

SAS MACROS FOR ECONOMETRIC ANALYSIS

Copyright(C) SSCNARS, IASRI (ICAR), NEW DELHI. ALL RIGHTS RESERVED.

These SAS Macros for application in the field of economics and other social sciences have been developed under the NAIP Consortium Strengthening Statistical Computing for NARS (SSCNARS). This group of macros is a free software, you can redistribute it and/or modify it under the terms of GNU General Public License as published by the Free Software Foundation. Basically all copies, modifications or derivative works must allow the user to freely use the Macro, to copy, modify and distribute, and must carry this same License for free use. Source code must be distributed, but distribution charges of any magnitude are permitted. This macro is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY, without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. A copy of the GNU General Public License can be obtained from Free Software Foundation, or <http://www.gnu.org/licenses/gpl.html>.

Developed by – Sivaramane N

Date: 13.09.2011

VERSION 1.0: 13.09.2011

Features:

This group of macros include (i) diversity indices such as Barry's Index, Normalized Barry's Index, Herfindahl's Index and Normalized Herfindahl's Index, (ii) Instability Index using Cuddy-Della Valle approach, (iii) Compound growth rate showing significance levels, (iv) Garret Scoring Technique and (v) Demand Analysis using LA-AIDS model. A brief review of the techniques along with the procedure for calling the SAS macros is given below. This material comes along with SAS catalogue file containing all the demonstrated macros and a model SAS dataset (test_61b).

General instructions for executing SAS macros

1. Create a folder (e.g. SAS_Econometric_macros) anywhere in desktop/laptop and copy the files (sasmacr and test_61b) into it. Here sasmacr.sas7bcat is the catalogue of macros and
2. Open SAS program (sas.exe) and create a new library (e.g. econlib) by specifying the path (e.g. C:\users\user\desktop\SAS_Econometric_macros)
3. Create data file in SAS or Import the datafile (xls, xlsx, dat) which you want to analyse into SAS dataset
4. Specify the library name, parameters and path in calling macro statement as described below in the steps for executing/calling macros. For template of the code for invoking/calling macros, please see the details given at the end of each section in the sequel. Further, a ready-reckoner is provided containing the executable syntax only in a separate file named as **(Ready Reckoner for Executing SAS Macros for Econometric Analysis.doc)**
5. Execute the code.
6. Check for errors (RED color lines) in LOG window. Use the output only if no error is present.

7. For output in WORD format, use the code `ODS RTF FILE="Outfile.doc";` (in lieu of outfile any name can be provided) before executing macro and use `ODS RTF CLOSE ;` after the macro statements.
8. For any queries or suggestions, write to sscnars@iasri.res.in or sivaramane@iasri.res.in

A brief outline of different procedures along with the steps for execution of the macros is presented below separately for each type of index and demand analysis.

1. DIVERSITY INDEX

There are several diversity indices developed which have their own pros and cons. These indices find application in several fields of disciplines including science, economics, literature, etc. It was used in the computation of biodiversity by counting the number of species in a habitat, diversity in field crops, say, diversity in rice cultivation (several varieties including local, high-yielding and hybrid varieties) in an agro-ecological zone, market competition, etc. It can also be used in knowing the diversity among people based on socio-economic factors.

The **Herfindahl index (HI)** also known as **Herfindahl–Hirschman Index**, named after economists Orris C. Herfindahl and Albert O. Hirschman, is an indicator to measure (market) concentration.

$$HI = \sum_{i=1}^N s_i^2$$

where s_i is the share of crop i in the a given space, and N is the number of items.

The Herfindahl Index (HI) ranges from $1 / N$ to one.

A HI index below 0.01 (or 100) indicates high diversity.

A HI index below 0.1 (or 1,000) indicates moderate diversity.

A HI index between 0.1 to 0.18 (or 1,000 to 1,800) indicates moderate concentration.

A HI index above 0.18 (above 1,800) indicates high concentration.

There is also a normalised Herfindahl index. Whereas the Herfindahl index ranges from $1/N$ to one, the **NORMALIZED HERFINDAHL INDEX** ranges from 0 to 1. It is computed as:

$$NHI = \frac{HI - 1/N}{1 - 1/N}$$

The **Barry's index (BI)** is simply one minus Herfindahl's index and similarly, normalized Barry's index (NBI) is one minus normalized Herfindahl's index.

$$BI = 1 - HI$$

$$NBI = 1 - NHI$$

The macro for diversity index computes Barry's index (BI), Normalized Barry's index (NBI), Herfindahl's index (HI) and Normalized Herfindahl's index (NHI).

Interpretation

The higher the Barry's index, higher the diversity, while the lower the Herfindahl index, the higher the diversity is. Normalized indices are generally preferred as it ranges from zero to one which facilitates better interpretation.

Macro Execution

The command for invoking the macro is explained in the context of the model data given below.

The model dataset

The model data set contains the acreage of different crops sown by 5 selected farmers in a particular region. The diversity index of cultivation of wheat and barley (i.e. among cereals) can be computed by calling the macro as illustrated below.

```

Data econlib.ex1;
Input paddy wheat barley gram peas;
Cards;
5 7 8 5 2
2 5 6 7 9
9 9 9 9 9
8 6 5 4 1
8 6 5 4 1
Run;

```

Invoking the macro

```

options mstored sasmstore=econlib;
%divindex (indata=econlib.ex1, outdata=newdata, var= wheat barley, outvar=cereal, opt=3)

```

Note

Indata=input data (specify with libname)

Outdata= output data (you can specify any name or use input data name to replace it)

Var= variables used in the computation of the selected index

Outvar= Variable name of newly created Indices (you can specify any name).

opt=1 for HI, 2 for NHI, 3 for BI ,4 for NBI or any number for all the indices.

Index	Highly diversified	Highly concentrated
Herfindahl's index (HI)	1/S	1
Normalized Herfindahl's index (NHI)	0	1
Barry's index (BI)	(S-1)/S	0
Normalized Barry's index (NBI)	1	0

Where S=No. of crops/species

Reference

http://en.wikipedia.org/wiki/Herfindahl_index

2. CUDDY DELLA VALLE INSTABILITY INDEX

Cuddy Della Valle Instability index (Cuddy and Della Valle, 1978 & Della Valle, 1979) is a modification of CV to accommodate for trend, which is commonly present in time series economic data. This method is superior over other scale dependent measures such as standard deviation or Root Mean Square of the residuals (RMSE) obtained from the fitted trend lines of the raw data, and hence suitable for cross comparisons..

The Cuddy Della Index (I_x) was calculated as follows:

$$I_x = \frac{SEE}{\bar{y}} \cdot 100$$

Where, I_x = Instability index

SEE = Standard error of the trend line estimates

\bar{y} = Average value of the time series data

Alternatively, I_x could be alternatively measured as:

$$I_x = CV \sqrt{(1 - \bar{R}^2)}$$

Where, CV = Coefficient of variation (σ/\bar{X})

\bar{R}^2 = Adjusted coefficient of multiple determination

Caution

The SAS macro is made to accommodate only the linear trend present in the data.

Macro execution

The command for invoking the macro is explained in the context of the model data given below.

Model dataset

The model dataset contains time series information on the marine exports from India in tonnes from 1999 to 2008. The Cuddy Della Valle Instability index which is suitable for time series data is computed by calling the CDI macro as shown in the illustration. In the following illustration, CDI is computed for the period 2000 (2nd obs) to 2006 (8th obs).

```
Data marine;
Input year m_exports;
Cards;
1999 1121977
2000 1233938
2001 1056658
2002 1190011
```

```
2003 1417853
2004 1249689
2005 1423707
2006 1311250
2007 1412334
2008 1597722
;
```

Invoking the macro

```
options mstored sasmstore=econlib;
%CDI (data=sashelp.air, var=air, pfrom=5, pto=30)
```

Note

Var=specify variable name (only one variable need to be specified)

pfrom=period from (mention the starting observation number)

Pto=period to (mention the final obs no. which need to be included in the analysis)

References

Cuddy,J.D.A, and P.A.Della Valle (1978). "Measuring the instability in time series data." *Oxford Bulletin of Economics and Statistics*. **40**(1): 79-85.

Della Valle,P.A. (1979) "On the instability index of time series data: A generalization", *Oxford Bulletin of Economics and Statistics*. **41**(3): 247-248.

3. COMPOUND ANNUAL GROWTH RATE (CGR)

Compound growth rate is simply a compounding of annual growth rates over period. It can be easily computed using two data points with constant returns as in case of Fixed deposits. However, in case of annual growth rates which are not constant, but for monotonically increasing or decreasing functions, the compound growth rate is computed based on its fit using non-linear models, especially, the exponential model. The exponential model is more commonly used in economic analysis.

Conventionally, the CGRs were estimated after converting the growth model to semilog form and estimated through Ordinary Least Square (OLS) technique assuming multiplicative error term. However, there are several problems associated with this including the difficulty in estimating Standard Error of Estimates of original parameters (Prajneshu and Chandran, 2005). Hence, a non-linear estimation technique for solving exponential model assuming additive error terms was used in this syntax.

$$y_t = \text{constant} * (1 + \text{CGR})^t + \epsilon_t$$

Note that Marquardt algorithm was used in the algorithm and moving average can be used to smoothen the data to remove bias induced by the outliers. The macro generates both growth rates (expressed in per cent terms and no need to convert them and secondly, it generates significance probability and standard errors useful for reporting).

Caution

The SAS macro is valid for the specified period of data where the function is increasing or decreasing monotonically. This has to be verified through graphs. The CGRs are represented as percentages.

Macro execution

The command for invoking the macro is explained in the context of the model data given below.

Model dataset

The model dataset contains time series information on the marine exports from India in tonnes from 1999 to 2008. The compound growth rate which is solved using non-linear estimation procedure and provides approximate standard error with the option of smoothing highly volatile data is computed by calling the CGR macro as shown in the illustration. In the following illustration, CGR is computed for the period from 1999 (1st obs) to 2006 (8th obs).

Data marine;	
Input year m_exports;	
Cards;	
1999	1121977
2000	1233938
2001	1056658
2002	1190011
2003	1417853
2004	1249689
2005	1423707

2006	1311250
2007	1412334
2008	1597722
;	

Invocation of the macro

<pre>options mstored sasstore=econlib; %cgr(data= marine, var= m_exports, ma=3, pfrom=1, pto=8)</pre>

Note

Data=specify dataset name in full with library

Var=specify variable name (only one variable need to be specified)

Ma=moving average (if moving average is not required you can remove the term)

pfrom=period from (mention the starting observation number)

Pto=period to (mention the final obs. no. which need to be included in the analysis)

Model dataset

The above dataset (marine exports) provided in the Cuddy Della Valle Instability Index chapter may be tried for applying this macro.

Reference

Prajneshu and Chandran, K.P. (2005) Computation of compound growth rates in agriculture: Revisited, *Agricultural Economic Research Review*. **18**: 317-324.

4. GARRET SCORING TECHNIQUE

This is also known Garret ranking technique. This technique is widely in social sciences for policy analysis. This method takes into account of the ranks of the respondents on certain choices based on the importance of the each choice to them. The rank 1 is assigned for the best choice. The procedure includes converting the ranks to percentage position and to Garrett scores. The Garrett scores are then averaged over various choices to find out the order and magnitude of importance of the choices to the overall respondents.

The percentage position is calculated as

$$\text{Percentage Position} = \left(\frac{R_{ij} - 0.5}{N_j} \right) * 100$$

Where, R_{ij} =Rank assigned for i^{th} choice by j^{th} individual and

N_j = Total number of ranks assigned by j^{th} individual.

The percentage position of each rank is converted to Garret scores using Garret table (Ref. Garret and Woodworth, 1969). Then the scores for each choice are averaged over all respondents. The higher the score is the higher is the importance of the particular choice.

SAS macro

The SAS macro can use any number of choices (options/problems) and the user may assign ranks to few or all the choices. Assign different numbers to the choices (only numerals or periods/dots are to be filled) and fill the data sheet against different ranks as variable names (say, rank1, rank2,...rank10)

Macro execution

The command for invoking the macro is explained in the context of the model data given below.

Model dataset

The model dataset contains information on ranks (4 ranks) assigned by 20 farmers in a selected village for 15 statements containing problems faced by them in water markets. Garret ranking technique helps in ranking all major problems (15 statements) faced by the farmers in the particular village on water markets.

Data garetdata;				
Input farmer	rank1	rank2	rank3	rank4;
Datalines;				
1	1	10	.	.
2	5	10	13	15
3	9	10	12	.
4	9	10	13	14


```

5    1    10    12    15
6    9    10    .     .
7    9    10    12    15
8    4    5     10    .
9    5    9     10    15
10   1    10    12    15
11   1    10    .     .
12   1    5     10    15
13   4    9     10    15
14   10   12    .     .
15   1    10    12    15
16   9    10    12    13
17   1    .     .     .
18   1    10    13    15
19   9    .     .     .
20   1    10    13    15
;run;

```

Invoking the macro

```

options mstored sasmstore=econlib;
%garette (data=garettdata, var=rank, no=0, ns=15, mr=4)

```

Note

Var=specify generic variable name (e.g. For rank1-rank10, specify var=rank)

no=correction factor (user can remove this or leave it as zero)

ns=no. of choices or statements

mr=maximum rank (specify the maximum no. of ranks used, it may or may not be equal to the no. of statements).

It is important to note that the dataset may contain missing values and one can rank all the statements also.

Reference

Garret, H.E. and R.S.Woodworth (1969). Statistics in psychology and education, Vakils, feffer and Simons Pvt. Ltd., Mumbai.

5. DEMAND MODEL

There are various models used in the estimation of demand for different articles for consumption. Presently the demand analysis is dominated by “almost ideal demand system“ (AIDS) which is an indirect demand estimation technique using system of equations. This model is considered to be ideal for it posses the following desirable properties.

- Gives an arbitrary first order approximation to any demand system; it satisfies the axioms of choice (e.g. transitivity or acyclicity) exactly;
- It aggregates perfectly over consumers without invoking parallel linear Engel curves;
- It has a functional form which is consistent with known household-budget data;
- It is simple to estimate largely avoiding the need for non-linear estimation; and
- It can be used to test the restrictions of homogeneity and symmetry through linear restrictions on fixed parameters.

The general form of the Almost Ideal Demand System (AIDS) with linear approximation (LA) is given by

$$w_i = \alpha_i + \sum_j \gamma_{ij} \ln P_j + \beta_i \ln \left(\frac{X}{P} \right) + v_i$$

Where,

w_i = expenditure share of the i^{th} commodity

P_j = Price of the j^{th} commodity

X = is the total expenditure

v_i = error term

P = is the price level

The theoretical demand restrictions in terms of adding up, homogeneity in prices and income, and the symmetry of cross effects of demand functions are given below

Adding up $\sum_{i=1}^n \alpha_i = 1, \sum_{i=1}^n \gamma_{ij} = 0$ & $\sum_{i=1}^n \beta_i = 0$

Homogeneity $\sum_j \gamma_{ij} = 0$

Symmetry $\gamma_{ij} = \gamma_{ji}$

Application of LA-AIDS macro

The linear approximate almost ideal demand system used for estimation of demand for selected food groups is given below. The data was extracted as a representative sample of 1000 from the 61st round national sample survey organisation. The LA-AIDS model has several variants and the macro is based on the basic one.

SAS macro

This macro is useful for fitting cross-section data or pooled data appropriately correcting for panel data effects. The variables should be prepared with the same suffices such as `_v` and `_lnp`. The model can accommodate only five consumption groups/articles and the price variables should be provided in log form. Identification number (`ident`) and weights (`wts`) can be used but optional. This model is not suitable for time series consumption data with stochastic trends and in datasets where consumption data have excess zeros. This model utilizes iterative seemingly unrelated regression method for solving the parameters.

Caution

This macro can be used only in case of system of equations with five dependent variables. However, more number of independent variables can be added to it. This macro accounts for linear relationship between dependent and independent variables. Multicollinearity may severely affect the results.

Model dataset

The model data set (test_61b) contains the data on consumption of different food commodities of sampled households in India during 2004-05. The dataset is a part of the 61st round of NSS household consumer survey conducted in all states of India during 2004-05. Since it is a huge dataset, a small part of it is reproduced here for clarity. To perform demand analysis using LA-AIDS approach, the macro code given below is executed.

ident	cereals_v	pulses_v	veg_v	fruits_v	ofood_v
1	166.607	24.66	41.201	4.21	32.99
2	169.099	38.666	30.417	4.984	65.03
3	108.265	20.931	28.871	0	49.97

ident	cereals_lnp	pulses_lnp	veg_lnp	fruits_lnp	ofood_lnp	lmpce30	hhdsz	wts
1	2.36677	3.57989	2.13434	2.03537	3.65581	6.2226	4	4081.1
2	2.32281	3.58597	2.2474	2.19109	3.9551	6.5485	7	775.09
3	2.48735	3.55207	2.00185	2.60334	4.0461	6.1106	5	3236.22

Abbreviations

_v = value of consumption of consumption article x,

_lnp = natural log of prices of articles x,

Hhdsz = household size in numbers

Wts = weights

Lmpce30 = natural log of monthly percapita consumption expenditure.

Invocation of the macro

```
options mstored sasstore=econlib;
%laaids (data=econlib.test_61b, dep1=cereals, dep2=pulses, dep3=veg, dep4=fruits,
dep5=ofood, indep_lninc=lmpce30, indep_z=lhdsz, weights=wts)
```

Note

Data = specify dataset name in full with library

Dep1-dep5 = names of the consumption groups/articles

indep_lninc = natural log of total expenditure of the household

indep_z= other independent variables like household size, education, etc. including dummies

weights = weights (optional)

References

Deaton, A. and J. Muellbauer (1980). An Almost Ideal Demand System. *American Economic Review*, 70:312-326. & 359-68.

<http://support.sas.com/rnd/app/examples/ets/aids/sas.htm>