****

**The ‘Vulnerability Targeting Model’ in Jordan**

***An Econometric Study of the Economic Vulnerability of Syrian Refugees in Jordan:***

***Re-estimating the Welfare Model & Introducing the Probit/Logit***

*Draft 2 – September 22, 2014[[1]](#footnote-1)*

**Marco Santacroce[[2]](#footnote-2)**

**Abstract**

Following the Syrian Uprising that began in the early spring of 2011, an overwhelming inflow of Syrians have entered the Kingdom of Jordan as refugees. As of today – September 2014 – north of 600,000 people are registered as refugees with the United Nations High Commissioner for Refugees (UNHCR). Considering the funding gap requirements faced by UNHCR, the objective of this study is to develop an advanced targeting methodology to identify high ‘vulnerability’ households amongst non-camp refugees. Vulnerability is defined as “expenditure per capita”. Empirical data is examined in order to develop the ‘Welfare Model’. The first draft of this study had analysed 700 records to develop the Welfare Model; this study has tested the validity of those results by re-estimated the model using 5000 records. The re-estimation has been greatly successful. The Welfare Model based on 700 records was composed of 9 variables and an R-squared of 49%. The Welfare Model based on 5000 records is composed of an equal amount of variables, but an R-squared of 57%. Furthermore, with just 7 indicators the Welfare Model 5000 can explain 56% of the variation in expenditures. The econometric analysis undertaken identifies the following 7 characteristics as the best predictors of expenditures: *house crowding, living with a host family, the debt-to-expenditure ratio, family members living in the house****,*** *coping strategies to meet basic food needs, the saving ratio, and income per capita*. The first four characteristics are negatively associated with expenditures, whilst the last three characteristics manifest a positive association. The identified economic vulnerability predictors are the same in both drafts.

1. **Table of Contents**
2. Introduction………………………………………………………………………………………………………………………….4
3. Methodology (Data Exploration)…… ……………………………………………………………………………….……4

STEP 1) Bivariate Regressions…………………………………………………………………….…………....4

STEP 2) Stepwise Regression…………………………………………………………………………..……..…5

STEP 3) Diagnostic Tests on Final Model………………………………………………………..………….5

1. Variable Generation………………………………………………………………………………………………………………5

3.1 – New Variables…………………………………………………………………………………………………..5

3.2 – Composite Indices……………………………………………………………………………………..….….6

1. Model Specification…………………………………………………………………………………………………………….…8

4.1 – OLS………………………………………………………………………………………………………………..…8

4.2 – Probit & Logit……………………………………………………………………………………………………9

1. The Vulnerability Targeting Model………………………………………………………………………………….……..9

5.1 – Digression: The Log Transformation……………………………………………………………..…10

5.2 – The Models………………………………………………………………………………………………….…11

Welfare Model 5000 (OLS)…………………………………………………………………….…..11

Probit & Logit Models (LDV)…………………………………………………………………….…13

Expenditure Model…………………………………………………………………………….….…..14

Models from Draft 1…………………………………………………………………………………..15

1. Diagnostic Tests…………………………………………………………………………………………………………………..15

6.1 – Multicollinearity…………………………………………………………………………………………....15

6.2 – F-tests…………………………………………………………………………………………………….….…..16

6.3 – Model Mis-specification……………………………………………………………………….………..16

6.4 – Heteroscedasticity……………………………………………………………………………..…….…….17

6.5 – Outliers……………………………………………………………………………………………….…….……17

1. Findings………………………………………………………………………………………………………………….…………..18

7.1 – OLS…………………………………………………………………………………………………………………18

7.2 – LDV…………………………………………………………………………………………………………….….18

7.3 – OLS and LDV…………………………………………………………………………………………….…….19

1. Conclusions………………………………………………………………………………………………………………………...20

Appendix…………………………………………………………………………………………………………………………….22

Appendix A……………………………………………………………………………………………………..……...22

Appendix B……………………………………………………………………………………………………………..25

Appendix C……………………………………………………………………………………………………………..27

1. Introduction**:**

For the past two months, the development of an advanced targeting methodology to identify high vulnerability amongst non-camp refugees has been the objective of this study. The analysis, that here by follows, relies on econometric tools and the VAF (Vulnerability Assessment Framework) data. The data comprises 4889 households and 700 variables. The goal is to develop the ‘Welfare Model’ (or the ‘Vulnerability Targeting Model’, henceforth terms are used interchangeably), a targeting methodology that is able to identify the most highly economically vulnerable households. The poverty line has been set at 50 JD per person per month.

Econometrics regression analysis is the development of statistical models that can predict and explain a phenomenon. In this particular work, the phenomenon that is wished to be explained is “expenditure per capita”. To do so, a series of Welfare Models have been developed; their objective is to identify the characteristics present in a household that can best predict their expenditures. Once these characteristics have been identified, the model can be employed to predict the expenditure of any household. The idea behind this analysis commences with a notion that economic vulnerability equals expenditure per capita and that expenditure per capita equals welfare. Defining welfare, and how best to estimate it, is a matter that requires an extensive discussion. There is however, a worldwide consensus – or better yet an acceptance – that “expenditure” is one way of expressing welfare. For this reason, “expenditure” has been used as the welfare aggregate for this analysis. Other proxy estimators exist, and these are being explored. For example, the first draft of this study presented a preliminary model that predicted food security as the welfare aggregate.

This paper documents the data exploration process, the methodology used, the diagnostic tests performed on the models, and the findings.

Parts of this paper are maths intensive. These were included only for completeness. It is not necessary to comprehend the mathematical computations included in order to understand the work, nor to understand the results and findings. The mathematics are included only for the pleasure of the reader. They are not to be interpreted or agonized upon unless familiar with these quantitative methods.

1. **Methodology (Data Exploration):**

The methodological process is here below documented. For clarity, the methods used for the purpose of data exploration have been listed into three ‘steps’ and subdivided in multiple bullet points.

**STEP 1) Bivariate Regressions:**

* The analysis of two variables: the independent variable X and the dependent variable Y to determine whether an empirical relationship between the two exists; estimated by examining their correlation
* Variable Generation: creating new variables and composite indices as these might have greater or better explanatory/predictive power than their individual components
* Expenditure per Capita used as dependent variable (must be generated first)
* Run about 800 bivariate regressions on all variables available
* Eliminate independent variables that are not statistically significant at 10% (90% Confidence Interval)
* Rank variables by , the Coefficient of Determination. The in a bivariate regression is the square of the sample correlation coefficient, thus why the is always positive and ranges from 0 to 1
* Check for non-linear relationship between X and Y

**STEP 2) Stepwise Regression:**

* Forward selection and Backward elimination (i.e. Bidirectional elimination): adding and subtracting explanatory variables from the model to examine how the model is improved or worsened by this process
* Run Preliminary Tests: Variance Inflation Factor, F-tests etc…
* Finalize model

**STEP 3) Diagnostic Tests on Final Model:**

* Multicollinearity (Variance Inflation Factor)
* Model Mis-specification (Ramsey RESET)
* Heteroskedasticity vs. Homoskedasticity (Breusch-Pagan; White)
* Decision whether OLS, WLS, GLS or FGLS is preferred
* F-tests (for fit; for linear restrictions)
* Outliers (Bonferroni; Cooks Distance; Leverage)
* Normality in Residuals (Jarque-Bera)

1. **Variable Generation:**

The creation of new variables and composite indices is valuable exercise in regression analysis. The variables listed below have been generated in order to achieve explanatory power that is greater than their individual components (and for parsimonious reasons). Of these indicators, some were found to be highly statistically significant and included in the final model, other instead, failed to produce meaningful inferences.

**3.1 - New Variables:**

* **Expenditure per Capita**: Total Expenditure / Family Size (in same file number)
  + Why: Expenditure components measured per family size (not per people living in a house both in the same and in another file number)
* **House Crowding**: Number of Family Members in House (both in the same file number or in another file) / Number of Rooms excluding Kitchen and WASH facilities
  + An Indicator of People per Room
* **House Crowding2**: Total Square Meter Area of House excluding Kitchen and WASH facilities / Number of Family Members in House (both in the same file number or in another file)
  + An Indicator of Square Area per Person
* **Family with School Aged Children BUT Not Sent to School**: Discrepancy Between ‘Do You Have School Aged Children’ and ‘Do You Have Children Attending School’
* **Log Expenditure per capita:** The natural log of ‘Expenditure per capita’
* **Savings Ratio in Family:** The ratio of Total Savings to Family Size
* **Debt-to-Expenditure Percentage:** Total Debt up to now divided by Total Expenditure per file Number multiplied by 100
* **Debt-to-Income Percentage:** Total Debt up to now divided by Total Income per file Number multiplied by 100
* **Income per capita:** Total income reported divided by family size
* **Income per capita Squared:** Total income reported divided by family size Squared
  + Generated to better fit non-linear relationships
* **Family Size in same File Number Squared:** Family Size in same file number Squared
  + Generated to better fit non-linear relationships
* **House Crowding Squared:** House Crowding variable Squared
  + Generatedto better fit non-linear relationships
* **Coping Strategies Used to Meet Basic Food Needs Squared:** Coping Strategies used to meet basic food needs variable Squared
  + Generatedto better fit non-linear relationships

**3.2 - Composite Indices:**

* **House Assets**: Kitchen Access + Sanitary Facilities Access + Ventilation + Electricity Access
  + MAX: 4
* **House Luxury Assets**: Floor Mattress + Sofas/Beds + Cabinets + Kitchen utilities + Freezer + Fridge + Television + Water heater + Others + Oven
  + MAX: 10
* **Type of Housing (Poor Conditions)**: Damp Walls + Leaking Roof + Hygienic Concerns + Privacy Concerns + Rodents + Poor Insulation + Broken Windows
  + MAX: 7
* **Total Amount of NFIs & NCIs Received in last 6 Months**: Mattresses + Blankets/Pillows + Detergents/Hygiene Items + Kitchen Equipment + Furniture + Stove + Gas Cylinder + Fan + Cloth/Shoes + Electrical Equipment
  + MAX: 10
* **Total Amount of NFIs & NCIs that are Regular Family Needs**: Mattresses + Blankets/Pillows + Detergents/Hygiene Items + Kitchen Equipment + Furniture + Stove + Gas Cylinder + Fan + Cloth/Shoes + Electrical Equipment
  + MAX: 10
* **No Water Coping Strategies (Total Amount of ‘What Did You Do’ in the Past Month When You Didn’t Have Water)**: Buy from Own Pocket + Borrow From Family or Borrow Money to Buy + Shop Credit + Stay Without + Other
  + MAX: 5
* **Coping Strategies Index (Total Number of Coping Strategies Taken in the Last 6 Months)**: Living together with host family (Jordanian & Syrian) + Sharing costs with host family (Jordanian & Syrian) + Support from family members (irregular remittances) + Support from host community (Jordanian & Syrian) + Humanitarian assistance (CBOs. personal donations. etc.) + Selling properties (jewelry. car. etc.) + Selling food vouchers + Selling household assets + Borrowing money + Buying against credit + Dropping children out from school + Child labor + Begging + Savings + Irregular Work + Have not paid the rent for the past months
  + MAX: 16
* **WFP Food Consumption Score**: Times Consumed Each Product in the Last Week multiplied by its Weight
  + PRODUCTS: Cereals (Weight: 2), White Tubers & Roots (Weight: 2), Vegetables & Leaves (Weight: 1), Fruits (Weight: 1), Meat (Weight: 4), Eggs (Weight: 4), Fish & Seafood (Weight: 4), Pulses, Nuts & Seeds (Weight: 3), Milk & Dairy Products (Weight: 4), Oils & Fats (Weight: 0.5), Sweets (Weight: 0.5), Spices & Condiment (Weight: 0)
* **WFP Coping Strategy Index**: 5 Coping Strategies Employed in the Last Week Multiplied by its Weight
  + COPING STRATEGIES: Rely on less preferred and less expensive food (i.e. cheaper lower quality food) (Weight: 1), Borrow food or relied on help from relative(s) or friend(s) (Severity Weight: 2), Reduce number of meals eaten a day (Severity Weight 2), Limit portion size at mealtime (different from above: i.e. less food per meal (Severity Weight: 1), Restrict consumption by adults in order for small children to eat (Severity Weight: 2)
* **Coping Strategies Used to Meet Basic Food Needs (In the Past 30 Days)**: Spent savings + Bought food on credit or borrowed money to purchase food + Reduced essential non-food expenditure such as education/health + Sell household goods (jewelry, phone, furniture, electro domestics, etc.) + Sell productive assets or means of transport (sewing machine, car, wheel barrow, bicycle, motorbike, etc.) + Since arriving in Jordan have you accepted high risk, illegal, socially degrading or exploitive temporary jobs + Sent adult family members to beg + Sent children (under 18) family members to beg
  + MAX: 8
* **Reasons Children Do Not Attend School**: Do not know + Not interested in school + Child marriage/engagement + Child labour + No resources + Distance to school + Issues at School + Safety fears + Other + Do not know if school registration is possible + Were not going to school in Syria + Waiting for return in Syria to register children in school + financial constraints + Expired asylum-seeker certificate + Big gap between last grade in home country and one supposed to attend in Jordan + Don’t have an MOI card + Psychological distress + Disability + Not comfortable with teachers/teaching methods/curriculum + Prefer teachers from same nationality + Difficult dialect + Moving from one house to another + Arrival in Jordan in the middle of academic year
  + MAX: 23
* **Chronic Diseases in Family**: Hypertension + Diabetes + Cardiovascular + Critical Medical Conditions + Other
  + MAX: 5
* **Vaccinations NOT received**: Measles + Polio + Routine Vaccines (EPI)
  + MAX: 3
* **Age & Disability Amongst Family Members in Same File Number**: Chronic diseases/impairments/dishabilles + Pregnant females with complications + Visual/Hearing Impairment + Other physical disability + Mental disability + Intellectual disability + Temporary Injured + Chronically ill or serious medical condition + People in need of support to do daily activities
  + MAX: 9

1. **Model Specification:**

**4.1 - OLS:**

The econometrical estimation is computed using Ordinary Least Squares (OLS) analysis. The regression model estimated will take the form:

In a multivariate regression:

The OLS method minimizes the sum of squared errors – the first difference between actual and predicted values derived from the linear estimation:

Idea: find values of that minimize RSS:

Solving these two equations together:

**4.2 – Probit & Logit:**

This study introduces Limited Dependent Variable (LDV) models, more specifically the Probit and the Logit models. These models are binary classification models and the outcome variable takes value 0 for ‘non-poor’ or 1 for ‘poor’. The poverty threshold is set at 50 JD per person per month. The estimated response probabilities are strictly between zero and one. Values above 0.5 denote ‘poor’. The Probit model is based on the standard normal cumulative distribution function (cdf), whilst the standard logistic cdf yields the Logit model. The cumulative standard normal of the Probit is:

And the cumulative standard logistic is:

For: the Probit and Logit require a function that satisfies the following properties:

The standard normal and the logistic normal cdf satisfy these criteria. They take ‘S-shaped’ curves for (see Appendix A).

1. **The Vulnerability Targeting Model:**

The econometric analysis undertaken in the first study produced a total of 6 Welfare models (see Appendix B). The first five were “monetary” models – the dependent welfare aggregate was expenditure per capita. The sixth model was a preliminary “non-monetary” model that predicts food security, an indicator that ranges from 1 (least vulnerable) to 4 (most vulnerable). In this second draft, the main Welfare Model from the former paper (‘Welfare Model 666’) is re-estimated on a larger sample size (producing the ‘Welfare Model 5000’). The current model has been estimated with 7 and 9 predictors. The Probit and the Logit models have also been estimated. Expenditure is net of UNHCR’s cash assistance in all models.

**5.1 - Digression: The Log Transformation**

A household of 4 has total expenditures of 100. Expenditure per capita is therefore:

= 25

The natural log of 25:

= 25

Because:

= 25

Thus:

≈ 3.2188

Therefore:

By employing the log transformed model, regression estimates must be interpreted differently than in the level model (the non-log transformed model). For the sake of an example if 55 JD denotes the vulnerability threshold, then we may utilize two methods for result interpretation:

**Method 1: Interpreting log scores**

Compute the natural log of 55:

If natural log of 55 equals 4.0073, then Euler’s number, , raised to the power of 4.0073 must equal 55:

Because:

The value 4.0073 is therefore equivalent to 55 JD – the chosen poverty threshold. Any values above 4.0073 are consistent with expenditures higher than 55 JD per capita and any values below 4.0073 are consistent with expenditures below 55 JD per capita. *Nota bene*: the Welfare Model expenditure per capita prediction estimates are at times expressed in natural logs.

**Method 2: Reverting log scores back into standard monetary terms**

If log-predicted household vulnerability score is 4.0073, then the exponential of it will revert the value into Dinars:

Thus, raising Euler’s constant to the power of the predicted welfare estimate will yield a monetary value in JD.

**5.2 – The Models:**

**Welfare Model 5000 (OLS):**

The Welfare Model 5000 is the principal regression estimation derived from Ordinary Least Squares (OLS) methods. It succeeds the Welfare Model 1 (from Draft 1) or Welfare Model 666 (as it is named in this second Draft). The findings from the Welfare Model 666 have not be supplanted, on the contrary, they have been tested and validated from the larger sample size. Expenditures are net of UNHCRs’ cash assistance. The Model 5000 has two forms: the *ennea*-variable and the *hepta*-variable form. The former model has an R-squared of 57%, whilst the latter of 56%. As a reminder, the Welfare Model 666 achieved an R-squared of 49%.

***Hepta*-Variable Form:**

+

***Ennea*-Variable Form:**

+

The regression estimates for the Welfare Model 5000 are provided below:

|  |  |  |
| --- | --- | --- |
| Welfare Model 5000  (7 Variable - OLS Regression Estimation) | | |
| Variable | **Coefficient** | **p-value** |
| Intercept | 1.009e+02 | < 2e-16 \*\*\* |
| House Crowding  (people per room) | -1.876e+01 | < 2e-16 \*\*\* |
| House Crowding Squared  (people per room) | 1.782e+00 | < 2e-16 \*\*\* |
| Coping Strategies  (to meet basic food needs) | 8.190e+00 | 1.08e-14 \*\*\* |
| Coping Strategies Squared  (to meet basic food needs) | -6.953e-01 | 0.000255 \*\*\* |
| Living with Host Family | -3.573e+01 | < 2e-16 \*\*\* |
| Saving Ratio in Family | 1.571e-02 | < 2e-16 \*\*\* |
| Debt to Expenditure Ratio | -3.168e-02 | < 2e-16 \*\*\* |
| Income per capita | 4.664e-01 | < 2e-16 \*\*\* |
| Income per capita Squared | 9.531e-05 | 3.83e-15 \*\*\* |
| Family Members in House | -9.231e+00 | < 2e-16 \*\*\* |
| Family Members in House Squared | 3.338e-01 | < 2e-16 \*\*\* |
| Statistical Significance: “\*” = 0.05; “\*\*” = 0.01; “\*\*\*” = 0.001 | | |
| Observations: 4889 | | |
| Variables (excludes quadratic terms): 7 | | |
| R^2: 0.5587 (Adjusted R^2: 0.5577) | | |
| F-statistic: 561.3; p-value: < 2.2e-16 | | |

|  |  |  |
| --- | --- | --- |
| Welfare Model 5000  (9 Variable - OLS Regression Estimation) | | |
| Variable | **Coefficient** | **p-value** |
| Intercept | 7.934e+01 | < 2e-16 \*\*\* |
| House Crowding  (people per room) | -1.907e+01 | < 2e-16 \*\*\* |
| House Crowding Squared  (people per room) | 1.974e+00 | < 2e-16 \*\*\* |
| Coping Strategies  (to meet basic food needs) | 7.727e+00 | 1.50e-13 \*\*\* |
| Coping Strategies Squared  (to meet basic food needs) | -6.950e-01 | 0.000213 \*\*\* |
| Living with Host Family | -3.480e+01 | < 2e-16 \*\*\* |
| Saving Ratio in Family | 1.543e-02 | < 2e-16 \*\*\* |
| Debt to Expenditure Ratio | -3.154e-02 | < 2e-16 \*\*\* |
| Income per capita | 4.547e-01 | < 2e-16 \*\*\* |
| Income per capita Squared | 9.644e-05 | 7.18e-16 \*\*\* |
| Family Members in House | -9.231e+00 | < 2e-16 \*\*\* |
| Family Members in House Squared | 3.359e-01 | < 2e-16 \*\*\* |
| Spices and Condiment Bought with Cash | 1.835e+01 | < 2e-16 \*\*\* |
| Type pf Occupancy: for Rent | 2.188e+01 | < 2e-16 \*\*\* |
| Statistical Significance: “\*” = 0.05; “\*\*” = 0.01; “\*\*\*” = 0.001 | | |
| Observations: 4889 | | |
| Variables (excludes quadratic terms): 9 | | |
| R^2: 0. 0.5713 (Adjusted R^2: 0.5702) | | |
| F-statistic: 499.8; p-value: < 2.2e-16 | | |

The regression estimates from the Welfare Model 666 are provided below. The characteristics identified as highly associated with economic vulnerability and economic resilience are largely unaltered with respect to the Welfare Model 5000. The main difference is the models in the exclusion of: *coping strategies used in the last 6 months* and *sharing a latrine toilet with no security issues.* The former failed to yield statistical significance, whilst the latter is no longer available in the data.

|  |  |  |
| --- | --- | --- |
| Welfare Model 666  (9 Variable - OLS Regression Estimation) | | |
| Variable | **Coefficient** | **p-value** |
| Intercept | 2.064e+00 | 9.69e-13 \*\*\* |
| House Crowding  (people per room) | -4.299e-01 | 6.04e-05 \*\*\* |
| House Crowding Squared  (people per room) | 4.157e-02 | 0.003276 \*\* |
| Coping Strategies  (total amount taken in last 6 months) | -4.437e-01 | 0.012284 \* |
| Coping Strategies  (to meet basic food needs) | 6.605e-01 | 5.85e-12 \*\*\* |
| Coping Strategies Squared  (to meet basic food needs) | -7.825e-02 | 2.88e-05 \*\*\* |
| Living with host family | -1.353e+00 | 3.22e-07 \*\*\* |
| Toilet shared up to 3 HH and no security issues | 6.185e-01 | 0.000558 \*\*\* |
| Total Savings | 2.470e-05 | 0.003734 \*\* |
| Debt to Expenditure Percentage | -2.076e-05 | 8.05e-14 \*\*\* |
| Income per capita | 1.458e-02 | < 2e-16 \*\*\* |
| Income per capita Squared | -1.813e-05 | 7.14e-05 \*\*\* |
| Family Size | 2.490e-01 | 0.006307 \*\* |
| Family Size Squared | -2.350e-02 | 0.023834 \* |
| Statistical Significance: “\*” = 0.05; “\*\*” = 0.01; “\*\*\*” = 0.001 | | |
| Observations: 666 | | |
| Variables (excludes quadratic terms): 9 | | |
| R^2: 0.4855 | | |
| F-statistic: 47.33; p-value: < 2.2e-16 | | |

Covariates for the 666 and the 5000 models manifest almost identical directional effects. The most notable difference is the sign reversal for *family size* (as correctly predicted in the earlier study; influential outliers or a bias towards a subset of the population was distorting the results for *family size*). The curvature of quadratic terms remains largely unaltered.

**Probit & Logit Models (LDV):**

The Probit and Logit Models take the same functional form of the Welfare Model 5000. The difference lies in their estimation method (see Section 4: Model Specification). Likewise to the OLS estimation, the Limited Dependent Variable (LDV) models have been estimated both in the 7 and the 9-variable form. The estimated coefficients are provided below:

|  |  |  |
| --- | --- | --- |
| Limited Dependent Variable Models:  Probit & Logit | | |
| Variable | **Coefficient:**  **Probit** | **Coefficient:**  **Logit** |
| Intercept | -5.518e-01 | -9.701e-01 |
| House Crowding  (people per room) | 4.956e-01 | 7.649e-01 |
| House Crowding Squared  (people per room) | -1.662e-02 | -3.856e-03 |
| Coping Strategies  (to meet basic food needs) | -2.207e-01 | -3.806e-01 |
| Coping Strategies Squared  (to meet basic food needs) | 2.051e-02 | 3.536e-02 |
| Living with host family | 1.268e+00 | 2.399e+00 |
| Saving Ratio in Family | -5.047e-04 | -8.756e-04 |
| Debt to Expenditure Ratio | 1.242e-02 | 2.249e-02 |
| Income per capita | -1.139e-02 | -2.090e-02 |
| Income per capita Squared | 5.380e-06 | 9.760e-06 |
| Family Members in House | 3.514e-01 | 6.629e-01 |
| Family Members in House Squared | -1.152e-02 | -2.357e-02 |
| Spices and Condiment Bought with Cash | -6.606e-01 | -1.127e+00 |
| Type pf Occupancy: for Rent | -1.072e+00 | -1.956e+00 |
| Observations: 4889 | | |
| Variables (excludes quadratic terms): 9 | | |

The Probit and Logit do not allow an analogous interpretation of regression coefficients as for OLS. The direction of effect is provided by the sign of the coefficient, but their marginal effect is not, as the latter is dependent on the entire model.

**Expenditure Model:**

Following the work of UNHCRs’ ‘Targeting Task Force’ in Lebanon the Welfare Model 5000 has been re-estimated using the natural log of expenditure per capita as an independent variable. The ‘Proxy Means Test’ model developed by the ‘Targeting Task Force’ takes two forms: a first form without expenditures and a second form with expenditures. The same approach is taken here. The results are presented below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Welfare Model 5000 with expenditure | | | | |
|  | **Model 1**  *with expenditures* | **Model 2**  *with expenditures* | **Model 3**  *with expenditures* | **Model 4**  *with expenditures* |
| R-Squared | 0.76 | 0.76 | 0.74 | 0.74 |
| Parsimony/Number of Variables | 9 | 7 | 4 | 3 |
| Sample Size | 4889 | 4889 | 4889 | 4889 |

The variables included in Model 4 are: *log expenditures per capita*, *house crowding* and *income per capita*. *House crowding* by itself can explain 10% of the variation in household expenditures.

**Models from Draft 1:**

A re-cap on the Models developed and presented in Draft 1 are provided below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Recap on the Models (from Draft 1) | | | | | | |
|  | | **Model 1**  *without*  *expenditure* | **Model 2**  *with*  *expenditure* | **Model 3**  *parsimonious*  *with expenditure* | **Model 4**  *super*  *parsimonious* | **Model 5**  *without*  *expenditure* | **Model 6**  *food security* |
| R-Squared | | 0.49 | 0.71 | 0.64 | 0.60 | 0.50 | 0.44 |
| Parsimony/  Number of Variables | | 9 | 7 | 3 | 3 | 10 | 10 |
| Sample Size (HH) | | 666 | 666 | 666 | 666 | 634 | 666 |
| n.b.: the functional form of the above 6 models is provided in Appendix B | | | | | | | |

1. **Diagnostic Tests:**

**6.1 - Multicollinearity:**

No perfect collinearity is a condition of OLS estimation. However, imperfect multicollinearity[[3]](#footnote-3) remains an issue. Multicollinearity will not reduce the predictive power or reliability of the model as a whole, but it will reduce the precision of estimators. This potential loss of precision has to be tackled. Formally, multicollinearity arises due to:

Rho represents the correlation amongst the two regressors. In multivariate analysis multicollinearity occurs as so:

The latter term of the equation is the Variance Inflation Factor (VIF). is the for the regression of against all other covariates. If a regressor is not orthogonal to the other independent variables then its standard error will be inflated. As a rule of thumb VIF levels of 5 and 10 have been determined as serious cases of multicollinearity to be examined. No instances of multicollinearity were detected, except in cases of quadratic specification. In these instances the predictors have been mean centered to reduce their correlation.

**6.2 - F-tests:**

The F-test of fit has been performed on all models. All models had a highly statistically significant goodness of fit.

Where k is number of explanatory variables, n the number of observations, and the confidence level. All models failed to accept the null.

Joint hypothesis F-tests for multiple exclusion have been run. For example:

Where q is the number of restrictions – in this case 2. The F-test accepted the alternative hypothesis for all Welfare models. Tests for joint significance have been run as well.

**6.3 - Model Mis-specification:**

The Ramsey RESET is a test for functional form mis-specification. The limitation of the test is that it does not state how to re-specifty the model (whether to exclude or include extra variables). Model mis-specification appears to be present at times. The mis-specification does *not* occur due to the inclusion of quadratic form.

**6.4 - Heteroscedasticity:**

One of the OLS assumptions is homoscedasticity[[4]](#footnote-4). However, when analyzing empirical data, homoscedasticity is at many times not present. The Breusch-Pagan test and the White test for heterscedasticity are methods to estimate the presence of constant variance. Heteroscedasticity reduces estimate precision, but not predictive power of the model.

The Breusch-Pagen test regresses Y on all the X variables through OLS. It then then obtains the residual values for each observation:

And regresses these onto the covariates:

To then obtain an R-squared value for the residuals:

The Breusch-Pagan LM test statistic is:

The null hypothesis of the test is homoscedasticity. A small p-value signifies a failure to accept the null, and it is therefore recommended to use heteroscedastic robust estimates.

The White test is an extension of the BPLM test that accommodates for nonlinearities. It is an extended form of the BP test that investigates greater forms of heteroscedasticity. However, this test can fail due to model mis-specification rather than heteroscedasticity (must perform Ramsey RESET first).

The tests computed indicate the presence of heteroscedasticity. Estimations using robust standard errors or estimation through GLS, WLS, FGLS should be considered.

**6.5 - Outliers:**

The Bonferroni, Cooks Distance and Leverage tests have identified possible cases of influential outliers. These cases have been thoroughly inspected. Some of these have been excluded from the final model.

1. **Findings:**

**7.1 – OLS:**

One of the findings of the Welfare Model 666 was that expenditures appeared to be under-reported. This is consistent with the findings of the ‘Targeting Task Force’ in Lebanon. The results from the Welfare Model 5000 are alike. The current model predicts less ‘poor’ than reported ‘poor’. Reported expenditures and predicted expenditures using Model 666 were almost identical for low and moderate vulnerability, whilst high and extreme vulnerability cohorts manifested a 9-to-10-percent discrepancy between declared and predicted values. If the results of the estimates were correct, then expenditures tend to be under-reported in the poorer cohorts. Given that the range of values that define extreme and high vulnerability are smaller, than for moderate and low vulnerability (extreme has a range of 40; high of 30; moderate of 80; and low is unbounded), it is more likely that misreporting occurs in the poorer cohorts.

|  |  |  |
| --- | --- | --- |
| Predictions of the Welfare Model 666 | | |
| Vulnerability as Expenditure | **Reported** | **Predicted by Model 1** |
| Low (150+ JD) | 14.11% | 15.32% |
| Moderate (70 – 150 JD) | 20.42% | 19.22% |
| High (40 – 70 JD) | 25.08% | 34.98% |
| Extreme (<40 JD) | 40.39% | 30.48% |

Model 666 predicts 30% extreme vulnerability, 35% high vulnerability, 20% moderate vulnerability and 15% living in low vulnerability (see Appendix C or table above). The Welfare model predictions based on governorate indicate that Irbid and Mafraq have the highest ratios of vulnerability, whilst Amman, Balqa and Madaba are subject to more constant ratios of vulnerability (see Appendix C).

Welfare Model 5000 identifies 2484 cases of ‘poor’ out of 2986 self-reported cases of ‘poor’. The correctly predicted rate is 83-percent.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Welfare Model 5000: OLS (9 variables) | | | | |
| True = 0 | **True = 1** | **Pred. = 0** | **Pred. = 1** | **Percentage Poor Correctly Predicted** |
| 1903 | 2986 | 2405 | 2484 | 83% |
| Total = 4889 | | | | |

**7.2 – LDV:**

The Probit and Logit models predict very similarly. However, these models appear to over-predict ‘expenditure per capita’. The LDV models report more cases of ‘poor’ than self-reported cases of ‘poor’. The tables that follow illustrate the discrepancies between predicted and reported values of expenditures (reported scores are the “True” values, however these might not necessarily be true in reality). The LDV models are employed to identify households living with no more than 50 JD per person per month. Prediction = 1 is therefore ‘poor’.

|  |  |  |
| --- | --- | --- |
| Probit/Logit Predictions (template) | | |
|  | **Pred. = 0** | **Pred. = 1** |
| True = 0 | Correct | *Incorrect* |
| True = 1 | *Incorrect* | Correct |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Probit (7 variables) | | | | |
|  | **Pred. = 0** | **Pred. = 1** | **Total True** | **Percentage Correctly Predicted** |
| True = 0 | 1259 | 644 | 1903 | 66% |
| True = 1 | 349 | 2637 | 2986 | 88% |
| Total Pred. | 1608 | 3281 | = 4889 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Probit (9 variables) | | | | |
|  | **Pred. = 0** | **Pred. = 1** | **Total True** | **Percentage Correctly Predicted** |
| True = 0 | 1287 | 616 | 1903 | 68% |
| True = 1 | 344 | 2642 | 2986 | 88% |
| Total Pred. | 1631 | 3258 | = 4889 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Logit (7 variables) | | | | |
|  | **Pred. = 0** | **Pred. = 1** | **Total True** | **Percentage Correctly Predicted** |
| True = 0 | 1283 | 620 | 1903 | 67% |
| True = 1 | 362 | 2624 | 2986 | 88% |
| Total Pred. | 1645 | 3244 | = 4889 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Logit (9 variables) | | | | |
|  | **Pred. = 0** | **Pred. = 1** | **Total True** | **Percentage Correctly Predicted** |
| True = 0 | 1301 | 602 | 1903 | 68% |
| True = 1 | 352 | 2634 | 2986 | 88% |
| Total Pred. | 1653 | 3236 | = 4889 |  |

The Logit and Probit estimators, for both the *hepta* and the *ennea*-variable models, predict almost identically. The discrepancy of the predictions amongst the models is almost negligible. The sample data includes 1903 reported cases of ‘non-poor’ and 2986 reported cases of ‘poor’. The *ennea*-Probit model predicts 1631 ‘non-poor’ and 3258 ‘poor’. The *ennea*-Logit model predicts 1653 ‘non-poor’ and 3236 ‘poor’. These statistics indicate that the Logit reduces the margin of predictive error. However, observing the diagonals we notice that the Logit reduces the margin of error for prediction equals 0, but not for prediction equals 1.

**7.3 – OLS and LDV:**

Below are the tabulated results from the three models. The sample ‘poor’ rate is 61-percent. OLS predicts 51-percent, the Logit predicts 66-percent and the Probit predicts 67-percent. The percentages rounded to two significant figures are identical in the 7 and the 9 variable model. However, when rounded to three or four significant figures, the 9-variable model performs better.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Logit vs. Probit vs. OLS  (7 Variables) | | | | | | | | |
| Logit | | **Probit** | | | **OLS** | | **True**  **(Reported in Sample)** | |
| Pred. = 0 | **Pred. = 1** | **Pred. = 0** | | **Pred. = 1** | **Pred. = 0** | **Pred. = 1** | **True = 0** | **True = 1** |
| 1645 | 3244 | 1608 | 3281 | | 2404 | 2493 | 1910 | 2987 |
| Percent Poor: 66% | | Percent Poor: 67% | | | Percent Poor: 51% | | Percent Poor: 61% | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Logit vs. Probit vs. OLS  (9 Variables) | | | | | | | | |
| Logit | | **Probit** | | | **OLS** | | **True**  **(Reported in Sample)** | |
| Pred. = 0 | **Pred. = 1** | **Pred. = 0** | | **Pred. = 1** | **Pred. = 0** | **Pred. = 1** | **Reported = 0** | **Reported = 1** |
| 1653 | 3236 | 1631 | 3258 | | 2411 | 2486 | 1910 | 2987 |
| Percent Poor: 66% | | Percent Poor: 66% | | | Percent Poor: 51% | | Percent Poor: 61% | |

1. **Conclusions:**

This second econometric study has identified the characteristics statistically associated with economic resilience and economic vulnerability amongst refugee households in Jordan. The results from this empirical study corroborate the findings of the former research. This second draft on the ‘Vulnerability Targeting Model’ has provided a validation of both the ‘Welfare Model’ and of the identified vulnerability-related characteristics. The large sample size provides confidence in the direction and magnitude of the marginal effect of each vulnerability predictor. The re-estimation of the model has been greatly successful. The Welfare Model 5000 has an R-squared of 57%.

The econometric analysis undertaken suggests that: *house crowding* (measured as “people per room*”), living with a host family,* the *debt-to-expenditure ratio*, and *family size* are all negatively associated with expenditure per capita. The presence of the aforementioned characteristics amongst refugees exerts a downward relation on expenditure per capita. In contrast, the remaining characteristics: *coping strategies to meet basic food needs* (from a list of 8 possible strategies), the *level of saving in the family,* and *income per capita* are positively associated with expenditure per capita; their presence thereby exerts an upward effect on expenditure per capita. The *ennea*-variable Welfare Model identifies an extra two characteristics that positively relate to expenditures: *having purchased spices or condiments with cash in the last 7 days* and *living in a rent type of occupancy.* The additional two variables result in a 1.25 percentage point rise in the R-squared. The main identified characteristics are the same in both studies.

The findings of this study indicate that the former *four*-variables instigate economic vulnerability amongst refugees, whilst the latter *three*-characteristics predispose households into economic resilience by inducing higher expenditures. However, only the summation of these variables establishes the predicted economic vulnerability of a household. The presence of more negative-affecting variables does not imply a higher predisposition to vulnerability, as the marginal effect of impact varies across characteristics, and it is this in turn that determines how susceptible individuals are to economic vulnerability.

Targeted humanitarian assistance by employment of the Welfare Models developed in this quantitative empirical research analysis could be effectuated in five ways: 1. Apply the OLS model to predict the expenditures of households and provide assistance based on the eligibility criteria; 2. Utilize the Probit or the Logit model to dichotomize households between ‘poor’ and ‘non-poor for the same poverty threshold; 3. Employ both models, eligibility for assistance occurs when the OLS predictions are consistent with the LDV predictions, and cases of mismatch require further investigation; 4. Provide assistance purely based on the LDV predictions and apply the OLS model only for appeal cases; 5. Provide assistance purely based on the OLS predictions and apply the LDV model only for appeal cases. A final alternative is a mixed usage of the econometrical models, and sector based rules that flag specific non-monetary vulnerabilities. The development of a comprehensive re-inclusion policy is also necessary.

**Appendix:**

**Appendix A:**

Figure : Quadratic Specification

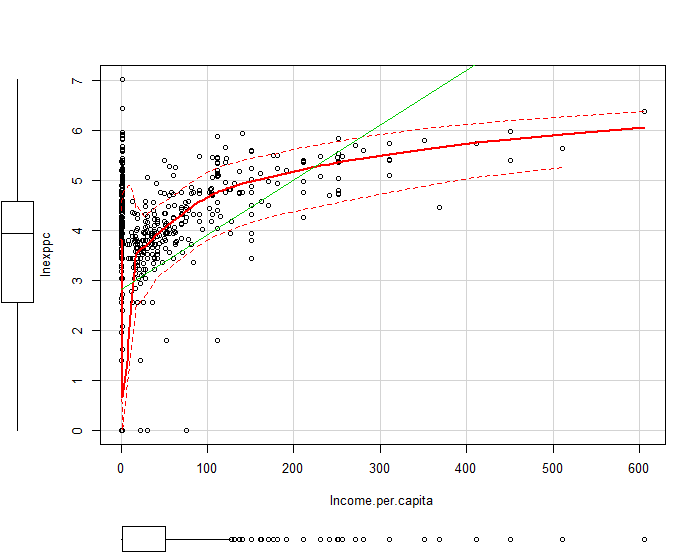
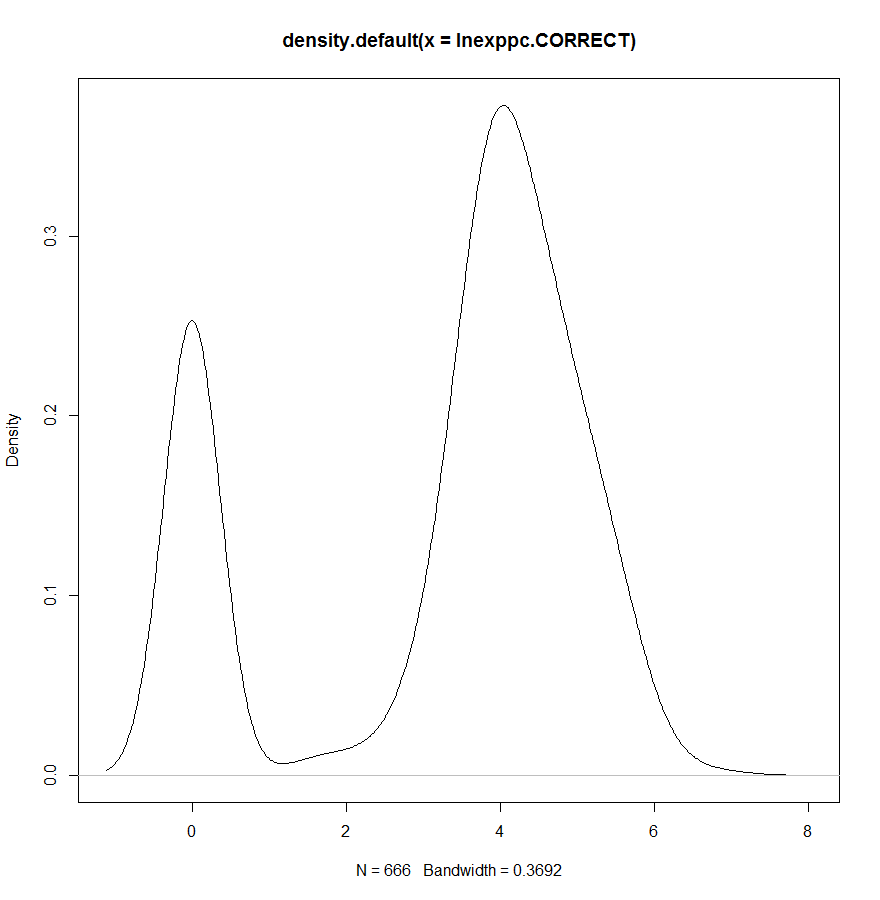
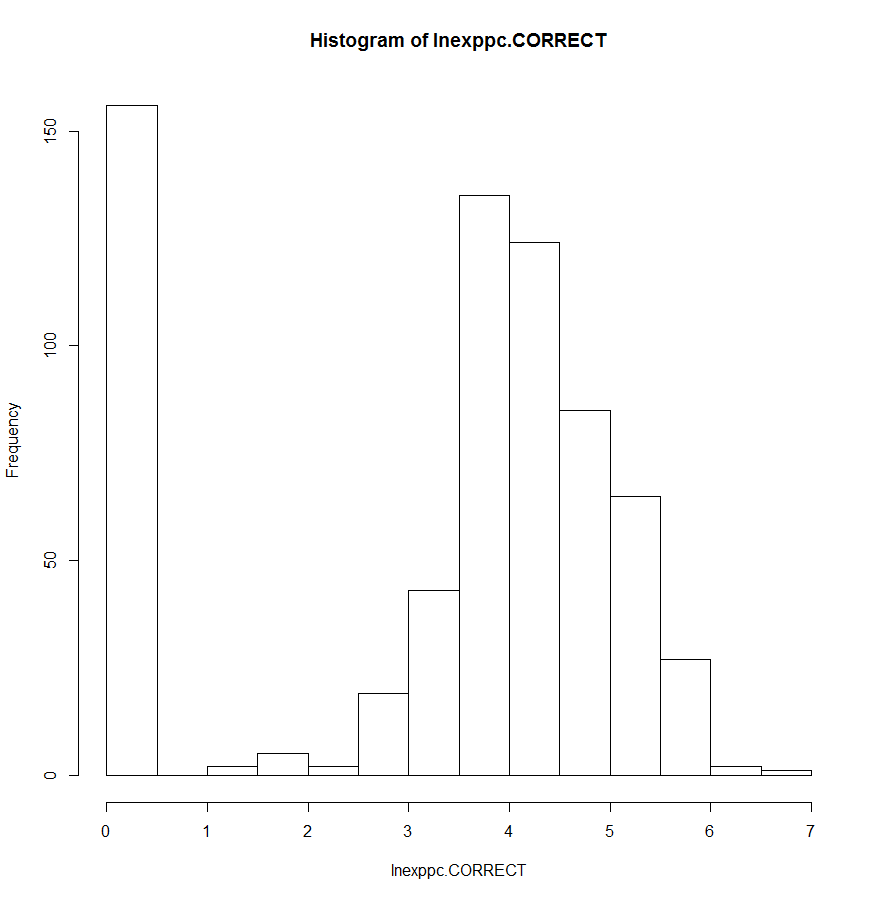


Figure 2: Density Log Expenditure (666)

Figure 3: Histogram Log Expenditure (666)



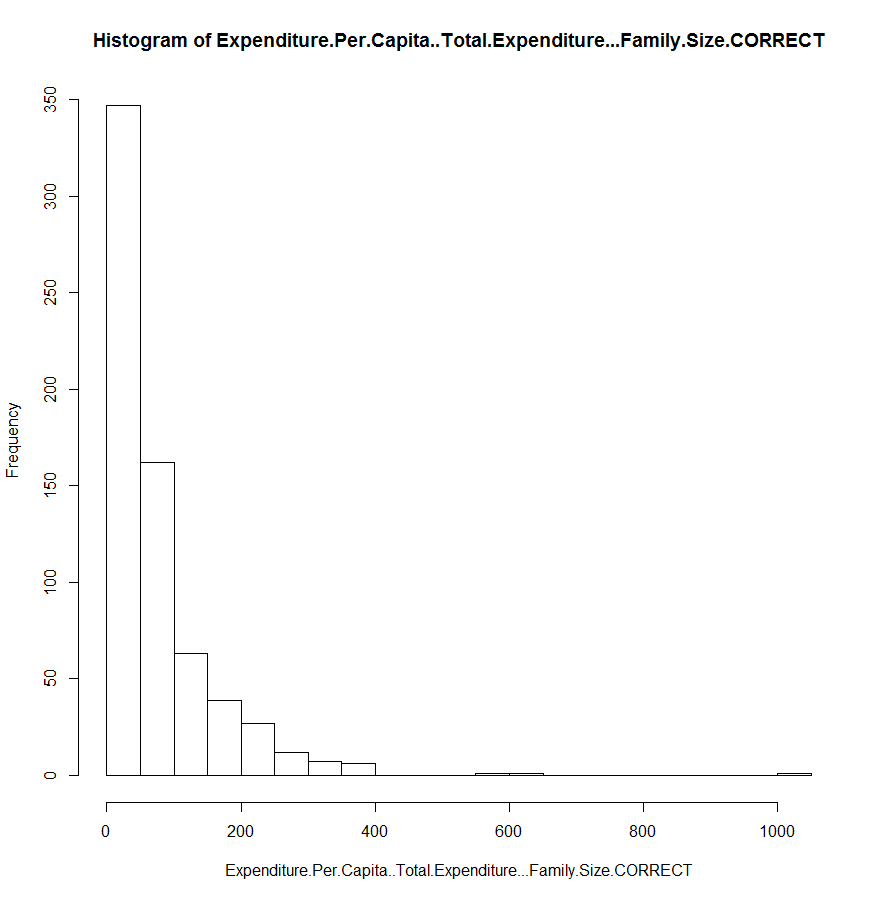


Figure 2: Density Level Expenditure (666)

Figure 5: Histogram Level Expenditure (666)

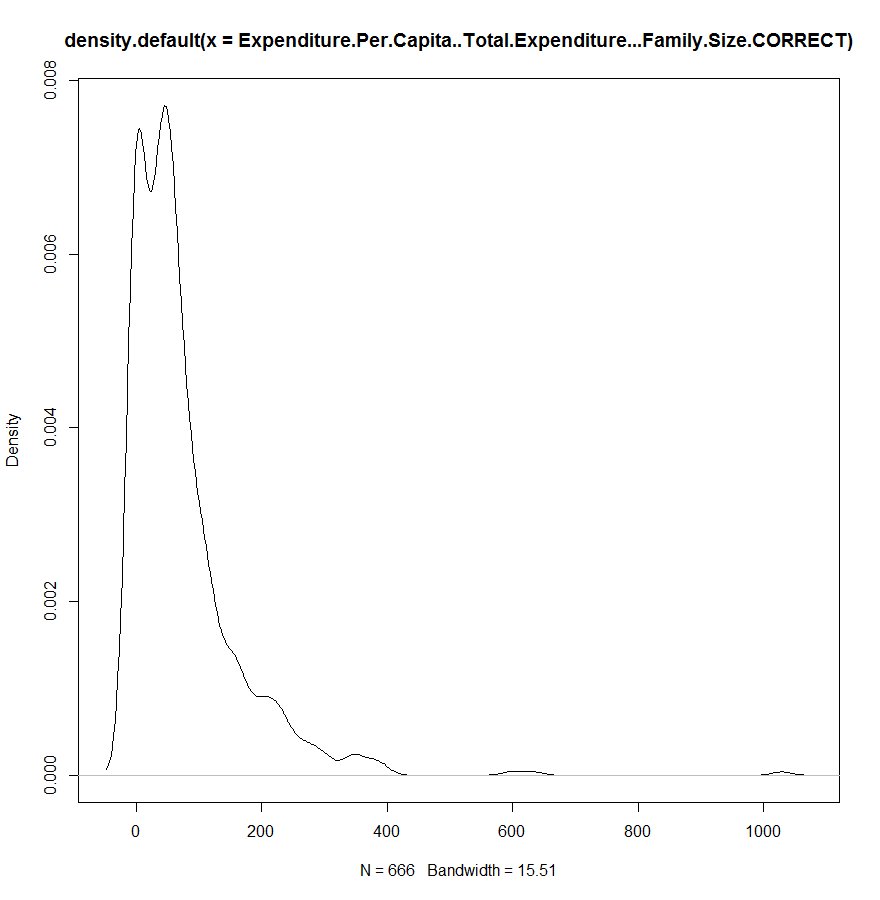


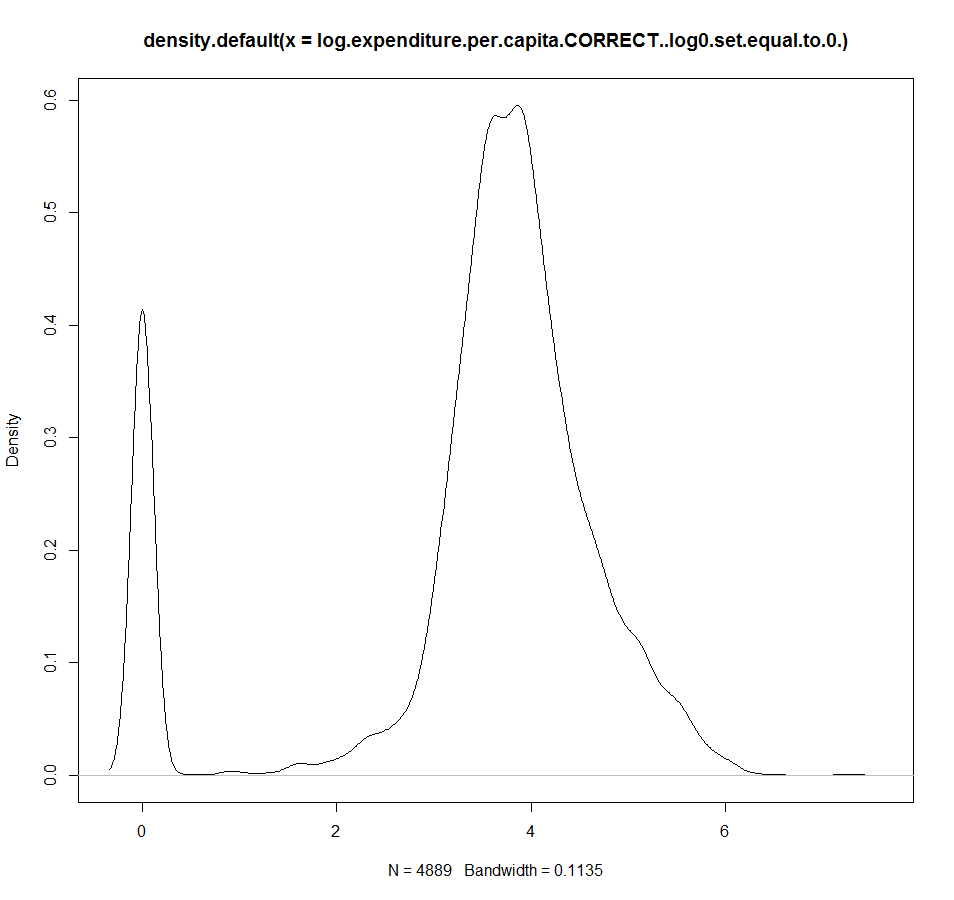
Figure 6: Density Log Expenditure (5000) 

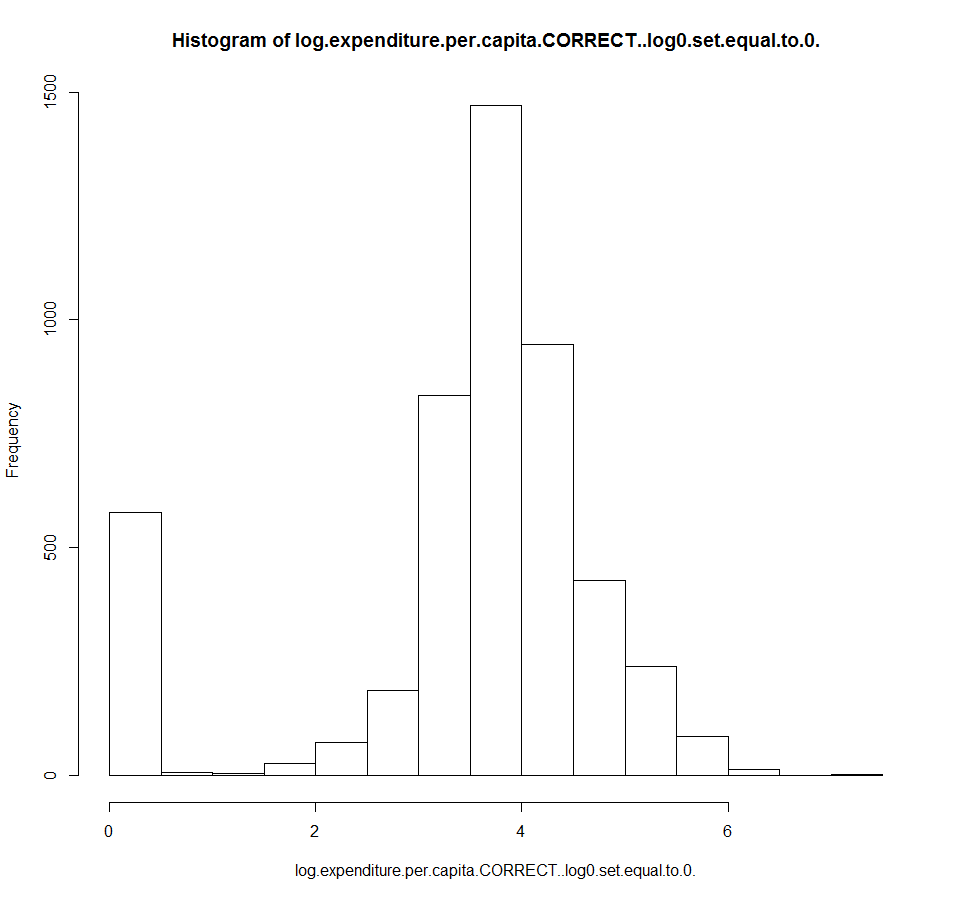
Figure 7: Histogram Log Expenditure (5000) 

Figure 8: Density Level Expenditure (5000)

Figure 9: Histogram Level Expenditure (5000)

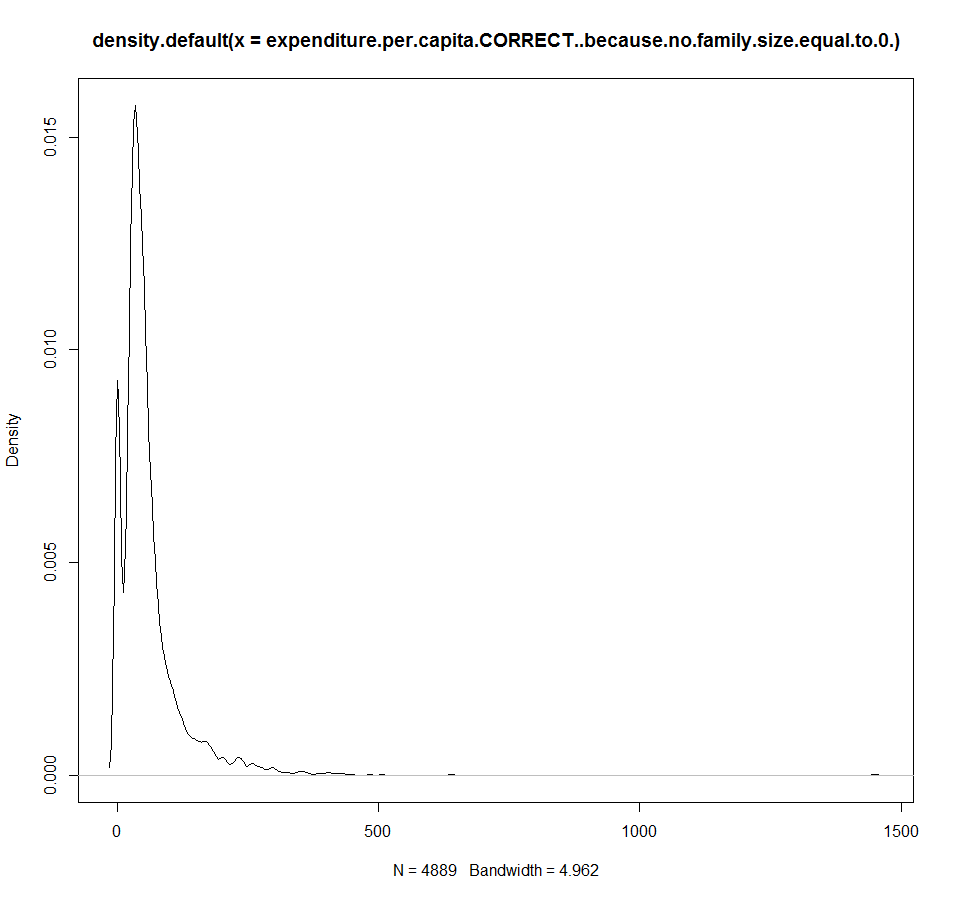
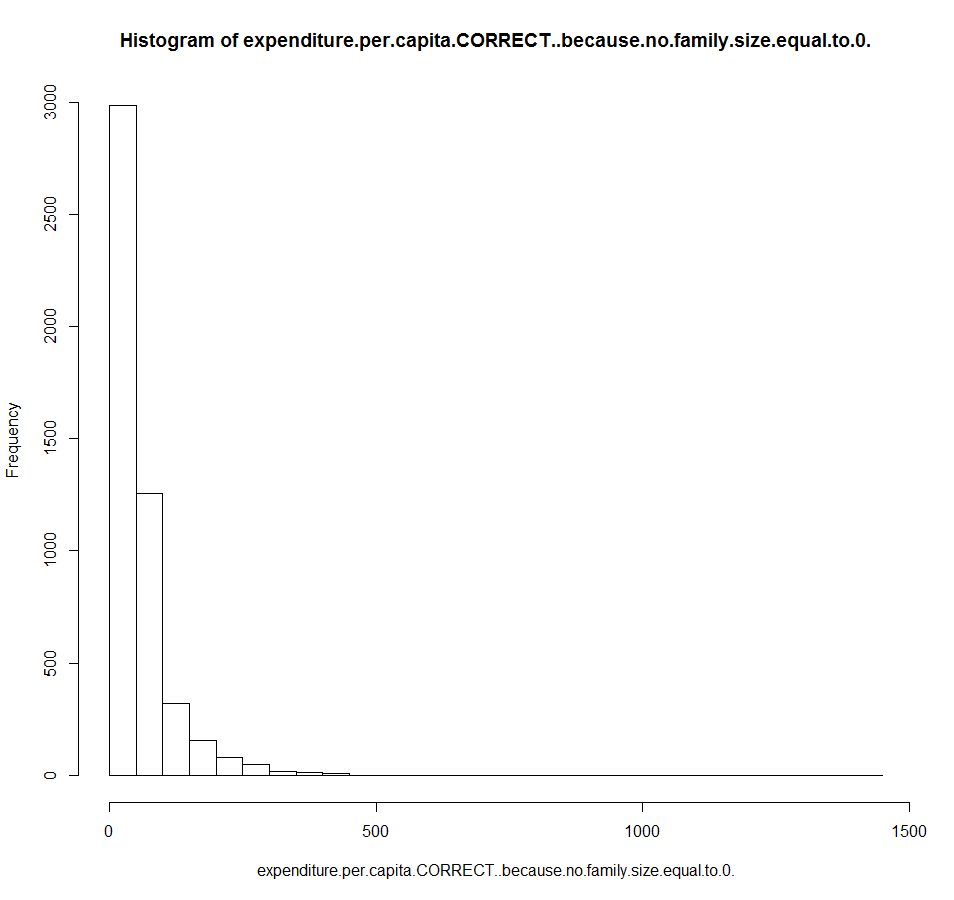
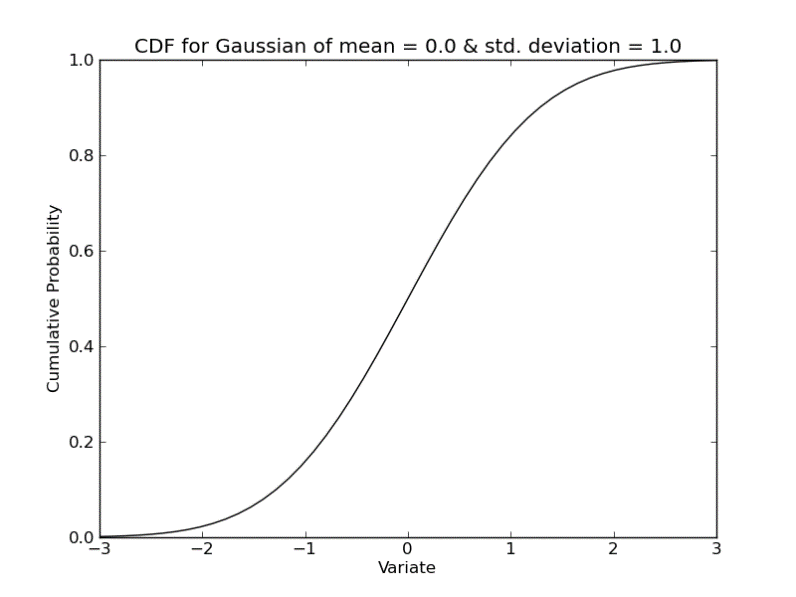


Figure 10: the Cumulative Distribution Function

****

**Appendix B:**

**Welfare Model (1): without expenditure**

Welfare model 1 is log-transformed, includes 9 different variables (quadratic terms are not new variables), has a sample size of 666 and *excludes* expenditure per capita as a dependent variable. Its R-squared stands at 49%.

+

**Welfare Model (2): with expenditure**

Welfare model 2 is log-transformed, includes 7 different variables, has a sample size of 666 and *includes* expenditure per capita as a dependent variable. Its R-squared stands at 71%.

+

**Welfare Model (3): parsimonious with expenditure**

Welfare model 3 is the so-called “parsimonious model[[5]](#footnote-5)”. It is a level model consisting of only 3 different variables, has a sample size of 666 and *includes* the natural log of expenditure per capita as its dependent variable. Its R-squared stands at 64%.

+

**Welfare Model (4): super parsimonious**

Welfare model 4 is the simplest, yet a highly predictive model. It is a level model consisting of only 3 variables and no quadratic terms, it has a sample size of 666 and *includes* ln expenditure per capita as a dependent variable. Its R-squared stands at 60%.

**Welfare Model (5): without expenditure**

Welfare model 5 is an exact replica of Welfare model 1 with the additional variable “number of working age people in household”. It is log-transformed, includes 10 different variables, has a sample size of 634 and *excludes* expenditure per capita as a dependent variable. Its R-squared is 50%.

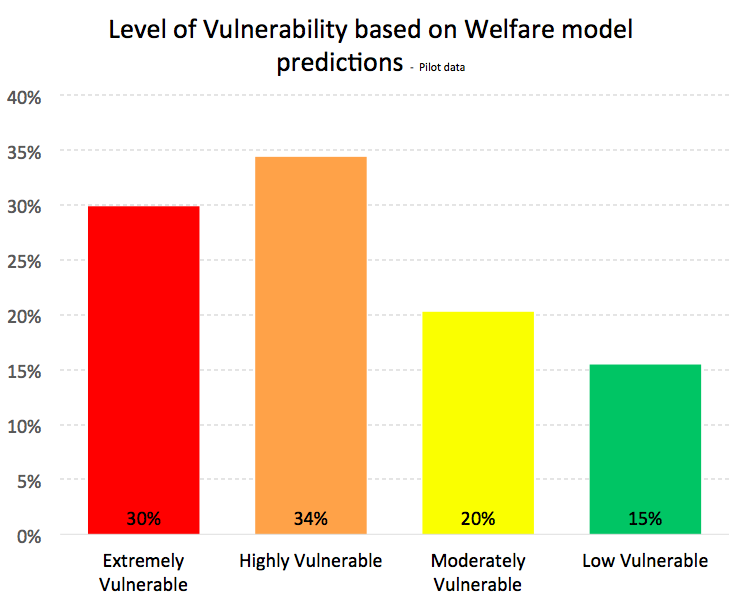
+

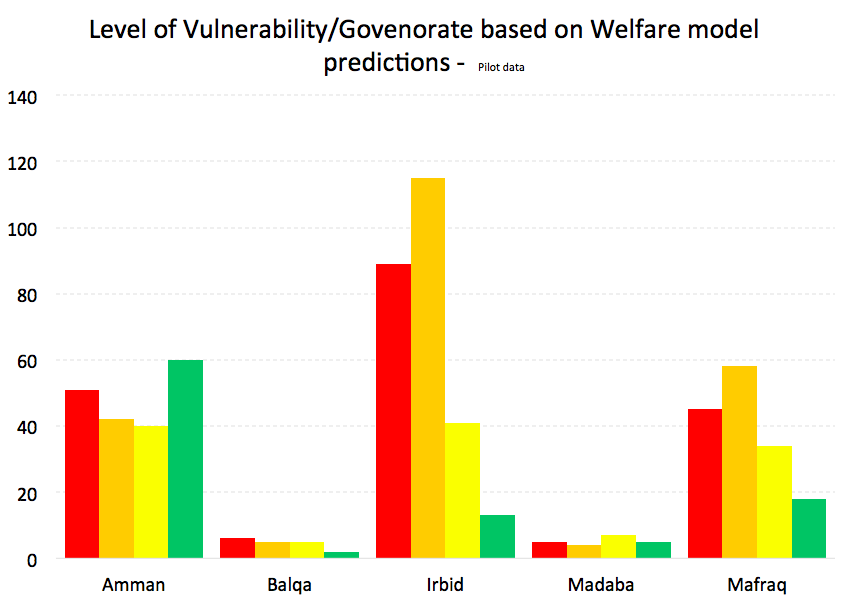
**Welfare Model (6): food security**

Welfare model 6 is a preliminary non-monetary vulnerability predicting model. It is a level model, consisting of 10 variables, has a sample size of 666 and *includes* coping strategies used to meet basic food needs as a dependent variable. Its R-squared stands at 44%.

+

**Appendix C:**





1. Draft 1 – September 09, 2014 [↑](#footnote-ref-1)
2. UNHCR email: [santacro@unhcr.org](mailto:santacro@unhcr.org) ; Private email: [marco.santacroce@cantab.net](mailto:marco.santacroce@cantab.net) [↑](#footnote-ref-2)
3. Correlations amongst regressors. [↑](#footnote-ref-3)
4. Constant variance of residuals. [↑](#footnote-ref-4)
5. A model that achieves a desired level of explanatory power using as few predictors as possible to attain that. [↑](#footnote-ref-5)