WWW.JAQM.RO

JOURNAL OF APPLIED QUANTITATIVE METHODS

Quantitative Methods Inquires

Vol. 9
No. 2
Summer
2014

ISSN 1842–4562

JAQM Editorial Board

JAQM Advisory Board

Contents

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

TOWARD FITS TO SCALING-LIKE DATA,BUT WITH INFLECTION POINTS & GENERALIZED LAVALETTE FUNCTION

Marcel AUSLOOS

E-mail: ausloosm@fastmail.fm

ABSTRACT

Experimental and empirical data are often analyzed on log-log plots in order to find some scaling argument for the observed/examined phenomenon at hands, in particular for rank-size rule research, but also in critical phenomena in thermodynamics, and in fractal geometry. The fit to a straight line on such plots is not always satisfactory. Deviations occur at low, intermediate and high regimes along the *log(x)*-axis. Several improvements of the mere power law fit are discussed, in particular through a Mandelbrot trick at low rank and a Lavalette power law cut-off at high rank. In so doing, the number of free parameters increases. Their meaning is discussed, up to the 5 parameter free super-generalized Lavalette law and the 7-parameter free hyper-generearalized Lavalette law. It is emphasized that the interest of the basic 2-parameter free Lavalette law and the subsequent generalizations resides in its "noid" (or sigmoid, depending on the sign of the exponents) form on a semi-log plot; something incapable to be found in other empirical law, like the Zipf-Pareto-Mandelbrot law. It remained for completeness to invent a simple law showing an inflection point on a log-log plot. Such a law can result from a transformation of the Lavalette law through $x \rightarrow log(x)$, but this meaning is theoretically unclear. However, a simple linear combination of two basic Lavalette law is shown to provide the requested feature. Generalizations taking into account two super-generalized or hyper-generearalized Lavalette laws are suggested, but need to be fully considered at fit time on appropriate data.

Keywords: graphs, plots, nonlinear laws.

1. INTRODUCTION

In recent years, following the rise in the understanding of critical phase transitions [1] through the notion of critical exponents, many results have been presented on log-log graphs. It should be emphasized at once that the search for a straight line fit on such a graph is of interest when the hypothesis of scaling is appropriate for the examined property or effect. Then, the slope on the plot gives some indication of some characteristic exponent at the phase transition because the underlying analytical function, the excess free energy [1], has a homogeneity property. Two other major scientific concepts, related to some underlying scaling hypothesis, have also led to examining log-log plots for various quantities: one is the notion of fractal dimension [2], the other is the rank-size relationship through so called Zipf plots [3].

It is often discussed whether the scaling law should hold over many decades of the *x*-axis variable, -whatever the *x*-axis (reduced temperature ε, bin size *n*, rank *r*, ...). Officially, this "many decades validity" should be the case, if a scaling law fully holds. However, phenomena for which (quasi) straight lines are seen on a log-log plot are rarely found, - outside laboratories or computer simulations. Yet, there is no harm in recognizing that such a straight line existing on a small *x*-axis range indicates the presence of a specific regime; see for example the case of the population size of large italian cities, as illustrated in Fig. 1, for which two regimes rather than a single one can be imagined. Therefore, weak scaling can be accepted as physically suggestive within finite *x*-axis ranges.

Nevertheless, the data can often present convex or concave shapes, and often gaps, jumps, drops (see Fig. 1) or shoulders. Such a large variety of basic shapes demands to pursue some systematic inquiry of the simplest appropriate analytical forms representing complicated data. Much difficulty resides in (interpreting and) theoretically manipulating inflection points, - often visible when a line is drawn through the data "for the eye".

The 2-parameter free power law (using thereafter the discrete variable *r* for the *x*-axis)

$$y_r = \frac{a}{r^{\alpha}} \qquad (1)$$

on a log-log plot is referred to Zipf's plot. Zipf had thought that the particular case α= 1 represents a desirable situation, in which forces of concentration balance those of decentralization [3, 4]. Such a case is called the rank-size *rule* [4]-[8]. Thus the scaling exponent *a* can be used to judge whether or not the size distribution is close to some optimum (equilibrium) state.

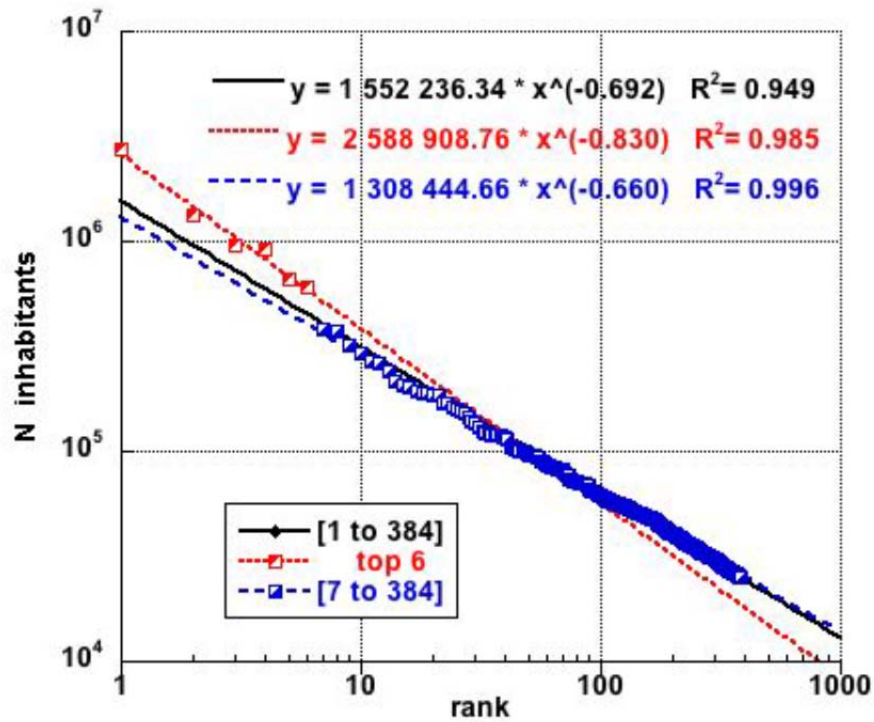


Figure 1: The 384 largest Italian cities ranked by decreasing order of their population size, pointing to a drop after the main 6; different power law fits for the whole range (black line) or when distinguishing two regimes (red and blue line) are indicated with their corresponding correlation coefficient R^2

The pure power-law distribution, for a continuous variable, reads

$p(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$ (2)

where k is a positive integer usually measuring some variable of interest; $p(k)$ is the probability of observing the value k; γ is the power-law exponent; and $\zeta(\gamma) \equiv \sum_{k=1}^{\infty} k^{-\gamma}$ is the Riemann zeta function; note that γ, in Eq.(2) must be greater than 1 for the Riemann zeta function to be finite.

However, the fit to a straight line on a log-log plot is not always truly perfect, as any reader has surely had the experience considering various data with expected scaling. The error bar (e.g., on γ) can be very large for a R^2 or χ^2 test point of view. Moreover, broadly used methods for fitting to the power-law distribution provide biased estimates for the power-law exponent [9].

The deviations occur in various regimes along the *log(x)*-axis.

When the data crushes at high x-axis value, Lavalette suggested [10] to use the 2-parameter free (κ, χ) form

$$y(r) = \kappa \left[\frac{N\, r}{N-r+1}\right]^{-\chi} \qquad (3)$$

in which the role of *r* as the independent variable, in Eq.(1), is taken by the ratio *r/(N — r +1)* between the descending and the ascending ranking numbers; *N* is the number of data points on the x-axis, and $\chi \geq 0$; the +1 role in *(N — r + 1)* is easily understood. Other ways of writing this 2-parameter Lavalette form function are of interest

$$y(r) = \kappa\, (N\, r/(N - r + 1))^{-\chi} \equiv k\, N^{-\chi}\, (r/(N - r + 1))^{-\chi} \qquad (4)$$

$$\equiv k\, (N\, r)^{-\chi}\, (N - r + 1)^{+\chi} \qquad (5)$$

$$\equiv \hat{k}\, r^{-\chi}\, (N - r + 1)^{+\chi}. \qquad (6)$$

in order to be emphasizing a power law decay with a power law cut-off. The interest in such a function which is strictly decreasing, Fig.2, from infinity at r = 0 under a $r^{-\chi}$ law to a zero value at *r* = N *+1* as $(N - r + 1)^{+\chi}$, best appears on a semi-log plot, Fig. 3: observe the inflection point presence at *r* = N/2. The slope s at such a point is equal to $-\,4\,\chi\,\frac{N+1}{N(N+2)}$ which for "large *r*" ~ – 4 χ *(1/N)(1 — 1/N)*. In some sense, it is realistic to reproduce this intermediary regime as y ~ e^{-sr}.

When $\chi \leq 0$, - not a rank-size rule case, the function is increasing, - it is a flipped Lavalette function. Both functions, i.e. with $\chi \geq 0$ or $\chi \leq 0$, are shown in Fig. 2 on a log-log plot, - where the shape is apparently simple, i.e. a power law followed by a sharp cut-off indeed, and on a semi-log plot in Fig. 3, where the shape is "more trivial". On a semi-log plot, Eq.(3) with $\chi \leq 0$, gives a flat N-shape "noid" function (which could be called a "reverse sigmoidal") near its inflection point, which with the correspondingly flat S-shape, but nevertheless called "sigmoid" function, allows to cover various convex and concave data display shapes[1].

[1] Recall that these functions/shapes are found in laboratory when measuring the (I,V) characteristics of junctions or diodes; they present an *N* or *S* shape, beside the Ohm law. The sigmoid or noid form are also describing speculator's different strategies on the stock market [11] .

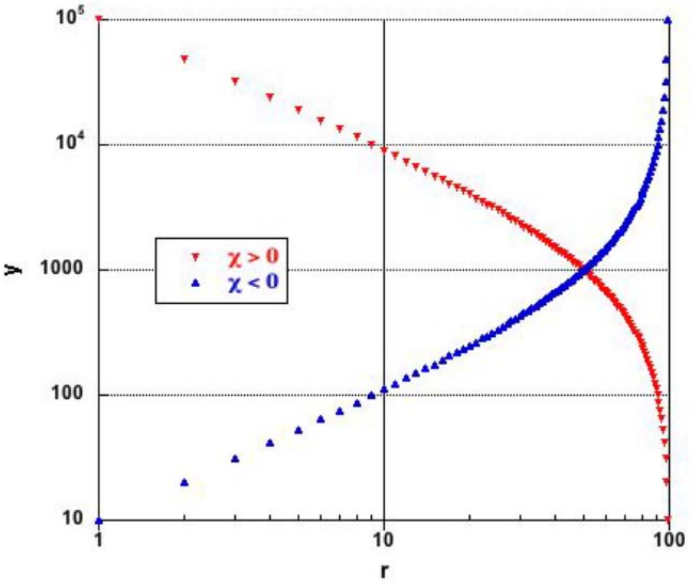


Figure 2: Lavalette function, Eq.(4) with either χ >0 (red dots) or < 0 (blue dots) on a log-log plot, for N = 100 and $\hat{K}$ = 10^6

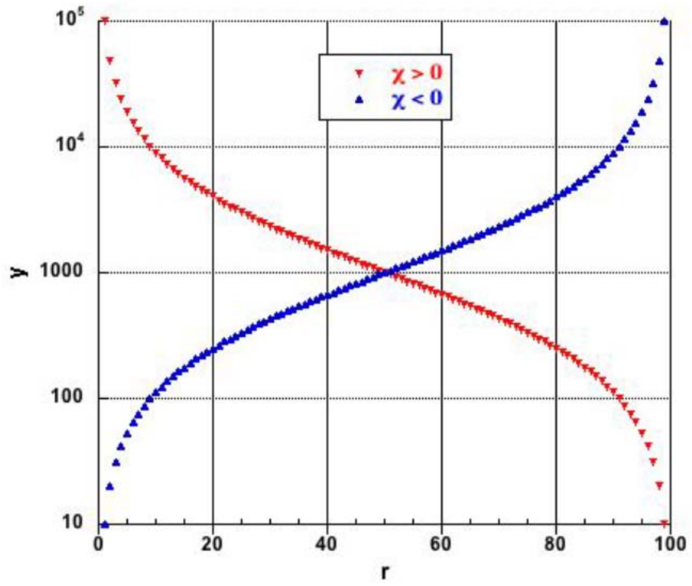


Figure 3: Lavalette function, Eq.(4) with either χ >0 (red dots) or < 0 (blue dots) on a semi-log plot, for N = 100 and $\hat{K}$ = 10^6, emphasizing the inflection points at $r = N/2$

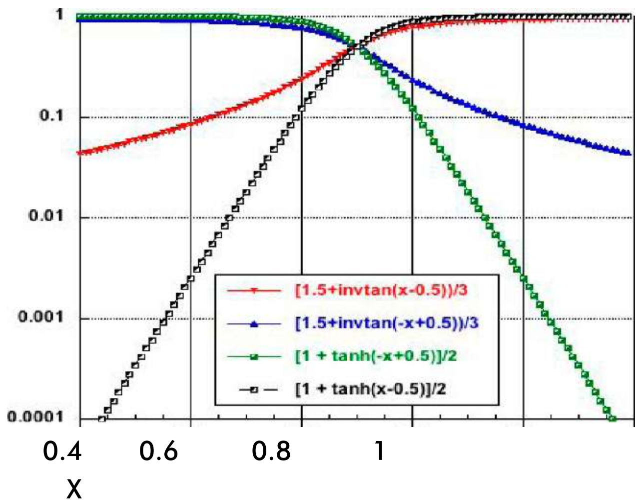


Figure 4: Display of types of sigmoid functions (*invtan(x)* and *tanh(x)*) on semi-log axes

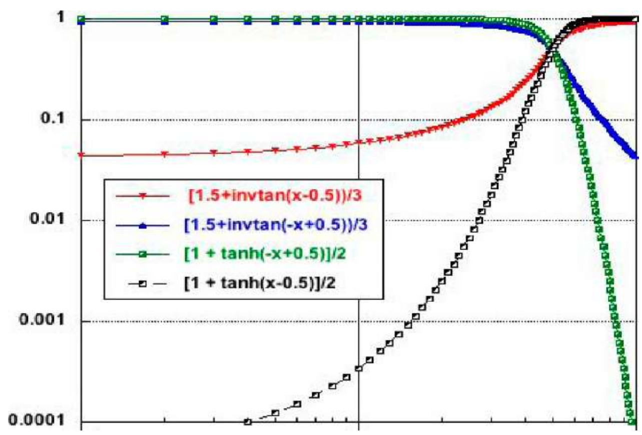


Figure 5: Display of types of sigmoid functions (*invtan(x)* and *tanh(x)*) on log-log axes

No need to recall that other often seen (or used) 2-parameter free (amplitude and slope at inflection point) have a sigmoid shape; they are *tanh(γx)* and and *invtan(γx)*. There is of course no need to represent such well known functions on classical graphs. They are rarely seen, thus shown on semi-log and log-log plots in Figs. 4-5 respectively. The functions have been adapted and scaled in order to read them on appropriate graphs, for comparison with other functions[2]. A technical point is in order here. Note that *N* (as a factor of *r*, e.g. in Eqs.(3-4) is not really needed. In fact it is more usefully replaced, at fit time, by some simple factor having the order of magnitude of *y(N/2)*. This was made in Fig.6, for example. The Aggregated Income Tax of the 43 cities in the province of Agrigento (AG) in Italy was ranked in decreasing order, for each available year in [2007-2011], from the Italian Minister of Economy, and fitted by an adapted simple Lavalette law, i.e. $\kappa 10^7\ [r/(43 - r\ +\ 1)]^{-\chi}$. Note the high regression coefficient values, but a not so visually pleasing fit at high rank.

Finally, considering cut-offs at high rank, there is on the contrary not much discussion in the literature on the wide flattening of the data at high rank, - although such

[2] It should be obvious to the reader that all these *S* or *N* shape functions can occur on different types of plots. The question is whether it can be trivially made *x* → *log*(*x*), whether this "transformation" has any impact on data analysis, and whether some theoretical hypothesis can sustain/justify such a transformation.

cases are encountered, e.g. in co-author ranking [12, 14, 15, 16, 17], and in other "very long flat tail" cases.

For completeness, other 2-parameter free simple functions are recalled in Sect. 5.

2. A FEW 3-PARAMETER FREE FUNCTIONS

Having, introduced well known 2-parameter free functions, to represent complicated data, let us turn on functions with 3 (or more, see below) free parameters, toward elaborating an attempt on how to take into account deviations from simple data approximations by power law-like lines.

2.1. Logistic or Verhulst function

For completeness, recall that the 3-parameter (σ, y_M, and $r_M/_2$) sigmoid forms are well represented through the usually called Verhulst logisitic [18]

$$y(r) = \frac{y_M}{1+e^{-\sigma*(r-\frac{r_M}{2})}} \qquad (7)$$

based on the exponential (growth) function, but invented for limiting the maximum value which such a growth function can reach. This well known function does not need to be shown on an ordinary scale graph. The function is topologically similar to tang(γ x) and invtan(γ). However it is unusual to see this sigmoid function represented or n log-log plot or on semilog plots, whence this is shown in Fig. 7 and Fig. 8, for different σ^* values (with r_M =100), pointing to non-trivial shapes, -also different to those on Fig. 4-5, as the reader can usefully observe by him/ herself.

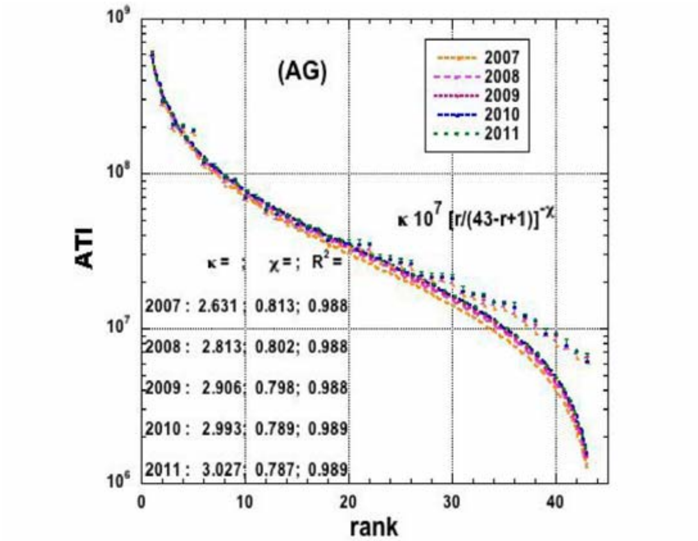


Figure 6: Basic 2-parameter free fit Lavalette law to the Aggregated Income Tax (ATI) of the *N*=43 cities, ranked in decreasing ATI order, in the province of Agrigento, IT, for recent years. Note the high regression coefficient, but not the visually pleasing fit at high rank ($r \geq$ 22)

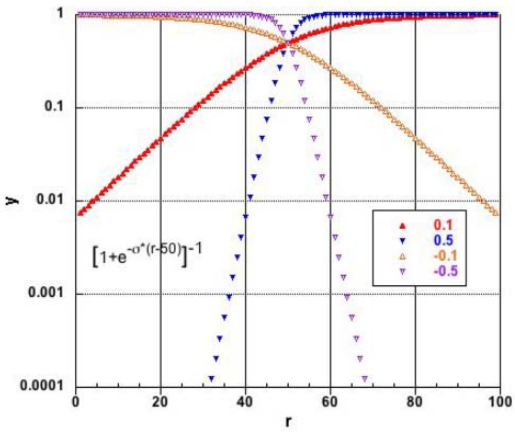


Figure 7: Logistic function, Eq.(7), on semi-log axes

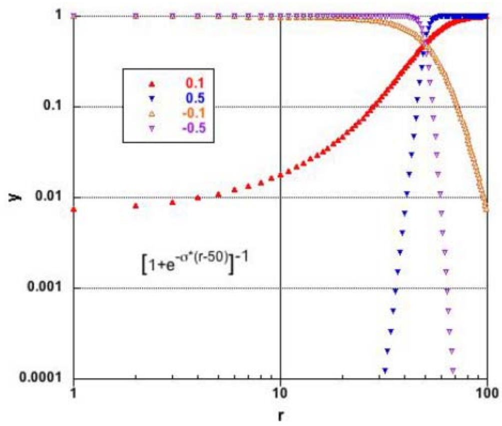


Figure 8: Logistic function, Eq.(7), on log-log axes similar to *tanh(γx)* and *invtan(γx)*. However, it is unusual to see this sigmoid function represented on a log-log plot or on a semi-log plots; whence this is shown, in Fig. 7 and Fig.8, for different σ^* values (with r_M=100), pointing to non trivial shapes, - also different from those on Figs. 4-5, as the reader can usefully observe by himself..

Interestingly, and "obviously", it can noted that some data which could be represented by the Verhulst logisitic, Eq.(7), can be transformed through a simple combination, $[y(r)/(y_M - y(r))]$, into some *Y(r)* which is ≡e^(-σ*(r-r_M/2)). Therefore a semi-log plot of *Y(r)* vs. *r* expectedly leads to a graph with a straight slope from which parameters can be easily deduced [19]; practically, y_M can be used as an appropriate input parameter to optimize the fit.

2.2. Zipf-Mandelbrot function

When the data upsurges at low rank ($r \sim 1$), on a log-log plot, as in [20], one mentions a "king effect" [21], apparently first emphasized in city population size distributions [20]. When the data flattens, below the expected straight line, at low r values, when a so called "queen effect" occurs [12], it is best to modify Eq.(1) into a 3-parameter free form, called the Zipf-Mandelbrot-Pareto (ZMP) law [13], which reads

$$y(r) = \hat{c}/(\eta + r)^{\zeta} \equiv [c/(\eta + r)]^{\zeta}, \quad (8)$$

since obviously *y(0)* takes a finite value. The value η is understood as a measure of the "harem" [14], - as seen in co-authors of papers distributions.

2.3. Generalized 2-exponent Lavalette function

There is no reason for which the behavior near the crushing point be of (analytically) identical type as the vertical asymptotic behavior at low rank. The basic 2-parameter Lavalette form Eq.(3) can be generalized as a 3-parameter free form [22]

- e.g. allowing two exponents (χ and ξ):

$$y_N(r) = \kappa \frac{(N\,r)^{-\chi}}{(N-r+1)^{-\xi}} \quad (9)$$

which is emphasizing the number of data points as in Eq.(3), but can be simply written

$$y(r) = \Lambda \frac{[r]^{-\phi}}{[N-r+1]^{-\psi}} \equiv \Lambda [r]^{-\phi} [N+1-r]^{+\psi} \quad (10)$$

$$\equiv \Lambda [N+1]^{\psi-\phi} \left[\frac{r}{N+1}\right]^{-\phi} \left[1 - \frac{r}{N+1}\right]^{+\psi} \quad (11)$$

$$\equiv \hat{\Lambda} u^{-\phi} (1-u)^{+\psi} \quad (12)$$

In fact, the case $\phi > 0$ and $\psi < 0$ is the Feller-Pareto function. The case $\phi = -1$ and $\psi = +1$ is the Verhulst function introduced in the right hand side of the (logistic) evolution differential equation.

However, interestingly, in Eq.(10), both exponents, among the 3-parameters, can take several signs, whence graphical forms can be quite different, as seen in Figs.9-11 shown on the three types of plots.

- but also admitting the same exponent χ, on both tails, but changing the range, leaving free N_1 instead of imposing a predetermined ($N + 1$), - of course imposing $N_1 - r > 0$, i.e.

$$y_N(r) = \kappa \left[\frac{N\,r}{N_1 - r}\right]^{-\chi} \equiv k [N\,r]^{-\chi} [N_1 - r]^{+\chi}, \quad (13)$$

thus somewhat in the sense of Mandelbrot modification of Zipf law, but at high rank here. In analogy with the theory of critical phenomena [1], one would consider N_1 as the "critical range", - analogous to a "critical temperature". One variant of Eq.(13) is merely equivalent to a simple redefinition of κ: $\hat{k} \equiv \kappa N^{-\chi}$. Note again that the role of *N* as a factor of *r* makes "no practical sense". Technically, for optimizing the data fits, it is better to scale the right hand side of such relations, e.g., by a factor 10^m, m obtained, in terms of the order of magnitude of *y*.

3. GENERALIZED 4-PARAMETER FREE LAVALETTE FUNCTION

The modification made in Eq.(13) suggests to apply the Mandelbrot modification also at low rank, in Eq.(9), when there is some flattening of the data at low rank, i.e., one introduces the a similar ZMP trick, as in Eq.(8) on Lavalette function. such that

- combining Eq.(8) idea with the form of Eq.(3), (note that it is different from Eq.(13)), - here keeping the same "names" for the parameters:

$$y_N(r) = k\frac{N^{-\chi}(m+r)^{-\chi}}{(N-r+1)^{-\xi}} \equiv \hat{k}[m+r]^{-\chi}[N-r+1]^{+\xi} \qquad (14)$$

- another 4-parameter free generalized Lavalette function would be

$$y_N(r) = k\frac{N^{-\chi}(r)^{-\chi}}{(N-r+m)^{-\xi}} \equiv \hat{k}[r]^{-\chi}[N-r+m]^{+\xi} \qquad (15)$$

- still a 4-parameter free generalized Lavalette function would be

$$y_N(r) = k\frac{N^{-\chi}(m+r)^{-\chi}}{(N-r+m)^{-\xi}} \equiv \hat{k}[m+r]^{-\chi}[N-r+m]^{+\xi} \qquad (16)$$

Data 23basicFellerPareto

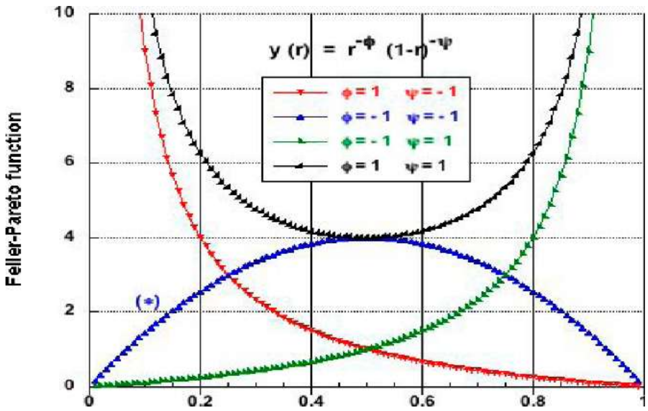


Figure 9: Feller-Pareto function, *y(r)* = $r^{\phi}(1-r)^{-\psi}$, but extended to allow different signs (and possible values) for ϕ and ψ; for readability the amplitude of the ϕ = -1 and ψ = +1 case has been multiplied by a factor 16 as pointed out by (*).

These differ from a generalization [23, 24] based on a Zipf-Mandlebrot function.

4. GENERALIZED 5-PARAMETER FREE LAVALETTE FUNCTION

"Finally", and rather generally a 5-parameter free function is "obviously" in order:

$$y_N(r) = k\frac{N(m+r)^{-\chi}}{(N-r+n)^{-\xi}} \equiv \hat{k}[m+r]^{-\chi}[N+n-r]^{-+\xi} \qquad (17)$$

No graph illustrates this super-generalization; a simple combinatory calculation indicates that one would ask for ten of them. It is better to suggest to envisage such a form when those with a lower number of free parameters do not lead to satisfactory or successful fits. It seems that one can rather easily understand the effect of the new parameters when examining the functions.

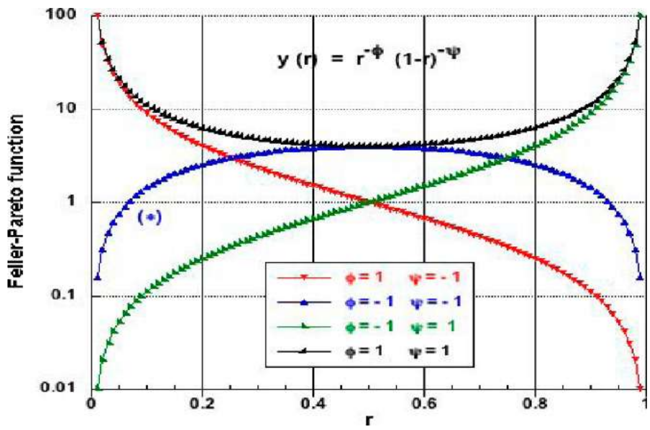


Figure 10: Feller-Pareto function, *y(r)* = $r^{\phi}(1-r)^{-\psi}$, on a semi-log plot, but extended to allow different signs (and possible values) for ϕ and ψ; for readability the amplitude of the ϕ = -1 and ψ = +1 case has been multiplied by a factor 16 as pointed out by (*)

5. A FEW OTHER FORMULAE FOR FITS

For completeness, recall a few other often used formulae for fitting data (often) on log-log plots.

5.1. 2 parameters

Beside the power law, Eq.(1) and the basic 2-parameter Lavalette form Eq.(3), one should mention

- the (2 parameter) exponential case

$$y(r) = b\, e^{-\beta r} \tag{18}$$

- a law suggested by Tsallis and de Albuquerque[3] (for ranking paper citations) [25]

$$y(r) = \frac{\phi}{[1+(\psi'-1]\ln(r))^{\psi}} \tag{19}$$

with $\psi' \equiv \psi$, although there does not seem any reason why it should be so.

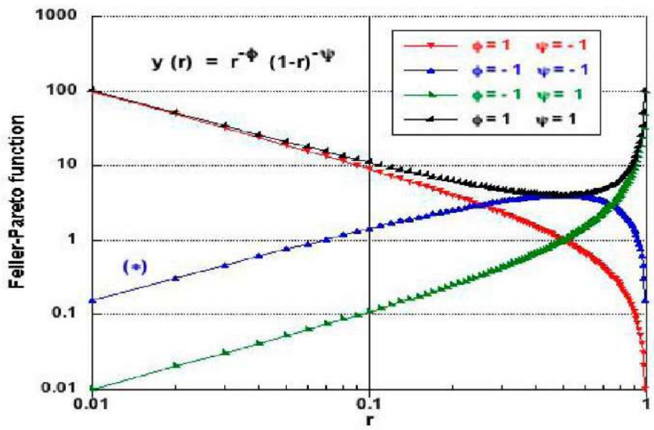


Figure 11: Feller-Pareto function, *y(r)* = $r^{\phi}(1-r)^{-\psi}$ on a log-log plot, but extended to allow different signs (and possible values) for ϕ and ψ; for readability the amplitude of the ϕ = -1 and ψ = +1 case has been multiplied by a factor 16 as pointed out by (*)

[3] correcting a misprint in [23].

- the log-normal distribution [26],

$$y(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right) \quad (20)$$

where $x > 0$, μ and σ are the parameters, mean and standard deviation of the log of "variable" in the data distribution.

5.2. 3 parameters

Beside the Verhulst logistic form, Eq.(7) and the Zipf-Mandelbrot-Pareto (ZMP) law [13], Eq.(8), other often used 3-parameter statistical distributions, generalizing the power and/or exponential law are to be examined :

- the Yule-Simon distribution, i.e. a power law with exponential cut-off [27] (the free parameters are: d, α, and λ)

$$y(r) = d\, r^{-\alpha} e^{-\lambda r} \quad (21)$$

- the stretched exponential [21] (the free parameters are: θ, μ and ν)

$$y(r) = \theta\, r^{\mu-1} e^{-\nu\, r^{\mu}} \quad (22)$$

- the Gompertz double exponential [28] (the free parameters are: g_1, r_2, and g_3)

$$y(r) = g_1 e^{-e^{-(r-r_2)/g_3}} \quad (23)$$

These functions also bend in convex form on a log-log plot.

5.3. 4 parameters

There are several possible generalizations of the above, often introducing the Mandelbrot trick, at low rank, i.e. $r \rightarrow r + \rho$, with a possibly different ρ at high and low ranks, but they do not seem of major interest. Indeed, look at

- a ZMP4 form, e.g.,

$$y(r) = m_3/(m_2 + m_4\, r)]^{\zeta}, \quad (24)$$

which obviously reduces to Eq.(8) by a trivial change in the parameter notations, e.g. $\widehat{m_3} \rightarrow m_3/m_4^{\zeta} \equiv c$, and $m_2/m_4 \equiv \eta$,

- or

$$y(r) = m_3(r - m_4)^{-m_1} e^{-m_2(r-m_4)} \quad (25)$$

with $m_4 \equiv$ to some r_0, which it is nothing else that

$$y(r) = \widehat{m_3}(r - m_4)^{-m_1} e^{-m_2 r} \quad (26)$$

Usually such functions reproduce one tail but not the other. Technically, such improvements do not change in a dramatic way the regression coefficient, since the high rank tail does not have a great impact upon this coefficient, - because of the change in the order of magnitude between the low and high rank regions.

6. HYPERGENERALIZED (LAVALETTE) FIT FUNCTIONS

It might be reminded that the modification of Keynes differential growth equation by Verhulst through a *(1 - x)* term was purely a mathematical *ad hoc* mean to avoid a full exponential growth. There is no economic or demographic argument to use a linear *(1 - x)* term; a quadratic term $(1 - x^2)$ or any other polynomial decaying near $x = 1$ or many more complicated terms could be used. Therefore,considering that the basic phenomena might not necessarily depend linearly on *r*, but the rank-size rule should (or could) contain higher order terms, other generalizations may come in mind within the present considerations. One such a case was found in considering city sizes (in Bulgaria, e.g. [29, 30]), but might occur more frequently than "expected", - however are likely not reported because of missing framework. Therefore, a hypergeneralization of Lavalette function can be imagined:

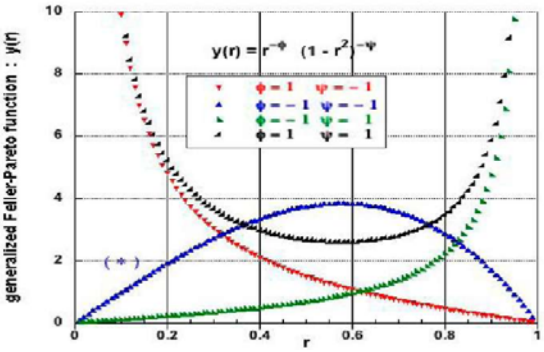


Figure 12: Hypergeneralized Feller-Pareto function, y(r) = $r^{\phi}(1-r^2)^{-\psi}$, on ordinary axes; (*) indicates that the function has been multiplied by a factor 16 for better readability

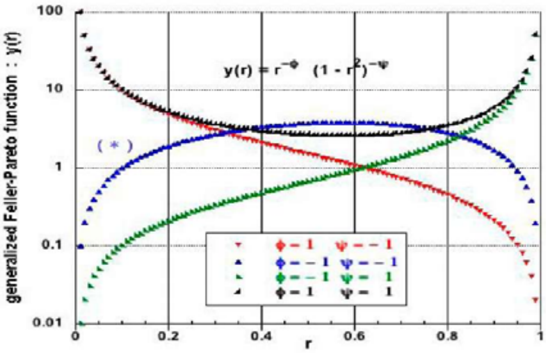


Figure 13: Hypergeneralized Feller-Pareto function, y(r) = $r^{\phi}(1-r^2)^{-\psi}$, on a semi-log plot; (*) indicates that the function has been multiplied by a factor 16 for better readability

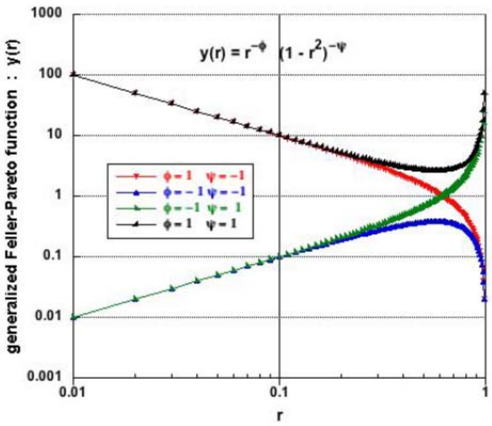


Figure 14: Hypergeneralized Feller-Pareto function, y(r) =$r^{\phi}(1-r^2)^{-\psi}$, on a log-log plot

• the 3-parameter generalized Lavalette form [22] can be hypergeneralized into

$$y(r) = \frac{[\Lambda\, r^n]^{-\phi}}{[N+1-\ r^m]^{-\psi}} \quad or \quad = \Lambda \frac{[r^n]^{-\phi}}{[N+1-\ r^m]^{-\phi}} \tag{27}$$

• the 4-parameter generalized Lavalette form [24] can be hypergeneralized into

$$y(r) = \frac{[\Gamma\,(r^n+\nu)]^{-\eta}}{[N-r^m+\nu]^{-\zeta}} \quad or \quad = \Gamma \frac{[r^n+\nu]^{-\eta}}{[N-\ r^m+\nu]^{-\zeta}} \tag{28}$$

• the 5-parameter supergeneralized Lavalette form (also) can be hypergen - eralized into

$$y(r) = \frac{[\Gamma\,(r^n+\mu)]^{-\eta}}{[N-r^m+\nu]^{-\zeta}} \quad or \quad = \Gamma \frac{[(r^n+\mu)]^{-\eta}}{[N-\ r^m+\nu]^{-\zeta}} \tag{29}$$

Note that variants : $[(r^n + \nu)] \rightarrow [(r + \nu)^n]$, and $[(r^n + \mu)] \rightarrow [(r+\mu)^n]$, with or without $[(r^m - \nu)] \rightarrow [(r - \nu)^m]$, can be written. The writing choice is left for fit optimization

7. ON INFLECTION POINTS ON LOG-LOG PLOTS

Finally, not the least, the above formulae have much emphasized possible fits which indeed allow inflection points on semi-log graphs, but have left opened the case of inflection points on log-log graphs. Let it be understood that such a case occurs when some power law decay ("from infinity") at low rank is followed by another intermediary regime before some cut-off occurs at high rank. A trivial transformation x → *log(x)* of all the above formulae is possible, but demands much reflection. Indeed, one could transform the basic Lavalette equation to read

$$y(r) \simeq \left[\frac{N\log(r)}{N+1-\log(r)}\right]^{-\chi} \tag{30}$$

and similarly all others. But it remains to be done some interpretation and much theoretical work !

Another possibility comes from realizing that if there is an inflection point, the slope has the same (negative) sign for the whole *r* range, but the derivative of the slope has some structure, i.e.allowing for a concave to a convex shape of the approximation to the data. The intermediary regime can also be considered in a first approximation to be a scaling law. The high rank regime can be either a Lavalette cut-off or an exponential cut-off. Therefore the following functions can be appropriately imagined

- in its most generalized form, with power law cut-off

$$y(x) = [A(x+m_5)^{-m_1} + B(x+m_6)^{-m_2}](N+m_4-x^{m_7})^{m_3} \tag{31}$$

- or with an exponential cut-off

$$y(x) = [A(x+m_5)^{-m_1} + B(x+m_6)^{-m_2}]e^{-m_3(x+m_4^{m_7}} \tag{32}$$

A few of such cases are shown in Figs.21-22 demonstrating the interest of such forms in order to discuss inflection points on log-log plots.

8. APPLICATIONS

This section serves as an illustration of a few cases discussed above, displaying some data on either semi-log or log-log plots for comparison. However the data pertains to some empirical study requesting a brief introduction. In so doing, it is hoped that the "universality" of the approach receives a positive argument.

Consider the following investigation. In Italy, 638 cities contain a saint or an angel name, as counted after translating the names into italian, from french, german, or local dialects (like Santu Lussurgiu = Santo Lussorio, or Santhia who is Santa Agata), Note that Sant'Angelo (24 times), San Salvatore (5 times) or Santa Croce (7 cities), and similar "concepts" (Sansepolcro) are not counted. Some distinction can be made between male and female saints. Note that two cities have a name with two saints. The name of the saints can be ranked according to their frequency [31] and an appropriate statistical analysis can follow for the rank-frequency distribution.

However, one can also ask, as did Pareto in 1896, how many times one can find an "event" greater than some size *n*, i.e. study the *size-frequency relationship*. Pareto found out that the cumulative distribution function (CDF) of such events follows an inverse power of *n*, or in other words, *P [N > n]* $\sim n^{-\omega}$., - whence the frequency *f* of such events of size *n*, (also) follows an inverse power of *n*.

Thus, one can count how many cities have a happax hagionym, how many cities have a name with a saint occurring only twice, etc. up to how many cities have a name associated to the "most popular" (= most frequent) saint (San Pietro). This counting is normalized and turned into a probability distribution, i.e. *CDF(n)*. The data is illustrated in Figs. 15-20, either with semi-log or log-log plots, and fits with a Zipf-Mandelbrot or Lavalette function.

Short final comments: (i) two "queen effects" and a "king effect" are well seen on Fig. 16; (ii) the *CDF* shows a pronounced cut-off at high *n* in all cases. Therefore, it could be argued that the *CDF* is less pertinent to observe minute effects. This is understandably true, since the *CDF* results from an integration scheme. However, again understandably, the *CDF* fits are much more stable. No need to say that one should not report too precise parameter values, since these are non linear fits; a final technical information: the Levenberg-Marquardt algorithm was used.

9. CONCLUSIONS

It has been shown that semi-log plots are of interest in order to analyze whether experimental or empirical data are underlined by some scaling argument for the observed/examined phenomenon at hands. The fit to a straight line on log-log plots is not always satisfactory indeed. Deviations occur at low, intermediate and high regimes along the x-axis. Several improvements of the mere power law fit have been discussed, in particular through a Mandelbrot trick at low rank and a Lavalette power law cut-off at high rank.

In so doing, the number of free parameters increases. Their meaning has been discussed, up to the 5 parameter free super-generalized Lavalette law and the 7-parameter free hyper-generealized Lavalette law[4] . It has been emphasized that the interest of the basic 2-parameter free Lavalette law and the subsequent generalizations resides in its "noid" (or sigmoid, depending on the sign of the exponents) form on a semi-log plot; something incapable to be found in other empirical law, like the Zipf-Pareto-Mandelbrot law. The connection with other laws, e.g. Feller-Pareto and Verhulst logistic laws, as been pointed out.

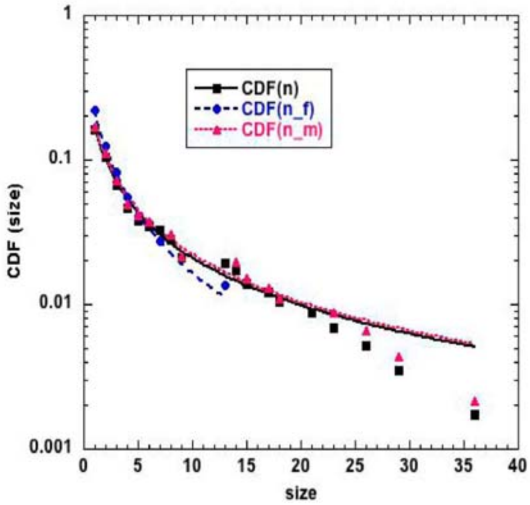


Figure 15: Semi-log plot of the cumulative distribution function (CDF) of the frequency of Italian cities containing a saint name n-times, so called "size", given according to the Zipf-Mandelbrot-Pareto function, like Eq.(8), distinguishing between male (n_m) and female (n_f) saint names; the fit parameter values are given in Fig.16. Observe the need for a cut-off at high rank/size.

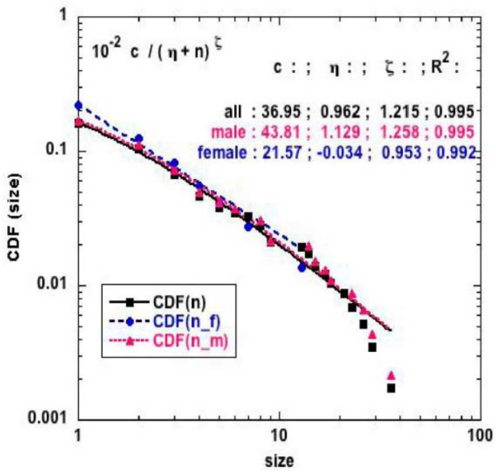


Figure 16: Log-log plot of the cumulative distribution function (CDF) of the frequency of Italian cities containing a saint name n-times given according to the Zipf-Mandelbrot-Pareto function, like Eq.(8), distinguishing between male (n_m) and female (n_f) saint names; observe that η is negative for the female case, pointing to a king effect (Santa Maria), and queen effects, since $\eta \geq 0$, for the males and the overall distribution. Observe the need for a cut-off at high rank/size.

[4]In this conclusion, one could recall that 6 or 7-parameter free functions are also used for fitting data like in financial market crash predictions [32, 33, 34, 35, 36] or in earthquake predictions [37]

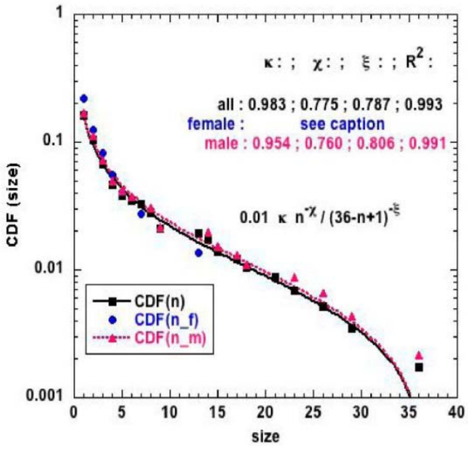


Figure 17: Semi-log plot of the cumulative distribution function (CDF) of the frequency of Italian cities containing a saint name, given *n*-times, so called "size"; fit according to a Lavalette function with 3 free parameters, Eq.(10), for the distribution of all such 36 cities(black line) or only those 36 with a male saint name (n_m; red line); the parameter values for the female case are given in Fig.18, with the corresponding fit. Observe the interest of leaving the high rank/size value be a free parameter, as on Fig.19

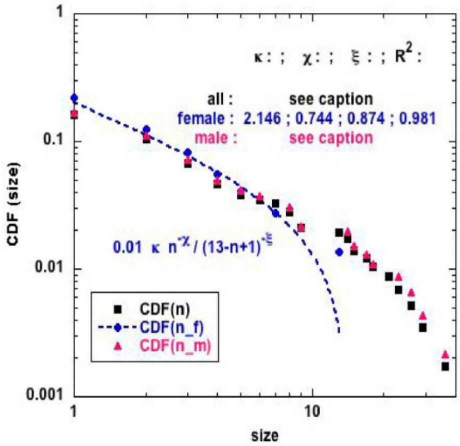


Figure 18: Log-log plot of the cumulative distribution function (CDF) of the frequency of Italian cities containing a saint name given *n*-times, so called "size"; fit according to a Lavalette function with 3 free parameters, Eq.(10) is shown for the distribution of only those 13 cities with a female saint name (n_f; blue line); the parameter values for the male case and the whole distribution are given in Fig.17, with the corresponding fits. Observe the interest of leaving the high rank/size value be a free parameter, as on Fig.20.

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

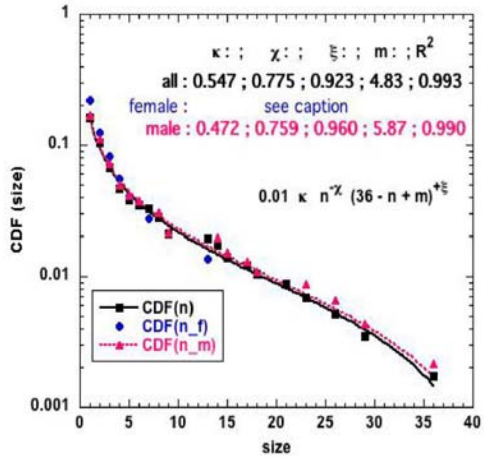


Figure 19: Semi-log plot of the cumulative distribution function (CDF) of the frequency of Italian cities containing a saint name given *n*-times, so called "size"; fits with a 4 parameter free Lavalette function, Eq.(15) are shown for the distribution of all such 36 cities (black line) or only those 36 with a male saint name (n_m; red line); the parameter values for the female case are given in Fig.20, with the corresponding fit.

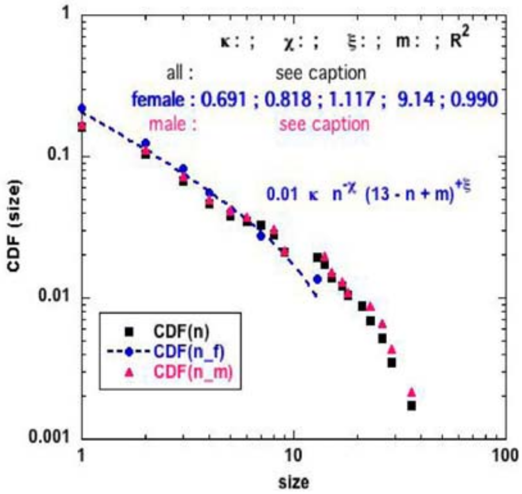


Figure 20: Log-log plot of the cumulative distribution function (CDF) of the frequency of Italian cities containing a saint name given *n*-times, so called "size"; fit according to a Lavalette function with 4 free parameters, Eq.(15) shown for the distribution of only those 13 cities with a female saint name (n_f; blue line); the parameter values for the male case and the whole distribution are given in Fig.19, with the corresponding fits.

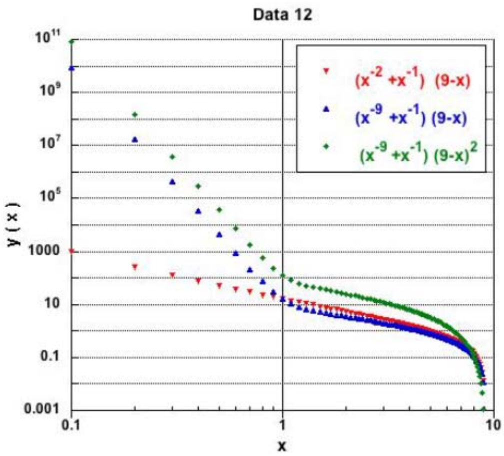


Figure 21: Display of a "simple" function with inflection point on a log-log plot, allowing for fit to data with large king or queen effect and power law cut-off, i.e. with an inflection point in the middle range, as approximated by a simple function for which the general form is Eq.(31).

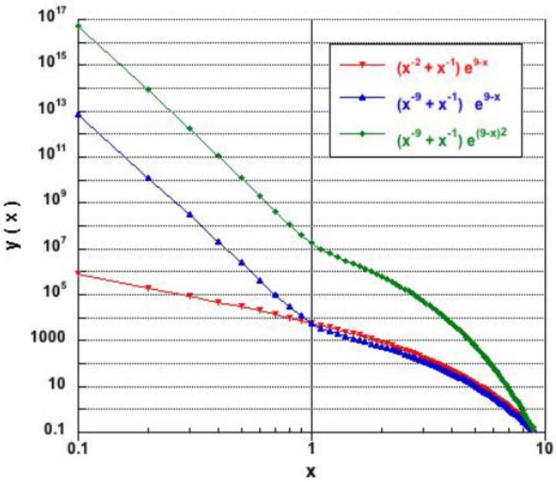


Figure 22: Display of a "simple" function with inflection point on a log-log plot, allowing for fit to data with large king or queen effect and exponential cut-off, i.e. with an inflection point in the middle range, as approximated by a simple function for which the general form is Eq.(32).

It has been shown that the additional parameters introduced into the basic Lavalette function, Eq.(3), facilitates a rather good reproduction of rank-probability distribution in the ranges of small and high rank values. Indeed, each parameter or ratio involved in the suggested modification of Lavalette function, Eq.(3), enhances the fit in different ranges of *r*.

It has remained for completeness to invent a simple law showing an inflection point on a log-log plot. Such a law could have been the result of a transformation of the Lavalette law through $x \to log(x)$, but this meaning is theoretically unclear. It has been shown that a simple linear combination of two basic Lavalette law provides the requested features.

Generalizations taking into account two super-generalized or hyper-generalized Lavalette laws are suggested, but need to be fully considered at fit time on appropriate data.

A few examples are used for illustrating various points, like deviations or visually unattractive fits, - though the regression coefficient R^2 is often quite satisfactory looking. Examples have been taken mainly for rank-size rule research. However, in order to demonstrate a larger validity of generalizing the usual fit formulae, and some interest for generalizing the basic concepts, some short analysis has been presented of the cumulative distribution function (CDF) of the city names in Italy containing a (male or female) saint name.

REFERENCES

1. H. E. Stanley, **Phase Transitions and Critical Phenomena**, Oxford Univ. Press, Oxford, 1971
2. B.B.Mandelbrot, **The Fractal Geometry of Nature**, W.H.Freeman, New York, 1982
3. Zipf, G.K. **Human Behavior and the Principle of Least Effort : An Introduction to Human Ecology**. Cambridge, Mass.: Addison Wesley, 1949.
4. X. Gabaix, **Zipf's law for cities: An explanation**, The Quarterly Journal of Economics **114,** 739-767, 1999.
5. G. Brakman, H. Garretsen, C. van Marrewijk, and M. van den Berg, **The Return of Zipf: Towards a Further Understanding of the Rank-Size Distribution**, Journal of Regional Science **39,** 182-213, 1999.
6. D.R. Vining, Jr., **The rank-size rule in the absence of growth**, Journal of Urban Economics **4,** 15-29,1977.
7. J.-C. Cordoba, **On the distribution of budget sizes**, Journal of Urban Economics **63,** 177-197, 2008.
8. Yanguang Chen, **The rank-size scaling law and entropy-maximizing principle**, Physica A **391,** 767-778, 2012.
9. Michel L. Goldstein, Steven A. Morris, and Gary G. Yen, **Problems with Fitting to the Power-Law Distribution**, *0402322f ittingtopowerlaw_v3?*
10. D. Lavalette, **Facteur d'impact: impartialite ou impuissance?**, Internal Report, INSERM U350, Institut Curie, France, Nov. 1966
11. M. Ausloos, **Gas-kinetic theory and Boltzmann equation of share price within an equilibrium market hypothesis and ad hoc strategy**, Physica A 284, 385-392, 2000.
12. Ausloos, M. **A scientometrics law about co-authors and their ranking**. The co-author. *Scientometrics,* 95(3), 895-909, 2013.

13. R. A. Fairthorne, **Empirical hyperbolic distributions (Bradford-Zipf- Mandelbrot) for bibliometric description and prediction**, Journal of Documentation 25, 319-343, 1969.

14. Ausloos, M. **Binary Scientific Star Coauthors Core Size**. Scientometrics, in press, 2014

15. Bougrine, H. **Subfield Effects on the Core of Coauthors**. Scientometrics, 98(2), 1047-1064, 2014.

16. Miskiewicz, J. **Effects of Publications in Proceedings on the Measure of the Core Size of Coauthors**. Physica A, 392(20), 5119-5131, 2013.

17. G. Rotundo, **Black-Scholes-Schrdinger-Zipf-Mandelbrot model framework for improving a study of the coauthor core score**, Physica A, in press DOI information: 10.1016/j.physa.2014.02.011

18. Verhulst, P.F., **Recherches mathematiques sur la loi d'accroissement de la population**, Nouveaux Memoires de l'Academie Royale des Sciences et Belles-Lettres de Bruxelles **18,** 1-38, 1845.

19. Montroll, E.W., **Social dynamics and the quantifying of social forces**, Proc. Nat. Acad. Sci. USA **75**, 4633-4637, 1978.

20. Jefferson, M. **The law of primate city**. Geographical Review, 29(2), 226-232, 1939.

21. Laherrere, J. & Sornette, D. **Stretched exponential distributions in nature and economy fat tails with characteristic scales**. European Physics Journal B, 2(4), 525-539, 1998.

22. R. Mansilla, E. Koppen, G. Cocho, P. Miramontes**, On the behavior of journal impact factor rank-order distribution**, Journal of Informetrics **1,** 155-160, 2007.

23. I. Popescu, **On a Zipf's Law Extension to Impact Factors**, Glottometrics **6,** 83-93, 2003.

24. I.A. Voloshynovska, **Characteristic Features of Rank-Probability Word Distribution in Scientific and Belletristic Literature**, Journal of Quantitative Linguistics **18,** 274-289, 2011.

25. C. Tsallis, M.P. de Albuquerque, **Are citations of scientific papers a case of nonextensivity?**, Eur. Phys. J. B **13,** 777-780, 2000.

26. E.W. Montroll, M.F. Shlesinger, **Maximum entropy formalism, fractals, scaling phenomena, and 1/f noise: a tale of tails**, J. Stat. Phys. 32, 209-230, 1983.

27. C. Rose, D. Murray, D. Smith, **Mathematical Statistics with Mathematica**, Springer, New York, p. 107, 2002.

28. Gompertz, R., **On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies**, Philos. Trans. R. Soc. London, **115**, 513-585, 1825.

29. Dimitrova, Zlatinka and Ausloos, Marcel. **Primacy analysis of the system of Bulgarian cities**, arXiv preprint: arXiv: 1309.0079, 2013

30. Nikolay K. Vitanov, Zlatinka I. Dimitrova, Marcel Ausloos, **City population sizes and power laws. The case of Bulgarian cities**, submitted

31. Marcel Ausloos and Roy Cerqueti, **Local and regional disparities of urban hagiotoponyms in Italy**, in preparation

32. D. Sornette, A. Johansen, and J.P. Bouchaud, **Stock Market Crashes, Precursors and Replicas**, J. Phys. I (France) 6, 167-175, 1996.

33. N. Vandewalle and M. Ausloos, **How the financial crash of October 1997 could have been predicted**, Eur. J. Phys. B 4, 139-141, 1998.

34. A. Johansen and D. Sornette, **The NASDAQ crash of April 2000: Yet another example of log-periodicity in a speculative bubble ending in a crash**, Eur. J. Phys. B 17, 319-328, 2000.

35. M. Ausloos, K. Ivanova, and N. Vandewalle, **Crashes: symptoms, diagnoses and remedies, in Empirical sciences of financial fluctuations. The advent of econophysics**, H. Takayasu, Ed. Springer Verlag, Berlin, pp. 62-76, 2002.

36. J. Kwapien and S. Drozdz, **Physical approach to complex systems**, Physics Reports 515, 115-226, 2012.

37. A. Johansen, D. Sornette, H. Wakita, U. Tsunogai, W.I. Newman, and H. Saleur, **Discrete Scaling in Earthquake Precursory Phenomena: Evidence in the Kobe Earthquake, Japan**, J. Phys. I (France) 6, 1391-1402, 1996

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

A SVAR ANALYSIS OF THE RELATIONSHIP BETWEEN ROMANIAN UNEMPLOYMENT RATES AND THE SIZE OF THE SHADOW ECONOMY

Adriana AnaMaria DAVIDESCU[1]
PhD.Lecturer, Department of Statistics and Econometrics
Bucharest Academy of Economic Studies

E-mail: adrianaalexandru@yahoo.com

Ion DOBRE[2]
PhD.Professor, Department of Informatics and Cybernetics
Bucharest Academy of Economic Studies

E-mail: dobrerio@ase.ro

ABSTRACT

The paper analyses the relationship between shadow economy and unemployment rates using a Structural VAR approach for quarterly data during the period 2000-2010. The size of Romanian shadow economy is estimated using the currency demand approach based on VECM models, stating that its size is decreasing over the analyzed period, from 36.5% at the end of 2000 to about 31.5% of real GDP at the middle of 2010.

The relationship between the variables is tested by imposing a long-run restriction in the Structural VAR model to analyze the impact of the shadow economy to a temporary shock in unemployment. The accumulated responses generated by a positive supply shock (unemployment rate) confirms that in the short-run, a rise in both registered and ILO unemployment rates in formal sector will lead to a decrease in the number of people who work in the shadow economy in the second quarter following the initial shock and to an smaller increase in the size of the Romanian shadow economy in the third quarter following the initial shock.

Keywords: *shadow economy, unemployment rates, Structural VAR, Romania*
JEL classification: C32, E41, O17

1. INTRODUCTION

The paper aims to investigate the relationship between the size of the shadow economy (SE) and unemployment rates for the case of Romanian data using SVAR analysis for quarterly data covering the period 2000-2010. The size of Romanian shadow economy is estimated using currency demand approach based on vector error correction models (Davidescu and Dobre (2013)).

The empirical results of currency demand approach based on VECM models emphasizes that there is a general downward trend in the size of the shadow economy as % of official GDP for the period 2000-2010 with an highlight on two low periods, 2003Q1 and 2008Q4.Thus, the size of the shadow economy as % of official GDP measures approximately

36.6% in 2000Q1 and follows a downward trend after registering the value of 31% by 2008. For the past few quarters, there is a slightly upward trend in the size of Romanian shadow economy.

The results are consistent with studies of Schneider (2007) and Albu (2007, 2010, 2011) which show a mainly downward trend of shadow economy in Romania.

It is important to note that because of its undetectable nature and character, it is nearly impossible to measure precisely the size of economic activities taking place in the informal economy of any country in the world, whether developed or less developed. Given this, any theoretical or empirical inference derived from these results should always be regarded as an approximation. In the face of these difficulties, the results drawn from these estimates should be interpreted with due reserve, given the limitations of the methods.

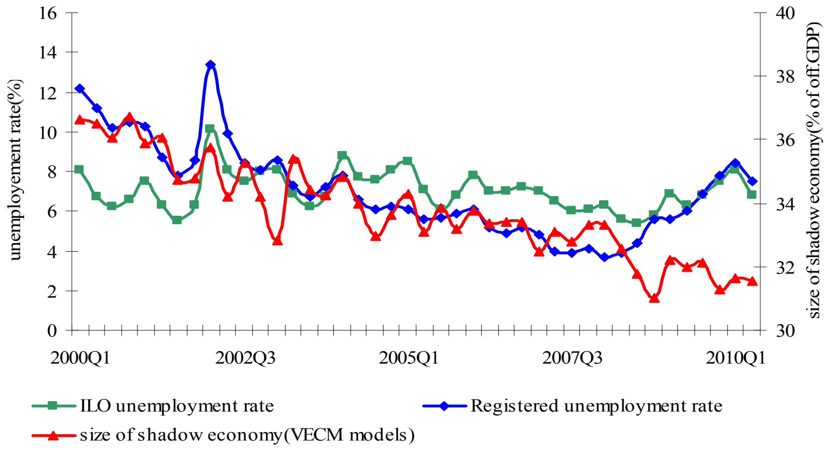
2. THE RELATIONSHIP BETWEEN UNEMPLOYMENT RATES AND SHADOW ECONOMY IN ROMANIA. A SVAR ANALYSIS

According to Giles and Tedds (2002), two opposing forces determine the relationship between unemployment and the informal economy. On the one hand, an increase in the unemployment rate may involve a decrease in the informal economy because it is positively related to the growth rate of GDP and eventually negatively correlated with unemployment (Okun's law). On the other hand, increase in unemployment leads to an increase in people working in the informal economy because they have more time for such activities.

Dell'Anno and Solomon (2007) stated that there is a positive relationship in the short-run between unemployment rate and U.S. shadow economy for the period 1970-2004. Using SVAR analysis, they investigate the response of the shadow economy to an aggregate supply shock (impact of the shadow economy to a temporary shock in unemployment). The empirical results show that in the short-run, a positive aggregate supply shock causes the shadow economy to rise by about 8% above the baseline.

Regarding the Romanian unemployment data, there are two measures available for unemployed persons: the first is the registered unemployment rate, who is calculated by National Agency for Employment (NAE) and based on statements of people who pass by employment agencies and said that they are unemployed and the ILO unemployment rate, who is published quarterly by the National Institute of Statistics and is based on labour force survey (LFS).

Fig.1. Shadow economy vs. unemployment rates in Romania



Source: Size of the shadow economy (% of official GDP); Tempo database, National Institute of Statistics, Monthly Bulletins 2000-2010, National Bank of Romania.

The graphical evolution of the shadow economy versus unemployment rates reveal the existence of a positive relationship between variables, low for the case of ILO unemployment rate, quantified by a value of about 0.22 of correlation coefficient and strong for the case of registered unemployment rate, quantified by a value of 0.67 of correlation coefficient.

The aim of the paper is to investigate the nature of the relationship between unemployment rates and the size of the Romanian shadow economy using SVAR approach.

2.1. Methodology and data

The data used in the research covers the period 2000:Q1- 2010Q2; the number of observation is 42. The variables used are as follows: the size of the Romanian shadow economy expressed as % of official GDP (SE) obtained using the VECM approach; ILO unemployment rate (ILO_UR) and registered unemployment rate(R_UR). The unemployment rates were seasonally by means of tramo seats method. The main source of the data for unemployment rates is the National Institute of Statistics (Tempo database) and the National Bank of Romania.

The SVAR approach (also called the analysis of disturbances) has been developed over the last decade to interpret business cycle fluctuations and to help identify the effects of different economic policies. It is an extension on the traditional theoretic VAR approach in that it combines economic theory with time-series analysis to determine the dynamic response of economic variables to various disturbances. The main advantage with SVAR analysis is that the necessary restrictions on the estimated reduced form model, required for identification of the underlying structural model, can be provided by economic theory. These restrictions can be either contemporaneous or long-run in nature depending on whether the underlying disturbances are considered to be temporary or permanent in nature. Once the identification is achieved it is possible to recover the structural shocks. These shocks can then be used to generate impulse response and variance decomposition functions to assess the dynamic impacts on different economic variables.

In the development of SVAR approach, the contributions of Sims (1986), Bernanke(1986) and Blanchard and Watson(1986) should be remembered, since they use the economic theory to impose restrictions on the observed values of the estimated residuals (et) to recover the underlying structural disturbances (t). Instead of the arbitrary method of restriction imposition used in traditional VARs, the SVAR approach estimates the structural parameters by imposing contemporaneous structural restrictions based on economic theory. These can be considered as short-run restrictions in that the shocks are considered to have temporary effects.

An alternative SVAR approach, advanced by Blanchard and Quah (1989), is to consider the shocks as having permanent effects. This would imply that the variables are non-stationary since the shocks continue to accumulate through time given they are permanent. The presence of unit roots in the variables can give rise to spurious regression if the VAR is estimated in levels. Therefore it is necessary to use first differences[1] to ensure stationarity in the case of shocks that have permanent effects.

Therefore, a Structural VAR is a standard VAR where the restrictions needed for identification of the underlying structural model are provided by economic theory. These can be either contemporaneous or long-run restrictions depending on whether economic theory suggests the shocks are either temporary or permanent in nature (McCoy, 1997).

We aim to investigate the existence of a structural relationship between shadow economy and unemployment rate, in order to extract information on underlying aggregate supply and demand disturbances using VAR decomposition. We recover the underlying demand and supply disturbances using the Structural Vector Autoregression technique developed by Blanchard and Quah[2] (1989).

The basic idea is that an economy is hit by two types of shocks, demand and supply shocks. Demand shocks are identified with the help of the restriction that their long-term impact on output is zero. Only supply shocks can have a permanent effect on output.

The procedure proposed by Blanchard and Quah (1989) decomposes permanent and temporary shocks to a variable using a VAR model. The structural VAR methodology with long-run restrictions proposed by Blanchard and Quah (1989) does not impose restrictions on the short-run dynamics of the permanent component of output, but incorporates a process for permanent shocks that is more general than a random walk.

Blanchard and Quah (1989) provide an alternative way to obtain a structural identification. Instead of associating each disturbance (t) directly with an individual variable, they consider the shocks as having either temporary or permanent effects. The objective is to decompose real GNP into its temporary and permanent components. Economic theory is used to associate aggregate demand shocks as being the temporary shocks and aggregate supply shocks as having permanent effects[3].

So, they develop a macroeconomic model such that real GNP is affected by demand-side and supply-side disturbances. In accordance with the above mentioned theoretical framework, the demand-side disturbances have no long run effect on real GNP. On the supply side, productivity shocks are assumed to have permanent effect on output.

---

[1] Alternatively, a cointegrated framework can be used to avoid the loss of information about the equilibrium relationships in the model that can result from first differencing. The stationary linear combinations of the non-stationary variables can be constructed prior to estimation (Keating, 1992). This cointegration constraint can then be imposed using a vector error correction model (VECM).

[2] A detailed presentation of this topic is provided in Enders, W.(1995). Applied Econometric Time Series, Wiley, New York.

[3] Long run restrictions are imposed to identify the aggregate demand and aggregate supply disturbances.

Using a bivariate VAR, Blanchard and Quah show how to decompose real GNP and recover the two pure shocks that cannot otherwise be quantified. They assume that there are two kinds of disturbances, each uncorrelated with the other and that neither has a long run effect on unemployment. They assume however, that the first has a long run effect on output while the second does not. These assumptions are sufficient to just identify the two types of disturbances and their dynamic effects on output and unemployment.

In the same manner, we consider a Vector Autoregression representation of a system composed by two variables that are the first differences of the shadow economy (SE) and unemployment rates (R_UR and ILO_UR)(we have considered both registered unemployment rate and ILO unemployment rate). The Blanchard - Quah technique requires that both variables must be stationary.

Thus, the two variables that compose VAR are:

$$\mathrm{X_t} = \begin{bmatrix} \Delta \mathrm{SE_t} \\ \Delta \mathrm{UR_t} \end{bmatrix} \quad (1)$$

The classical VAR can be writing as:

$$\Delta SE_t = b_{10} - b_{12}\Delta UR_t + \gamma_{11}^1 \Delta SE_{t-1} + \gamma_{12}^1 \Delta UR_{t-1} + ... + \gamma_{11}^p \Delta SE_{t-p} + \gamma_{12}^p \Delta UR_{t-p} + \varepsilon_{dt} \quad (2)$$

$$\Delta UR_t = b_{20} - b_{21}\Delta SE_t + \gamma_{21}^1 \Delta SE_{t-1} + \gamma_{22}^1 \Delta UR_{t-1} + ... + \gamma_{21}^p \Delta SE_{t-p} + \gamma_{22}^p \Delta UR_{t-p} + \varepsilon_{st} \quad (3)$$

We can re-write the above equations in a matrix form:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} \Delta SE_t \\ \Delta UR_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11}^1 & \gamma_{12}^1 \\ \gamma_{21}^1 & \gamma_{22}^1 \end{bmatrix} \begin{bmatrix} \Delta SE_{t-1} \\ \Delta UR_{t-1} \end{bmatrix} + ... + \begin{bmatrix} \gamma_{11}^p & \gamma_{12}^p \\ \gamma_{21}^p & \gamma_{22}^p \end{bmatrix} \begin{bmatrix} \Delta SE_{t-p} \\ \Delta UR_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{dt} \\ \varepsilon_{st} \end{bmatrix} \quad (4)$$

Furthermore, in general form it becomes:

$$\mathrm{BX_t} = \Gamma_0 + \Gamma_1 \mathrm{X_{t-1}} + ... + \Gamma_p \mathrm{X_{t-p}} + \varepsilon_t \quad (5)$$

where:

$\mathrm{X_t}$ is a vector of the two considered variables, Γ_t are the matrices of coefficients, p lags are considered and ε_t is the vector of error terms.

By multiplying with the inversion of B matrix $(1 - \mathrm{b_{12}b_{21}} \neq 0)$ we obtain:

$$\mathrm{X_t} = \mathrm{B}^{-1}\Gamma_0 + \mathrm{B}^{-1}\Gamma_1 \mathrm{X_{t-1}} + ... + \mathrm{B}^{-1}\Gamma_p \mathrm{X_{t-p}} + \mathrm{B}^{-1}\varepsilon_t \quad (6)$$

Re-writing the VAR model, we obtain:

$$\mathrm{X_t} = \mathrm{A_0} + \mathrm{A_1 X_{t-1}} + ... + \mathrm{A_p X_{t-p}} + \mathrm{e_t} \quad (7)$$

$$\mathrm{X_t} = \mathrm{A(L)LX_t} + \mathrm{e_t}$$

which can be re-written as follows:

$$\begin{bmatrix} \Delta SE_t \\ \Delta UR_t \end{bmatrix} = \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} \Delta SE_{t-1} \\ \Delta UR_t \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \quad (8)$$

where:

$\mathrm{X_t} = \begin{bmatrix} \Delta \mathrm{SE_t} \\ \Delta \mathrm{UR_t} \end{bmatrix}; e_t = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}; A(L) =$ the 2×2 matrix with elements equal to the polynomials $A_{ij}(L)$ and the coefficients of $A_{ij}(L)$ are denoted by $a_{ij}(k)$.

Since the demand-side and supply-side shocks are not observed, the problem is to recover them from VAR estimation. The residuals from this estimated VAR are composites of the structural disturbances ε_{dt} and ε_{st}.

If we ignore the intercept terms, in the particular bivariate moving average form, the VAR can be written:

$$\Delta SE_t = \sum_{k=0}^{\infty} b_{11}(k)\varepsilon_{dt-k} + \sum_{k=0}^{\infty} b_{12}(k)\varepsilon_{st-k} \tag{9}$$

$$\Delta UR_t = \sum_{k=0}^{\infty} b_{21}(k)\varepsilon_{dt-k} + \sum_{k=0}^{\infty} b_{22}(k)\varepsilon_{st-k} \tag{10}$$

or

$$\begin{bmatrix} \Delta SE_t \\ \Delta UR_t \end{bmatrix} = \sum_{i=0}^{\infty} L^i \begin{bmatrix} b_{11i} & b_{12i} \\ b_{21i} & b_{22i} \end{bmatrix} \begin{bmatrix} \varepsilon_{dt} \\ \varepsilon_{st} \end{bmatrix} \tag{11}$$

The vector $\varepsilon_t = \begin{bmatrix} \varepsilon_{dt} \\ \varepsilon_{st} \end{bmatrix}$ contains the two structural shocks, the demand one and the supply one. The elements b_{11i} and b_{21i} are the impulse responses of an aggregate demand shock on the time path of the shadow economy and unemployment rate. The coefficients b_{12i} and b_{22i} are the impulse responses of an aggregate supply shock on the time path of shadow economy and unemployment rate respectively.

The key to decomposing the SE_t sequence into its trend and irregular components is to assume that one of the shocks has a temporary effect (no long-run effect) on the SE_t. According to Blanchard and Quah, the key is to assume that one of the structural shocks has a temporary effect on ΔSE_t. We assume that an aggregate supply (unemployment rate) shock has no long-run effect on shadow economy. In the long-run, if the shadow economy is to be unaffected by the supply shock, it must be the case that the cumulated effect of a ε_{st} shock on the ΔSE_t sequence must be equal to zero. In other words, we impose a long-run restriction on the relationship between the observed data (SE) and the unobserved structural shock (ε_{st}) such that: $\sum_{k=0}^{\infty} b_{12}(k)\varepsilon_{st-k} = 0$ (12)

Equation (12) is an Aggregate Supply Shock stating that the second structural shock (aggregate supply) has no long-run effect on shadow economy.

2.2. Empirical results

In order to analyze the nature of the relationship between shadow economy and unemployment rate (registered or ILO unemployment rate), we use the Structural VAR approach, for Blanchard and Quah methodology. In order to identify supply and demand shocks, we start by running two bivariate VAR models (first for shadow economy and registered unemployment rate and second for shadow economy and ILO unemployment rate).

The variables included in the VARs are suspected to have a unit root. To verify this, ADF and PP unit root tests were applied revealing that the variables are non-stationary at their levels but stationary at their first differences, being integrated of order one, I(1).

Table 1. ADF and PP Tests for Unit Root

		Shadow economy(SE)			Registered unemployment rate(R_UR)			ILO unemployment rate(ILO_UR)		
		T&C	C	None	T&C	C	None	T&C	C	**None**
Level	ADF	-6.29*	-1.05	-3.28	0.24	-1.58	-0.73	-3.03	-2.70	**-0.36**
	lag	(0)	(6)	(6)	(4)	(4)	(4)	(1)	(0)	**(0)**
	PP	-6.29*	-1.74	-1.38	-0.68	-2.13	-1.34	-2.88	-2.70	**-0.32**
	lag	(1)	(3)	(1)	(3)	(1)	(1)	(1)	(0)	**(5)**
First diff.	ADF	-10.63*	-10.74*	-10.45*	-3.51***	-2.83***	-2.89*	-6.13*	-6.22*	**-6.30***
	lag	(0)	(0)	(0)	(3)	(3)	(3)	(0)	(0)	**(0)**
	PP	-11.34*	-9.90*	-8.86*	-7.14*	-6.17*	-6.19*	-6.59*	-6.72*	**-6.85***
	lag	**(3)**	**(2)**	**(3)**	**(7)**	**(1)**	**(1)**	**(6)**	**(6)**	**(6)**

Note:
T&C represents the most general model with a drift and trend; C is the model with a drift and without trend; None is the most restricted model without a drift and trend. Numbers in brackets are lag lengths used in ADF test (as determined by SCH set to maximum 12) to remove serial correlation in the residuals. When using PP test, numbers in brackets represent Newey-West Bandwith (as determined by Bartlett-Kernel). Both in ADF and PP tests, unit root tests were performed from the most general to the least specific model by eliminating trend and intercept across the. *, ** and *** denote rejection of the null hypothesis at the 1%, 5% and 10% levels respectively. Tests for unit roots have been carried out in E-VIEWS 6.0.

Because the variables are integrated of the same order, I(1) we will difference the variables and we introduce the first difference in the VAR analysis[4]. Including a sufficient number of lags to eliminate serial correlation from the residuals is crucial as using a lag structure that is too parsimonious can significantly bias the estimation of the structural components.

Table 2. Optimal lag length

Models	Sequential LR	AIC	SC	HQ	FPE	Chosen[5]
d(SE) and d(R_UR)	-	1	1	1	1	**1**
d(SE) and d(ILO_UR)	**-**	**1**	**1**	**1**	**1**	**1**

Note: LR is the sequential modified LR test statistic; FPE is the Final Prediction Error; AIC is the Akaike Information Criterion; SBC is the Schwarz Information Criterion; HQ is the Hannan-Quinn Information Criterion.

[4] According Blanchard and Quah(1989), we estimate the VARs models without intercept.
[5] Given the small size of our series, we preferred to choose the optimal lag as 1 based on the discussion of Mills and Prasad, 1992.

Table 2 offers the optimal lag length for each model according to the five criteria. It can be observed that the optimal lag length is found to be one. The number of lags for each VAR was chosen according with the information criteria above and by taking into consideration other information from VAR analysis. At the same time, the autocorrelation of residuals was analyzed to be sure that through the number of chosen lags the residuals do not remain with autocorrelation. Further on, the both VARs verify **the stability condition**. Since each VAR represents a system of linear first-order difference equations, it is stable only if the absolute values of all eigenvalues of the system matrix lie inside the unit circle. This condition is fulfilled by both VARs. Furthermore, we inspect the diagnostic concerning non-autocorrelation[6], homoskedasticity[7] and normality[8] of the residuals. These hypotheses were verified by the residuals of both estimated VARs.

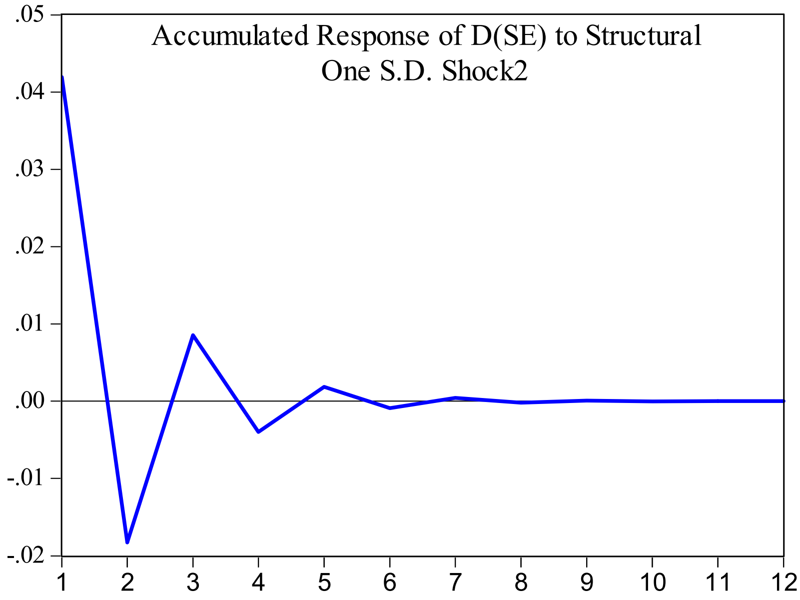
We have estimated the VAR models with one lag who verifies the stability condition[9]. Furthermore, we impose on this VAR a long-run restriction which specifies that the long run effect of the supply shocks on the shadow economy is null.

According to Blanchard and Quah, the key is to assume that one of the structural shocks has a temporary effect on ΔSE. Following Dell'Anno and Solomon[10] (2006) we assume that an aggregate supply (unemployment rate) shock has no long-run effect on shadow economy. The long-run restriction on the relationship between the observed data (SE) and the unobserved structural shock (ε_{st}) is: $\sum_{i=0}^{\infty} b_{12i} = 0$ (13)

The restriction in (13) implies that the cumulative effect of ε_{st} on ΔSE_t is zero and consequently the long-run effect of ε_{st} on the level of SE_t itself is zero. The supply shock (ε_{st}) has only short-run effects on the shadow economy. Starting from this model, we analyze the impulse response function for the structural version of the model.

[6] The presence of autocorrelation in the residuals was tested using the LM test.
[7] For the homoskedasticity it was applied the White test.
[8] The normality of the residuals was tested using Cholesky normality test.
[9] Since each VAR represents a system of linear first-order difference equations, it is stable only if the absolute values of all eigenvalues of the system matrix lie inside the unit circle.
[10] Dell'Anno and Solomon (2006) investigate the relationship between shadow economy and unemployment rate for he case of United States using SVAR approach for the period 1970-2004. They that an aggregate supply (unemployment rate) shock has no long-run effect on shadow economy.

Fig 2. Accumulated responses of shadow economy to a positive aggregate supply shock (registered unemployment rate)



In the short-run, a positive aggregate supply shock (registered unemployment rate) causes a decrease in the shadow economy by about 1.8% below the baseline. This occurs in the second quarter following the initial shock. In the third quarter, the size of the shadow economy will begin to increase by about 0.6% above the baseline.

The variance decomposition using the actual ε_{st} and ε_{dt} sequence allow assessing the relative contributions of demand and supply shocks to forecast error variance of the shadow economy.

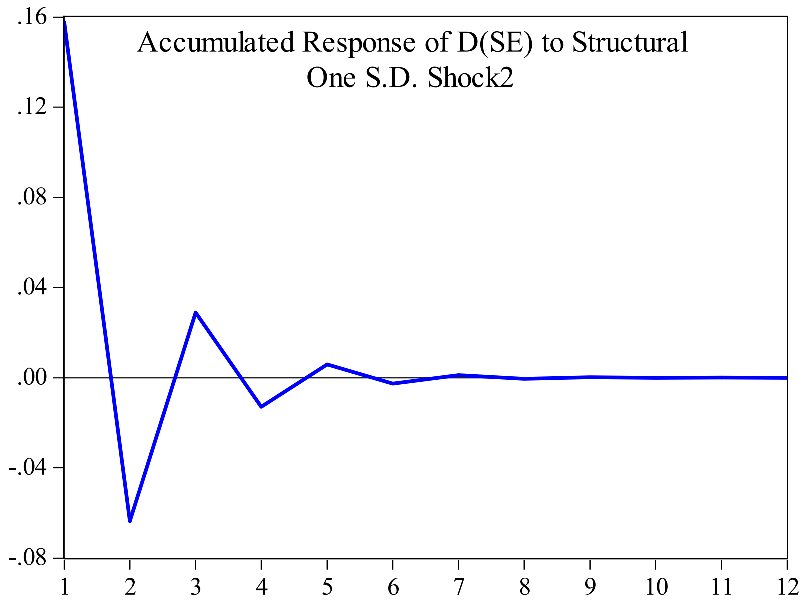
Table 3. Variance decomposition of D(SE) due to supply-side shock(registered unemployment rate)

	Percent of Forecast Error Variance due to:	
Period	Shock1(demand-side shock)	Shock2(supply-side shock)
1	99.71260	0.287399
3	99.22069	0.779312
6	99.20373	0.796272
9	99.20355	0.796446
12	99.20355	0.796448

Factorization: Structural

As is immediately evident, demand-side shocks explain almost all the forecast error variance of the shadow economy at any forecast horizon. Hence, the demand shocks are responsible for movements in shadow economy.

Fig 3. Accumulated responses of shadow economy to a positive aggregate supply shock (ILO unemployment rate)



In the short-run, a positive aggregate supply shock (ILO unemployment rate) causes a decrease in the shadow economy by about 6.3% below the baseline. This occurs in the second quarter following the initial shock. In the third quarter, the size of the shadow economy will begin to increase by about 2.8% above the baseline.

Table 4. Variance decomposition of D(SE) due to supply-side shock(ILO unemployment rate)

	Percent of Forecast Error Variance due to:	
Period	Shock1(demand-side shock)	Shock2(supply-side shock)
1	95.79529	4.204706
3	89.48092	10.51908
6	89.31066	10.68934
9	89.30929	10.69071
12	89.30928	10.69072
Factorization: Structural		

As is immediately evident, demand-side shocks explain almost all the forecast error variance of the shadow economy at any forecast horizon. Hence, the demand shocks are responsible for movements in shadow economy.

CONCLUSIONS

In this paper, the SVAR methodology with long-run restrictions has been applied to analyze to relationship between shadow economy as % of official GDP and unemployment rate for the case of Romania using quarterly data covering the period 2000-2010. The size of the shadow economy as % of official GDP was obtained using the currency demand approach based on VECM models. Its size is estimated to be decreasing over the sample period from 37% to 31% of official GDP.

The relationship between the variables is tested by imposing a long-run restriction in the Structural VAR model to analyze the impact of the shadow economy to a temporary shock in unemployment.

The accumulated responses generated by a positive supply shock (unemployment rate) confirms that in the short-run, a rise in both registered unemployment rate and ILO unemployment rate in formal sector will lead to a decrease in the number of people who work in the shadow economy in the second quarter following the initial shock and to an increase in the size of the Romanian shadow economy in the third quarter following the initial shock.

References

1. Albu, L.L.Iorgulescu, R. and Stanica, C. **Estimating Hidden Economy and Hidden Migration: The Case of Romania**, Journal for Economic Forecasting, Institute for Economic Forecasting, vol. 0(2), 46-56, 2010.

2. Albu, L.L., Ghizdeanu, I., and Stanica, C. **Spatial Distribution of the Informal Economy. A Theoretical and Empirical Investigation,** SCIENZE REGIONALI, FrancoAngeli Editore, vol. 0(1), 63-80, 2011.

3. Alexandru A. A., Dobre I. and Ghinararu C. **A SVAR approach of the relationship between shadow economy and unemployment rate: the case of United States**, WSEAS Transactions on Business and Economics, issue 2, no.5, 2011, pg.324-333, ISSN 1109-9526.

3. Blanchard O.J. and Watson, M.W. **Are Business Cycles All Alike?** in R. Gordon (ed) The American Business Cycle: Continuity and Change, NBER and University of Chicago Press, 123-156, 1986.

4. Blanchard, O.J. and Quah, D. **The Dynamic Effects of Aggregate Demand and Supply Disturbances**, American Economic Review, 79 (4), 655-73, 1989.

5. Boeri, T. and Garibaldi, P. **Shadow Activity and Unemployment in a Depressed Labor Market**, *CEPR Discussion papers,* n.3433, 2002.

6. Davidescu, A.A and Dobre, I. **The Impact of Unemployment Rate on the Size of Romanian Shadow Economy**, Public Finance Review, Special Issue: The

Shadow Economy, Tax Evasion, and Money Laundering, September 2013 41(5): 578-607, ISSN: 1552-7530, 2013.

7. Davidescu A. A, and Dobre I. **Revisiting the Relationship between Shadow Economy and the Level of Unemployment Rate. A SVAR Empirical Ivestigation for the Case of United States,** Journal of Applied Quantitative Methods, volume 8, issue 3, september 30, 2013, pg.11-19, B+, indexed in DOAJ, ISSN 1842-4562.

8. Dell'Anno, R. and Solomon, O. **Shadow economy and unemployment rate in USA. Is there a structural relationship?**, Annual Meeting of the European Public Choice Society, Finland, April 20-23, 2007.

9. Evans, G.W. **Output and Unemployment Dynamics in the United States: 1950 – 1985**, Journal of Applied Econometrics*, 4*/213-238, 1989.

10. Enders, W. and Hurn, S. **Identifying aggregate demand and supply shocks in a small open economy**, Oxford Economic Papers, No. 59, 2007, pp. 411-429.

11. Enste, D.H. **Shadow Economy and Institutional Change in Transition Countries in Boyan Belev (eds.), The Informal Economy in the EU Assessment Countries: Size, Scope, Trends and Challenges of the Process of EU-enlargement**, Center for Study of Democracy, 2003, Sofia, 81-114.

12. Giles, David E.A. and Tedds, L.M. **Taxes and the Canadian Underground Economy**. Canadian Tax paper n.106. Canadian Tax Foundation. Toronto, 2002.

13. Giles David E.A. **Measuring the hidden economy: Implications for econometric modelling**, The Economic Journal, 109/456, pp.370-380, 1999.

14. Gottshalk, J. and Willem, V.Z. **Do Bivariate SVAR Models with Long-Run Identifying Restrictions Yield Reliable Results? The Case of Germany**, Kiel Working Paper No. 1068, 2001.

15. Johnson, S., Kaufmann, D. and Zoido-Lobatón, P. **Regulatory discretion and the unofficial Economy**, The American Economic Review, 88/ 2, pp. 387-392, 1998a.

16. Johnson, S., Kaufmann, D. and Zoido-Lobatón, P. **Corruption, public finances and the unofficial economy**, American Economic Review, vol. 88, no. 2, pp. 387-392, 1998b.

17. McCoy, D. **How useful is Structural VAR Analysis for Irish economics?**, Eleventh Annual Conference of the Irish Economic Association in Athlone, April 4-6th, 1997.

18. Schneider, F. **Shadow economies around the world: What do we really know**, European Journal of Political Economy 21, 598-642, 2005.

19. Schneider, F. and Enste, D.H. **Shadow economies: size, causes and consequences**, Journal of Economic Literature 38, 77-114, 2000.

20. Schneider, F. **Shadow Economies and Corruption all over the world: New estimates for 145 Countries,** Economics 2007/9, 1-47, 2007.

21. Sims, C.A. **Are Forecasting Models Usable for Policy Analysis**, Federal Reserve Bank of Minneapolis Quarterly Review, 2-16, 1986.

22. The Economist, **American unemployment**, august 14th 2010, pg.56.

23. *** **Quarterly National Accounts database**, Eurostat.

24. *** **Quarterly Labor Force Survey database**, Eurostat.

25. *** **Quarterly Government Finance Statistics database**, Eurostat.

26. *** **Quarterly Interest Rates database**, Eurostat.

27. *** **Quarterly Monetary and Financial Statistics database**, Eurostat.

28. *** **Employment and Unemployment database**, Eurostat.

29. *** **Tempo database**, National Institute of Statistics, www.insse.ro

30. *** **Monthly Bulletins of National Bank of Romania**, 2000-2010, www.bnr.ro.

31. *** Eviews 6.0 software

---

[1] Adriana AnaMaria Davidescu is lecturer at the Department of Statistics and Econometrics of Bucharest Academy of Economic Studies and scientific researcher at the National Scientific Research Institute for Labour and Social Protection. Scientific Research Interests: Time Series Analysis, Macroeconometrics

[2] Ion Dobre is professor at the department of Cybernetic and Economic Informatics of Bucharest Academy of Economic Studies. Scientific Research Interests: Time Series Analysis, Macroeconometrics, Decision Theory, Simulation of economic processes

JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS

INTUITIONISTIC FUZZY MULTICRITERIA GROUP DECISION-MAKING APPROACH TO QUALITY CLAY-BRICK SELECTION PROBLEMS BASED ON GREY RELATIONAL ANALYSIS

Kalyan MONDAL[1]
Birnagr High School (HS), Birnagar, Ranaghat, Nadia, West Bengal, India

E-mail: kalyanmathematic@gmail.com

Surapati PRAMANIK[2]
Department of Mathematics, Nandalal Ghosh B.T. College, Panpur, PO-Narayanpur, and District: North 24 Parganas, West Bengal

E-mail: sura_pati@yahoo.co.in

ABSTRACT

This paper presents quality Clay-Brick selection process based on intuitionistic fuzzy multi criteria group decision making through grey relational analysis. Brick plays an important role in construction field. Intuitionistic fuzzy weighted averaging operator is used to aggregate individual opinions of decision makers into a group opinion. Six criteria are considered for selection process and the criteria are obtained from expert opinions. The criteria are namely solidity, color, size and shape, strength of Brick, cost of Bricks and carrying cost. Weights of the criteria are obtained from domain experts by using a questionnaire. The rating of an alternative with respect to certain criterion offered by decision maker is represented by linguistic variable that can be expressed by intuitionistic fuzzy sets. An intuitionistic fuzzy set, which is characterized by membership function (degree of acceptance), non-membership function (degree of rejection) and the degree of indeterminacy or the degree of hesitancy, is a more general and suitable way to deal with imprecise information. Grey relational analysis is used for ranking and selection of alternatives. An illustrative numerical example for quality Brick selection is solved to show the effectiveness of the proposed model.

Keywords: *Bricks, Bricks field, Fuzzy set, Grey relational analysis, Grey relational grade, Grey system theory, Group decision making, Intuitionistic fuzzy set, Linguistic variables, Multi-criteria decision making.*

INTRODUCTION

Multi criteria decision-making (MCDM) problem generally consists of finding the most satisfactory alternative from all the feasible alternatives. Classical MCDM [20] deals with crisp numbers i.e. the ratings and the weights of criteria are measured by crisp numbers. Fuzzy MCDM [19] deals with fuzzy or intuitionistic fuzzy numbers i.e. the ratings and the weights are expressed by linguistic variables. In 1965, Zadeh [31] published his seminal paper studying with fuzzy sets. In 1986, Atanassov [2] extended the concept of fuzzy sets to intuitionistic fuzzy sets (IFSs).

Brick selection process from various Brick fields is a special case of material selection to construct a building structure. In traditional way, we select Bricks roughly based on its color, size and total cost of Brick, without considering other attributes of Brick. In that case the building construction may have some problems regarding low rigidity, longevity, etc., which be dangerous. So it is necessary to formulate a scientific based selection method. In order to select the most suitable Brick to construct a building, the following criteria of Bricks will have to consider. The criteria are namely, solidity, color, size and shape, strength of Brick, cost of Brick and carrying cost.

Liang and Wang [21] presented a fuzzy MCDM algorithm for personnel selection. Karsak (18) developed a fuzzy MCDM approach based on ideal and anti-ideal solutions for the selection of the most suitable candidate. Gibney and Shang [14] and Günör et al. [15] presented the use of the analytical hierarchy process (AHP) in the personnel selection process, respectively. Dağdeviren (9) proposed a hybrid model, which employs analytical network process (ANP) and modified technique for order preference by similarity to ideal solution (TOPSIS) for supporting the personnel selection process in the manufacturing systems. Dursun and Karsak [12] presented a fuzzy MCDM approach by using TOPSIS with 2-tuples for personnel selection. Robertson and Smith [24] presented good reviews on personnel selection studies. They investigated the role of job analysis, contemporary models of work performance, and set of criteria used in personnel selection process. A comprehensive survey of the state of the art in MCDM can be found in the book authored by Ehrgott and Gandibleux [13].

In this study, we present an intuitionistic fuzzy multi criteria group decision-making model with grey relational analysis for quality Brick selection for constructional field. A good quality Brick should be regular in shape and size, with smooth even sides and no cracks or defects. Normally poor quality Bricks are a result of using poor techniques when making the Bricks but these errors can often be easily corrected. Poor quality Bricks, must have not enough strength to carry the weight of the roof. If Bricks have been well- made and well-fired, you will hear a metallic sound or ring when they are knocked together. If they make a dull sound, it could mean that they are either cracked or under fired [26, 27, 28]. Rest of the paper is constructed in the following manner. Section 2 presents preliminaries of fuzzy sets. Section 3 describes the conversion of linguistic variables into intuitionistic fuzzy numbers. Section 4 presents the definition of operation terms. Section 5 explains grey relational analysis. Section 6 is devoted to present intuitionistic fuzzy multi-criteria group decision making based on grey relational analysis for Brick selection process. Section 7 illustrates example for Brick selection process. Sensitivity analysis of weight structure and ranking of Bricks are presented in Section 8. Section 9 describes the advantage of the proposed approach. Section 10 presents conclusion and future direction of research work.

2. PRELIMINARIES OF FUZZY SETS

In 1965, Zadeh [31] first introduced the concept of fuzzy sets as a mathematical form for representing impreciseness.

Definition 2.1: Fuzzy set: A fuzzy set $\tilde{A}$ in a universe of discourse X is defined as the following set of pair $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) : x \in X\}$. Here, $\mu_{\tilde{A}}(x) : x \to [0,1]$ is a mapping called the membership value of x∈X in a fuzzy set $\tilde{A}$.

Definition 2.2: Intuitionistic Fuzzy set: An intuitionistic fuzzy set (IFS) $\tilde{A}$ = $\{\langle x, \mu_{\tilde{A}}(x), v_{\tilde{A}} \rangle : x \in X\}$, where the functions $\mu_{\tilde{A}}(x)$: X→ $[0,1]$ and $v_{\tilde{A}}(x)$:X→ $[0,1]$ define the degree of membership and degree of non-membership respectively of the element x∈X to the set $\tilde{A}$ that is a subset of X, and every x∈X, $0 \le \mu_{\tilde{A}}(x) + v_{\tilde{A}}^{i}(x) \le 1$.

Definition 2.3: The value of $\pi_{\tilde{A}^i}$ (x) = 1- $(\mu_{\tilde{A}}(x) + v_{\tilde{A}}(x))$ is called the degree of non-determinacy (or uncertainty or hesitancy) of the element x∈X to the intuitionistic fuzzy set.

Definition 2.4: Hamming distance is defined as $H\left(\tilde{A}, \tilde{B}\right) = \delta$

$$\frac{1}{2}\sum_{x \in E}\left(\left|\mu_{\tilde{A}}(x) - \mu_{\tilde{B}}(x)\right| + \left|v_{\tilde{A}}(x) - v_{\tilde{B}}\right| + \left|\pi_{\tilde{A}}(x) - \pi_{\tilde{B}}\right|\right)$$

3. CONVERSION BETWEEN LINGUISTIC VARIABLES AND INTUITIONISTIC FUZZY NUMBERS (IFNs)

The description of linguistic variable is more realistic when we discuss a problem in intuitionistic fuzzy environment. For example, the ratings of alternative with respect to qualitative criteria could be expressed using linguistic variables such as very bad, bad, good, fair, very good etc. Linguistic variable can be converted into IFNs.

4. OPERATIONAL DEFINITIONS OF THE CRITERIA OF BRICK

i) **Solid clay Brick**: Loam soil is used to prepare rigid Brick. An ideal extended solid rigid body whose size and shape are definitely fixed and remain unaltered when forces are applied. The distance between any two given points of a rigid body remains constant in time when external forces applied on it. Basically solid Brick implies more and more rigidity of Brick. If we soap a Brick in water and fall downwards from 3 or 4 feet heights [3, 27], it remains unbroken.

ii) **Color:** Quality Brick has a color reddish or light maroon.

iii) **Size and shape**: All Bricks will be more or less same size and shape. A Brick has a length, width and height. The size or dimensions of a Brick are determined by how it is used in construction. Standard size of a Brick is about $190mm \times 90mm \times 40mm$.

Width:

The width of a Brick should be small enough to allow a Bricklayer to lift the Brick with one hand and place it on a bed of mortar. If the Brick was wider, the Bricklayer would have to put down the trowel while building the wall to pick up the Brick with two hands and as a result, time would be wasted. In addition, a wider Brick would weigh more and therefore tire the mason more quickly. For the average Brick, the width should not be more than 115 mm.

Length:

There is a very important relationship between the length of a Brick and its width because of how we use Bricks to build a wall. The length of a Brick should be equal to twice its width plus 10 mm (for the mortar joint). A Brick with this length will be easier to build with because it will provide an even surface on both sides of the wall. For example, if an individual follows the rule of the length and width of the Brick, if breadth is set as 115 mm, then the ideal length would be 240 mm.

Height:

The height of a Brick, though of less importance, also has a relationship with the length of the Brick. The height of three Bricks plus two 10 mm joints should be equal to the length of a Brick. This allows a Bricklayer to lay Bricks on end (called a soldier course) and join them into the wall without having to cut the Bricks. To determine the height of a Brick, subtract 20 mm (the thickness of the two 10 mm mortar joints) from the length and divide the result by three (this represents the three Bricks).

Possible Brick Sizes:

Therefore, using these rules, the largest size Brick that would still permit a Bricklayer to comfortably pick it up with one hand, would be 240 mm in length, 115 mm in width and 73 mm in height. A Brick of this size would weight about 3.5 kg. Every country in the world seems to have a different size of Brick. The sizes are a result of centuries of tradition or custom but almost all use the same rules and lie within the limits mentioned. No one size is better than the other. In India the standard Brick size is 190 mm x 90 mm x 40 mm while the British standard is 215 mm x 102.5 mm x 65 mm. To choose your Brick size, first contact the local public works department to see if your country has a standard size. If not, you will have to choose according to need and desires and practical utility of Bricks. Possible Brick sizes are shown in the Table 1.

Table1. Possible sizes of Bricks

Length	Width	Height
240mm	115mm	73mm
230mm	110mm	70mm
220mm	105mm	67mm
215mm	102mm	65mm
210mm	100mm	63mm

iv) **Well dried and burnt (Strength of Brick):** Raw Brick must be well dried in sunshine and then properly burnt. If any two Bricks from a blend are touched with some force then it occur material sound [3]. Bricks must have enough strength to carry the weight of the roof. If Bricks have been well- made and well-fired, you will hear a metallic sound or ring when they are knocked together. If they make a dull sound, it could mean that they are either cracked or under fired. A simple test for strength is to drop a Brick from a height of 1.2 meters (shoulder height). A good Brick will not break. This test should be repeated with a wet Brick (a Brick soaked in water for one week). If the soaked Brick does not break when dropped, the quality is good enough to build structures.

v) **Brick Cost:** Decision makers will try to purchase Brick at minimum cost. The Brick must be in reasonable cost as for as possible. Quality Brick with reasonable price must be more acceptable.

vi) **Carrying Cost**: There must be a reasonable distance between Brick field and construction site for maintaining reasonable carrying cost.

5. GREY RELATIONAL ANALYSIS

The calculation process for grey relational analysis (GRA) [10] can be presented as follows:

Suppose G is a factor set of grey relation, $G = \{G_0, G_1, \cdots, G_m\}$ where $G_0 \in G$ represents the referential sequence; $G_i \in G$

Denotes the comparative sequence G_i ,i = 1, 2, …, m. G_0 and G_i consist of n elements and can be expressed as follows:

$G_0 = (G_0(1), G_0(2), \cdots, G_0(k), \cdots, G_0(n))$,

$G_i = (G_i(1), G_i(2), \cdots, G_i(k), \cdots, G_i(n))$, where i =1, 2, …, m; k = 1, 2, …, n;

$n \in N$, and $G_0(k)$ and $G_i(k)$ are the numbers of referential sequences and comparative sequences at point k, respectively. In practical applications, the referential sequence can be an ideal objective and the comparative sequences are alternatives. The best alternative corresponds to the largest degree of grey relation. If the grey relational coefficient (GRC) of the referential sequences and comparative sequences at point k is $\tau(G_0, G_i)$ subject to the four conditions:

1. $0 \le \tau(G_0, G_i) \le 1$

$\tau(G_0, G_i) = 1 \Leftrightarrow G_0 = G_i$

$\tau(G_0, G_i) = 0 \Leftrightarrow G_0, G_i \in \Phi$, where Φ is empty set.

2. $G_0, G_i \in G$, then

$\tau(G_0, G_i) = \tau(G_i, G_0) \Leftrightarrow G = \{G_0, G\}$

3. $\tau(G_0, G_i) \overset{\text{often}}{\neq} \tau(G_i, G_0)$

4. If $|G_0(k) - G_i(k)|$ is large, $\tau(G_0(k), G_i(k))$ becomes smaller. The essential condition and quantitative model for grey relation are produced based on the above four prerequisites. The grey relational coefficient of the referential sequences and comparative sequences at point k can be expressed as follows:

$$\tau(G_0(k), G_i(k)) = \frac{\min_i \min_k |G_0(k) - G_i(k)| + \rho \max_i \max_k |G_0(k) - G_i(k)|}{|G_0(k) - G_i(k)| + \rho \max_i \max_k |G_0(k) - G_i(k)|} \quad (1)$$

The symbol ρ in equation (1) represents “contrast control,” “environmental coefficient”, or the “distinguishing coefficient”. This coefficient is a free parameter. Its value, over a broad appropriate range of values, does not affect the ordering of the grey relational grade values, but a good value of the contrast control is needed for clear identification of key system factors. For the end points 0 and 1, i.e. for $\rho = 1$, the comparison environment is unaltered and for $\rho = 0$, the comparison environment disappears. In cases where data variation is large, ρ usually ranges from 0.1 to 0.5 for reducing the influence of extremely large $\max_i \max_k |G_0(k) - G_i(k)|$. Generally, $\rho = 0.5$ is considered for decision-making situation.

6. INTUTIONISTIC FUZZY MULTI-CRITERIA GROUP DECISION-MAKING BASED ON GREY RELATIONAL ANALYSIS

We present here intuitionistic fuzzy multi-criteria group decision-making (MCGDM) using grey relational analysis. For a MCGDM problem, let $A = \{A_1, A_2, \cdots, A_m\}$ $(m \geq 2)$ be a finite set of alternatives, $D = \{D_1, D_2, \cdots, D_t\}$ $(t \geq 2)$ be a set of decision makers, $C = \{C_1, C_2, \cdots, C_n\}$ be a set of criteria. The weight information of the criteria and the decision maker are completely unknown. Let us denote $M = \{1, 2, \cdots, m\}$, $T = \{1, 2, \cdots, t\}$, $N = \{1, 2, \cdots, n\}$. The proposed approach is based on the following steps.

Step1. **Formation of Intuitionistic fuzzy decision matrices**:

We assume that the rating of alternative A_i $(i \in M)$ with respect to the criteria C_j $(j \in N)$ offered by the t-th decision maker $(t \in T)$ is linguistic variable x_{ij}^t that can be expressed by IFSs. A MCGDM problem can then be expressed by the following decision matrix:

$$G = \left(G_{ij}^t\right) = \begin{bmatrix} G_{11}^t, G_{12}^t, \cdots, G_{1n}^t \\ G_{21}^t, G_{22}^t, \cdots, G_{2n}^t \\ \cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots \\ G_{m1}^t, G_{m2}^t, \cdots, G_{mn}^t \end{bmatrix}, \quad t \in T \tag{2}$$

Here, $\left(G_{ij}^t\right) = \left(\mu_{ij}^t, v_{ij}^t, \pi_{ij}^t\right)$.

Step2. Determination of the weights of the decision makers:

Case1. We assume that the decision making group consists of T decision makers. The importance of the decision makers in the selection committee may not be equal. The importance of decision makers are considered as linguistic variables expressed by intuitionistic fuzzy numbers (IFNs). Let $D_t = (\mu_t, v_t, \pi_t)$ be an intuitionistic IFN that represents the rating of the t-th decision maker. Then the weight of the t-th decision maker can be determined as:

$$\gamma_t = \frac{\left(\mu_t + \pi_t\left(\frac{\mu_t}{\mu_t + v_t}\right)\right)}{\sum_{t=1}^{T}\left(\mu_t + \pi_t\left(\frac{\mu_t}{\mu_t + v_t}\right)\right)}, \text{ where } \sum_{t=1}^{T}\gamma_t = 1 \tag{3}$$

The linguistic variables for the importance of the decision makers are provided in the Table 2.

Case 2. If the importance of all the decision makers is same namely extremely importance, the rating of the t-th decision maker can be expressed as (1, 0, 0). Then the weight of each decision maker will be 1/T.

Step3.

Formation of the aggregated intuitionistic fuzzy decision matrix based on the opinions of decision makers: Let G= $\left(G_{ij}^{t}\right)_{m\times n}$ be an intuitionistic fuzzy decision matrix of the t-th decision maker. $\gamma=\left(\gamma_1,\gamma_2,\cdots,\gamma_T\right)$, be the weight set of the decision-makers and $\sum_{t=1}^{T}\gamma_t=1$, $\gamma_t\in[0,1]$. In the group decision-making process, all individual decisions needed to be fused into a group opinion to construct an aggregate intuitionistic fuzzy decision matrix. In order to do, we use intuitionistic fuzzy weighted average (IFWA) operator due to Xu [30] as follows:

$$G_{ij}=IFGA_{\gamma}\left(G_{ij}^{(1)},G_{ij}^{(2)},\cdots,G_{ij}^{(t)}\right)=\gamma_1G_{ij}^{(1)}\oplus\gamma_2G_{ij}^{(2)}\oplus\cdots\oplus\gamma_tG_{ij}^{(t)}$$

$$=\left(1-\prod_{t=1}^{T}\left(1-\mu_{ij}^{(t)}\right)^{\gamma_t},\coprod_{t=1}^{T}\left(v_{ij}^{(t)}\right)^{\gamma_t},\coprod_{t=1}^{T}\left(1-\mu_{ij}^{(t)}\right)^{\gamma_t}-\coprod_{t=1}^{T}\left(v_{ij}^{(t)}\right)^{\gamma_t}\right) \quad (4)$$

The aggregate intuitionistic fuzzy decision matrix then can be written as:

$$G=\begin{bmatrix}(\mu_{11},v_{11},\pi_{11})\,(\mu_{12},v_{12},\pi_{12})\cdots(\mu_{1n},v_{1n},\pi_{1n})\\(\mu_{21},v_{21},\pi_{21})\,(\mu_{22},v_{22},\pi_{22})\cdots(\mu_{2n},v_{2n},\pi_{2n})\\\cdots\cdots\cdots\cdots\cdots\cdots\\\cdots\cdots\cdots\cdots\cdots\cdots\\(\mu_{m1},v_{m1},\pi_{m1})\,(\mu_{m2},v_{m2},\pi_{m2})\cdots(\mu_{mn},v_{mn},\pi_{mn})\end{bmatrix}=\begin{bmatrix}G_{11}G_{12}\cdots G_{1n}\\G_{21}G_{22}\cdots G_{2n}\\\cdots\cdots\cdots\\\cdots\cdots\cdots\\G_{m1}G_{m2}\cdots G_{mn}\end{bmatrix} \quad (5)$$

Here $G_{ij}=\left(\mu_{ij},v_{ij},\pi_{ij}\right)$ $\left(i=1,2,\cdots,m\right)$ and $\left(j=1,2,\cdots,n\right)$ is an element of the aggregate intuitionistic fuzzy decision matrix. $\mu_{ij}=1-\coprod_{t=1}^{T}\left(1-\mu_{ij}^{(t)}\right)^{\gamma_t}$, $\nu_{ij}=\prod_{t=1}^{T}\left(\nu_{ij}^{t}\right)^{\gamma^t}$, $\pi_{ij}=\prod_{t=1}^{T}\left(1-\mu_{ij}^{(t)}\right)^{\gamma_t}-\prod_{t=1}^{T}\left(v_{ij}^{(t)}\right)^{\gamma_t}$, $i\in M, j\in N$

Step4. Determination weights of the criteria:

In the decision making situation, decision makers may feel that all criteria of Bricks are not equal importance. Here the importance of the criteria is obtained from expert opinion through questionnaire method i.e. the weights of the criteria are previously determined such that the sum of the weights of the criteria is equal to unity. Incidental sampling is employed to collect data. Data was collected from 10 constructional engineers, 2 earth quake specialist of Nadia district and 10 Brick fields of surrounding areas. After extended interviews and discussions with the experts, the list of criteria of Brick that are identified as playing important role in model formulation of the problem. After structured procedures of filling in specific questionnaire by our domain experts, the criteria for Brick selection are identified as solidity, color, size, strength of Brick, cost of Bricks and carrying cost. We have average weight of each criteria w_j j=1, 2,... 6 as

$w_1 = .275, w_2 = .175, w_3 = .2, w_4 = .1, w_5 = .05, w_6 = .2$ and $\sum_{j=1}^{6} w_j = 1$. Alternately, the entropy weights of the criteria may be used. In order to obtain weight, a set of grades of importance, intuitionistic fuzzy entropy may be used due to Vlachos & Sergiadis [29] as follows:

$$E_j = -\frac{1}{n\ln 2}\sum_{i=1}^{m}\left[\mu_{ij}\ln\mu_{ij} + v_{ij}\ln v_{ij} - (1-\pi_{ij})\ln(1-\pi_{ij}) - \pi_{ij}\ln 2\right] \quad (6)$$

The entropy weight of the j-th criteria is defined as follows:

$$w_j = \frac{1-E_j}{n - \sum_{j=1}^{n} E_{ij}} \quad (7)$$

Step5. Determination of the reference sequence based on IFNs:

$$\tilde{G}^+ = \left((\mu_1^+, v_1^+, \pi_1^+), (\mu_2^+, v_2^+, \pi_2^+), \cdots, (\mu_n^+, v_n^+, \pi_n^+)\right) \quad (8)$$

Here $\tilde{G}_{ij}^+ = (\mu_j^+, v_j^+, \pi_j^+) = \left(\max_i \mu_{ij}, \min_i v_{ij}, \min_i \pi_{ij}\right)$, j = 1, 2,…,n (9)

Reference sequence should be the optimal sequence of the criteria values. Since the aspired level of the membership value, non-membership value and indeterminacy value are 1, 0, 0 respectively, the point consisting of highest membership value, minimum non-membership value and minimum indeterminacy value would represent the reference value or ideal point or utopia point. In the intuitionistic fuzzy decision matrix, the maximum value $\tilde{G}_j^+ = (1,0,0)$ can be used as the reference value. Then the reference sequence is expressed as $\tilde{G}^+ = \{(1,0,0),(1,0,0)\cdots,(1,0,0)\}$ (10)

Step6.

Calculation of the grey relational coefficient []of each alternative from positive ideal solution (PIS) using the following equation:

$$\tau_{ij} = \frac{\min_{1\le i\le m}\min_{1\le j\le n}\delta\left(\tilde{G}_{ij}, \tilde{G}_{ij}^+\right) + \rho\max_{1\le i\le m}\max_{1\le j\le n}\delta\left(\tilde{G}_{ij}, \tilde{G}_{ij}^+\right)}{\delta\left(\tilde{G}_{ij}, \tilde{G}_{ij}^+\right) + \rho\max_{1\le i\le m}\max_{1\le j\le n}\delta\left(\tilde{G}_{ij}, \tilde{G}_{ij}^+\right)} \quad (11)$$

τ_{ij} is the grey relational coefficient between $\tilde{G}_{ij}$ and $\tilde{G}_{ij}^+$. $\rho \in [0,1]$ is the distinguishing coefficient or the identification coefficient. Smaller value of distinguishing coefficient will yield in large range of grey relational coefficient. ρ is used to adjust the range of the comparison environment, and to control level of differences of the relation coefficients.

Step7.

Calculation of the degree of grey relational coefficient of each alternative from PIS using the following equation:

$$\eta_i = \sum_{j=1}^{n} w_{ij}\tau_{ij}, i = 1,2,\ldots,n. \quad (12)$$

Step8.

Ranking all the alternatives: We rank all the alternatives A_i (i = 1, 2, ..., m) according to the decreasing order of their grey relational grades η_i $(i=1,2,...,m)$. Greater value of η_i reflects the better alternative A_i.

Step9.

End.

7. ILLUSTRATIVE EXAMPLE FOR BRICK SELECTION PROCESS

Suppose that the administration of an authority is going to construct a building. For this purpose it is necessary to collect quality Bricks from various Brick Fields. After initial screening, five types of Bricks (alternatives) A_1, A_2, A_3, A_4, A_5 remain for further selection. A selection committee is formed with five decision makers or experts D_1, D_2, D_3, D_4, D_5. Six criteria of Bricks obtained from expert opinions, namely, Solidity C_1 color C_2, size and Shape C_3, strength of Brick C_4, cost of Bricks C_5 and carrying cost C_6 are considered for selection criteria. Decision maker D_t (t = 1, 2, 3, 4, 5) uses linguistic variable to evaluate the ratings of the five types of Bricks A_t (t = 1, 2, 3, 4, 5) with respect to the criterion C_j (j= 1, 2, 3, 4, 5, 6). They construct the decision matrix $G^{(t)} = \left(G_{ij}^{t}\right)_{5\times 6}$ (t= 1, 2, 3, 4, 5) shown in the tables 4, 5, 6, 7, & 8.

Table2. Conversion between linguistic variables and IFNs

Linguistic variables(Brick quality)	IFNs
Extreme high(EH)	$(0.95,0.05,0.00)$
Very high(VH)	$(0.85,0.10,0.05)$
High(H)	$(0.75,0.15,0.10)$
Medium high(MH)	$(0.65,0.25,0.10)$
Medium(M)	$(0.50,0.40,0.10)$
Medium low(ML)	$(0.35,0.55,0.10)$
Low(L)	$(0.25,0.65,0.10)$
Very low(VL)	$(0.15,0.80,0.05)$
Extreme low(EL)	$(0.05,0.95,0.00)$

Table3. Linguistic variable for the importance of the experts or decision makers

Linguistic variables	IFNs
Very important	$(1,0,0)$
Important	$(0.75,0.20,0.05)$
Medium	$(0.50,0.40,0.10)$
Unimportant	$(0.25,0.60,0.15)$
Very unimportant	$(0.10,0.80,0.10)$

TABLE 4: Decision matrix $G^{(1)}$

A_i	C_1	C_2	C_3	C_4	C_5	C_6
A_1	EH	VH	VH	H	H	M
A_2	EH	H	VH	VH	M	ML
A_3	VH	H	H	H	MH	ML
A_4	VH	H	H	VH	M	H
A_5	VH	VH	H	H	M	H

Table 5: Decision matrix $G^{(2)}$

A_i	C_1	C_2	C_3	C_4	C_5	C_6
A_1	EH	VH	VH	H	M	MH
A_2	VH	H	EH	H	H	M
A_3	VH	VH	EH	H	H	ML
A_4	EH	EH	VH	VH	M	ML
A_5	VH	VH	H	H	H	H

Table 6: Decision matrix $G^{(3)}$

A_i	C_1	C_2	C_3	C_4	C_5	C_6
A_1	VH	H	VH	MH	MH	M
A_2	VH	H	H	MH	H	ML
A_3	H	VH	H	H	M	ML
A_4	H	H	H	H	M	H
A_5	VH	VH	H	H	H	H

Table 7: Decision matrix $G^{(4)}$

A_i	C_1	C_2	C_3	C_4	C_5	C_6
A_1	VH	VH	H	H	M	ML
A_2	H	VH	H	MH	M	ML
A_3	VH	H	VH	MH	MH	H
A_4	VH	H	VH	H	MH	M
A_5	H	VH	H	MH	M	M

Table 8: Decision matrix $G^{(5)}$

A_i	C_1	C_2	C_3	C_4	C_5	C_6
A_1	EH	H	VH	H	H	M
A_2	VH	H	MH	H	MH	H
A_3	VH	VH	MH	MH	M	MH
A_4	H	H	H	MH	M	M
A_5	VH	H	VH	MH	MH	M

The selection process is done based on following the steps

1. Construct the intuitionistic fuzzy decision matrices of each decision maker. Convert the linguistic evaluation shown in Table 4-8 into IFNs by using Table 2. Then, the intuitionistic fuzzy decision matrix G^t (t = 1, 2, 3, 4, 5) of each decision maker is constructed.

2. Determine the weight of the decision makers. The importance of the decision makers in the group decision-making process is shown in Table 3. These intuitionistic fuzzy linguistic variables can be converted into IFNs. Here, importance of decision maker is considered as very important i.e. (1, 0, 0). Using equation (3), we obtain the weights of the decision makers $\gamma_t = 0.2$, (t =1, 2, 3, 4, 5).

3. Construct the aggregated intuitionistic fuzzy decision matrix based on the opinions of decision makers. Utilize the IFWA operator given by the equation (4) to aggregate the intuitionistic fuzzy decision matrices G(t) (t = 1, 2, 3, 4, 5) into a complex intuitionistic fuzzy decision matrix X .

4. Consider the weights of the criteria obtained from expert opinions. We have average weight of each criterion $w_j, (j = 1,2,\cdots,6)$ as $w_1 = 0.275, w_2 = 0.175, w_3 = 0.2, w_4 = 0.1, w_5 = 0.05, w_6 = 0.2$ such that $\sum_{j=1}^{6} w_j = 1$.

5. Determine the reference sequence based on IFNs. The reference sequence can be presented as $\tilde{G}^+ = [(1,0,0),(1,0,0),(1,0,0),(1,0,0),(1,0,0)(1,0,0)]$.

6. Calculate the grey relational coefficient of each alternative from PIS using the equation (11).

$$(\tau_{ij})_{5\times6} = \begin{bmatrix} 1.0000 & 0.7826 & 0.8182 & 0.7347 & 0.5625 & 0.4675 \\ 0.8571 & 0.7200 & 0.7500 & 0.6923 & 0.6667 & 0.4444 \\ 0.8182 & 0.7826 & 0.7500 & 0.6429 & 0.5455 & 0.4615 \\ 0.8182 & 0.7500 & 0.7500 & 0.7200 & 0.4865 & 0.5143 \\ 0.8182 & 0.8182 & 0.7200 & 0.6429 & 0.6102 & 0.5806 \end{bmatrix}$$

Calculation of $\min \delta_{ij}$ and $\max \delta_{ij}$

	C_1	C_2	C_3	C_4	C_5	C_6	$\min \delta_{ij}$	$\max \delta_{ij}$
δ_{1j}	.09	.19	.17	.22	.37	.50	.09	.50
δ_{2j}	.15	.23	.21	.25	.27	.54	.15	.54
δ_{3j}	.17	.19	.21	.29	.39	.51	.17	.51
δ_{4j}	.17	.21	.21	.23	.47	.43	.17	.47
δ_{5j}	.17	.17	.23	.29	.32	.35	.17	.35
δ_{min}							.09	
δ_{max}								.54

7. Calculate the degree of grey relational coefficient of each alternative from PIS using the following equation:

$$\eta_i = \sum_{j=1}^{n} w_j \tau_{ij}\ ,\ \ j = 1.2,3,4,5,6\ \ i = 1,2,3,4,5$$

$\eta_1 = 0.7707, \eta_2 = 0.7032, \eta_3 = 0.6957, \eta_4 = 0.7187, \eta_5 = 0.7231$

Greater the value of η_i implies the better alternative A_i.

Here, the relationship between grey relational grades is as follows:

$\eta_1 > \eta_5 > \eta_4 > \eta_2 > \eta_3$

Then, the five Bricks are ranked as:

$A_1 \succ A_5 \succ A_4 \succ A_2 \succ A_3$

Therefore, the most appropriate Brick is A_1.

8. SENSITIVITY ANALYSIS

With the change of weights of the criteria, it is observed that ranking order is sensitive with weight vectors (see table 9). We consider another weight structure of the criteria as follows:

$w_1 = 0.05 \quad w_2 = 0.22 \quad w_3 = 0.14 \quad w_4 = 0.19 \quad w_5 = 0.17 \quad w_6 = 0.23$

Then using the equation (12), we obtained $\eta_1 = 0.6795 \quad \eta_2 = 0.6533 \quad \eta_3 = 0.6391$ $\eta_4 = 0.6487 \quad \eta_5 = 0.6811$

The relationship between grey relational grades is as follows:

$\eta_5 \succ \eta_1 \succ \eta_2 \succ \eta_4 \succ \eta_3$

Therefore, the most appropriate Brick is A_5.

So, it is observed the change in weights of the criteria will produce change in the ranking order.

Note1: From grey relational coefficient matrix we have seen that for more priority level of the first four Brick criteria (Solidity, Color, Shape and Size, Strength), A_1will be the best alternative.

TABLE 9: Sensitivity analysis of weight structure and ranking of Bricks

Serial No.	Weight Structure: $w_1, w_2, w_3, w_4, w_5, w_6$	Grey Relational Grades: $\eta_1, \eta_2, \eta_3, \eta_4, \eta_5$	Ranking of Bricks	Best Alternative
1	.25, .2, .2, .15, .1, .1	.7833,.7232,.7082, .7126,.7277	$A_1 \succ A_5 \succ A_2 \succ A_4 \succ A_3$	A_1
2	.167, .167, .167, .167, .165, .167	.7276,.6884,.6668, .6732,.6984	$A_1 \succ A_5 \succ A_2 \succ A_4 \succ A_3$	A_1
3	.2, .2, .3, .2, .05, .05	.8004,.7344,.7241, .7327,.7314	$A_1 \succ A_2 \succ A_4 \succ A_5 \succ A_3$	A_1
4	.15, .1, .1, .45, .1, .1	.7437,.6982,.6660, .6968,.6849	$A_1 \succ A_2 \succ A_4 \succ A_5 \succ A_3$	A_1

9. ADVANTAGES OF THE PROPOSED APPROACH

The proposed approach is very flexible. New criteria could easily be incorporated in the model based on the need, desire and new situations. In this paper, we showed how the proposed approach could provide a well-structured, coherent, and justifiable selection practice.

10. CONCLUSION

Grey relational analysis based intuitionistic fuzzy multi-criteria group decision-making approach is a practical, versatile and powerful tool that identifies the criteria and offers a consistent structure and process for selecting Bricks by employing the concept of acceptance, rejection and indeterminacy of Intuitionistic fuzzy sets simultaneously. In this study, we demonstrated how the proposed approach could provide a well-structured, coherent, and scientific selection practice. Therefore, in future, the proposed approach can be used for dealing with multi-criteria decision-making problems such as project evaluation, supplier selection, manufacturing system, and many other areas of management decision problems. After emergence of fuzzy sets, the paradigm shift occurred in decision-making arena. In intuitionistic fuzzy sets, although degree of rejection (non membership) is independent of degree of acceptance (membership) but degree of indeterminacy (hesitancy) is dependent on degree of acceptance and rejection. However, in reality degree of indeterminacy may be independent of degree of acceptance and rejection. Therefore, the researchers feel that the degree of indeterminacy with independent characteristics should be incorporated in the selection process. In this sense, the concept of neutrosophic set due to Smarandache [25] appears to be a promising one to deal with realistic selection process. Some studies due to Ye [16, 17], and Biswas et al. [5, 6] open up new avenue of research in the field of multiple attribute decision making in neutrosophic environment. It is hoped that if grey system theory and neutrosophic logic are used simultaneously, new area of research may be opened. Although this paper has shown the effectiveness of the proposed approach, many areas need to be explored and developed in neutrosophic environment [5, 6, 16, 17] which is very realistic in nature.

11. ACKNOWLEDGMENTS

The authors are grateful to all the running stuffs and workers of various Brick fields in Santipur, Nadia, West Bengal for their encouragements and constant support during the present study. Nabin Sing (Brick specialist), Ananda Das (constructional engineer), Subhashis Banik (Senior teacher in Physics), Anupam saha (Earth quake specialist) spent their specious time and intellect in providing us their sincere and valuable suggestions in conducting the study.

REFERENCES

1. Arvey, R. D., Campion, J. E., **The Employment Interview: A Summary and Review of Recent Research.** Personnel Psychology, 35(2), 281–322, 1982.
2. Atanassov, K. T., **Intuitionistic Fuzzy Sets.** Fuzzy Sets and Systems**,** 20(1), 87–96, 1986.
3. Bawa, N. S., **Manufacture of Burnt Clay-**Bricks In Wood-Fired Clamps, Technical Paper No. 6, Building and Road Research Institute, Kumasi, Ghana, 25p, 1985.
4. Beamish, A., Donovan. W., **Village-Level Brickmaking. A Publication of the Zentrum für Entwicklungstechnologien-GATE** in: Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH, 1993.

5. Biswas, P., Pramanik, S., Giri, B. C., **Entropy Based Grey Relational Analysis Method for Multi-Attribute Decision Making under Single Valued Neutrosophic Assessments**. Neutrosophic Sets and Systems (NSS) 2, 102-110, 2014.

6. Biswas, P., Pramanik, S, Giri, B. C., **A New Methodology for Neutrosophic Multi-Attribute Decision-Making With Unknown Weight Information**. Neutrosophic Sets and Systems (NSS) 3, 2014. (In Press), 2014.

7. Boran, F. E., Genc, S., Kurt, M., Akay, D., **A Multicriteria Intuitionistic Fuzzy Group Decision Making for Supplier Selection With TOPSIS Method.** Expert Systems with Applications 36(8), 11363–11368, 2009.

8. Campion, M. A., Pursell, E. D., Brown, B. K., **Structured Interviewing; Raising The Psychometric Properties of The Employment Interview. Personnel Psychology,** 41(1), 25–42, 1988.

9. Dağdeviren, M. A., **Hybrid Multi-Criteria Decision Making Model for Personnel Selection in Manufacturing Systems**. Journal of Intelligent Manufacturing, 21, 451–460, 2010.

10. Deng, J. L., **Introduction to Grey System Theory.** The Journal of Grey System,1(1),1-24, 1989.

11. Donovan, W., **Brickmaking Field Notes - Abyei Sudan**, Unpublished Field Notes For the Harvard Institute for International Development, Cambridge, Mass., April 1981.

12. Dursun, M., Karsak, E. E., **A Fuzzy MCDM Approach for Personnel Selection.** Expert Systems with Applications**,** 37, 4324–4330, 2010.

13. Ehrgott, M., Gandibbleux, X., **Mulitple Criteria Optimization: State of The Art Annotated Bibliography Survey.** Kluwer Academic Publishers, Boston, 2002.

14. Gibney, R., Shang, J., **Decision Making in Academia: A Case of the Dean Selection Process.** Mathematical and Computer Modelling 46(7-8), 1030–1040, 2007.

15. Günör, Z., Serhadlıog˘lu, G., Kesen, S. E., **A fuzzy AHP Approach to Personnel Selection Problem.** Applied Soft Computing, 9, 641-646, 2009.

16. Ye, J., **Multicriteria Decision-Making Method Using the Correlation Coefficient Under Single-Valued Neutrosophic Environment.** International Journal of General Systems, 42(4), 386-394, 2013.

17. Ye, J., **Single Valued Neutrosophic Cross Entropy for Multicriteria Decision Making Problems**, Applied Mathematical Modelling, 38(3), 1170-1175, 2014.

18. Karsak, E. E., **Personnel Selection Using a Fuzzy MCDM Approach Based on Ideal and Anti-Ideal Solutions**. Lecture Notes in Economics and Mathematical Systems, 507, 393–402, 2001.

19. Lin, L., Yuan, X. H., Xia, Z. Q., **Multicriteria Fuzzy Decision-Making Based on Intuitionistic Fuzzy Sets**. Journal of Computers and Systems Sciences, 73(1), 84-88, 2007.

20. Liu, H. W., Wang, G. J., **Multi-Criteria Decision Making Methods Based on Intuitionistic Fuzzy Sets**. European Journal of Operational Research, 179, 220-233, 2007.

21. Liang, S. L., Wang, M. J., **Personnel Selection Using Fuzzy MCDM Algorithm**. European Journal of Operational Research, 78, 22–33, 1994.

22. Parry, J. P. M., **Brickmaking in Developing Countries**. Garston, Watford**,** British Research Institute, 88pp, 1979.

23. Pramanik, S., Mukhopadhya, D., **Grey Relational Analysis Based Intuitionistic Fuzzy Multi-Criteria Group Decision-Making Approach for Teacher Selection in Higher Education**, International Journal of Computer Applications, 34(10), 21-29, 2011.

24. Robertson, I. T., Smith, B., **Personnel Selection.** Journal of Occupational and Organizational Psychology, 74, 441–472, 2001.

25. Smarandache, F., **Neutrosophic Set. A Generalization of Intuitionistic Fuzzy Sets.** International Journal of Pure and Applied Mathematics, 24, 287-297, 2005.

26. Smith, R. C., **"Improved Moulding Devices for Hand-made Bricks", in Appropriate Technology, London,** Intermediate Technology Publications, 7(4), 1981.

27. Svare, T. I., **Better Burnt Bricks, Technical Pamphlet No. 1, Dar-es-Salaam**. National Housing and Building Research Unit, 1971.

28. Thomas, D. W., **Small-Scale Manufacture of Burned Building Brick, Arlington, Virginia, Volunteers in Technical Assistance**, 1977.

29. Vlachos, I. K., Sergiadis, G. D., **Intuitionistic Fuzzy Information-Applications to Pattern Recognition.** Pattern Recognition Letters, 28, 197–206, 2007.

30. Xu, Z. S., **Intuitionistic Fuzzy Aggregation Operators**. IEEE Transaction of Fuzzy Systems, 15(6), 1179–1187, 2007.

31. Zadeh, L. A**., Fuzzy Sets.** Information and Control, 8, 338-353, 1965.

32. Zhang, S. F., Liu, S. Y., **A GRA-Based Intuitionistic Fuzzy Multi-Criteria Group Decision Making Method for Personnel Selection**. Expert Systems with Applications, 38, 11401–11405, 2011.

---

[1] **Kalyan Mondal** (M.Sc., B. Ed.) passed B. Sc Honours and M. Sc in Mathematics in 2003 and 2005 respectively in University of Calcutta and University of Kalyani. Currently, he is an assistant Teacher of mathematics at Birnagar High School (HS), Birnagar, Ranaghat, Nadia, Pin Code: 741127, West Bengal, India. His field of research interests includes fuzzy goal programming, intuitionistic Fuzzy decision making and neutrosophic decision making. Kalyan Mondal may be reached at **kalyanmathematic@gmail.com**

[2] **Dr. Surapati Pramanik** (Ph. D., M.SC., M. Ed.) is an Assistant Professor of Mathematics, Nandalal Ghosh B. T. College, Panpur, P.O.-Narayanpur, West Bengal, India. He received Ph. D. in Mathematics in 2010 from Bengal Engineering and Science University (BESU) Shibpur, India. He is a life member of Operation Research Society of India, Indian Statistical Institute (ISI, Kolkata), Calcutta Mathematical Society, All India Association for Educational Research, Indian Association for Teacher Educators, Global Educational Research Association, Netaji Subhas Foundation, Kolkata. He is acting as Coordinator of Netaji Subhas Open University. He presented research papers in various International Conferences/ Seminars (11) and National Seminars (11) and State Level Seminars (10). He acted as reviewer for different reputed International Journals. His publication consists of more than 45 research papers published in reputed peer reviewed journals. His coauthors with him won best paper awards in State Level Seminars, West Bengal State Science and Technology Congress 2008, 2011in Mathematics respectively. He and his

colleagues published five books for B.Ed. students. Dr. Surapati Pramanik may be reached at **sura_pati@yahoo.co.in**

OIL SEEDS AREA AND PRODUCTION VARIABILITY IN BANGLADESH

Muhammad Abdul Baker CHOWDHURY[1]
Senior Research Officer, Siam Health Care, Dhaka-1205, Bangladesh

E-mail: bakerchowdhury@yahoo.com

M. Taj UDDIN[2]
Professor, Department of Statistics, Shahjalal University of Science and Technology (SUST), Sylhet-3114, Bangladesh

E-mail: mtajstat@yahoo.com

M. Jamal UDDIN[3]
Assistant Professor, Department of Statistics, Shahjalal University of Science and Technology (SUST), Sylhet-3114, Bangladesh

E-mail: mjamalu@yahoo.com

ABSTRACT

Oil seed is one of the most important sources of vegetable oil. It plays a vital role in agricultural sectors of Bangladesh. However, the production of oilseed cannot meet up its annual demand of Bangladesh. The objective of this study was to measure the change and instability of oil seeds in Bangladesh in the context of area, production, and yields. Data were extracted from the statistical year books of Bangladesh Bureau of Statistics (BBS) and the study period was 1987 to 2010. Our analysis revealed that the production and yield of oil seeds were increased sharply though the cultivable areas were decreased. The growth rate in production and yield of oil seeds were satisfactory over the study period although they were not stable. Moreover, the results showed that it is not sufficient to fulfil the present demand of vegetable oil in Bangladesh. We recommend policy makers and stakeholders can give due attention to improve this sector for the betterment of the vegetable sector sustainable crop security in Bangladesh.

Keywords: *change and volatility; production; area; oilseed; growth rate*

INTRODUCTION

Agriculture is the single most important sector of the economy in Bangladesh. Bangladesh has suitable climate and soil conditions for the production of a variety of oilseed species all the year round. But since her independence to current date there is continuous decline in both acreage and total production of oilseeds except some exceptional years (Rahman, Chowdhury et al. 2009). A comparison between the acreage and production of oilseeds in 2005-06 compared to 2000-2001 shows that reduction in acreage was 18.59%, However, the production was got increased by more than 55% of the same of 2000-2001 (Rahman, Chowdhury et al. 2009). The oil production in Bangladesh is decreasing due to the replacement of oil crop area by HYV Boro rice and high population pressure (Rahman, Akhtaruzzaman et al. 2001, Amin, Jahan et al. 2009). Less than 2.37% of the total cultivatable lands were used to cultivate oil seeds (Program 2011). The present annual production of oilseed is about 716 thousand metric tons (Bangladesh Bureau of Statistics 2011) this can't satisfy the present demand of consumption. Therefore the country met up it through import. Bangladesh imported 116833.767 metric tons of oilseeds in the fiscal year of 2010-11 which valued Tk. 4960126 (Bangladesh Bureau of Statistics 2011).

In recent year, the department of agricultural extension (DAE) and Ministry of Agriculture, Government of Bangladesh have considered pulses and oilseed as a high priority subsector and have taken a plan titled "Pulses and Oil crops Research and Development Vision: 2030" (Rahman, Chowdhury et al. 2009) to increase the oilseed and pulse production. To fulfill the increasing demand of the country, it is necessary to increase the production of this important cereal crop. The rate of increase in production and yield of oil seeds should be increased and kept stable. However, there is a lack of research to improve this sector and it requires support with effective approach for adoption of the technology along with appropriate market linkage for good price to the producers. Therefore this study aimed to assess the change and instability of oil seed relationship among area, production, and yield as well as the growth rate of oil seeds in Bangladesh.

MATERIALS AND METHODS

Data: We considered secondary data on area and production of oil seeds for a period of 24 years (from 1987 to 2010). Data were collected from different issues of the Statistical Yearbook of Bangladesh published by Bangladesh Bureau of Statistics (BBS) (Bangladesh Bureau of Statistics 2011). The study period was divided into two period's viz., first period from 1987 to 1998 and second period from 1999 to 2010 to compare in area, production and yield of oil seeds.

Statistical analysis: To examine the nature of change, instability and degree of relationship in area, production and yields of oil seeds, various descriptive statistical tools (mean, correlation coefficient, simple linear regression technique, semi log growth model and t-test) were used. These statistics were suggested by Hasan et al (Hasan, Miah et al. 2008) as a better measurement to measure the change and instability. The statistical data analysis was performed by using SPSS 17.0.

REGRESSION ANALYSIS

To estimate the parameter, simple linear regression models were fitted to examine the change of production by the change of area.
The model can be expressed as:

$$y = \alpha + \beta x + e$$

Where, $e \sim N(0, \sigma^2)$, y is the production (in ton), x is the area (in acre), α is the intercept and β is the regression coefficient of the model.

MEASUREMENT OF GROWTH RATE

The growth rates of area, production and yield of oil seed were worked out by fitting a semi-log function of the type: $\log y = \alpha + \beta t$, where, y is the area (in acre) or production (in ton) or yield (ton/acre) and t is the time period (in year).

MEASUREMENT OF INSTABILITY

An index of instability was computed for examining the nature and degree of volatility in area, production, and yield of oilseed and pulse in Bangladesh. The co-efficient of variation (CV) was worked out for area, production, and yield to measure of variability. However, simple CV does not explain properly the trend component inherent in the time series data. Alternatively, the Coefficient of variation around the trend (CV_t) rather than co-efficient of variation around the mean (CV) was suggested by Cuddy and Della[3] as a better measure of variability.

A linear trend $y = \alpha + \beta t + e$ was fitted to the indices of area, production and yield for the period 1987-2010 and trend co-efficient 'β' was tested for significance. Whenever the trend co-efficient was found significant, the index of instability was constructed as follows:

$$cv_t = (cv) \times \sqrt{1 - R^2}$$

Where, $cv = \frac{s}{\bar{x}} \times 100$, $\bar{x}$ and s are the mean and standard deviation of the sample period.

In words, co-efficient of variation (CV_t) around the mean was multiplied by the square root of the proportion of the variation, which was unexplained by the trend equation, $y = \alpha + \beta t + e$.

RESULTS

Change in area, production and yield

The amount of land area of oil seed has decreased in the second period while the production has increased significantly. But interestingly, though the cultivable area of oil seeds has reduced, the yield of oil seeds has increased significantly (Table 1). From the analysis, it is evident that there is a significant change in area, production and yield of oil seed in Bangladesh.

Table 1. Change in area, production and yield of oil seeds in Bangladesh

Field of Measurement	Mean Value 1st Period (1987-98)	2nd Period (1999-2010)	t-Value	P(T<t) two tail
Area (in acre)	1317667	916083[a]	5.05	0.000
Production (in ton)	440750	562500	-1.77	0.100
Yield (ton/acre)	0.3365	0.6577 [b]	-3.24	0.008

a and b represents significant at 1% and 5% level of significance

A frequently used technique for measuring the changing attitude of area and production of any crop is correlation. Table 2 shows that the area and production of oil seed is strongly correlated (r= 0.932) for the first period which is significant at 1% level. On the other hand, it is found that the relation between production and area of oil seed is negative for the whole period and second period. The negative value of correlation between production and area may be due to other quantitative and qualitative reason which is beyond this analysis.

Table 2. Relationship between area and production of oil seed in Bangladesh

Criteria	Value of Correlation (r)		P(T<t) two tail
Area Vs production	Whole Period	-0.361^{b}	0.042
	1st Period	0.932^{a}	0.000
	2nd Period	-0.526	0.079

a and b represents significant at 1% and 5% level of significance

The simple linear regression models were fitted for estimating the response of production of oil seeds due to the change of the respective area. From the regression analysis we found the average production of oil seed has increased 0.276 times for a unit change in area for the first period, but it is in decreasing rate for the second period and whole period by -0.661 and -0.227 times respectively (Table 3).

Table 3. Testing dependency of production on area of oil seed in Bangladesh

Period	Constant Value	Reg. Coefficient	t- Value	P(T<=t) two tail
Whole Period	755138.77	-0.227	-1.814	0.083
1st Period	76619.35	0.276	8.123^{a}	0.000
2nd Period	1168000	-0.661	-1.955	0.079

a and b represents significant at 1% and 5% level of significance

GROWTH RATE OF AREA, PRODUCTION AND YIELD

The growth rate of area, production and yield provides a good measure of change in past and acceptable indication of change in future. The semi log model is used to measure the growth rate. From the analysis we found that the growth rates of areas of oil seeds for the whole period and second period is negative which is highly significant at 1% level. In spite of reduction of area, the growth rates of the production and yield of oil seeds are positive. It is also observed that the growth in yield of oil seeds has improved rapidly in second period (Table 4). Therefore it is clear that the yield of oil seed has increased though the cultivable area is reduced.

Table 4. Growth rate of area, production and yield of oil seed in Bangladesh

Field of Measurement	Sample period	Growth Rate (%)	P(T<=t)
Area	Whole Period	-2.5[a]	0.000
	1st Period	2.4	0.172
	2nd Period	-4.5[a]	0.000
Production	Whole Period	2.[3b]	0.003
	1st Period	3.0[b]	0.032
	2nd Period	6.1[b]	0.030
Yield	Whole Period	4.9[a]	0.000
	1st Period	0.60	0.291
	2nd Period	10.6[b]	0.003

a and b represents significant at 1% and 5% level of significance

INSTABILITY OF AREA, PRODUCTION AND YIELD

The agricultural production of Bangladesh is fluctuated by many factors like natural calamities: floods, droughts, cyclon etc. It is more or less common scenery of the country. Our analysis supports this hypothesis. The area and the production of oil seeds during the whole period and second period showed the highest degree of instability, which is significant at 1% level. The area of oil seeds during the second period also fluctuated significantly. The yield of oil seeds during whole period is also showed variation significantly at 1% level (Table 5). Therefore it is clear from the analysis that oil seed has showed instability during the study period.

Table 5. Instability in area, production and yield of oil seed in Bangladesh

Field of Measurement	Measurement Statistics	Whole Period (1987-2010)	1st Period (1987-1998)	2nd Period (1999-2010)
Area	CV	25	15.57	20
	R-square	0.429	0.157	0.739
	P(T<=t) two tail	0.000[a]	0.203	0.000[a]
	D-W	0.692	1.173	0.975
	CV around trend line	17.81	14.30	10.22
Production	CV	35	13.82	41
	R-square	0.273	0.432	0.254
	P(T<=t) two tail	0.009[b]	0.020[b]	2.192
	D-W	1.195	1.558	0.095
	CV around trend line	29.92	10.22	35.4
Yield	CV	8.71	1.76	15.06
	R-square	0.468	0.098	0.414
	P(T<=t) two tail	0.000[a]	0.322	0.024b
	D-W	1.563	0.867	2.074
	CV around trend line	6.35	1.67	11.52

a and b represents significant at 1% and 5% level of significance

DISCUSSION AND CONCLUSION

Our analysis shows that the overall production of oil seed in Bangladesh is satisfactory and the average production and yield of oil seed has increased from first period to second period though the average area has decreased over the study period. The yield has increased almost double from first period to second period where as the area reduced approximately 30%; which gives the evidence of increasing tendency of production and yield of oil seed. This may happen for the development of good quality of seed over time and/or modern cultivations system.. While testing the dependency of oilseed production on cultivatable land area a negative trend was observed for the whole period and the second period. It may be the caused by damage of crops by unfavourable weather (like, floods, droughts, storms etc.)

It can be concluded that the oil seed production in Bangladesh has bright future to mitigate the increasing demand. Through the production trend of oil seed in Bangladesh is satisfactory, it is not sufficient for the country's demand. Therefore, researchers, policy makers, and farmers should give proper attention to increase oil seed production.

References

1. Amin, A. K. M. R., et al. **Growth Dynamics of Soybean (Glycine max L.) As Affected by Varieties and Timing of Irrigation,** American-Eurasian Journal of Agronomy 2(2): 95-103, 2009.

2. Bangladesh Bureau of Statistics, B**. Statistical Yearbook of Bangladesh**. Dhaka, Bangladesh Bureau of Statistics (BBS), 2011.

3. Bangladesh Bureau of Statistics, B. **Year Book of Agricultural Statistics of Bangladesh**. Dhaka, Bangladesh Bureau of Statistics (BBS), 2011.

4. Hasan, M. N., et al. **Change and instability in area and production of wheat and maize in Bangladesh,** Bangladesh Journal of Agricultural Research 33(3): 409-417, 2008.

5. Program, W. F. **Agricultural Production & National Food Balance**, 2011. Retrieved December 22, 2012, from http://www.foodsecurityatlas.org/bgd/country/availability/agricultural-production.

6. Rahman, M., et al. **Prospects of Safflower as a Minor Oilseed Crop in Bangladesh**. Proceedings of the 5th International Safflower Conference, Williston, North Dakota and Sidney, Montana, USA, 23-27 July, 2001. Safflower: a multipurpose species with unexploited potential and world adaptability., Department of Plant Pathology, North Dakota State University.

7. Rahman, M. L., et al. **Agricultural Research Priority: Vision-2030 and Beyond,** Bangladesh Agricultural Research Council, 2009

APPENDIX TABLE. Area, production and yield of oil seeds from 1987 to 2010

Year	Cultivable area (in acre)	Production (ton)	Yield (ton/acre)
1987	678000	253500	0.373893805
1988	1351000	448500	0.331976314
1989	1451000	434000	0.299104066
1990	1418000	438000	0.308885755
1991	1407000	448000	0.31840796
1992	1399000	462000	0.330235883
1993	1319000	449000	0.340409401
1994	1381000	471000	0.341057205
1995	1307000	453000	0.346595256
1996	1370000	471000	0.34379562
1997	1344000	478000	0.355654762
1998	1387000	483000	0.348233598
1999	1364000	449000	0.329178886
2000	1078000	406000	0.376623377
2001	1009000	385000	0.381565907
2002	955000	376000	0.393717277
2003	988000	368000	0.372469636
2004	960000	406000	0.422916667
2005	798000	1180000	1.478696742
2006	747000	595000	0.796519411
2007	742000	625000	0.842318059
2008	777000	642000	0.826254826
2009	773000	602000	0.778783959
2010	802000	716000	0.89276808

[1] **Muhammad Abdul Baker Chowdhury** is Senior Research Office at Siam Health Care (SHC) formerly he was Researcher Officer at SHC. He obtained his academic education in the area of statistics from the Shahjalal University of Science & Technology; Bangladesh. Currently he is doing a specialized master's program in Population Sciences at University of Dhaka. Mr. Chowdhury has received training and various skills building workshops from International Center for Dirrahoeal Disease Research, Bangladesh (icddr,b), National Academy for Educational Management (NAEM) and University of Dhaka. He is a life member of Bangladesh Statistical Association (BSA). He has a good command in statistical data analysis package SPSS, STATA, R.
Mr. Chowdhury is closely working on preparing research proposals and implementing the research projects. By this time he has accomplished several research projects successfully. Moreover he has published some scholarly research articles in some renowned journals.
[2] **Mohammed Taj Uddin** is currently serving as professor of Statistics at Shahjalal University of Science & Technology, Sylhet. He is working in the academic arena for more than 20 years. Mr. Uddin's research interest is statistical modeling, time series analysis, population aging, and biostatistics. He has already published more than forty articles in different journals conferences etc.
[3] **Md. Jamal Uddin,** Assistant Professor of Statistics, Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh. His research interests are in epidemiological study (causal inference), Statistical modeling, and Demography. Currently he is working on several epidemiological studies.

QUANTITATIVE RISK MANAGEMENT TECHNIQUES USING INTERVAL ANALYSIS, WITH APPLICATIONS TO FINANCE AND INSURANCE

Silvia DEDU
Ph.D., Bucharest University of Economic Studies, Department of Applied Mathematics;
Romanian Academy, Institute of National Economy

E-mail: silvia.dedu@csie.ase.ro

Florentin ŞERBAN
Ph.D. candidate, University of Bucharest, Faculty of Mathematics and Computer Science
Ph.D., Bucharest University of Economic Studies, Department of Applied Mathematics

Ana TUDORACHE
Ph.D., Bucharest University of Economic Studies, Department of Applied Mathematics

ABSTRACT

In this paper we study some risk management techniques using optimization problems under uncertainty. In decision making problems under uncertainty, the parameters of the models used can not be exactly described by real numbers, because of the imprecision of the data. In order to overcome this drawback the uncertainty of the parameters can be modeled by using interval numbers and interval random variables. Concepts of interval analysis are introduced in this article. Computational results are provided.

Keywords: *risk management, optimization, uncertainty, interval analysis*

1. INTRODUCTION

Interval analysis was introduced by Moore [4, 5]. The growing efficiency of interval analysis for solving various real life problems determined the extension of its concepts to the probabilistic case. Thus, the classical concept of random variable was extended to interval random variables, which has the ability to represent not only the randomness character, using the concepts of probability theory, but also imprecision and non-specifficity, using the concepts of interval analysis. The interval analysis based approach provides mathematical models and computational tools for modeling data and for solving optimization problems under uncertainty.

The occurrence of randomness in the model parameters can be formulated using stochastic programming models. Stochastic programming is widely used in many real decision making problems which arise in economy, engineering, social sciences and many other domains. It has been applied to a wide variety of fields such as manufacturing product and capacity planning, electrical generation capacity planning, financial planning and control, supply chain management, dairy farm expansion planning, macroeconomic modeling and planning, portfolio selection, transportation, telecommunications, banking and insurance. Recently a lot of papers investigate different methods for solving the portfolio selection problem [2,3, 6-10].

In this paper we study some quantitative risk management techniques based on interval analysis which can be used to model the uncertainty of data. The fundamental concepts of interval analysis are presented in Section 2. The theoretical basis provided is illustrated in Section 3 to model some financial data from Bucharest Stock Exchange. The results can be used to solve decisiona making problems under uncertainty. In Section 4 some conclusions are presented.

2. INTERVAL ANALYSIS

2.1. INTERVAL NUMBERS

Let x^L, x^U be real numbers, $x^L \leq x^U$.

Definition 2.1. An *interval number* is a set defined by:

$X = \{x \in \mathbf{R} | x^L \leq x \leq x^U;\ x^L, x^U \in \mathbf{R}\}$.

Remark 2.1. We will denote by $[x]$ the interval number $[x^L, x^U]$, with $x^L, x^U \in \mathbf{R}$.

We will denote by $\mathbf{IR}$ the set of all interval numbers.

Remark 2.2. If $x^L = x^U$, then the interval number $[x^L, x^U]$ is said to be a *degenerate* interval number. Otherwise, this is said to be a *proper* interval number.

Remark 2.3. A real number $x \in \mathbf{R}$ is equivalent with the interval number $[x, x]$.

Definition 2.2. An interval number $x = [x^L, x^U]$ is said to be:

- *negative*, if $x^U < 0$;
- *positive*, if $x^L > 0$;
- *nonnegative*, if $x^L \geq 0$;
- *nonpositive*, if $x^U \leq 0$.

2.2. THE INTERVAL ARITHMETICS

Many relations and operations defined on sets or pairs of real numbers can be extended to operations on intervals.

Let $x = [x^L, x^U]$ and $y = [y^L, y^U]$ be interval numbers.

Definition 2.3. The *equality* between interval numbers is defined by:

$[x] = [y]$ if and only if $x^L = y^L$ and $x^U = y^U$.

Definition 2.4. The *maximum* between two interval numbers is defined by:

$\max\{[x],[y]\} = [\max\{x^L, y^L\}, \max\{x^U, y^U\}]$.

Definition 2.5. The *minimum* between two interval numbers is defined by:

$\min\{[x],[y]\} = [\min\{x^L, y^L\}, \min\{x^U, y^U\}]$.

Definition 2.6. The *intersection* between two interval numbers is defined by:

$[x] \cap [y] = [\max\{x^L, y^L\}, \min\{x^U, y^U\}]$.

Definition 2.7. The *union* between two interval numbers is defined by:

$$[x]\cup[y]=\begin{cases}\left[\min\{x^L,y^L\},\max\{x^U,y^U\}\right], & \text{if } [x]\cap[y]\neq\varnothing\\ \text{undefined}, & \text{otherwise}\end{cases}.$$

Definition 2.8. The *median* of the interval number $x=[x^L,x^U]$ is defined by:

$$m(x)=\frac{x^L+x^U}{2}.$$

Definition 2.9. The *length* of the interval number $x=[x^L,x^U]$ is defined by:

$$l(x)=x^U-x^L.$$

Definition 2.10. The *absolute value* of the interval number $x=[x^L,x^U]$ is defined by:

$$|x|=\{|y|:x^L\le y\le x^U\}.$$

Let $x=[x^L,x^U]$ an interval number and $a\in\mathbf{R}$ a real number.

Definition 2.11. The *product between the real number a and the interval number* $[x]$ is defined by:

$$a\cdot[x]=\{a\cdot x|x\in[x]\}=\begin{cases}[a\cdot x^L,a\cdot x^U], & \text{if } a>0\\ [a\cdot x^U,a\cdot x^L], & \text{if } a<0\\ [0], & \text{if } a=0\end{cases}.$$

Let "$\circ$" be one of the basic operations with real numbers, denoted by "$+$", "$-$", "$\cdot$", "$:$".

Definition 2.12. The *operation* $[x]\circ[y]$ is defined by:

$$[x]\circ[y]=\{x\circ y|x\in X,y\in Y\}.$$

The *summation* of two interval numbers defined by:

$$[x]+[y]=[x^L+y^L,x^U+y^U].$$

The *subtraction* of two interval numbers defined by:

$$[x]-[y]=[x^L-y^U,x^U-y^L].$$

2.3. INEQUALITIES BETWEEN INTERVAL NUMBERS

We will extend the classical inequality relations between real numbers to inequality relations between interval numbers.

Let $x=[x^L,x^U]$ and $y=[y^L,y^U]$ be interval numbers, with $x^L,x^U,y^L,y^U\in\overline{\mathbf{R}}$.

Definition 2.13. We say that:

1) $[x]<[y]$ if $x^U<y^L$;
2) $[x]\le[y]$ if $x^L<y^L$ and $x^U<y^U$;
3) $[x]\le_I[y]$ if $\begin{cases}x^L\le y^L\\ x^U<y^U\end{cases}$ or $\begin{cases}x^L<y^L\\ x^U\le y^U\end{cases}$ or $\begin{cases}x^L\le y^L\\ x^U\le y^U\end{cases}$;
4) $[x]\prec[y]$ if $x^L<y^U$

2.4. INTERVAL RANDOM VARIABLES

Let (Ω, K, P) be a probability space and $\mathbf{IR}$ be the set of all interval numbers.

Definition 2.14. An *interval random variable* is an application $[X]:\Omega \to \mathbf{IR}$,

$x(\omega) = [x^L(\omega), x^U(\omega)]$,

where $X^L, X^U : \Omega \to \mathbf{IR}$ are random variables, with $X^L \le X^U$ almost surely.

Definition 2.15. The interval random variable $[X]:\Omega \to \mathbf{IR}$ is said to be a *discrete interval random variable* if it takes values in a discrete subset of the real numbers.

The *probability distribution* of the discrete interval random variable $[X]:\Omega \to \mathbf{IR}$ can be expressed as:

$$[X]:\begin{pmatrix} [x_i] \\ p_i \end{pmatrix}_{i\in I},$$

where I is a discrete set of real numbers, or as:

$$[X]:\begin{pmatrix} [x_i^L, x_i^U] \\ p_i \end{pmatrix}_{i\in I}.$$

If $I = \{1, 2, ...n\}$ is a finite set, then the probability distribution of the discrete interval random variable $[X]:\Omega \to \mathbf{IR}$ can be expressed as:

$$[X]:\begin{pmatrix} [x_1^L, x_1^U] & [x_2^L, x_2^U] & .... & [x_n^L, x_n^U] \\ p_1 & p_2 & .... & p_n \end{pmatrix}.$$

Definition 2.16. The *probability mass function* of the discrete interval random variable $[X]:\Omega \to \mathbf{IR}$ is the function defined by

$$f:\mathbf{R} \to [0,1],\ \ f(x) = \mathrm{P}([X] = [x]).$$

Definition 2.17. The *expectation* of the discrete interval random variable $[X]:\Omega \to \mathbf{IR}$ is the interval number defined by

$$\mathrm{E}([X]) = \sum_{i\in I} [x_i] p_i .$$

Definition 2.18. The *variance* of the discrete interval random variable $[X]:\Omega \to \mathbf{IR}$ is the interval number defined by

$$\mathrm{Var}([X]) = \sum_{i\in I} [x_i]^2 p_i - \left(\sum_{i\in I} [x_i] p_i\right)^2 .$$

3. CASE STUDY

In this section we will use the concepts defined in the previous section to model financial data. We have collected information concerning the values of the closing price ot a stock, TEL, listed at Bucharest Stock Exchange, between March 1, 2014 and June 25, 2014. The minimal value of the historical prices was 148400 and the maximal value was 228200, with an amplitude of 78800. **Table 1** presents the results obtained by organizing data into eight groups, with the amplitude of each group equal to 10000.

Table 1. The values of the historical prices

Interval	Absolute frequency
[148400; 158400)	50
[158400; 168400)	6
[168400; 178400)	8
[178400; 188400)	20
[188400; 198400)	16
[198400; 208400)	8
[208400; 218400)	4
[218400; 228400)	8

We model the data using a discrete interval random variable $[X]$, defined by:

$$[X]:\begin{pmatrix} [x_1^L,x_1^U] & [x_2^L,x_2^U] & \ldots\ldots\ldots\ldots & [x_8^L,x_8^U] \\ p_1 & p_2 & \ldots\ldots\ldots\ldots & p_8 \end{pmatrix},$$

with

$$[x_1^L,x_1^U]=[148400,158400];\ p_1=\frac{50}{120}=0.41666;$$

$$[x_2^L,x_2^U]=[158400,168400];\ p_2=\frac{6}{120}=0.05000;$$

$$[x_3^L,x_3^U]=[168400,178400];\ p_3=\frac{8}{120}=0.06667;$$

$$[x_4^L,x_4^U]=[178400,188400];\ p_4=\frac{20}{120}=0.16667;$$

$$[x_5^L,x_5^U]=[188400,198400];\ p_5=\frac{16}{120}=0.13333;$$

$$[x_6^L,x_6^U]=[198400,208400];\ p_6=\frac{8}{120}=0.06667;$$

$$[x_7^L,x_7^U]=[208400,218400];\ p_7=\frac{4}{120}=0.03333;$$

$$[x_8^L,x_8^U]=[218400,228400];\ p_8=\frac{8}{120}=0.06667.$$

$$[X]:10^5\cdot\begin{pmatrix} [1.484,1.584] & [1.484,1.584] & [1.484,1.584] & [1.484,1.584] & [1.484,1.584] & [1.484,1.584] & [1.484,1.584] & [1.484,1.584] \\ 0.41666 & 0.05000 & 0.06667 & 0.16667 & 0.13333 & 0.06667 & 0.03333 & 0.06667 \end{pmatrix}$$

The distribution of the discrete interval random variable can be used to compute relevant numerical values, such as expectation, standard deviation, skewness and kurtosis. The results obtained can be used to solve portfolio optimization problems.

4. CONCLUSIONS

The new approach based on interval analysis provides mathematical models and computational tools for modeling the imprecision of financial data and for solving decision making problems under uncertainty.

ACKNOWLEDGEMENT:
This paper has been financially supported within the project entitled ***"Horizon 2020 - Doctoral and Postdoctoral Studies: Promoting the National Interest through Excellence, Competitiveness and Responsibility in the Field of Romanian Fundamental and Applied Scientific Research",*** contract number POSDRU/159/1.5/S/140106. This project is co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013. **Investing in people!**

REFERENCES

1. Alefeld, G., Herzberger, J. **Introduction to Interval Computation**, Academic Press, 1983.
2. Barik, S.K, Biswal, M.P., Chakravartay, D. **Multiobjective Two-Stage Stochastic Programming Problems with Interval Discrete Random Variables**, Advances in Operations Research, Article ID 279181, 2012.
3. Hansen, E., Walster, G.W. **Global optimization using interval analysis**, Marcel Decker Inc., 2004.
4. Moore, R.E. **Interval Analysis,** Prentice Hall, 1966.
5. Moore, R.E. **Methods and Application of Interval Analysis**, SIAM Philadelphia, 1979.
6. Şerban, F., Ştefănescu, M.V., Dedu, S. **Building an Optimal Portfolio Using a Mean-VaR Framework**, Mathematical Methods and Techniques in Engineering and Environmental Science, 126-131, 2011.
7. Şerban, F., Ştefănescu M.V., Dedu, S. **The Relationship Profitability - Risk for an Optimal Portfolio Building with Two Risky Assets and a Risk-Free Asset**, International Journal of Applied Mathematics and Informatics, 5, 4, 299 – 306, 2011.
8. Toma, A., Leoni-Aubin, S. **Portfolio selection using minimum pseudodistance estimators**, Economic Computation and Economic Cybernetics Studies and Research, 47, 1, 97-113, 2013.
9. Toma, A., **Model selection criteria using divergences**, Entropy 16(5), 2686-2698, 2014.

10. Toma, A., Dedu, S. **Quantitative techniques for financial risk assesment: a comparative approach using different risk measures and estimation methods**, Procedia Economics and Finance 8, 712-719, Elsevier, 2014.