A methodological note on computing FGT indicators from CGE models

Dorothée Boccanfuso[+] and Luc Savard[*]

December 2001

[+] CREFA – Université Laval, BP 11007 CD Annexe, Dakar, Senegal ; tel : 221 644 7544 ;
Email : dboc@ecn.ulaval.ca.
[*] IDRC – WARO, BP 11007 CD Annexe, Dakar, Senegal ; tel : 221 864 0000 # 2072 ;
Email : lsavard@idrc.org.sn.

1. Introduction

This technical note present two methods to evaluate poverty with the Foster, Greer and Thorbecke (FGT) indices for a base year and after simulation with an endogenous poverty line. The first one is a non-parametric method and can be easily applied with free DAD software. The second is parametric. For the latter, we present some technical elements to help to choose a density function for the income distributions, to estimate the parameters of this distribution and finally to evaluate poverty with the Foster, Greer and Thorbecke (FGT) indices.

In the next section, the non-parametric method is presented with an empirical example. In Section 3, we present the parametric method in detail. First, the maximum likelihood (ML) estimation method is presented. This method is often pre-programmed in econometric or mathematical software. We present an example with Matlab[1] to estimate by ML the parameters of the Beta distribution. Generally, for the poverty analysis, the most frequent used probability density functions (pdf) are the Beta, Lognormal or truncated Normal or Pareto. But the Beta distribution is more flexible than others as it allows asymmetric forms (Poirier (1995) or Johnston and Kotz (1970)). After a short review of the FGT indices for continuous functions, we then propose a program to compute these indices. Before the conclusion, we present a complete program to estimate the parameters, calculate the FGT indices and to plot the distributions (base and simulated).

2. Non-parametric method: DAD's method

2.1. The Kernel method

The Kernel method is the most mathematically studied and commonly used non-parametric density estimation method. The Kernel estimation of *f(x)* is defined as:

$$\hat{f}(x)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K(\frac{x-x_i}{h}) \qquad (2\text{-}1)$$

where the Kernel function *K()* is generally unimodal, symmetric, bounded density function, for instance, the standard normal density function and the *h* is called smoothing parameter. Imagine it intuitively, a "bump" is placed on each data point, and the sum of all "bumps" reflects the overall distribution of all data points. The Kernel function determines the shape of each bump while the smoothing parameter determines the width of each bump. There are some properties such as:

- no need to know the data range in advance,
- $\hat{f}(x)$ itself forms a density function which inherits all the continuity, differentiability and integrability properties of the Kernel function,
- K and h are two factors affecting the accuracy but essentially by the smoothing parameter.

The estimation consists of measuring and minimizing the global error between the density estimation and the real underlying density function such as:

$$Mean\ Integrated\ Squared\ Error\ (\hat{f},f)=E\int(\hat{f}(x)-f(x))^2dx \quad \rightarrow \quad 0 \qquad (2\text{-}2)$$

The software DAD uses precisely the non-parametric method of Gaussian Kernel type[2].

---

[1] Matlab 5.1, Student version is used.

[2] However, Livny, Ramakrisnhnan and Zhang (1996) found that the Kernel method does not scale up well for large datasets.

2.2. Application of the method

In this section, we present the methodological steps to measure FGT poverty indices with the non-parametric method and the free DAD Software.

i. With DAD, you first load your data file with the base year income vector and to obtain the simulated income vector, you need to multiply the base year vector by the variation obtained in Gams output. You can do this exercise with EXCEL or DAD 4.0. With DAD, you do *Edit* and *Compute column*. You choose *series 1 x number* where the *series 1* is the base year vector and the *number*, the variation obtained in Gams[3]. For the poverty line, you do this operation again. You choose *series 1+number* in the operation frame. The *series 1* will be your new vector then choose an empty vector. The *number* will be the value of the poverty line.

ii. You can now calculate the FGT's measures with DAD with *Poverty* and *FGT index*. The output is the following:

Table 1: *FGT (Poverty) for base year and after simulation*

Session Date	**Mon Dec 03 15:50:56 GMT+00:00 2001**	**Session Date**	**Mon Dec 03 15:58:32 GMT+00:00 2001**
Execution Time	**0.02 sec**	**Execution Time**	**0.02 sec**
FileName	**mrur.dad**	**FileName**	**mrur.dad**
OBS	**371**	**OBS**	**371**
Variable of interest	**revo (Base Year)**	**Variable of interest**	**revb (After Simulation)**
Weight variable	**Vector5**	**Weight variable**	**Vector5**
Group variable	**No Selection**	**Group variable**	**No Selection**
Group Number	**1**	**Group Number**	**1**
Normalised	**YES**	**Normalised**	**YES**
alpha	**0**	**alpha**	**0**
Index Value	**0.68463976**	**Index Value**	**0.69003062**
	(0.02412411)		**(0.02401101)**
EDE Value	**0.00000000**	**EDE Value**	**0.00000000**
	(0.00000000)		**(0.00000000)**
Poverty Line	**930000.00000000**	**Poverty Line**	**891126.00000000**
	(0.00000000)		**(0.00000000)**

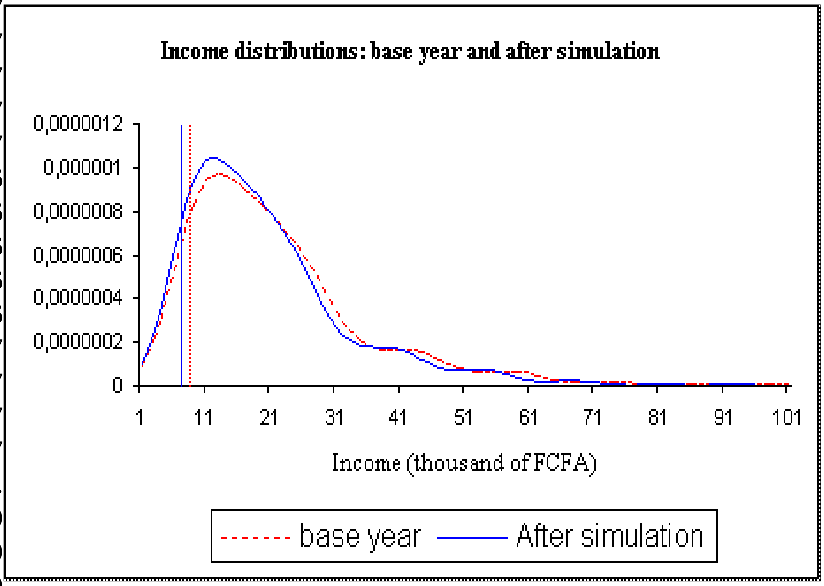
iii. To plot both distributions, you need to go to *Distribution* and *Density function* and graph the base year distribution and the simulated income distribution (do not forget to initialize the graph before).

iv. If you want to plot this graph, you need to save the coordinates in DAD (*Graph* and *save XY*) and to export them in EXCEL to plot de graphic. You can also add the poverty lines to your graph. The output will be the following:

[3] Do not forget to change the result case by an empty vector.

Table 2 : *The income distributions*

	Base year	After simulation
0	9,2625E-08	9,948E-08
40000	1,48247E-07	1,64436E-07
80000	2,2266E-07	2,52664E-07
120000	3,14922E-07	3,62462E-07
160000	4,21024E-07	4,87766E-07
200000	5,34232E-07	6,18929E-07
240000	6,46234E-07	7,44688E-07
280000	7,48719E-07	8,54623E-07
320000	8,34879E-07	9,41184E-07
360000	9,0041E-07	1,00074E-06
400000	9,43841E-07	1,03356E-06
440000	9,66313E-07	1,04318E-06
480000	9,71024E-07	1,03525E-06
520000	9,62438E-07	1,01607E-06
560000	9,45298E-07	9,91019E-07
600000	9,23563E-07	9,63285E-07
640000	8,99631E-07	9,33707E-07
680000	8,74225E-07	9,01722E-07
...	...	...
3840000	6,714E-09	2,728E-09
3880000	8,027E-09	1,488E-09
3920000	8,675E-09	7,2E-10
3960000	8,453E-09	3,08E-10
4000000	7,419E-09	1,17E-10



This method[4] allows to evaluate the FGT measures for the base year income vector and for the simulated income vector after transformation with the mean increase of the income of the representative household of the group (the output of GAMS). To apply this non-parametric method, you only need the software DAD, which is freely distributed on the CREFA web site (http://www.mimap.ecn.ulaval.ca/).

In the next section, a parametric method is developed. The main difference between these two approaches is that for the second, you need to postulate a distribution to estimate the parameters and to calculate the FGT indices. This approach requires you to choose a distribution that fits well the real distribution. The parametric approach is particularly useful when the primary household- or individual-level data is unavailable, but information is available on the underlying distribution (mean, minimum, etc.) However, the exercise is a bit more complicated.

3. A parametric method

3.1. Parameters estimation

3.1.1. The maximum likelihood estimator

The first step when you choice a continuous distribution is to estimate the parameters of this distribution. Most statistical software are programmed to do this estimation for commonly used

[4] This method is a (discrete) approximation of the true distribution.

distributions (Normal, Lognormal, Beta[5], …). But if this is not the case, you can program this estimation easily with the maximum likelihood estimation method, which is one popular criterion for goodness of fit. The likelihood has the same form as the pdf but the parameters are treated as known constants and the sample values (X) are the variables. The likelihood function thus reverses the roles. The sample values (x_i) are already observed then they are held constant and the variables become the unknown parameters. Maximum likelihood estimation (MLE) involves finding the values of the parameters that give the highest likelihood to the particular set of data.

In a random sample of N observations, the density of each observation is $f(x_i, \theta)$. We suppose that the N observations are independent then the joint density is :

$$f(x_1,x_2,...,x_N,\theta)=f(x_1,\theta).f(x_2,\theta).....f(x_N,\theta)=\prod_{i=1}^{N} f(x_i,\theta)=L(\theta/x_1,x_2,...,x_N) \quad (3\text{-}1)$$

The function $L(\theta/X)$ is the likelihood function for θ given the data X. The ML Estimator is the solution to the optimization problem:

$$\hat{\theta}_{ML} \equiv \underset{\theta}{\operatorname{argmax}} L(\theta/x_1,x_2,...,x_N) \quad (3\text{-}2)$$

<u>Example</u> :

Suppose $X = [x_1\ x_2 \ldots\ x_N] \sim \text{iid } N(\mu,\sigma^2)$ the vector of revenue. The log likelihood function is:

$$\ell(\theta/x_1,x_2,...,x_N)=\ln(\theta/x_1,x_2,...,x_N)=-\frac{N}{2}\ln(2\pi)-\frac{N}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2 \quad (3\text{-}3)$$

The score function contains two elements (2 parameters to estimate) :

$$\frac{\partial \ell(\theta/x_1,x_2,...,x_N)}{\partial \mu}=\frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu) \quad (3\text{-}4)$$

$$\frac{\partial \ell(\theta/x_1,x_2,...,x_N)}{\partial \sigma^2}=-\frac{N}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{N}(x_i-\mu)^2 \quad (3\text{-}5)$$

Equating (3-4) and (3-5) to zero and solving for $\hat{\mu}$ and $\hat{\sigma}^2$ yields the ML estimators:

$$\hat{\mu}=\overline{X} \quad (3\text{-}6)$$

$$\hat{\sigma}^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{X})^2 \quad (3\text{-}7)$$

In the following section, we present a Matlab code to estimate the pdf parameters. Note that if you use software that does not have the pre-programmed command, you only need to maximize the log-likelihood (or minimize the opposite log-likelihood). Most mathematical software (including Gams) can do this.

3.1.2. <u>The Matlab program for the estimation</u>[6]

In this section we will see how to program the estimation of parameters of the beta distribution with Matlab.

[5] See the section <u>Probability Distribution</u> for the characteristics of this different distributions

[6] The code for the beta distribution is given. The codes for Normal and Lognormal can be available on request.

```
clear all;
revo[7] = [...
   671934.77
   850703.94
   542814.04
        ∂
   364991.57
   164657.44
   279918.87
   425479.02];
 revo = sort(revo);     % This command sorts the income vector
ymin = min(revo);       % These two commands create the variables ymin (minimum income) and
                        ymax (maximum income) that are necessary for the following transformation.
ymax = max(revo);
n = length(revo);
revobeta = (revo - ymin)/(ymax-ymin);   %Here we transform the vector « revo » because
                                        Matlab uses the mathematical definition with the values
                                        between 0 and 1 (and strictly different of 0 and 1). This
                                        step is very important with the beta distribution but not
                                        necessary for estimation of Normal or Lognormal
                                        parameters.
revobeta = revobeta(2:n-1,:);   % With the transformation the first observation is 0 and the
                                last is 1 then you need to remove these two observations to
                                estimate.

[phat_base]= betafit(revobeta)  %This command computes the MLE of the parameters for the
                                beta distribution from the data in the vector revobeta. The
                                output phat_base, is the vector of the both estimated beta
                                parameters.
```

When the parameters of the chose distribution are estimated, it is relatively easy to compute the poverty indices.

3.2. FGT Measurements

3.2.1. The FGT indices

Suppose that *y* is the variable that measures the revenue of households or individuals and *F(y)* is the cumulative function of distribution (cfd):

$$F(y)=P(Y\leq z)=\int_0^z f(y)dy \qquad (3\text{-}8)$$

with *z* the poverty line.

Suppose that the vector of revenue, $y_1, y_2, \ldots, y_n$, is sorted by ascending order $y_1 \leq y_2 \leq \ldots \leq y_n$. The most popular money-metric poverty indices are those of Foster, Greer and Thorbecke (FGT) (1984) defined in the continuous case as:

[7] See the complete vector in the section Data in annex.

$$P_\alpha = \int_o^z \frac{(z-y)^\alpha}{z} f(y)\,dy \qquad (3\text{-}9)$$

where *z* is the poverty line and α represents the degree of aversion to poverty. When $\alpha = 0$, the FGT index is the standard head count ratio i.e. the share of households which are below the poverty line in the total population. This index does not give a measure of the extent of poverty for the poor. When $\alpha = 1$, the index measures the poverty gap, that is the level of income transfer needed to bring all poor to the poverty line. The shortcoming of this index is that it does not capture inequality among the poor, since a dollar of income gap for the extreme poor is given the same weight as a dollar of income gap for those who are just under the poverty line. Poverty severity, which is measured when $\alpha = 2$, addresses this issue.

3.2.2. The Matlab program for FGT measurement

Generally, the observed samples for statistical units such as households or individuals income have discrete distributions. However, for large samples, the exact distribution can be approximated with a limit distribution such as the Normal, the Lognormal or the Beta function.

The difficulty is only a technical one since there is an integral in the equation (3-9) and econometric or other software[8] do not always have the capacity to compute (or evaluate) integrals. The following code must be added to the previous program for the estimation:

z_a = 930000; %Create the variable “poverty line” and give a value to this variable.

p_a = phat_base(1,1); % Define *p_a* as the first element (first line, first row) of the output for the estimation of the beta parameters i.e. *phat_base*.

q_a = phat_base(1,2); % Define *q_a* as the second element (first line, second row) of the output for the estimation of the beta parameters i.e. *phat_base*.

for *alpha = 1:3;* % To calculate the FGT measures, a loop is programmed to repeat statements 3 times for $\alpha = 0, 1$ and *2*. The loop must begin by 1 then the formula of FGT will be modified in consequence.

if ymin > z_a % We use this conditional command because if the minimum income is greater than the poverty line, there is no poor, then the FGT0,1 and 2 are null.

FGT_baseza(alpha) = 0 ;

else % The command quad8 (or quad[9]) evaluates numerically the integral of the FGT formula. Quadrature is a numerical method of finding the area under the graph of a function. The command below returns the result of numerically integrating the function *fgtbeta* between the limit *,ymin* and *,z_a*. The two following arguments are empty matrix ([]) using the default values for TOL (Iterates until the relative error is less than TOL, 1e-3 by default) and traces (by default is zero. If this argument is nonzero, quad8 plots a graph showing the progress of the integration). *z_a,alpha,p_a,q_a,ymin,ymax* are coefficient allowed to be passed directly to the function *fgtbeta* such as F=*fgtbeta*(y,*z_a,alpha,p_a, q_a,ymin,ymax*). The function *fgtbeta* is in an other Matlab program (see below).

FGT_baseza(alpha)=quad8('fgtbeta',ymin,z_a,[],[],z_a,alpha,p_a,q_a,ymin,ymax);

end

end

---

[8] Gams does not offer this possibility.

[9] Quad uses an adaptative recursive Simpson’s rule and quad8 uses an adaptative recursive Newton-Cotes 8 panel rule.

FGT_baseza % This command puts the results on screen

The function program *fgtbeta* will be named *fgtbeta.m*. The syntax is the following:

```
function [y] = fgtbeta(revenu,z,alpha,p,q,ymin,ymax)

    if nargin==0, revenu = 0.5:.05:1;   % the function nargin indicates the number of input
                                         arguments and if there is no argument, suppose we take
                                         income between 0.5 and 1 with a jump of 0.5.
    end

y = (((-revenu+z)/z).^(alpha-1)).*(1/beta(p,q)).*((revenu-ymin).^(p-1)).*((ymax-revenu).^(q-1))/((ymax-ymin).^(p+q-1)) ; % This is the FGT formula (i.e. 3-9). The only difference is that
                                       since alpha is taken from 1 to 3, you must take (alpha-1) and not
                                       alpha.

%y = (((z-revenu)/(z-ymin)).^(alpha-1).*(1/beta(p,q)).*(revenu.^(p-1).*(1-revenu).^(q-1)));
                  % This is the same command if you prefer to work with transformed data i.e. if you
                  take the revenue between 0 and 1 (Here it is a comment since there is % in front).
```

3.3. Using Matlab to compute FGT indicators with various intra households distribution functions

In this section, a general program is presented to estimate the parameters and to evaluate the FGT measures for the base year and after simulation. We also present the Matlab commands to plot the distribution and the poverty line before and after the exogenous shock[10].

```
clear all
revo = [...
   671934.77
   850703.94
   542814.04
   158306.09
         ∂
   164657.44
   279918.87
   425479.02];
 revo = sort(revo);
ymin_a = min(revo);
ymax_a = max(revo);

% Before simulation
n = length(revo);
revobeta = (revo - ymin_a)/(ymax_a-ymin_a);
revobeta = revobeta(2:n-1,:);
```

% After Simulation. Note that in the case presented the variations of income are the same for all households or individuals so the vector after simulation can be calculated using the following

[10] In the section Other Matlab programs in annex, we propose the program to plot the real and the base distribution

syntax. The value *0.06891* is given by the output of GAMS after simulation[11]. We do the same transformation that for the base year.

```
revb = revo.*(1-0.06891);
revb =  sort(revb);
ymin_b = min(revb);
ymax_b = max(revb);
mb = length(revb);
revbbeta = (revb - ymin_b)/(ymax_b-ymin_b);
revbbeta = revbbeta(2:mb-1,:);
```

```
[phat_base]= betafit(revobeta)     % Estimation of distribution parameters with the base vector
[phat_simb]= betafit(revbbeta)     % Estimation of distribution parameters with the simulated vector.
```

```
z_a = 930000;     % Let z_a, the monetary poverty line for the base year.
z_b = 891126;     % Let z_b, the new monetary poverty line. To calculate this value, we have supposed that the commodity basket remains invariant from one simulation to another and we have multiplied with the prices vector obtained after the simulation.
```

```
%FGT Measurements
p_a = phat_base(1,1);
q_a = phat_base(1,2);

p_b = phat_simb(1,1);
q_b = phat_simb(1,2);

for alpha = 1:3;

if ymin_a > z_a
FGT_baseza(alpha) = 0 ;
else
FGT_baseza(alpha) = quad8('fgtbeta',ymin_a,z_a,[],[],z_a,alpha,p_a,q_a,ymin_a,ymax_a);
end

if ymin_b > z_b
FGT_simbzb(alpha) = 0
else
FGT_simbzb(alpha) = quad8('fgtbeta',ymin_b,z_b,[],[],z_b,alpha,p_b,q_b,ymin_b,ymax_b);
end
end
  FGT_baseza
  FGT_simbzb
```

% The following command saves the output (*betamrur phat_base phat_simb FGT_baseza FGT_simbzb*) in a ascii file.

```
save betamrur phat_base phat_simb FGT_baseza FGT_simbzb -ascii
```

[11] The consequence of this evaluation of the new vector of income is that it ignores the intra-group income distribution.

% The end of the program is to plot the base distribution and poverty line and the simulated distribution and poverty line[12].

*densite_a = ((((ymax_a-ymin_a)^(p_a+q_a-1)).*beta(p_a,q_a)).^(-1)).*((revo-ymin_a).^(p_a-1).*(ymax_a-revo).^(q_a-1));*
*densite_b = ((((ymax_b-ymin_b)^(p_b+q_b-1)).*beta(p_b,q_b)).^(-1)).*((revb-ymin_b).^(p_b-1).*(ymax_b-revb).^(q_b-1));*

% To put the poverty line on the graph, we need to transform the constant in a vector with the same length than the income vector.

```
for i = 1:length(revo)
        z_abeta(i) = z_a;
end;

for i = 1:length(revb)
   z_bbeta(i) = z_b;
end;
```

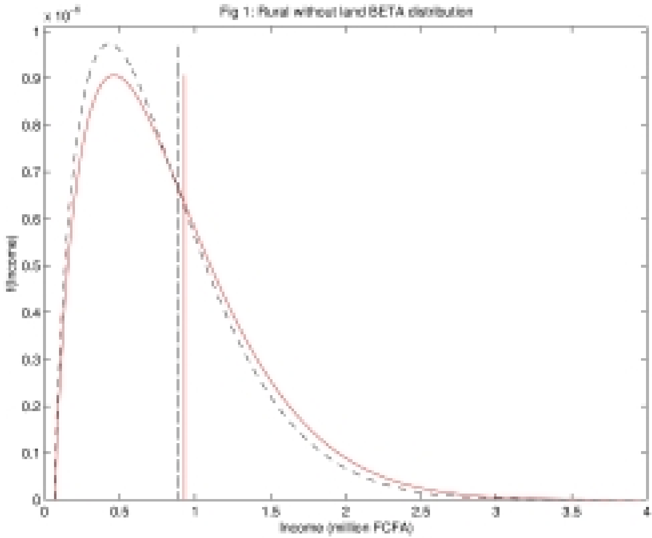
% The following command is used to plot both distributions and both poverty lines on the same graphic (*hold on* and *hold off*). We plot vectors *revo* (x) and *z_abeta* versus *densite_a* (y) with the color red ('*r*') and the vectors *revb* and *z_bbeta* versus *densite_b* (y) with the color black ('*k*')[13].

```
plot(revo,densite_a,'r',z_abeta,densite_a,'r')
hold on
plot(revb,densite_b,'k',z_bbeta,densite_b,'k')
hold off
axis([0 4000000 0 0.00000101])   % Provides an easy way to manipulate the most important
                                 % properties of axes. Here it sets the scaling for the x- and y-
                                 % axes on the current plot.
title('Fig X: ….’);              % Give a title at the top of the current plot.
xlabel('Income (FCFA)')          % Adds text on the current plot below the x- axis.
ylabel('f(Income)')              % Adds text on the current plot below the y- axis.

print basesim1 –depsc            % Saves the current plot with eps extension.
```

The output follows. The red solid line is the estimated Beta distribution for the base year and the back dashed line is the estimated Beta distribution after simulation.

[12] Note that as we work with the initial value of revenue, we cannot use the function beta pdf pre-programmed in Matlab since it uses the transformed data.
[13] There are various line types, plot symbols and colors. See the section Lines types, plot symbols and Colors.



4. Conclusion

In the first part of this methodological note, we present a non-parametric Kernel method to measure poverty for the base year and after simulation. The advantage of this method is that you only need the DAD software, which is well known to MIMAP researchers and freely distributed and that you are not required to impose a distribution. Furthermore, this method will give you a good approximation of the ‘true’ distribution and of the FGT’s measures.

In the second part, we present the main steps to estimate the parameters of the income distribution and to evaluate the FGT measures with continuous distributions (a parametric method). This method is particularly useful when the primary data is unavailable, but some information is available on the underlying distribution (mean, minimum, etc.) We provide an example to estimate parameters and compute indices with Matlab programs in the case of the Beta distribution. It should be relatively easy to modify these codes if you prefer to work with an another distribution function such as the Normal, the Lognormal or Pareto which are the functions usually used to approximate the income distributions. You should also use an another software with the same logic if this one can evaluate integrals. We illustrated this approach using the data provides by the 1995-1996 Household Income and Expenditure Survey of Cameroon.

5. Annex

5.1. Probability Distribution

5.1.1. Beta distribution

The Beta distribution describes a family of curves that are unique in that they are nonzero only on the interval [0 1]. A more general version of the function assigns parameters to the end-points of the interval.

There are two mathematical definitions of the Beta pdf :

a. $y = f(x/p,q) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} I_{(0,1)}(x)$ for **0 < x < 1**

Or

b. $y = f(x/p,q) = \frac{1}{B(p,q)} \frac{(x-\text{xmin})^{p-1}(\text{xmax}-x)^{q-1}}{(\text{xmax}-\text{xmin})^{p+q-1}}$ for **xmin < x < xmax**

where

$$B(p,q) = \int_{x\text{min}}^{x\text{max}} \frac{(x-\text{xmin})^{p-1}.(\text{xmax}-x)^{q-1}}{(\text{xmax}-\text{xmin})^{p+q-1}} dx$$

To estimate p and q, the parameters of the Beta pdf, most software works with data included between 0 and 1. If you have a income vector, you need to do the following transformation:

$$X_{0,1} = \frac{X-\text{xmin}}{\text{xmax}-\text{xmin}}$$

5.1.2. Normal distribution

The normal distribution is a two-parameter family of curves. The first parameter, μ, is the mean. The second, σ, is the standard deviation. The standard normal distribution (written Φ(x)) sets μ = 0 and σ = 1. The usual justification for using the normal distribution for modeling is the Central Limit Theorem which states that the sum of independent samples from any distribution with finite mean and variance converge to the normal distribution as the sample size goes to infinity.

The normal pdf is : $f(x/\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

Remark: Some software return likelihood estimators or the Minimum Variance Unbiased Estimator (MVUE), which is the statistic that has the minimum variance of all unbiased estimators of parameter. Matlab for instance, returns the MVUEs for μ and σ^2.

5.1.3. Lognormal distribution

The normal and lognormal distributions are closely related. If *X* is distributed lognormal with parameters θ and λ^2, then $y=\ln X$ is distributed normally with parameters μ and σ^2 where

$$\mu=\ln\theta^2-\frac{1}{2}\ln(\theta^2+\lambda^2) \qquad \text{and} \qquad \sigma^2=\ln\left(1+\frac{\lambda^2}{\theta^2}\right)$$

The lognormal distribution is applicable when the quantity of interest must be positive since lnX exists only when the random variable X is positive. When an economist wants to analyze poverty, they often model the distribution of income using a lognormal distribution.

The mathematical definition of the lognormal pdf is:

$$f(x/\mu,\sigma^2)=\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(\ln(x)-\mu)^2}{\sigma^2}} \qquad \text{for } \mathbf{x \geq 0}$$

Remark: This pdf is programmed in Matlab but not the fit. However it is easy to modify the normal or beta fit program.

5.2. Data

The illustrated approach uses data provided by the 1995-1996 Household Income and Expenditure Survey of Cameroon. The following vector is the complete vector of revenue (in FCFA per Household by year) for rural landless households of Cameroon for the base year:

```
revo = [...
671934.77
850703.94
542814.04
158306.09
576344.33
989909.28
728478.14
719755.38
790252.61
641760.77
245934.35
523387.72
1691507.64
1091157.46
1160355.45
1042680.93
1076362.74
1036527.65
1621830.15
388747.78
849717.10
279797.17
1801170.79
1137259.78
1274632.72
1286515.30
387490.43
261284.64
1017375.24
309858.83
364661.53
1051865.57
374489.52
231310.74
444331.45
```

918081.79
421872.73
959836.20
1134064.96
449752.35
807489.27
615326.80
1651857.50
365264.93
777098.93
764092.36
507928.14
661002.09
606747.41
771853.00
650864.59
1184092.73
904137.35
835145.54
1197478.79
274588.93
1198316.01
1091138.33
618037.25
491387.04
654427.50
1828204.97
1648739.75
1063835.40
566143.50
695716.18
837108.43
721237.85
793902.83
295096.90
728973.42
1390864.98
308075.76
1497463.87
876551.92
1046621.77
1952900.79
1708894.50
875753.77
1245011.06
1296311.30
841367.30
832070.65
750817.44
314961.87
372136.21
611920.08
818898.74
1726738.03
472643.07
1134893.21
789483.27
671571.97
964760.66
549932.99
690398.65
533524.08
1045166.03
198854.58
2876037.54
684037.09
699125.59
604951.72
1256488.75
453659.55
490104.58
197093.31
511897.15
621804.68
287864.42
141293.00
703349.27
460079.27
604481.59

393508.49
485970.42
463049.60
526797.65
282126.15
1552950.80
301161.37
340238.33
312167.21
384619.44
456067.59
457870.79
655550.47
1092112.94
703553.72
869010.86
885702.26
644501.96
702933.81
894511.14
445501.82
638347.96
573631.83
654234.65
705700.66
1048677.10
917509.13
990091.10
308398.31
826658.98
696298.09
1004859.19
1476466.20
625109.57
733810.80
752952.61
386987.95
353262.79
855181.57
1219625.15
258593.30
204074.00
510278.68
782975.88
1181796.23
1342078.45
1074997.65
852102.85
819765.39
847898.73
1355880.36
1741986.52
718164.67
711345.13
1428610.08
562009.45
1334227.99
228579.90
884759.84
479003.56
1906561.65
426047.54
244314.13
672204.20
1613671.31
2294911.80
1663045.70
914640.11
1631967.91
2093466.96
522037.41
1409261.36
1441743.18
753058.03
1071107.43
1003923.30
633679.63
963308.47
887043.71

1047666.09
1143772.94
860857.30
486233.08
1727353.96
510085.94
421174.54
1135133.20
683715.95
500897.38
961438.98
263206.36
1416948.73
1589525.31
2264872.58
711082.16
308539.35
508614.91
313288.15
2060266.73
573358.85
89913.09
288650.17
709826.43
164642.81
1122662.82
1770726.23
459609.42
769085.46
973465.50
400052.19
2490803.24
688844.06
730814.47
584955.98
508677.80
518425.82
722382.54
733762.16
313102.15
747279.34
2294738.74
824726.67
2802816.67
1792328.83
1071231.99
574152.88
412943.52
1465703.17
70856.19
599091.95
587188.40
418905.71
221407.41
352966.07
594923.07
878040.46
464715.10
495097.53
1916405.72
3442040.20
1224084.67
2351161.04
624287.97
1054352.72
2627184.71
704073.81
391430.79
453564.83
589773.02
6119029.39
420907.60
3930235.24
961628.73
198888.80
458950.10
192832.76
1052332.63
370834.72

501493.80
414474.11
136635.07
1225817.80
226637.10
386599.33
364873.00
294504.18
1302639.50
1875237.35
955746.39
3005453.22
2347611.47
2146448.32
951608.40
1161117.79
977988.96
978748.27
2379851.62
480400.08
380199.98
288982.30
909084.04
820022.32
763359.95
349591.09
499772.97
227787.06
219776.45
1145361.90
341835.80
559307.13
1434153.98
259964.35
137799.64
796974.63
193924.21
734537.09
927952.64
371195.06
1751884.06
779184.46
996297.19
1545609.49
535733.92
849739.18
1014919.47
1034266.52
1143649.44
547080.30
603537.21
187457.62
991402.24
365329.22
961549.93
2165657.70
961842.14
889800.77
697125.78
489532.13
345021.73
654934.06
585777.69
306490.34
2096720.98
263768.30
389293.39
381634.55
227696.99
522922.40
478576.93
891522.87
349728.27
1529903.72
502555.65
308136.54
472619.68
365358.37
341481.12

```
377644.22
832142.13
364491.03
547402.89
842322.38
579012.61
635496.13
884954.00
519141.78
423621.63
1039025.29
448897.71
179451.96
281153.42
272299.21
495958.09
364991.57
164657.44
279918.87
425479.02 ];
```

5.3. Other Matlab program

The following program allows to plot the real distribution i.e. this evaluates with the Kernel method in DAD for instance. The distribution of income is also approximated by a Beta pdf. To do this exercise you need to follow these steps :

a- With DAD 4.0, you plot the density (*Distribution* and *Density function*) of the income vector (with weights if you have them). You choose a range for the x-axis and you make the graph. You save the graph with a *txt* extension (*graph* and *save XY*). This saved file is a matrix (N x 2) where the first column is the income values and the second is the associated pdf values.

b- You open the *.txt* file and you change the format since Matlab does not recognize the scientific notations. For instance, if you have 1.56E-5, you must replace by 0.0000156.

c- You can now work with Matlab with the following program:

```
datafit        % you recall the program in which you estimate the beta distribution. This program
               will be in the same folder with the extension of Matlab i.e. .m.

densitereelledata = [...
0.0            0.00000009262473144001762
40000.0        0.0000001482466736767701
80000.0        0.00000022265976560418096
120000.0       0.0000003149222747057304
160000.0       0.00000042102365585761426
200000.0       0.0000005342319687760567
240000.0       0.0000006462342856074306
280000.0       0.0000007487185622161286
320000.0       0.0000008348785153702764
360000.0       0.0000009004098425842671
400000.0       0.0000009438407469401958
440000.0       0.0000009663132340999445
480000.0       0.0000009710240634436033
520000.0       0.0000009624381070211944
560000.0       0.0000009452981252212094
600000.0       0.0000009235631552498628
640000.0       0.0000008996309208295543
680000.0       0.0000008742246460111923
720000.0       0.000000846987442841215
760000.0       0.0000008173741010822906
800000.0       0.0000007852767365602202
840000.0       0.0000007510871423160605
880000.0       0.0000007153081653700752
920000.0       0.0000006780587163534125
960000.0       0.0000006388081810193755
1000000.0      0.0000005965329504250761
1040000.0      0.0000005502793411818379
1080000.0      0.0000004998831770236708
```

1120000.0	*0.0000004464596961941418*
1160000.0	*0.000000392372773436173*
1200000.0	*0.0000003406805301280453*
1240000.0	*0.0000002943233104821118*
1280000.0	*0.00000025539847900606154*
1320000.0	*0.00000022476824398333556*
1360000.0	*0.00000020208783653570174*
1400000.0	*0.00000018619122735900547*
1440000.0	*0.00000017563914053410042*
1480000.0	*0.0000001691570234360535*
1520000.0	*0.00000016574949588098066*
1560000.0	*0.0000001644885206814529*
1600000.0	*0.00000016420674308514058*
1640000.0	*0.00000016339852864882097*
1680000.0	*0.00000016047272159237284*
1720000.0	*0.00000015423808147580497*
1760000.0	*0.00000014433777045735736*
1800000.0	*0.00000013139033324257404*
1840000.0	*0.00000011678233311763315*
1880000.0	*0.00000010223923093900314*
1920000.0	*0.00000008936250750453118*
1960000.0	*0.00000007926993808748652*
2000000.0	*0.00000007240416891361575*
2040000.0	*0.0000000685406285804411*
2080000.0	*0.00000006699204871355875*
2120000.0	*0.00000006692112064531188*
2160000.0	*0.00000006758553327755698*
2200000.0	*0.0000000683742096070189*
2240000.0	*0.00000006867483594422212*
2280000.0	*0.00000006778318483335377*
2320000.0	*0.00000006503555887785757*
2360000.0	*0.00000006012197832604172*
2400000.0	*0.00000005333572894662004*
2440000.0	*0.00000004553605558176945*
2480000.0	*0.000000037822200096056227*
2520000.0	*0.00000003112904648123101*
2560000.0	*0.000000025981073461401316*
2600000.0	*0.000000022491493220351026*
2640000.0	*0.000000020516140971645905*
2680000.0	*0.000000019803454478901203*
2720000.0	*0.000000020046844982672685*
2760000.0	*0.00000002086384217218498*
2800000.0	*0.000000021791871852909962*
2840000.0	*0.000000022357556298437403*
2880000.0	*0.000000022191917103325682*
2920000.0	*0.000000021116682127272022*
2960000.0	*0.000000019156933498225652*
3000000.0	*0.000000016498238833237937*
3040000.0	*0.000000013432991503313493*
3080000.0	*0.000000010316370976044846*
3120000.0	*0.000000007521225060312*
3160000.0	*0.000000005381066374324682*
3200000.0	*0.000000004129963850319026*
3240000.0	*0.000000003855199151358549*
3280000.0	*0.000000004468676776893282*
3320000.0	*0.000000005699875379509166*
3360000.0	*0.0000000071266724921598446*
3400000.0	*0.000000008266367271469686*
3440000.0	*0.00000000872041859226543*
3480000.0	*0.000000008316177981912942*
3520000.0	*0.00000000716743044521876*
3560000.0	*0.000000005615777850204902*
3600000.0	*0.00000000408988159802877*
3640000.0	*0.0000000029649759806716923*
3680000.0	*0.0000000024841213273985677*
3720000.0	*0.000000002741658933511687*
3760000.0	*0.0000000036877719008637293*
3800000.0	*0.000000005125068534342541*
3840000.0	*0.0000000067135141514809705*
3880000.0	*0.000000008026620347379844*
3920000.0	*0.000000008675127309407812*
3960000.0	*0.000000008452675495675063*
4000000.0	*0.00000000741917876531281*
];	% It is the matrix saved with DAD and transformed in Matlab format (i.e. Output of steps a and b).

% See the section <u>Using Matlab to compute FGT indicators with various intra households distribution functions</u>

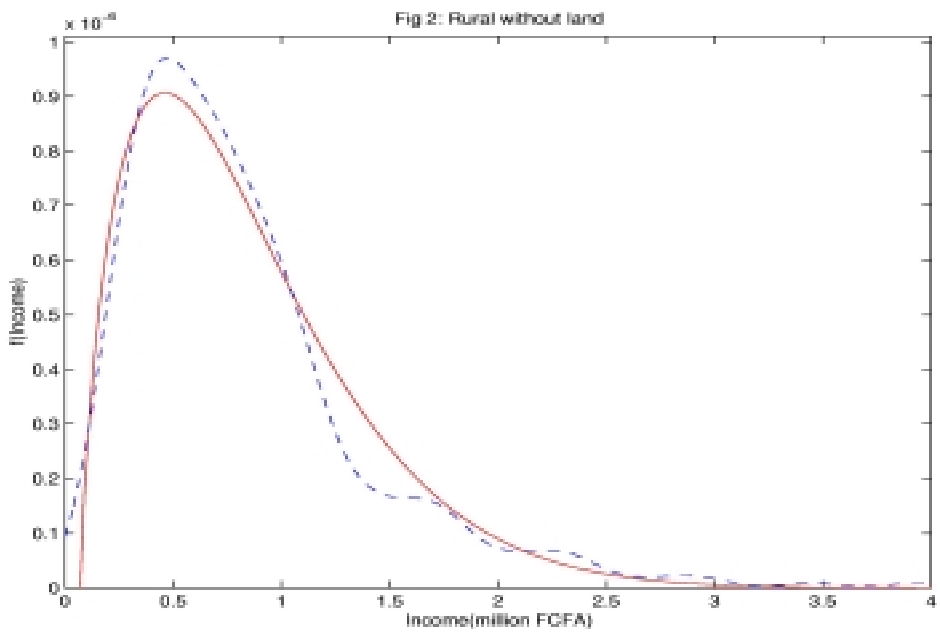
```
densite_a = ((((ymax_a-ymin_a)^(p_a+q_a-1)).*beta(p_a,q_a)).^(-1)).*((revo-ymin_a).^(p_a-1).*(ymax_a-revo).^(q_a-1));

plot(revo,densite_a,'r',densitereelledata(:,1),densitereelledata(:,2),'b--').
axis([0 4000000 0 0.00000101])

title('Fig XX:…');
xlabel('Income(FCFA)')
ylabel('f(Income)')

print base_reellebrute –depsc
```

The output follows. The red solid line is the estimated Beta distribution and the blue dashed line is the approximation by the Kernel method obtained with DAD Sofware.



Remark: Before choosing the distribution, it could be useful to do the both first steps with DAD to have an idea of what distribution is a better approximation of the real[14] distribution. Then you could have a better insight to choose the form of parametric distribution.

5.4. Lines types, plot Symbols and Colors in Matlab

Point	**.**	Yellow	**y**
Circle	°	Magenta	**m**
x-mark	**x**	Cyan	**c**
Plus	**+**	Red	**r**
Star	*****	Green	**g**
Solid line	**-**	Blue	**b**
Dotted line	**:**	White	**w**
Dashdot line	**-.**	Black	**k**
Dashed line	**--**		

6. Bibliography

De Janvry, A. and E. Sadoulet (1995), *Quantitative Development Policy Analysis*, The Johns Hopkins University Press, Baltimore and London.

Foster, J., J. Greer and E. Thorbecke (1984), "A Class of Decomposable Poverty Measures", Econometrica, 52(3), pp. 761-766.

Greene, W. (1997), *Econometric Analysis*, Third Edition, Prentice Hall, Upper Saddle River, New Jersey.

Johnsons, N. L. and S. Kotz (1970), *Continuous Univariate Distribution*, Vol. II, Wiley, New York.

Livny, M., R. Ramakrishnan and T. Zhang (1996), "Fast Density and Probability Estimation Using CF-Kernel Method for Very Large Databases" Computer Sciences Department, University of Wisconsin-Madison.

Matlab (1992), *High Performance Numeric Computation and Visualization Software: Reference Guide*, The MathWorks Inc, Massachusetts.

Matlab (1996), Statistics Toolbox for Use with Matlab, The MathWorks Inc, Massachusetts.

Poirier, D. (1995), *Intermediate Statistics and Econometrics: a Comparative approach*, The MIT Press, Cambridge, Massachusetts.

[14] As we see, this distribution is not the real one but the distribution approximated by a non-parametric method i.e. without estimation.