CPRC Toolkit Note

CREATING AND INTEPRETING CONTOUR PLOTS
USING DASP AND GNUPLOT

Bob Baulch and Karl Shutes[1]

October, 2008

Summary

This notes describes how contour plots, which are two dimensional representations of welfare distributions that can be regarded as a continuous analogue to transition matrices, can be creating using the Stata package DASP and the graphics package gnuplot. The procedure is illustrated with panel data from the Vietnam Household Living Standards Survey from 2002 to 2006.

1. Introduction

A well-established way of examining poverty dynamics is using transition matrices. Transition matrices are 2 x 2 cross tabulations which show the number (or percentage) of households (or individuals) who are poor and non-poor in two survey years. The 2 cells of the leading diagonal show households who are poor or non-poor in both survey years, while the off diagonal cells show the number of households who move into or out of poverty between the two survey years. Table 1 gives an example of a transition matrix contructed for the panel component of the 2002 and 2006 Vietnam Household Living Standards Surveys. The number in each cell show the number of households in each of the four poverty transtion categories.

Table 1: Poverty Transition Matrix for Vietnam, 2002-06

|		2006	
		Poor	**Non-Poor**
2002	**Poor**	203	418
	Non Poor	121	1409

There are a number of well known difficulties with transition matrices. These include: (i) households are classified as being poor or non-poor based on whether their incomes (or expenditures) are above or below a pre-determined poverty line (which may or may not vary between survey years. Therefore transition matrices do not indicate how poor or well-off a household is; and (ii) if incomes are measured with error, as is likely to be the case, some

---

households will be erroneously classified. This is likely to be a particular problem for households with incomes that are close to the poverty line in one or both survey years.

Contour plots, which can be regarded as the continous analogue of transition matrices are one way to circumvent these difficulties. Contour plots are diagrams which provide a two dimensional view of the income distribution, and resemble a topological maps of a mountain. [2] They can be interpreted in a similar way to the contours on an topological map, except the contours represent points of equal frequency rather than points of equal height. Once horizontal and vertical lines representing the poverty lines in two survey years are super-imposed on the contour plot, its relationship to the four categories in a standard transition matrix become clear: the four partitions of the contour plot correspond to the four cells of the transition matix. Figure 1 shows an example of a contour plot for the same panel data from Vietnam that was used to construct Table 1.

Figure 1: Contour Plot for Vietnam, 2002-2006
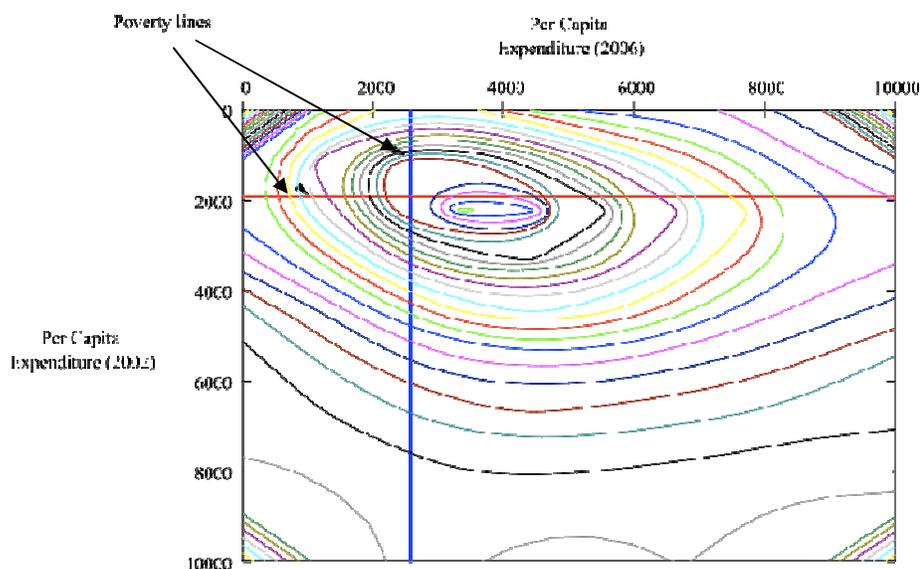
Note that if the peak of a contour plot is close to one or both of the poverty lines, this shows that a substantial number of households have incomes that are close to the poverty line, and that the numbers of households in the cells of the transition matrix are particularly susceptible to measurement error. Alternatively, if the peak of a contour plot is well inside one of the four partitions, this shows that a substantial number of households belong to the corresponding poverty dynamics category. In Figure 1, the peak of the expenditure distribution in 2002 can be seen to be particularly close to the poverty line for 2002.

While contour plots are a useful tool for examining poverty dynamics and economic mobility, they are not particularly easy to produce. To the authors' knowledge, there are currently no computer software packages with pre-programmed routines to produce contour plots. One option is to use the GAUSS routines described in Deaton (1997) and available for download from the World Bank's LSMS website (http://www.worldbank.org/html/prdph/lsms/tools/deaton/index.htm).

---

[2] See Chapter 3 (in particular pages 180-181) of Deaton (1997) for further information on the construction and interpretation of contour plots.

However, this requires access to and a reasonable familiarity with GAUSS. Another option, which we believe is easier for most users, is to use DAD/DASP to produce a surface plot of the bivariate income distribution, and then export this data to gnuplot to produce the contour plot.

The remainder of this note explains the procedure for producing contour plots using DAD/DASP and gnuplot. This is illustrated using the same data used to produce Tables 1 and Figure 1 above. Readers are invited to try these commands out for themselves using the Stata and gnuplot files VNexample.dta and VNexample.dat plus the gnuplot script file drawcplot.txt, all of which are posted on The Toolkit pages of the CPRC website. (www.chronicpoverty.org/toolkit).

2. Producing a Surface Plot using DASP

DASP is a package designed to assist research and policy analysts conduct distributional analysis using Stata, and was developed from the DAD (Distributional Analysis for Development) package. Stata itself needs to be purchased (www.stata.com) but DASP and DAD can be downloaded free-of-charge from http://132.203.59.36:83/index2.htm .

DASP contains a set of commands for producing 3-dimensional density and distribution plots, which are useful for visualising the shape of an income or expenditure distribution. Unfortunately DASP does not currently allow contour plots to be produced but the data from the density plot can be exported and then turned into a contour plot using gnuplot.

To use DASP to produce a density plot, use the following commands.

use 'Vnexample.dta'

sjdensity pcx02 pcxp06, lab1(2002) lab2(2006) sresg(VNexample.dat)

which produces the following on-screen 3 dimensional density plot

Figure 2: Joint Density Surface for Vietnam, 2002-2006
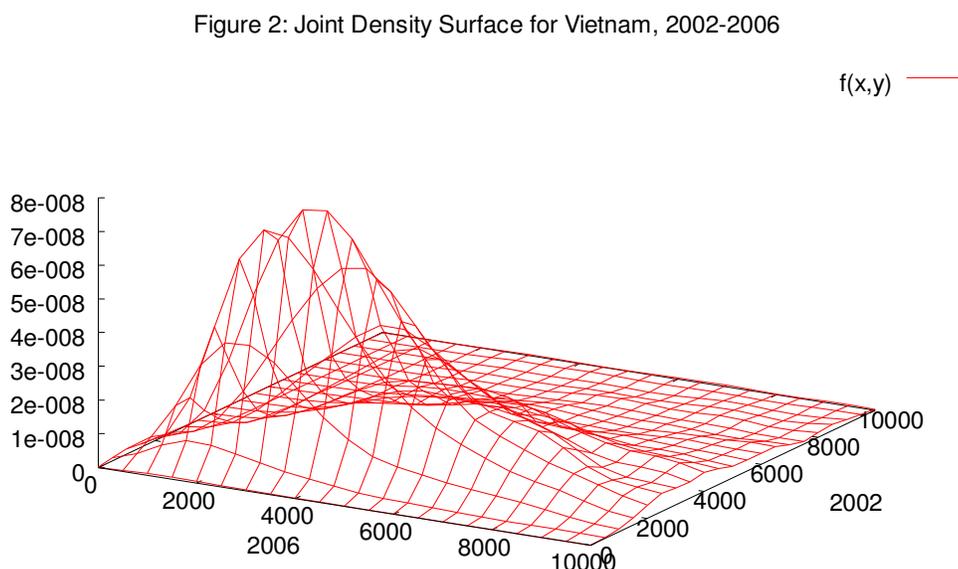
Note that this plot is produced by gnuplot. On screen, the surface plot can be examined from different angles by clicking on the corners and dragging them around using the mouse. To copy the plot for insertation into a document, click on the gnuplot graph icon in the top left-hand corner, select options, and then select the copy to clipboard option.

The above commands also save the data underlying the density plot is also saved in gnuplot format as Vnexample.dat. This first and last few rows of this file are:

```
0            0           7.51900000000e-11
0            500         8.95585000000e-10
0            1000        1.53765100000e-09
0            1500        9.16175000000e-10
0            2000        6.70142000000e-10
                    ....
10000        8000        1.44450000000e-09
10000        8500        1.85663000000e-09
10000        9000        1.30380000000e-09
10000        9500        8.51172000000e-10
10000        10000       2.77651000000e-10
```

The three columns of this file show the x,y and z coordinates of the density surface representing the 2002 expenditures, expenditures and height of the surface plot.

Alternatively, DASP menu-driven system can be used to produce the same plot using the following drop down menus:
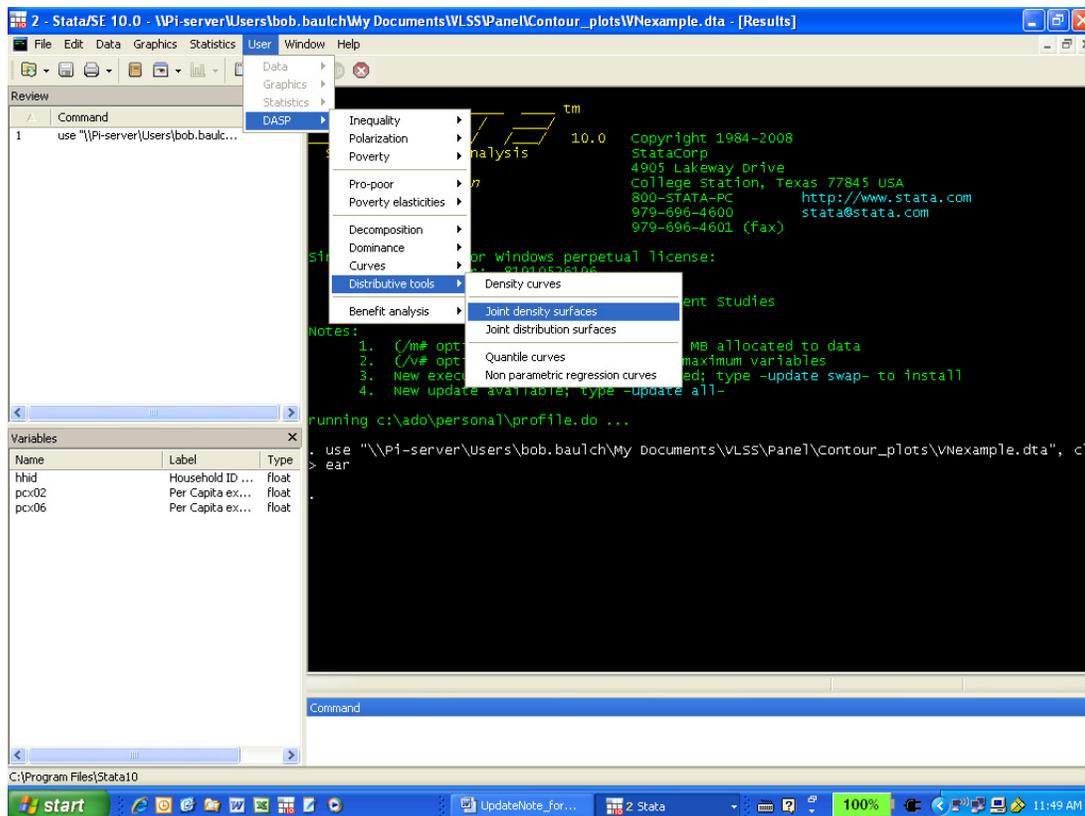
The dialogue box reproduced below then appears in which names of the expenditure variable (here pcx2006 and pcx2007) should be entered in the Dim. 1 variable and Dim. 2 variable boxes. The axes labels and title of the graph should also be completed. To prevent a few outliers dominating the appearance of the contour plot it is often useful to limit the range of the x and y axes, using the box below. Other options in this box allow the bandwidth and weighting of the graph to be specified.

To save the data in gnuplot format, click the Results tab in the top left-hand corner of the dialogue box, and specify a file name in the second row.

3. Using gnuplot to Produce a Basic Contour Plots

Gnuplot is a command-driven interactive function and data plotting program that can be downloaded free-of-charge from http://www.gnuplot.info .[3] Like DASP, gnuplot can be operating using either commands typed into the command line or using it drop down menus. This section describes the commands used to produce the basic contour plot in Figure 3, while the next sections describes some of the options for going beyond the basic graph. See the file drawcplot.txt in Appendix 1 for the script file used to plot the contour plot in Figure 8.

Figure 3: A Basic Contour Plot
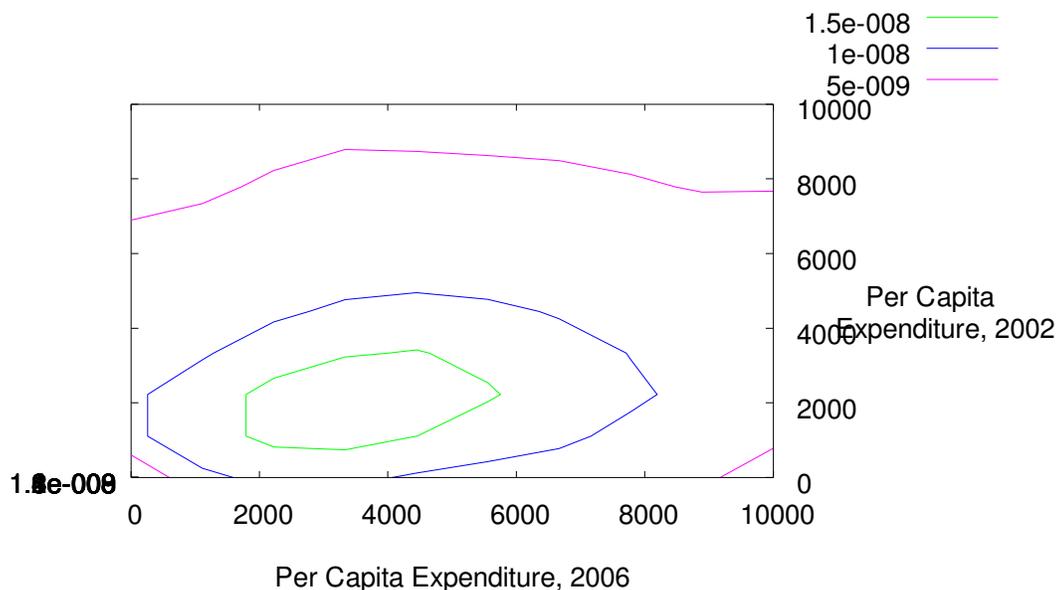
The basic script to produce this five contour plot is given below:

```
cd 'directory where the data_is_kept'
set dgrid3d
set style data lines
set contour
unset surface
set view 0,0
set xlabel 'Per Capita Expenditure, 2006'
set ylabel "Per Capita \n   Expenditure, 2002"
splot "VNexample.dat"
```

Once the working directory is set as the folder, lines two and three set up the initial data and graphic. Using set contour and unset surface will produce the basic contours without any smoothing. The unset surface command instructs gnuplot not to produce the 3-dimensional density surface. set view 0,0 instruct gnuplot to use the top down (aerial) view required for a contour plot. The x and y axes are labelled using the commands set x[y]label "Per Capita Expenditure" with the

---

relevant date added to the text. It is further possible to split these over two lines using \n inside the label which must use double inverted commas in this case. splot "VNexample.dat" then produces the plot on screen.

The easiest method of running this script is to open the script file in Notepad or an equivalent text editor, and then copy and paste it into the gnuplot command window using SHIFT+ INSERT. The script will then be run automatically, and the following basic contour produced:

The contour plot is a 2-dimensional representation of a 3-dimensional surface view from above. This can be seen by 're-setting the surface using set surface, changing the aerial orientation of the view using set view 30, 60 and then typing replot . This produces Figure 4 below:

Figure **4:** A Combined Contour and Surface Plot
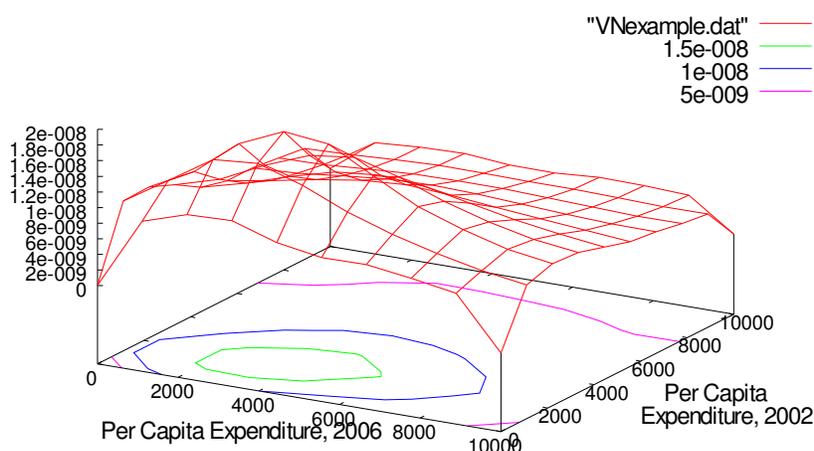
4. Beyond The Basic Plot

It is possible to adapt the code to increase the number of contours and to smooth the contours beyond those given in the data. The default number of contours in gnuplot is 5. In order to increase the number of contours the command set cntrparam levels 15 is used and then replot the graphic. This gives the Figure 5.

Figure 5: Contour Plot with 15 Contours
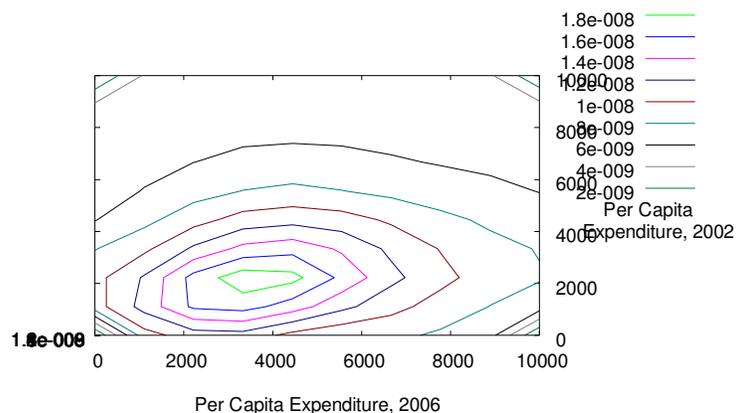
As can be seen the legend can get rather messy when there are many contours. The legend key can be removed using unset key. It is also possible to smooth the contour lines using different types of spline. Both cubic and B-splines are supported by gnuplot and can be set using set cntrparam cubic or set cntrparam bspline . The B-spline can be further refined using an order adjustment to a maximum level of 10. Examples of each of these are show in Figure 6 below. Both of these plots use a spline order of 9 using the command set cntrparam order 9 .

**Figure** 6: Contour Plots with **Cubic and Bc Spline**
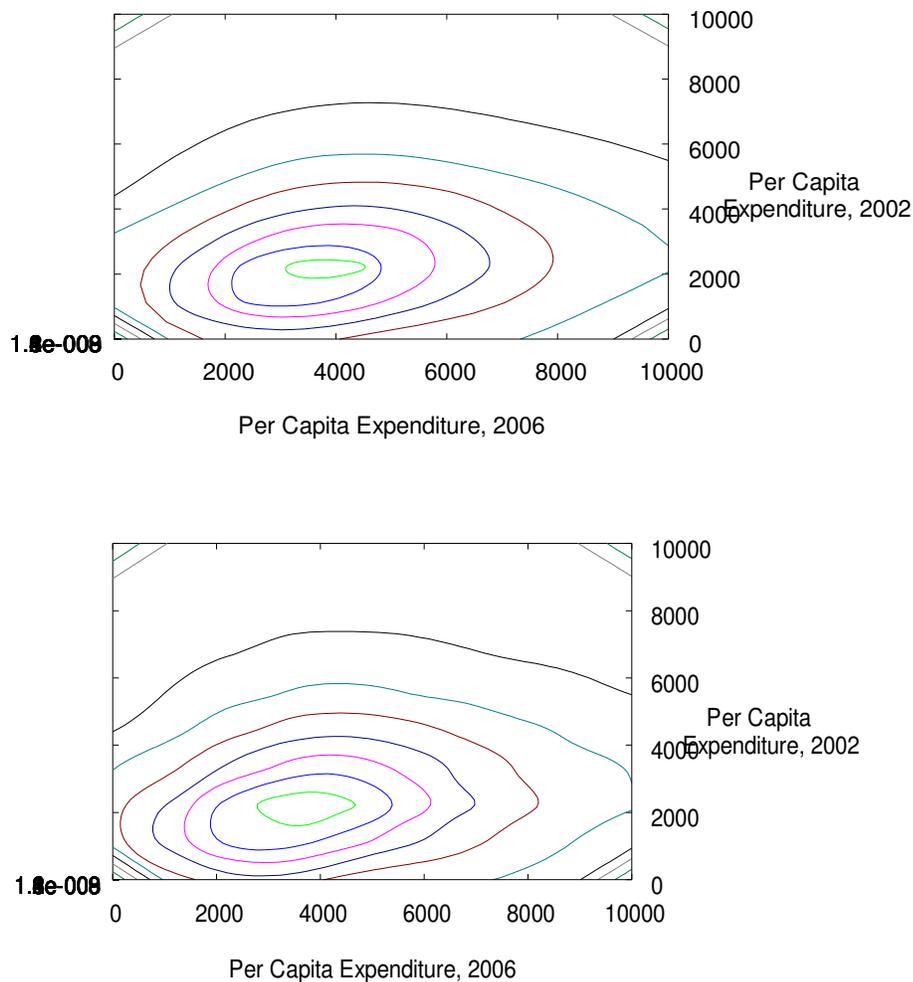
5. Further Modifications

To allow a clearer parallel with the transition matrix in Table 1, it is helpful to rotate the axes of the contour plot, remove the z-axis labels and add vertical and horizontal lines representing the poverty lines in each of the survey years. The set view 180,180 command puts the y-axis on the left hand side while set yrange [:] reverse and set yrange [:] reverse ensure that the axes both start together in the top left-hand corner at 0,0. The x and y range code with the [:] are included to ensure that the reversal of the axis is possible with an automatic range. For a

manual range the [:] can be replaced by [0:10000] for example. At present it is easiest to flip the whole graph over and reverse both axes rather than flip one and move the x axis up. unset ztics removes the (stacked) z axis tics The command set xtics offset 0,2 moves the relevant information out of the graph area. The final output of this code is given below in Figure 8. Using the. For a purely two dimensional view this is advisable, however it might be necessary to retain them if moving to a three dimensional plot. In this case it is necessary to insert a line break (\n) in the two labels. In order to do this the label text must be surrounded by double inverted commas rather than a single one.

To add lines to the plot, representing for example poverty lines, the set arrow command is use. This is because lines in gnuplot are actually arrows without heads! The command set arrow X from x1,y2 to x2,y2 nohead lt rgb "colour" adds vertical and horizontal lines to the contour plot. The start and end of the arrow are defined using the usual co-ordinate system, so that 2560,0 starts at x- co-ordinate 2560 and at the x-axis. The end of the line is set using the maximum value of the relevant range. For example, the Vietnamese expenditure poverty line for 2006 is VND 2560, so set arrow 1 from 2560,0 to 2560,10000 adds a vertical line to the contour plot corresponding to the poverty line in 2006. Adding the options nohead lt rgb "blue" lw 2 removes the head of the arrow, changes the colour of the line to blue, and increase its line width to ensure that the lines are visible and have sufficient contrast with the contours themselves. [4]

Figure 8: Rotated Contour Plot with Poverty Line
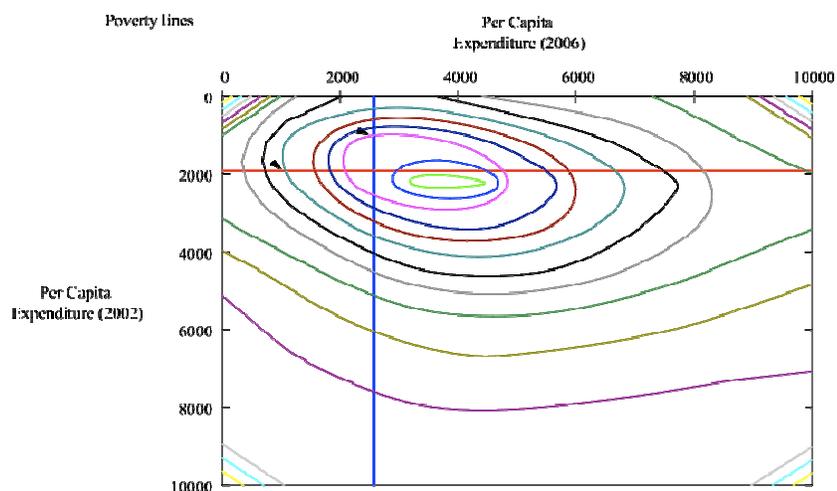
The final code for the contour plot is given in the Appendix 1 below and can also be downloaded as file drawcplot.txt from the CPRC website. Once the fonts are tweaked using the terminal settings, and some text is added to label the poverty lines (using set label Z "Label Text" at x,y, the option rotate will then rotate the orientation of the text), it is possible to produce a publication quality graphic such as in Figure 9.

---

[4] The full range of colours is given using the command show color palette. To remove any arrows or labels one uses the command unset arrow/label X.

Figure 9: Graphic Quality Contour Plot
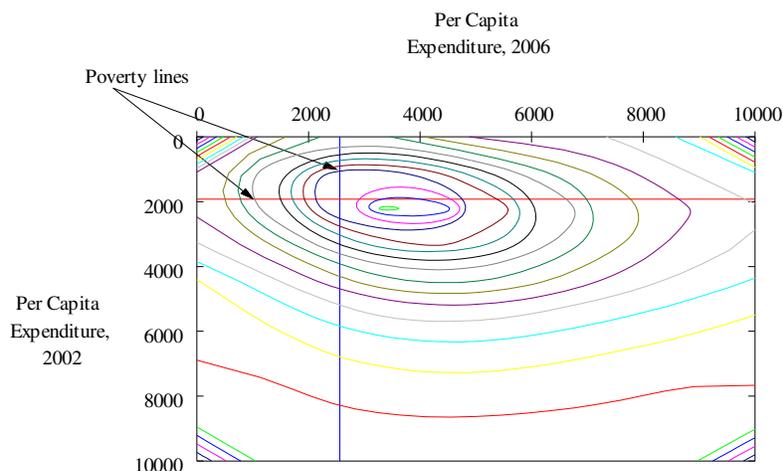
If the graphics are to be pasted into Microsoft Word or Powerpoint, it is also possible to copy from a windows terminal and paste the plot directly. This can be done right clicking-on the blue toolbar at the top of the gnuplot graph box, selecting options, and then copy to clipboard. Note that it is also possible to change the fonts in the graphics in a standard windows type terminal.. Alternatively, setting the terminal to emf and the output to "VNexample.emf" produces an EMF file for insertion in to documents. Other file types include png, jpeg, pdf, svg and latex[5]. [6]. The terminal needs to be set to the relevant type for the output to be saved in that format.

6. Additional Remarks and Caveats

For simplicity, the surface density and contour plots shown above have been produced without sampling weights. Whether or not to use sampling weights when working with panel data is a contentious issue (see Deaton, 1997:67-73) but since per capita measures are used here, there is certainly a case for weighting each households expenditure by household size. This can be accomplished easily by specifying the size variable in DASP joint density dialogue box or modifying the DASP sjdensity command as follows.

sjdensity pcx06 pcx02 , hsize(hhsize06) lab1(Per Capita Expenditure 2006) lab2(Per Capita Expenditure 2002)

Of course, there is still an issue whether to pick the household size for 2002 or 2006 as the weighting variable!

In some countries, there may be more than one poverty line to take account of regional or rural/urban price differentials. In this case, the contour plot can be modified to show the minimum and maximum poverty lines for each year (which will usually be quite close to each other).

In some countries, the income distribution may have a bi-modal structure. This is most commonly associated with significant differences between the rural and urban distributions. In such instances, the contour plot will have a "twin-peaks" structure.

---

5   Not all platforms seem to produce all terminals, using the code set terminal gives a list of available outputs.
6   Not all platforms seem to produce all terminals, using the code set terminal gives a list of available outputs.

References

Araar, A. And Duclos, J.Y. (2007) *DASP v 1.4 User Manual* , Universite Laval

Deaton, A (1997) *The Analysis of Households Surveys: A Microeconometric Approach to Development Policy,* Baltimore: John Hopkins University Press

Williams, T. and Kelley, C. *Gnuplot: An Interactive Plotting Program* (2008) available from http://www.gnuplot.info/docs/gnuplot.pdf

Appendix 1: gnuplot Code used to Produce Figure 9

```
##This can be run using load 'filename' from within gnuplot

#cd 'name of your directory'

set dgrid3d

set style data lines

set contour

unset surface

set cntrparam levels 15

##For Fig 1

##set cntrparam levels 33

set cntrparam order 9

set cntrparam bspline

unset key

set xlabel "Per Capita\n Expenditure, 2006" offset 0,6

set ylabel "Per Capita\n Expenditure,\n 2002"

set view 180,180

set tics in

unset ztics

set yrange [:] reverse

set xrange [:] reverse

set grid back

set xtics axis in

unset grid

set xtics offset 0,2

set arrow 1 from 2560,0 to 2560,10000 nohead lt rgb "blue" lw 2

set arrow 2 from 0,1920 to 10000,1920 nohead lt rgb "red" lw 2

set label 15 "Poverty lines" at -2000,-2000

set arrow 4 from -1500,-1500 to 2500,1000 lw 0.5 filled

set arrow 5 from -1500,-1500 to 1000,1900 lw 0.5 filled

splot "VNexample.dat"
```