**Homework 5 – Soütrik Banerjee**

**OECD data analysis**

**Introduction**

Data on different key socio-economic indicators for 17 OECD countries were collected during the years 1975, 1977, 1979, 1981 for all countries.

**Material and Methods**

Data of countries were repeated four times. Summary statistics were performed on the averaged (over the four measurements) 11 socio-economic indicators for the 17 countries. In the table below, the 95% Confidence Interval (C.I.) for the mean can be constructed from the Student’s *t*-distribution using the formula [Xbar ± {*tα/2;n – 1*\*(S.E.M.)}], where the value of *t*-statistic (for 17 – 1 d.f. = 16 d.f.) = 2.112, and Xbar = mean, S.E.M. = standard error of mean (= σhat / n0.5), where σhat = standard deviation in Table 1 below.

**Results**

Table 1 shows the summary statistics of the socio-economic indicators.

**S-E Indicators | Mean SD S.E.M. Median IQR Min Max CV**

**-----------------+---------------------------------------------------------------------------------**

**birthRate | 140.12 27.29 6.62 129.25 29.50 97.00 213.75 0.19**

**unemploy | 53.88 22.68 5.50 60.00 31.50 18.00 83.75 0.42**

**prWorkPerc | 107.15 72.17 17.50 90.75 67.00 27.00 293.50 0.67**

**secWorkPerc | 351.29 43.06 10.44 350.25 53.00 288.50 449.25 0.12**

**GDP | 7,574.35 2,648.48 642.35 8,527.75 3,577.50 1,922.00 11,066.25 0.35**

**asset | 227.69 41.18 9.99 213.00 51.00 174.50 316.25 0.18**

**priceIncr | 109.56 41.11 9.97 99.00 58.25 49.00 198.75 0.38**

**income | 403.25 94.26 22.86 403.75 63.75 251.25 580.50 0.23**

**infantMort | 135.88 63.45 15.39 132.75 57.00 76.50 354.25 0.47**

**animProteinDiet | 60.50 10.90 2.64 62.75 12.75 35.50 74.00 0.18**

**energyConsum | 425.00 207.97 50.44 375.50 205.50 99.50 898.25 0.49**

**---------------------------------------------------------------------------------------------------**

SD: Standard Deviation; S.E.M.: Standard Error of Mean; IQR: Inter-Quartile Range; CV: Coefficient of Variation.



Figures 1A-1K shows the evolution of socio-economic indicators.

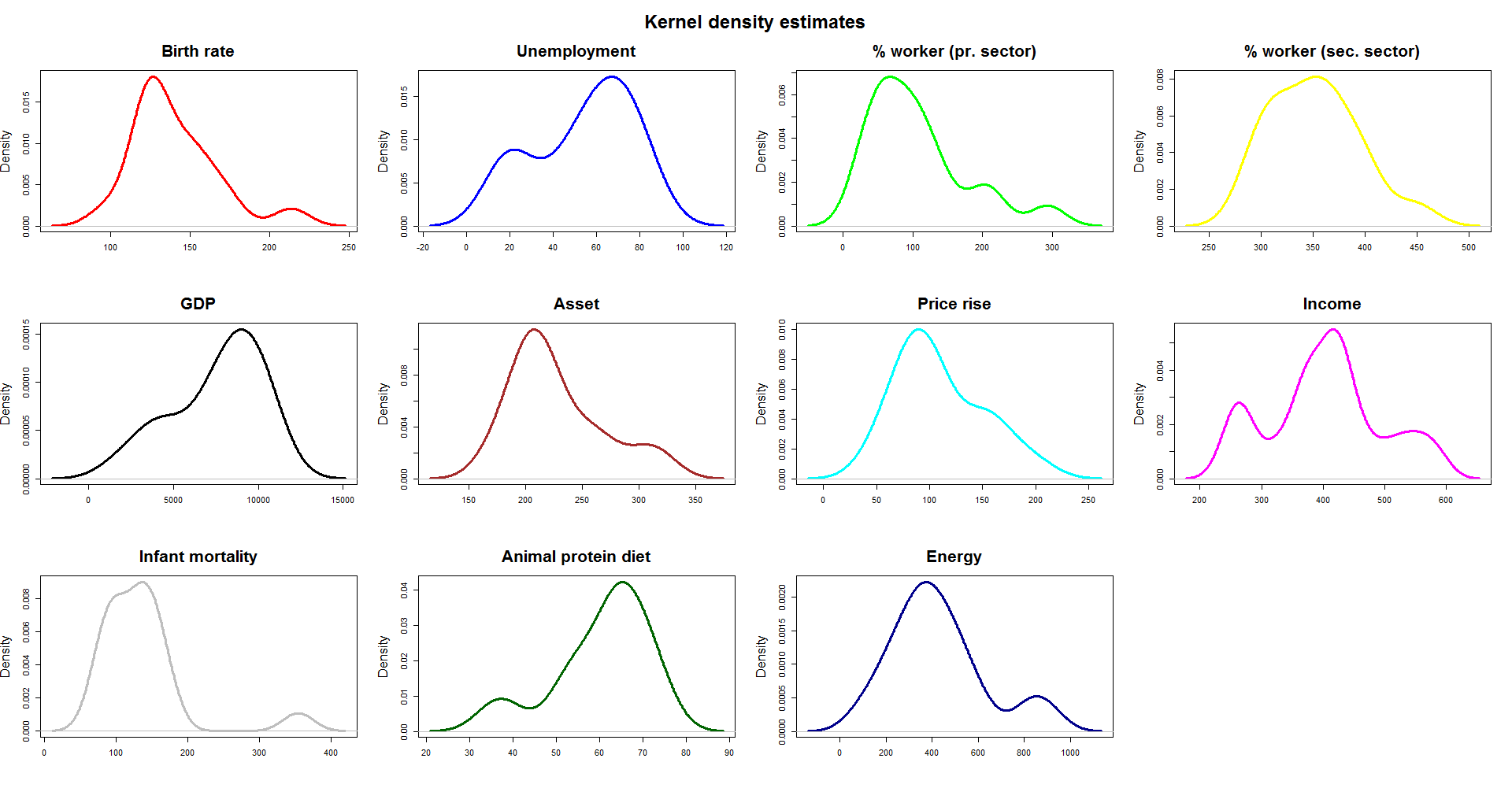


Figure 2A – 2K shows boxplots of the socio-economic indicators varying between years (that is within countries).



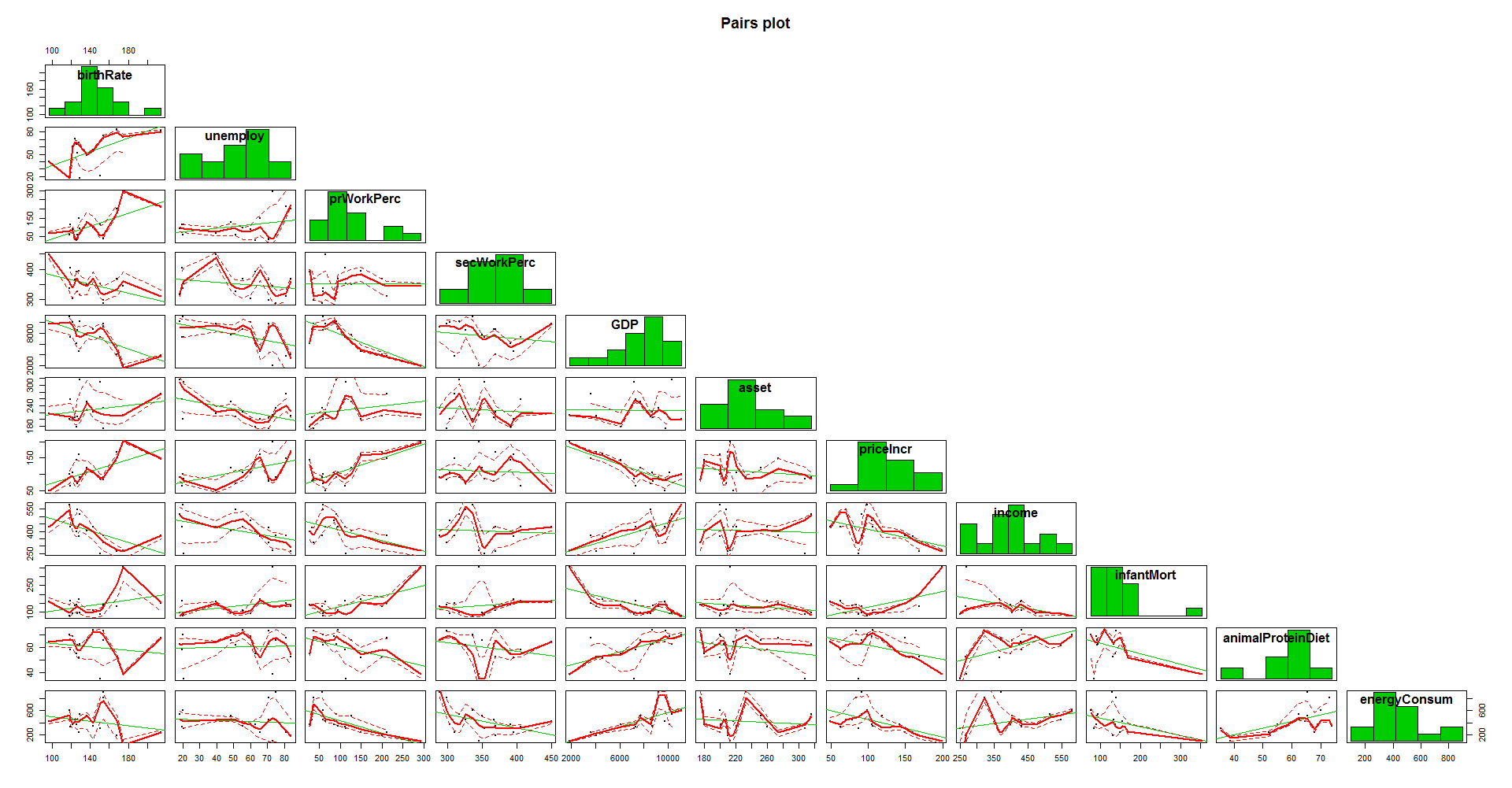


Figure 3A – 3B shows boxplot of the socio-economic indicators varying between countries.



Estimated bandwidths have been omitted in the x-axes.

Figure 4 shows kernel densities for the socio-economic indicators.



Green line: linear regression (between two S.E. indicators); red line: linear spline with confidence interval.

Figure 5 shows scatterplot matrix (with histograms in the diagonal) of the socio-economic indicators.

**Discussions**

Basic exploratory analyses show that at least a few of the densities of the socio-economic indicators could be a mixture of different densities (in different proportions) coming from different distributions.

Shapiro-Wilk’s test of Normality shows that unemployment (borderline), primary sector worker %, asset, infant mortality (strongly) reject the assumption of Normality.

There exists both between and within country variability of the socio-economic indicators.

There also exists some effect of time (year) for the evolution of different socio-economic indicators, however the variabilities (for different S.E. indicators) across the years appear to be “constant” over time from this preliminary analysis.

There are some extreme values for some S.E. indicators, which lie above the 95th percentile of the distributions. Bonferroni outlier’s test showed significant outlier(s) [*p* < 0.05] for birth rate, infant mortality.

**References**

1. Cook, R. D. and Weisberg, S. (1982) Residuals and Influence in Regression. Chapman and Hall.
2. Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) Graphical Methods for Data Analysis. Wadsworth & Brooks/Cole.
3. Patrick Royston (1982) An extension of Shapiro and Wilk's *W* test for normality to large samples. Applied Statistics, **31**, 115–124.
4. Sheather, S. J. and Jones M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. J. Roy. Statist. Soc. **B**, 683–690.