by Cook Distance. a absolute value of all others ≤ 0.596 b absolute value of all others ≤ 0.562 c all other values ≤ 0.042 d all other values ≤ 0.08 e absolute value of all other values ≤ 0.1062
Studentized Residuals	-8.83 12.2311 -2.36 b	Absolute value < 3	18 22 26	
h_{ii} (Diagonal Elements of Hat matrix)	0.967 c	$>3\frac{p}{n} = 0.19$ or $>2\frac{p}{n} = 0.129$	22	
$CD^2(i)$ (Cook Distance)	0.473 d 363.13	Near 1	18 22	
DFFITS(i)	-1.8992 66.6946 -0.4372 e	$>2\left(p/n\right)^{\frac{1}{2}} = 0.508$	18 22 26	

Table 4.3 Results for the values of AER, PPP, Fitted LS, Fitted LMS and Fitted MM

In	Count.	AER	PPP	FitLS	LMS	MM	D.Hat	WLS	WMM	WLMS
1	Argentin	1.00	1.57	-118.7	.92	1.87	.03	10.79	1.78	.92
2	Australia	1.42	1.07	-119.6	.46	1.41	.03	10.29	1.32	.48
3	Austra	12.00	14.8	-96.91	13.24	14.03	.03	23.85	13.96	13.24
4	Belgiam	35.2	47.39	-43.09	43.6	44.01	.03	56.03	43.96	43.6
5	Brazil	949	652	955.37	606.7	600.03	.03	6529	600.5	606.7
6	Britain	1.46	1.27	-119.3	.64	1.59	.03	10.5	1.5	.64
7	Canada	1.39	1.24	-119.3	.62	1.56	.03	10.5	1.48	.62
8	Chile	414	412	559.03	383.1	379.3	.03	416	379.6	383.2
9	China	8.7	3.91	-114.9	3.1	4.02	..03	13.1	3.93	3.1
10	Cech.R	29.7	21.7	-85.52	19.67	20.38	.03	30.7	20.31	19.67
11	Denmk	6.69	11.2	-102.8	9.89	10.72	.03	20.3	10.64	9.89
12	France	5.83	8.04	-108.0	6.95	7.82	.03	17.2	7.74	6.95
13	Germ.y	1.71	2.0	-118.0	1.32	2.26	.03	11.2	2018	1.32
14	Greece	251	270	324.53	25.92	248.7	.03	276	248.8	250.9
15	Holland	1.92	2.37	-117.4	1.67	2.6	.03	11.6	2.52	1.67
16	Hongko.	7.73	4.0	-114.7	3.19	4.1	.03	13.2	4.02	3.19
17	Hungary	103	73.48	-.01	67.9	68.0	.03	81.78	67.97	67.9
18	Italy	1641	1978	3145.1	1841.	1819.5	.04	1962	1821.	1842
19	Japan	104	170	159.39	157.7	156.76	.03	177.1	156.8	157.8
20	Malaysia	2.69	1.64	-118.7	.99	1.93	.03	10.86	1.84	.99
21	Mexico	3.36	3.52	-115.5	2.74	3.66	.03	12.71	3.57	2.74
22	Poland	22433	13478	2213.6	12552	12395	.97			
23	Portugal	174	191	194.07	177.3	176.08	.03	197.8	176.1	177.4
24	Russia	1775	1261	1961.1	1173	1160	.03	1254	1161.	1178
25	Singapor	1.57	1.3	-119	.67	1.62	.03	10.52	1.53	.67
26	S.Korea	810	1000	1530.1	930.8	920	.03	996.5	920.8	930.8
27	Spain	138	150	126.36	139.2	138.4	.03	157.3	138.4	139.2
28	Sweden	7.97	11.1	-103	9.8	10.63	.03	20.2	10.55	9.8
29	Swit.land	1.44	2.48	-117	1.77	2.7	.03	11.69	2.62	1.77
30	Taiwan	26.4	26.96	-76.83	24.57	25.22	.03	35.85	25.15	24.57
31	Thailand	25.30	20.87	-86.89	18.9	19.62	.03	29.84	19.54	18.9

5. Conclusion

Finally we reach on a conclusion that LS is totally fails due to outliers. We must have depends on robust regression and/or regression diagnostics methods when outliers present in the data set or error distribution is non-normal and we can use bootstrap technique to infer from the estimates of robust regression methods.

6. References

1. Belsley, D. A., Kuh, E., and Welsch, R.E. (1980). Regression diagnostics, identifying influential data and sources of collinearity, Wiley and Sons, NY.
2. Cookley, C.W. and Hettmansperger, T.P. (1993). A bounded influence, high breakdown efficient regression estimator, JASA, 88, pp. 872-880.
3. Cook, R. D. (1977). Detection of influential observations in linear regression, Technometrics, 19, pp. 65-68.
4. Colin C., SAS Institute Inc, Cary. NC, Paper 265-27; Robust Regression and Outlier Detection with the ROBUSTREG Procedure.
5. Davies, P.L. (19993). Aspects of robust linear regression, The Annals of statistics, 21, 4, pp. 1843-1899.
6. Gujarati, D.N. (1995). *Basic Econometrics,* McGraw-Hill, 3rd Edition. N.Y.
7. Efron, B. (1979). Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1-26.
8. Efron, B. (1992), Jackknife-after-bootstrapping Standard Errors and Influence Functions. J.Roy.Stat.Soc., Ser-B, 54, pp. 83-127.
9. Efron, B. and Tibishirani, R.J. (1993). An Introduction to the Bootstrap. Chapman & Hall. N.Y.
10. Hampel, F.R. (1975). Beyond Location Parameters, Robust Concepts and Methods, Bull Stat. Ins. 46, pp. 375-382.
11. Hampel, F.R., Ronchetti, E. M., Rousseeuw, P.J. and Stahel, W. A. (1986*). Robust Statistics: The Approach Based on Influence Function*. John Willey and Sons, Inc., N.Y.
12. Hossjer, O. (1994). Rank-based estimates in the linear model with high breakdown point, JASA, 89, pp. 149-158.
13. Huber, P.J. (1964). Robust estimation of a location parameter, Ann, Math. Statist, 35, pp. 73-101.
14. Huber, P.J. (1991). Between Robustness and Diagnostics. in *Directions in Robust Statistics and Diagnostics* Part-I, Stahel, W. and Weisberg, S, Editors Spriger-Verlag, N.Y., pp. 121-130.

15. Huber, P.J. (1973). Robust regression: Asymptotics, Conjectures, and Monte Carlo, *Annals of Statistics*, 1, pp.799-821.

16. Huber, P.J. (1981). *Robust Statistics*, Wiley and Sons, NY.

17. Mallows, C.L. (1975). On Some Topics in Robustness, Unpublished *Bell Telephone Lab. Report*, Murrey Hill, NJ.

18. Rahmatullah Imon, A.H.M. (2003). Regression Residuals, Moments and Their Use in Tests for Normality, *Communications in Statistics: Theory and Methods*.Vol.32, No.5, pp. 1021-1034.

19. Resampling Techniques, *Guide to Statistics*, Vol-2; S-PLUS Professional Release 1, Math Soft Inc. p. 542.

20. Rousseeuw, P.J. (1984). Least Median of Squares Regression, *Journal of American Statistical Association*, Vol-79, pp. 871-880.

21. Rousseeuw, P. J. and Leroy, A.M. (1987), Robust Regression and Outlier Detection. John Willey and Sons, Inc. N.Y.

22. Rousseeuw, P.J., and Yohai, V. (1984). Robust regression by means of S-estimators, in robust and non-linear time series analysis, edited by J. Franke, W. Hardle, and R.D. Martin, Lecture notes in Statistics No. 26, Springer-Verleg, NY, 256-272.

23. Rousseeuw, P.J., and Driessen K.V. (1999). Dept. of Mathematics, Univ. Anwerp,Available at : fttp://win-ftp.uia.ac.be./pub.

24. Shao, J. (1990). Bootstrap Estimation of the Asymptotic Variances of Statistical Functionals, *Ann. Inst. Statist. Math*..Vol. 42, No. 4, pp. 737-752.

25. Simpson, D. G., Ruppert, D. and Carroll, (1992). On One Step GM-Estimates and Stability of Inference in Linear Regression, JASA, 87, pp. 439-450.

26. S-PLUS 4. (1997). *Guide to Statistics*, Math Soft Inc., Seattle, Washington..

27. Stahel, W. and Weisberg, S. Editors Spriger-Verlag. (1991*). Directions in Robust Statistics and Diagnostics*, Part-I, N.Y.

28. Staudte, R. G. and Sheather, S.J. (1990). Robust Estimation and Testing, A Wiley-Interscience Publication., N.Y.

29. Willems, G. and Stefan, V. A. (2004). Fast and robust bootstrap for LTS, Elsevier Science.

30. Yohai, V., Stahel, W. A. and Zamar, R. H. (1991). A Procedure for Robust Estimation and Inference in Linear Regression, in *Directions in Robust Statistics and Diagnostics*, Part-II.

31. Stahel, W. and Weisberg, S. Editors Spriger-Verlag. (1991). *Directions in Robust Statistics and Diagnostics*, Part-I, N.Y., pp. 365-374.

Appendix

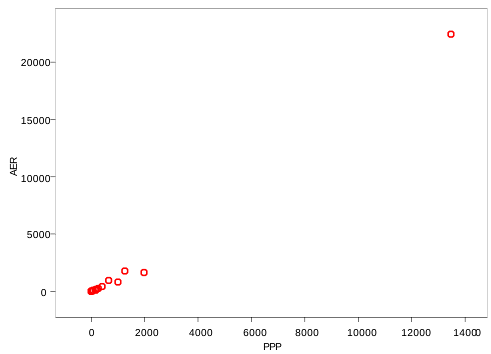

Fig. 4.1 Scatter plot for AER vs PPP.

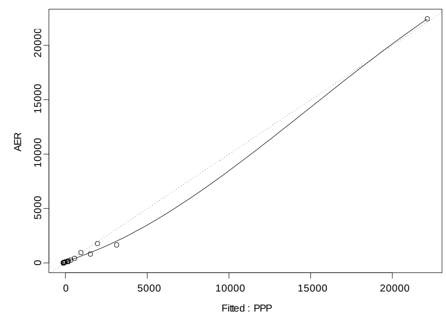

Fig. 4.2 LS fitted line with smooth plot.

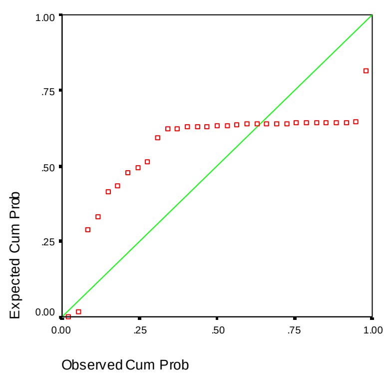

Fig. 4.3 Residuals normal PP plot.

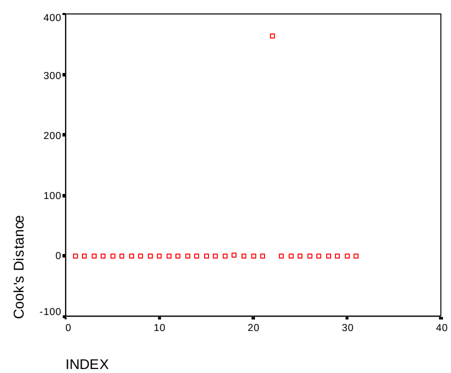

Fig. 4.4 Index Plot of Cook Distance.

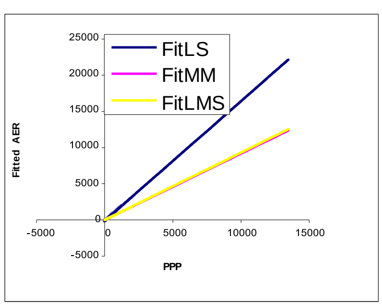

Fig. 4.5 Fitted line from LS, MM and LMS.

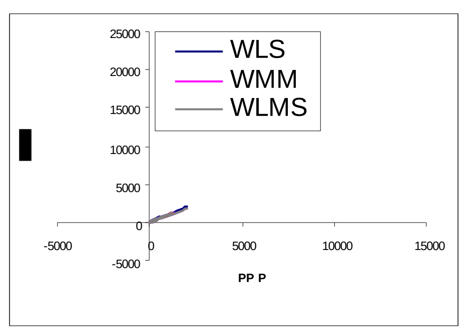

Fig. 4.6 Fitted line without 1 leverage point.