Variation of Poverty Index over Household Size through Estimated Engel Curve for Chhattisgarh (Economic Statistics End Semester Project - 1)

Srijan Chattopadhyay (BS2126)

Fall 2023

Introduction:

In the realm of economic statistics and policy formulation, the Consumer Survey serves as an invaluable tool to comprehend and assess the economic dynamics within a nation. The 68th round of the Consumer Survey in India, conducted with meticulous precision and expansive reach, presents a wealth of data that is crucial for understanding the consumption patterns, preferences, and economic well-being of Indian households. This comprehensive survey, undertaken by the Central Statistics Office (CSO) under the aegis of the Ministry of Statistics and Programme Implementation, serves as a cornerstone for empirical analysis and policy-making in the Indian economic landscape.

Goal:

In this project, we will analyze the consumer survey 68th round data for the state of Chhattisgarh and will provide an estimate of Engel curve for different item groups for the state with appropriate justification. Then we will see the variance of the poverty index (slightly modified version of [2]) over different household sizes. We will analyze all the findings also as much as possible.

Theoretical Terms:

Engel's Law:

Definition:

As defined in [1], Engel's Law is an economic theory that describes the relationship between household income and a particular good or service expenditures. It states that as family income increases, the percentage of income spent on food decreases. The theory was introduced by Ernst Engel, a German economist and statistician, in 1857.

Engel's Law is an observation in economics. It states that as the income of a family increases, the proportion of income spent on food decreases, although the absolute dollar expenditures on food are still increasing.

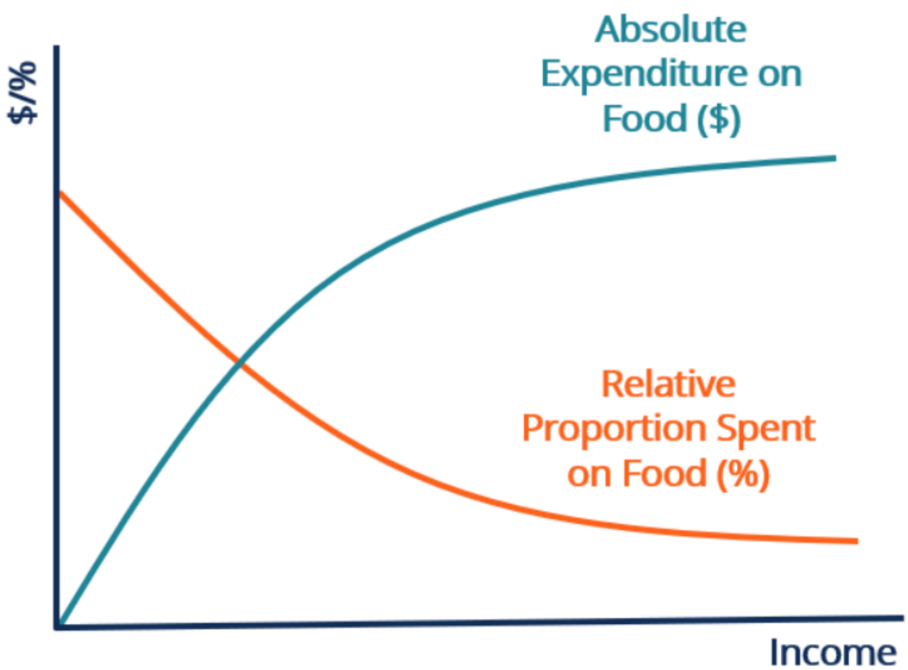


Figure 1: Engel Curve according to Engel's Law

Example:

Engel's Law does not assume that the spending on food remains constant as family income increases. For example, a family with a $5,000 monthly income spends $2,000 on food, which makes up 40% ($2,000/$5,000) of its budget. If the income of this family increases by 40% to $7,000, it will spend $2,500 on food.

Although its absolute expenditures on food increase by 25%, the relative proportion to income decreases to 35.7% ($2,500/$7,000). The decrease in proportion can be explained by the lower rate of increase in food expenditures relative to the rate of increase in income.

Assumption:

In this project, we will assume that the income is equal to Total expenditure, or, everybody keeps a fixed amount of the income with him, because in the Consumer Survey 68th Round data, the income of a particular household was not collected, and as it is a secret information for many people in a developing country like India, we rely on the Total expenditure of a household to get an idea about the total income of that household. Hence, from now on we use the convention that income means the total expenditure.

Engel Curve:

Definition:

A derivative concept is the Engel curve, which is based on Engel's Law. The Engel curve describes how the spending on a certain good varies with household income by either proportion or absolute dollar amount. The shape of an Engel curve is impacted by demographic variables, such as age, gender, and educational level, as well as other consumer characteristics.

The Engel curve also varies for different types of goods. With income level as the x-axis and expenditures as the y-axis, the Engel curves show upward slopes for normal goods, which have a positive income elasticity of demand. Inferior goods, with negative income elasticity, assume negative slopes for their Engel curves. In the case of food, the Engel curve is concave downward with a positive but decreasing slope.

Our Case:

68th round Consumer Survey data consists of Household level data, and not individual-level data, hence age, gender, etc can't be taken into the studies for estimating the Engel curve. We will be using household size and land owned by that household (which are properties of the entire household) instead in place of age, gender, etc.

Modelling of Engel Curve:

We will attempt to fit a model (based on our justification) that links total consumption to consumption of specific item groups, possibly under heteroscedasticity assumption. For that, we considered household size, Amount of land owned, MLT, MPCE, etc as our predictors. Although the later two didn't play significant role in modelling, hence we removed those eventually.

The steps for modelling we used for our purpose are as follows:

- First we take the best box-cox transformation possible for the data.
- Then, based on R square, we check which powers of the covariate are fitting most appropriately with the boxcox transformed data.
- Lastly, in the final model, bad covariates are removed based on some prior knowledge or justification.
- If after all these, the best model we got, has R square value < 0.5, we declare that based on this amount of data, a good model fitting is not possible, else we declare that model to be the best.

So, the final model for item group J will be something like this:

$$Y_{i,J} = (c_T T_{i,J}^{\lambda_1} + c_H H_{i,J}^{\lambda_2} + c_L L_{i,J}^{\lambda_3})^{1/\lambda} + \epsilon_i \tag{1}$$

where $Y_{i,J}$ is the expenditure for Jth item for ith household, $T_{i,J}$ is the total expenditure, $H_{i,J} = H_i$ is the size of household, and $L_{i,J} = L_i$ is the amount of land owned by household.

Use of Engel Curve:

As in [2], we define the following commodity-specific consumption deprivation index (poverty index):

$$D = \int_0^\infty \left(\frac{C^* - C(y)}{C^*} \right) f(y) dy \tag{2}$$

where C^* is the maximum value of consumption expenditure, $C(y)$ denotes the actual mean consumption at a given level of total expenditure y, and $f(y)$ is the pdf of total expenditure. As the deprivation function is a monotonic decreasing function in y, the above infinite integral converges and has a finite value.

As we are estimating D from our sampled data, we use the following modification by Riemann sum and Plug-in estimates with combined multiplier:

$$\hat{D} = \frac{1}{n} \sum_{y \in S_n} \frac{\hat{C^*} - \hat{C}(y)}{\hat{C^*}} \hat{f}(y) \tag{3}$$

where, $\hat{C^*}$ and $\hat{c}(y)$ are estimated for the population from the sample using combined multiplier and $\hat{f}$ is the non-parametric kernel density estimator of f (used in base package of R), and S_n is equi-partition in the sample range of y. The following lemma guarantees that $\hat{D}$ is a reasonably good estimator of D asymptotically.

Lemma 1. *If $\hat{C}$ is uniformly convergent, then as $n \to \infty$ and sample size increases, $\hat{D} \overset{a.s.}{\to} D$*

Proof.

$$\begin{aligned}
\hat{D} &= \frac{1}{n} \sum_{y \in S_n} \frac{\hat{C}^* - \hat{C}(y)}{\hat{C}^*} \hat{f}(y) \\
&= \frac{1}{n} \sum_{y \in S_n} \hat{f}(y) - \frac{1}{n\hat{C}^*} \sum_{y \in S_n} \hat{C}(y)\hat{f}(y) \\
&\overset{a.s.}{\to} \frac{1}{n} \sum_{y \in S'_n} f(y) - \frac{1}{nC^*} \sum_{y \in S'_n} C(y)f(y) \quad \text{[as sample size increases, by consistency of the estimators]} \\
&\to \int_0^\infty f(y)dy - \int_0^\infty \frac{1}{C^*} C(y)f(y)dy = D
\end{aligned}$$

□

Results with R Codes:

```
library(haven)
library(dplyr)
library(forecast)
library(nlme)
library(psych)
library(ggplot2)
library(GGally)
library(leaps)
library(olsrr)
library(glmnet)
library(EnvStats)
library(stats)
```

```
data1 = read_dta("F:/B3_resources/ass_or_hw/Eco Stat/Data/Only_SrChattishgarh/data1_Srijan.dta")
data2 = read_dta("F:/B3_resources/ass_or_hw/Eco Stat/Data/Only_SrChattishgarh/data2_Srijan.dta")
data3 = read_dta("F:/B3_resources/ass_or_hw/Eco Stat/Data/Only_SrChattishgarh/data3_Srijan.dta")
data4 = read_dta("F:/B3_resources/ass_or_hw/Eco Stat/Data/Only_SrChattishgarh/data4_Srijan.dta")
data5 = read_dta("F:/B3_resources/ass_or_hw/Eco Stat/Data/Only_SrChattishgarh/data5_Srijan.dta")
data6 = read_dta("F:/B3_resources/ass_or_hw/Eco Stat/Data/Only_SrChattishgarh/data6_Srijan.dta")
data7 = read_dta("F:/B3_resources/ass_or_hw/Eco Stat/Data/Only_SrChattishgarh/data7_Srijan.dta")
data8 = read_dta("F:/B3_resources/ass_or_hw/Eco Stat/Data/Only_SrChattishgarh/data8_Srijan.dta")

data1_sr1 = na.omit(data1[, c(6, 19, 23, 28, 30, 31)])
data2_sr1 = na.omit(data2[, c(6, 19, 21, 25, 27, 28)])
data3_sr1 = na.omit(data3[, c(6, 20, 28, 32, 34, 35)])
data4_sr1 = na.omit(data4[, c(6, 19, 20, 24, 26, 27)])
data5_sr1 = na.omit(data5[, c(6, 19, 20, 24, 26, 27)])
summary = na.omit(data6[, c(6, 19, 20, 24, 26, 27)])
colname = colnames(data1_sr1)
colname[3] = "TCV"
colnames(data1_sr1) = colnames(data2_sr1) = colnames(data3_sr1) =
```

```r
colnames(data4_sr1) = colnames(data5_sr1) = colnames(summary) = colname
total = subset(summary, Item_Code == "40")
final_0 = bind_rows(data1_sr1, data2_sr1, data3_sr1, data4_sr1, data5_sr1)
repeatation = c(
  129,139,159,169,179,189,199,219,239,249,269,279,289,299,309,319,329,349,379,389,399,
  409,419,429,439,449,459,479,499,519,529,549,559,569,579,599,609,619,629,639,649,659
)
colnames(total)[3] <- "Total"
condition <- as.numeric(final_0$Item_Code) %in% repeatation
final = final_0[condition,]
final <- final %>%
  left_join(dplyr::select(total, HHID, Total), by = "HHID")
final = final[,-1]
colnames(final)[2] = "TCV"
colnames(final)[6] = "Total"
engel_function = function(itemno){
  item = subset(final, Item_Code == itemno)
  hh11 = subset(data7, HHID %in% sort(unique(as.numeric(item$HHID))))
  hh12 = subset(data8, HHID %in% sort(unique(as.numeric(item$HHID))))
  item <- item %>% arrange(HHID)
  hh11 <- hh11 %>% arrange(HHID)
  hh12 <- hh12 %>% arrange(HHID)
  item$size = hh11$hh_size
  item$earn = hh12$Regular_salary_earner
  item$mpce = hh12$MPCE
  item$mlt = hh12$MLT
  item$land = hh11$Land_Owned
  item$land[is.na(item$land)] <- 0
  df = cbind(
    as.numeric(item$TCV),
    as.numeric(item$Total),
    as.numeric(item$size),
    as.numeric(item$mpce),
    as.numeric(item$mlt),
    item$land
  )
  df = data.frame(df)
  colnames(df) <- c("TCV", "Total", "hhsize", "mpce", "mlt", "land")
  ggpairs(df)
  model_frame = model.frame(TCV~.,data = df)
  boxcox_result = EnvStats::boxcox(lm(TCV ~ ., data = model_frame), optimize = T)
  lambda = boxcox_result$lambda
  df$TCV = boxcoxTransform(df$TCV, lambda)
  degree1 = seq(0, 2, length = 100)
  degree2 = seq(0, 1, length = 5)
  degree3 = seq(0, 1, length = 3)
  r_squared_values <-
    array(0, dim = c(length(degree1), length(degree2), length(degree3)))
  for (i in 1:length(degree1)) {
    for (j in 1:length(degree2)) {
      for (k in 1:length(degree3)) {
        X_i <- df$Total ^ (degree1[i])
        Y_i <- df$hhsize ^ (degree2[j])
```

```
        Z_i <- df$land ^ (degree3[k])

        design_matrix <- cbind(1, X_i, Y_i, Z_i, df$mpce, df$mlt)

        model <- lm(df$TCV ~ ., data = as.data.frame(design_matrix))
        r_squared_values[i, j, k] <- summary(model)$r.squared
      }
    }
  }
  best_combi = which(r_squared_values == max(r_squared_values, na.rm = T), arr.ind = T)

  model1 <- lm(TCV ~ I((Total)^(degree1[best_combi[1]]))+
I((hhsize)^(degree2[best_combi[2]]))+I((land)^(degree3[best_combi[3]]))+mpce+mlt,df)
  return_list1 = list(pairplot = ggpairs(df), Error =
"Error!! Maximum R squaredis also very less, A good model is not possible to fit!!")
  return_list2 = list(pairplot = ggpairs(df), best_model = model1, best_powers = best_combi,
boxcoxlambda =lambda)  if(summary(model1)$r.squared < 0.5)  return(return_list1)
else(return(return_list2))
}
```

Now, we will only present those items for which the model fitting is good, i.e. the best model can explain at least 50% variability of the data.

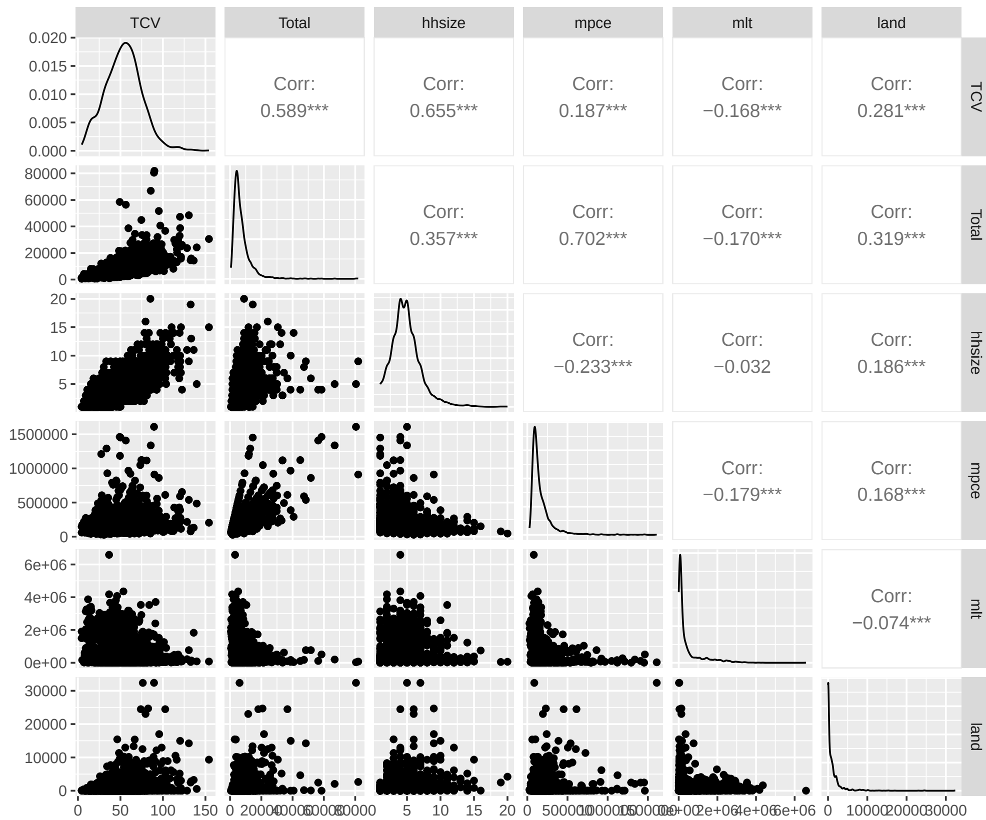
```
for (i in 1:length(repeatation)) {
len[i] = length(engel_function(repeatation[i]))
  }
```

Following is a brief discussion on the selected set of item groups......

Cereals:

```
engel_function(129)
```

Pairplots:



Summary of the Best Model:

```
## Call:
## lm(formula = TCV ~ I((Total)^(degree1[best_combi[1]])) +
##  I((hhsize)^(degree2[best_combi[2]])) +
##  I((land)^(degree3[best_combi[3]])) + mpce + mlt, data = df)

## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.055  -8.077   0.359   8.174  60.159

## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          -1.699e+02  7.194e+00 -23.623  < 2e-16 ***
## I((Total)^(degree1[best_combi[1]]))   8.844e+01  3.481e+00  25.407  < 2e-16 ***
## I((hhsize)^(degree2[best_combi[2]]))  2.759e+00  2.034e-01  13.563  < 2e-16 ***
## I((land)^(degree3[best_combi[3]]))    6.742e-02  1.203e-02   5.603 2.38e-08 ***
## mpce                                 -2.258e-05  3.435e-06  -6.573 6.17e-11 ***
## mlt                                  -8.168e-07  3.858e-07  -2.117   0.0344 *
## ---
```

```
## Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

## Residual standard error: 12.39 on 2129 degrees of freedom
## Multiple R-squared:  0.6695,Adjusted R-squared:  0.6687
## F-statistic: 862.6 on 5 and 2129 DF,  p-value: < 2.2e-16
```

Engel Curves:

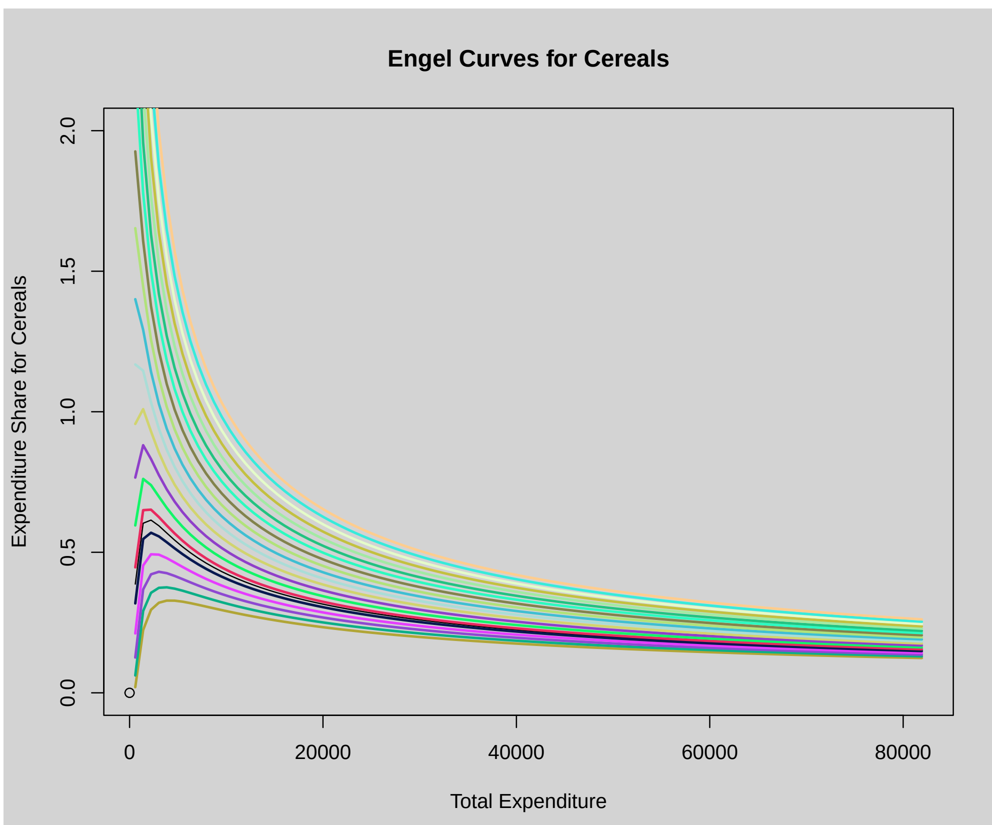
Hence, the fitted model for Cereals is:

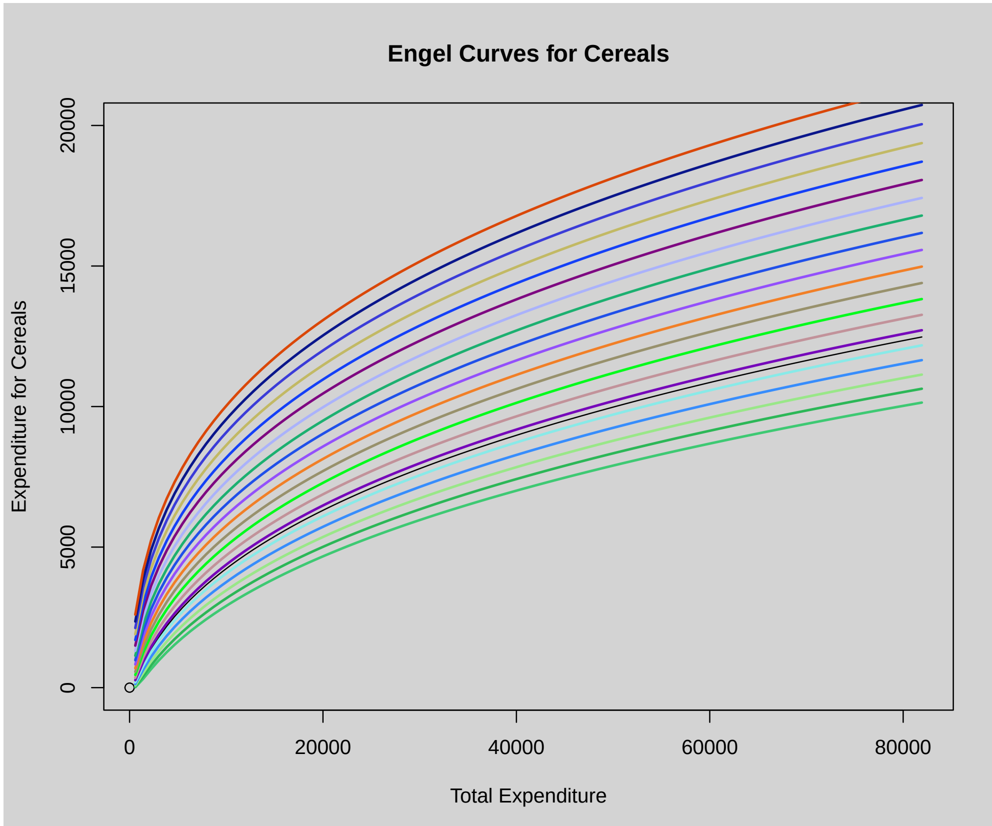
$$Y^{0.512} = 88.442T^{0.101} + 2.759H + 0.067L^{0.5} - 169.940 \tag{4}$$

where Y is the consumption for cereals, T is total consumption, H is size of Household, L is amount of land owned, MP is mpce, M is mlt. Hence, varying only T and fixing the rest, we can get the Engel curves for cereals at different levels. So, the expenditure share is:

$$\frac{Y}{T} = \frac{(88.442T^{0.101} + 2.759H + 0.067L^{0.5} - 169.940)^{1/0.512}}{T} \tag{5}$$

We traversed through all the household sizes, and Engel curves for all varying levels are in the following plot: And, the following is the plot for Total consumption of Cereal vs Total Expenditure:





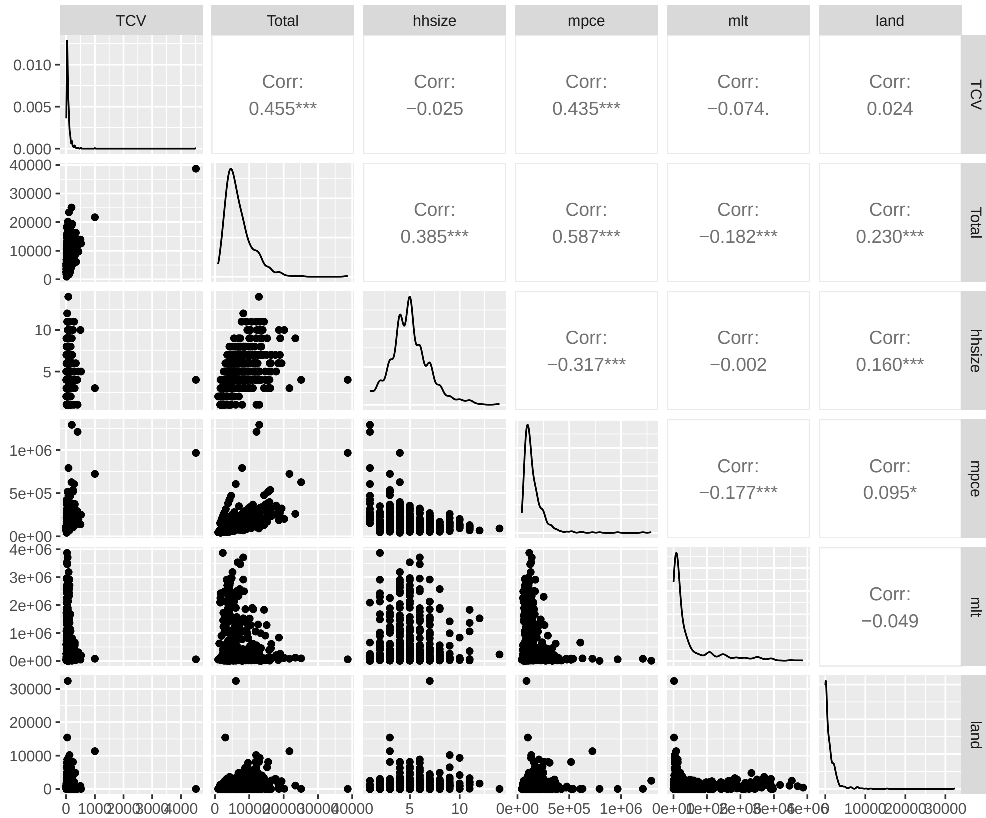
Discussion:

As we can see from the engel curves of cereals, consumption vs expenditure curve is always increasing, where the share vs expenditure is first increasing for some time and then decreasing. This may be due to lower income families, who don't even manage to fulfill basic requirements out of the total expenditure. Hence, on increasing the monetary income, it is increasing for some time, and then after a saturation, it is again decreasing.

Intoxicants:

```
engel_function(329)
```

Pairplots:



Summary of the Best Model:

```
    ## Call:
## lm(formula = TCV ~ I((Total)^(degree1[best_combi[1]])) + I((hhsize)^(degree2[best_combi[2]])) +
##     I((land)^(degree3[best_combi[3]])) + mpce + mlt, data = df)

## Residuals:
##     Min      1Q  Median      3Q     Max
## -838.92  -32.69    4.98   44.68 2013.28

## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           1.527e+02  2.549e+01   5.989 3.84e-09 ***
## I((Total)^(degree1[best_combi[1]]))   1.799e-06  8.993e-08  20.004  < 2e-16 ***
## I((hhsize)^(degree2[best_combi[2]])) -2.665e+01  3.943e+00  -6.758 3.61e-11 ***
## I((land)^(degree3[best_combi[3]]))   -7.411e-01  2.754e-01  -2.691  0.00734 **
## mpce                                 -2.359e-04  7.728e-05  -3.053  0.00237 **
## mlt                                   7.501e-06  8.257e-06   0.909  0.36401

## Significance codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

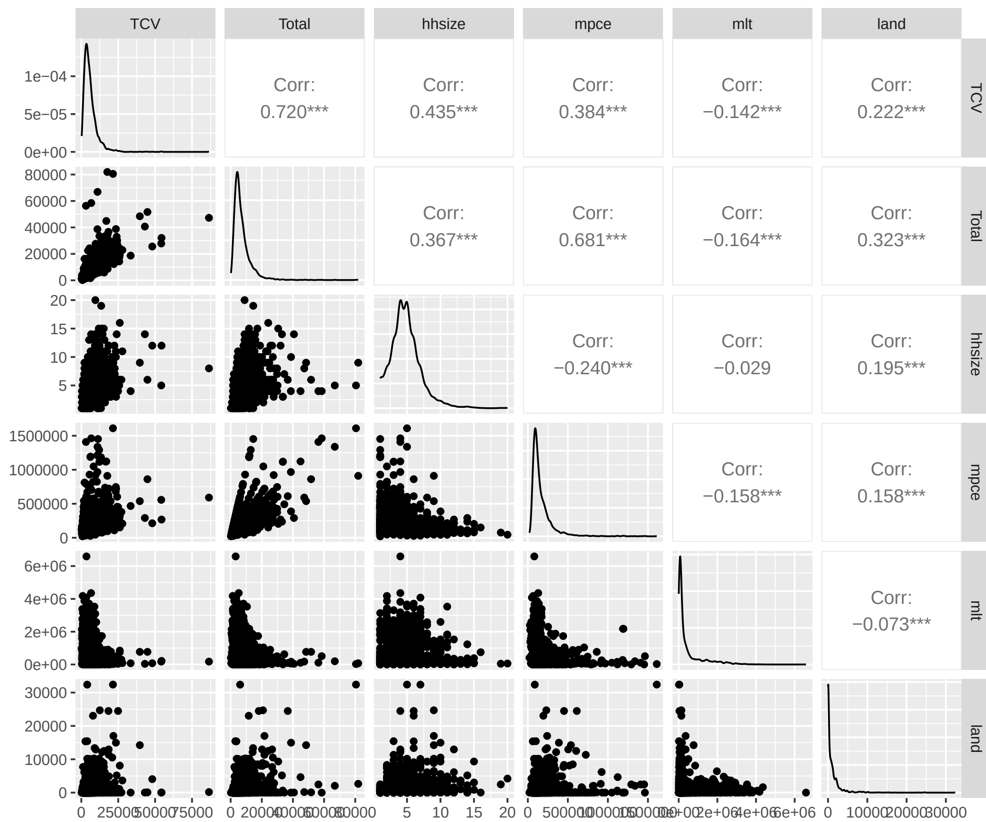
```
## Residual standard error: 139 on 543 degrees of freedom
## Multiple R-squared:  0.5458,Adjusted R-squared:  0.5416
## F-statistic: 130.5 on 5 and 543 DF,  p-value: < 2.2e-16
```

Discussion:

As per the coefficient estimates, the model fitting is good except for the predictor 'total', which has a very marginal effect on the income.

Clothing:

Pairplots:



Summary of the Best Model:

```
    ## Call:
## lm(formula = TCV ~ I((Total)^(degree1[best_combi[1]])) + I((hhsize)^(degree2[best_combi[2]])) +
##     I((land)^(degree3[best_combi[3]])) + mpce + mlt, data = df)

## Residuals:
##    Min      1Q  Median      3Q     Max
## -21150   -1294    -134     941   59391

## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                            -4.104e+02  9.098e+02  -0.451   0.6520
```

```
## I((Total)^(degree1[best_combi[1]]))   4.181e+01  1.538e+00  27.192  < 2e-16 ***
## I((hhsize)^(degree2[best_combi[2]])) -1.597e+03  7.358e+02  -2.170   0.0301 *
## I((land)^(degree3[best_combi[3]]))   -7.143e+00  3.108e+00  -2.298   0.0217 *
## mpce                                 -8.642e-03  1.100e-03  -7.857 6.15e-15 ***
## mlt                                   6.128e-06  9.823e-05   0.062   0.9503

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 3203 on 2164 degrees of freedom
## Multiple R-squared:  0.5854,Adjusted R-squared:  0.5845
## F-statistic: 611.2 on 5 and 2164 DF,  p-value: < 2.2e-16
```

Engel Curves:

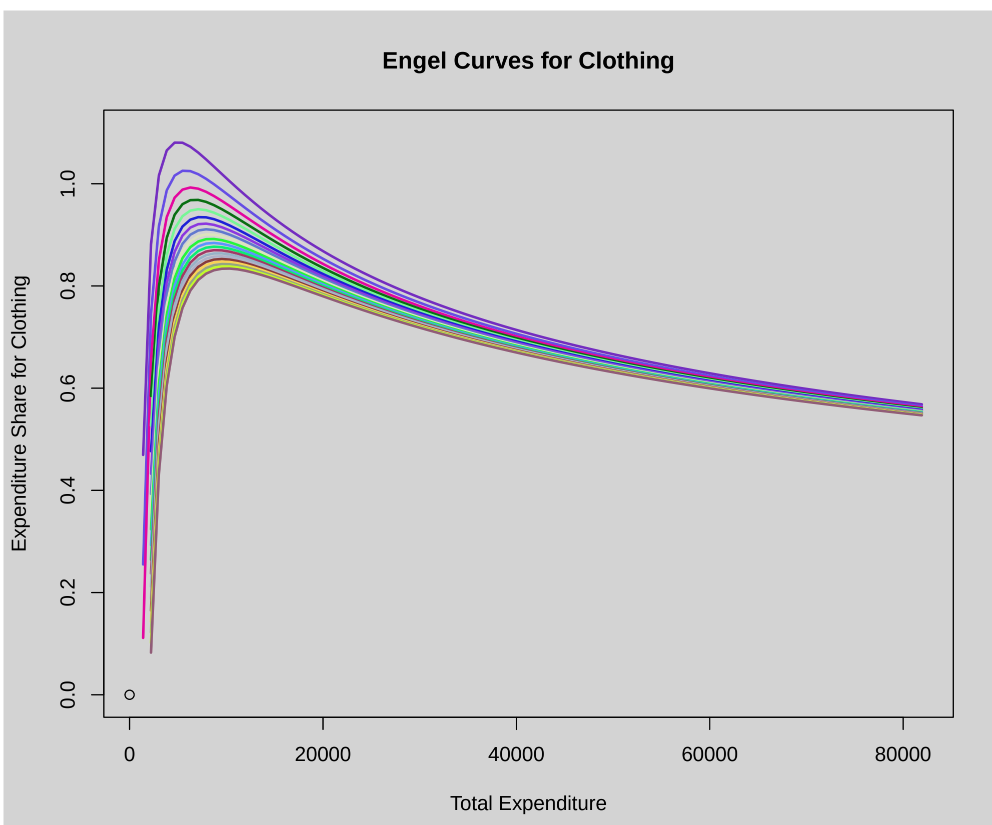
Hence, the fitted model for Clothing is:

$$Y^{1.000647} = 42T^{0.626} - 1597H^{0.25} - 7L^{0.5} - 410 \tag{6}$$

where Y is the consumption for clothing, T is total consumption, H is size of Household, L is amount of land owned, MP is mpce, M is mlt. Hence, varying only T and fixing the rest, we can get the Engel curves for cereals at different levels. So, the expenditure share is:

$$\frac{Y}{T} = \frac{(42T^{0.626} - 1597H^{0.25} - 7L^{0.5} - 410)^{1/1.000647}}{T} \tag{7}$$

We traversed through all the household sizes, and Engel curves for all varying levels are in the following plot:

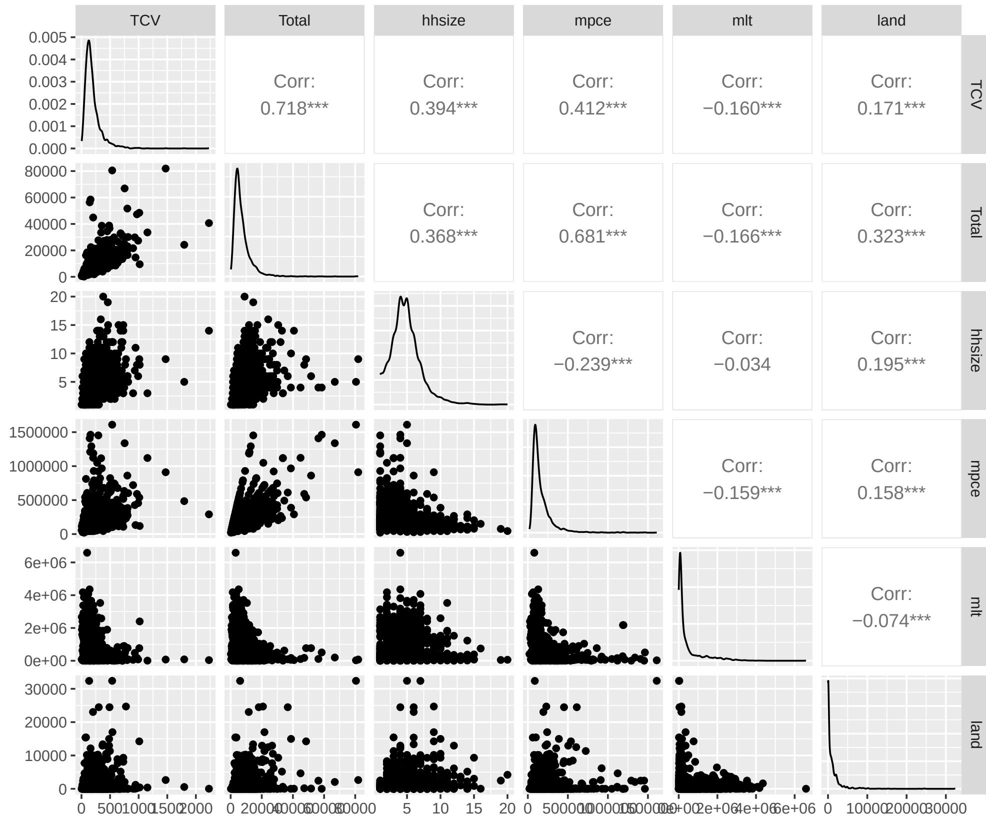


Discussion:

Looking at the curves, we can observe that as income increases, people tend to spend more on clothing after fulfilling their basic food needs. This trend continues until a saturation point is reached, which causes the increase to stabilize. Therefore, we see a significant increase in the first half, followed by a stabilization in the latter half.

Toilet Articles:

Pairplots:



Summary of the Best Model:

```
## Call:
## lm(formula = TCV ~ I((Total)^(degree1[best_combi[1]])) + I((hhsize)^(degree2[best_combi[2]])) +
##     I((land)^(degree3[best_combi[3]])) + mpce + mlt, data = df)

## Residuals:
##     Min      1Q  Median      3Q     Max
## -636.27  -45.32   -8.59   31.41 1423.58

## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          4.281e+01  2.758e+01   1.552 0.120804
## I((Total)^(degree1[best_combi[1]]))   1.623e+00  5.868e-02  27.663  < 2e-16 ***
## I((hhsize)^(degree2[best_combi[2]])) -8.652e+01  2.260e+01  -3.829 0.000132 ***
## I((land)^(degree3[best_combi[3]]))   -5.581e-01  9.494e-02  -5.878 4.79e-09 ***
## mpce                                -2.613e-04  3.364e-05  -7.768 1.22e-14 ***
## mlt                                 -2.315e-06  2.992e-06  -0.774 0.439021

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 97.84 on 2165 degrees of freedom
## Multiple R-squared:  0.578,Adjusted R-squared:  0.5771
```

```
## F-statistic: 593.2 on 5 and 2165 DF,  p-value: < 2.2e-16
```

Engel Curves:

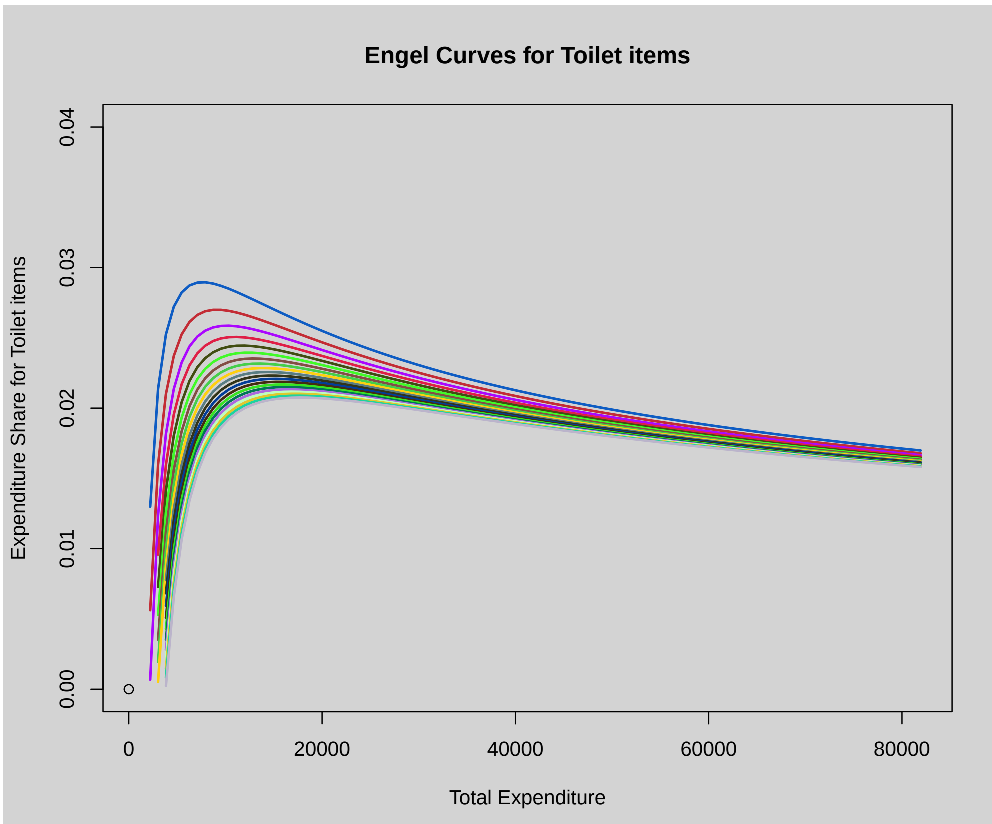
Hence, the fitted model is:

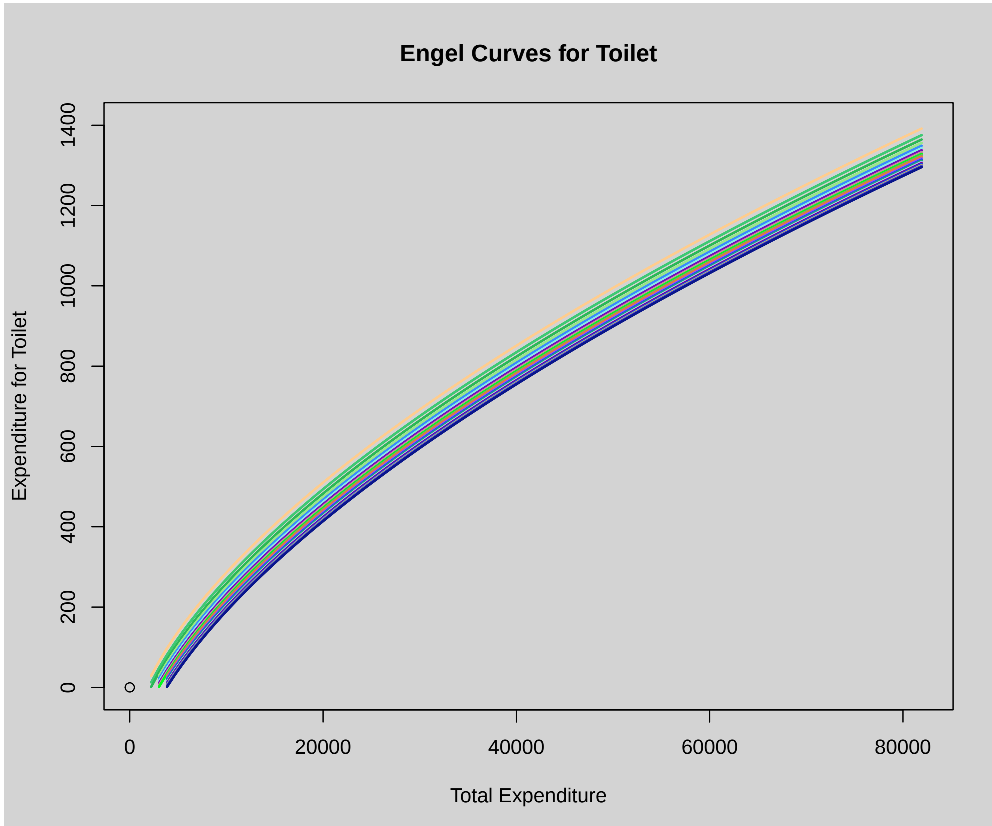
$$Y^{1.000647} = 1.623T^{0.626} - 86.524H^{0.25} - 0.558L^{0.5} + 42.807 \tag{8}$$

where Y is the consumption for this item, T is total consumption, H is size of Household, L is amount of land owned, MP is mpce, M is mlt. Hence, varying only T and fixing the rest, we can get the Engel curves for cereals at different levels. So, the expenditure share is:

$$\frac{Y}{T} = \frac{(1.623T^{0.626} - 86.524H^{0.25} - 0.558L^{0.5} + 42.807)^{1/1.000647}}{T} \tag{9}$$

We traversed through all the household sizes, and Engel curves for all varying levels are in the following plot:



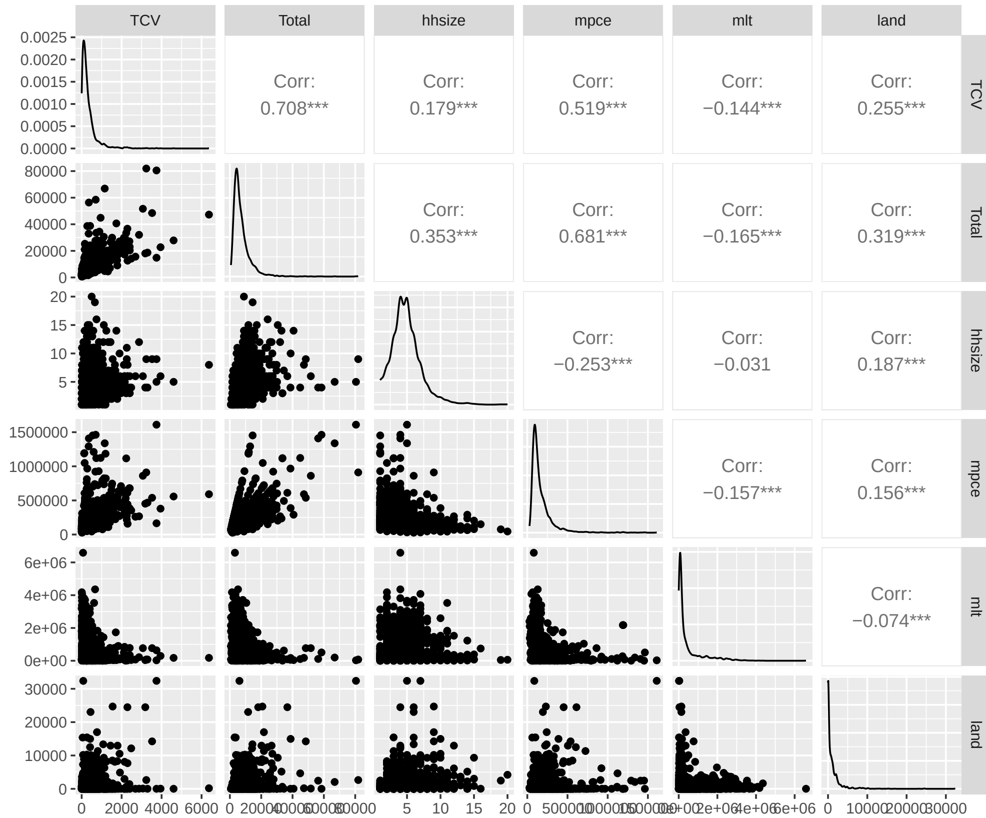


Discussion:

Similar consumption patterns are observed in the case of clothing and personal care products. When people have lower incomes, they may not have the means to purchase high-quality items or indulge in luxury products like high-end toiletries. However, once their basic needs are met, they are more likely to invest in these items and splurge as much as possible.

Consumer Services Excluding Conveyance:

Pairplots:



Summary of the Best Model:

```
## Call:
## lm(formula = TCV ~ I((Total)^(degree1[best_combi[1]])) + I((hhsize)^(degree2[best_combi[2]])) +
##     I((land)^(degree3[best_combi[3]])) + mpce + mlt, data = df)

## Residuals:
##     Min      1Q  Median      3Q     Max
## -2004.3  -112.4   -22.8    67.6  4185.3

## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           5.015e+02  9.101e+01   5.511 4.01e-08 ***
## I((Total)^(degree1[best_combi[1]]))   1.459e+00  6.002e-02  24.304  < 2e-16 ***
## I((hhsize)^(degree2[best_combi[2]])) -5.858e+02  6.955e+01  -8.423  < 2e-16 ***
## I((land)^(degree3[best_combi[3]]))    7.467e-03  3.003e-03   2.486    0.013 *
## mpce                                 -4.704e-04  1.036e-04  -4.542 5.89e-06 ***
## mlt                                   4.432e-10  9.288e-06   0.000    1.000

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 299.8 on 2115 degrees of freedom
## Multiple R-squared:  0.5281,Adjusted R-squared:  0.527
```

```
## F-statistic: 473.3 on 5 and 2115 DF,  p-value: < 2.2e-16
```

Engel Curves:

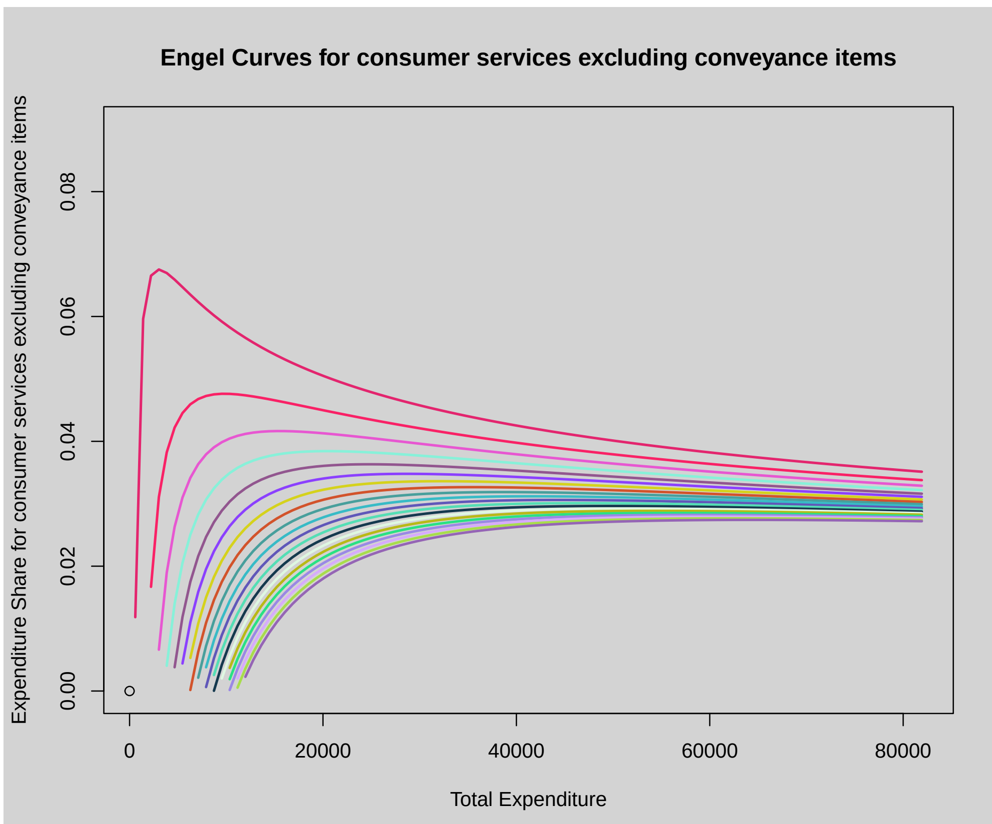
Hence, the fitted model is:

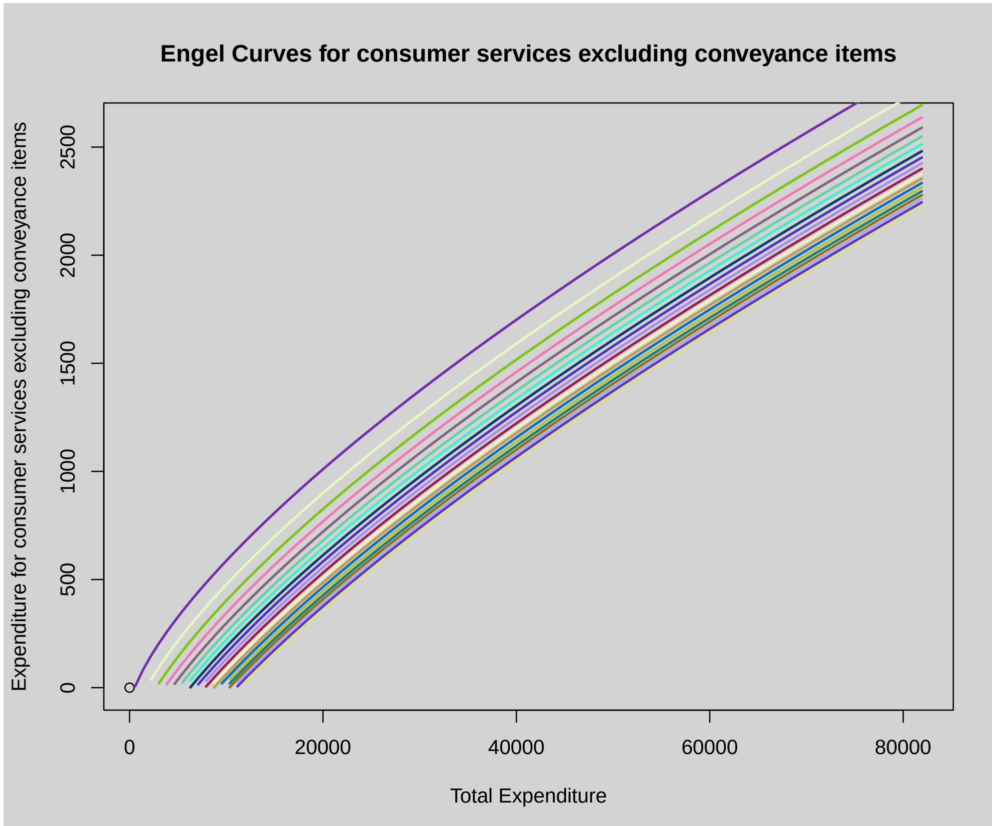
$$Y^{1.000647} = T^{0.707} - 86.586H^{0.25} + 502 \tag{10}$$

where Y is the consumption for this item, T is total consumption, H is size of Household, L is amount of land owned, MP is mpce, M is mlt. Hence, varying only T and fixing the rest, we can get the Engel curves for cereals at different levels. So, the expenditure share is:

$$\frac{Y}{T} = \frac{(T^{0.707} - 86.586H^{0.25} + 502)^{1/1.000647}}{T} \tag{11}$$

We traversed through all the household sizes, and Engel curves for all varying levels are in the following plot:



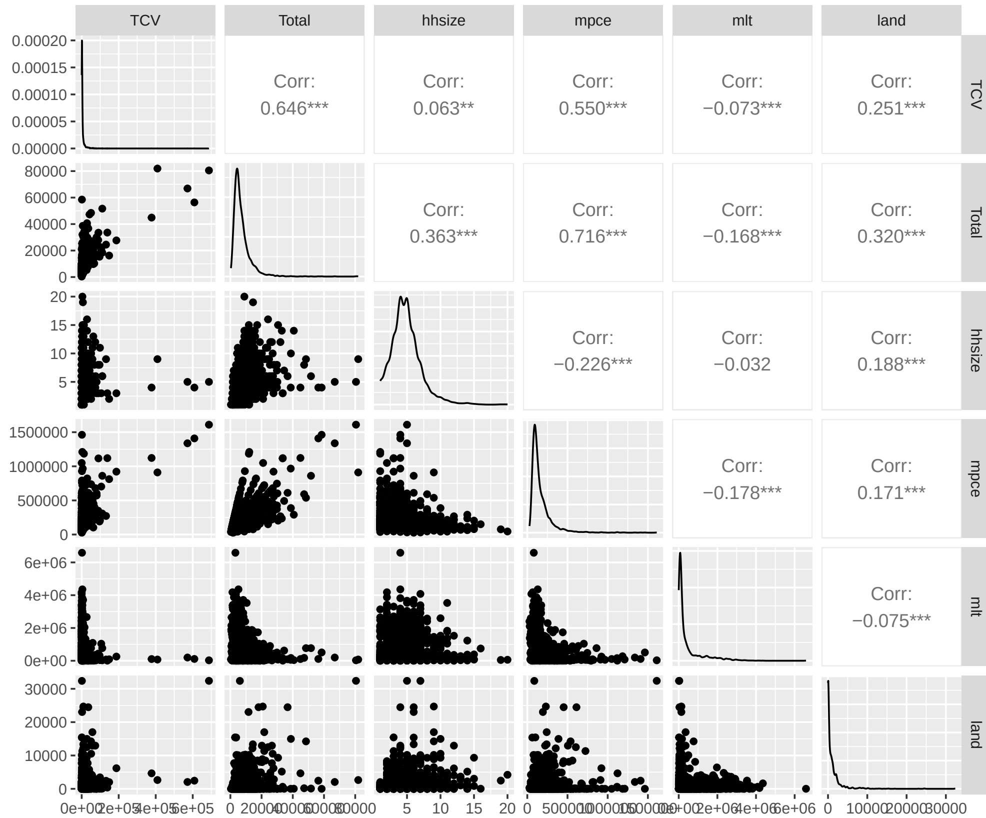


Discussion:

It can be observed that there is a significant increase in consumption for larger households, as well as for smaller ones. This is a natural occurrence as people tend to hire domestic help such as cooks and cleaners, aim to acquire a reliable internet connection and may even keep pets. Such consumption patterns are often observed once the basic needs are met.

Durable Goods:

Pairplots:



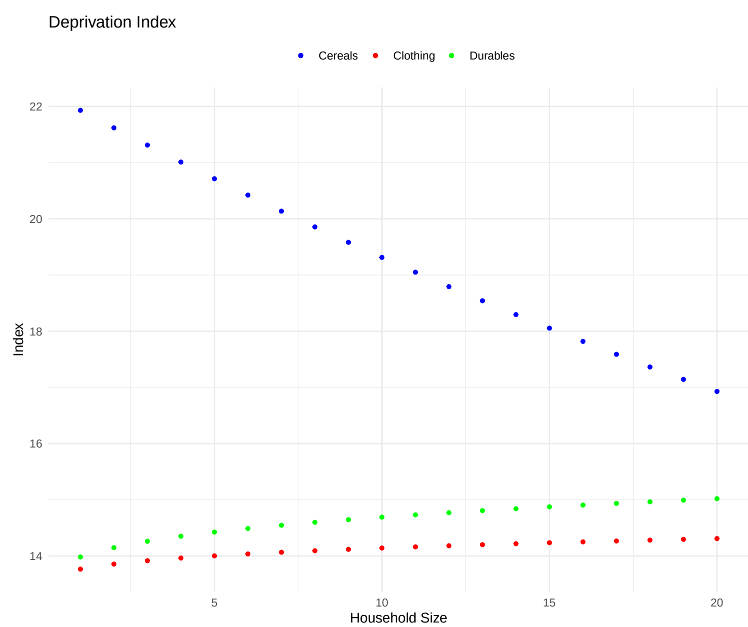
Summary of the Best Model:

```
## Call:
## lm(formula = TCV ~ I((Total)^(degree1[best_combi[1]])) + I((hhsize)^(degree2[best_combi[2]])) +
##     I((land)^(degree3[best_combi[3]])) + mpce + mlt, data = df)

## Residuals:
##     Min      1Q  Median      3Q     Max
## -271954   -2403      -1    2048  350892

## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          1.735e+03  1.339e+03   1.296   0.1952
## I((Total)^(degree1[best_combi[1]]))  7.663e-05  1.762e-06  43.502  < 2e-16 ***
## I((hhsize)^(degree2[best_combi[2]])) -1.009e+03  1.916e+02  -5.264 1.56e-07 ***
## I((land)^(degree3[best_combi[3]]))   7.734e+01  1.646e+01   4.700 2.77e-06 ***
## mpce                                 9.246e-03  3.829e-03   2.414   0.0158 *
## mlt                                 -1.976e-04  5.252e-04  -0.376   0.7068

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 16980 on 2119 degrees of freedom
## Multiple R-squared:  0.6568,Adjusted R-squared:  0.656
```

```
## F-statistic: 811.2 on 5 and 2119 DF,  p-value: < 2.2e-16
```

Deprivation Index variation with Household size

Now, we calculated the deprivation index for different household sizes based on estimated Engel curve as described earlier. We scaled it, as we are interested to see only the changes with respect to the change in household size. We observed a very interesting pattern. For cereals, the poverty index decreases sharply upon increasing the household size, which can be described as follows, many poor illerate people tend to have more babies but then don't satisfy the basic requirements, and these of type of things are mostly seen in poor families. As reported in [3], Given the lacklustre economic conditions and progress of the state, it is not surprising that Chhattisgarh tops the list of states in terms of poverty rate. Hence due lot of poor peoples, the index goes down. Whereas, for low household size, there is a variation between rich and poor due to the similar reason. But for clothing and toilet luxuries, we saw the stabilizing pattern and very less poverty index.



References

[1] Corporate Finance Institute. Engel's Law.

[2] T Krishna Kumar, Jayarama Holla, and Puja Guha. Engel curve method for measuring poverty. *Economic and Political Weekly*, pages 115–123, 2008.

[3] mint. Spatial poverty in chhattisgarh. Accessed: 24/11/23.