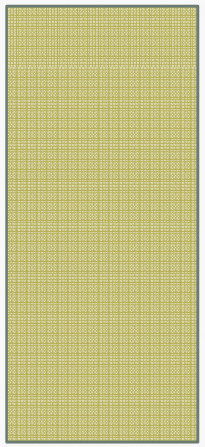
THE ROBUSTNESS OF THE MONETARY COST OF CHILDREN TO DATA PROBLEMS

STEFAN NORRBIN, FLORIDA STATE UNIVERSITY
DAVID MACPHERSON, TRINITY UNIVERSITY
TOM MCCALEB, FLORIDA STATE UNIVERSITY



MOTIVATION

- Federal guidelines require each U.S. state to examine their child support guidelines every four years.
- We have done the child support evaluation for Florida for 2004, 2008 and 2012
- We have noted that child support guidelines are sensitive to theoretical estimation and data issues
  - Why are they so sensitive?
  - Maybe use Mahalanobis grouping techniques to stabilize estimates!

SURVEY DATA ARE SENSITIVE

- In the Indian National Sample Survey switching from a 30-day recall period to a 7-day recall period for a number of items cut the Indian poverty rate by half!!!
- 200 million people moved up above the dollar-a-day definition

Source: Deaton, Angus, "Household Surveys, Consumption, and the Measurement of Poverty,"*Economic Systems Research*, Vol. 15, No. 2, June 2003.

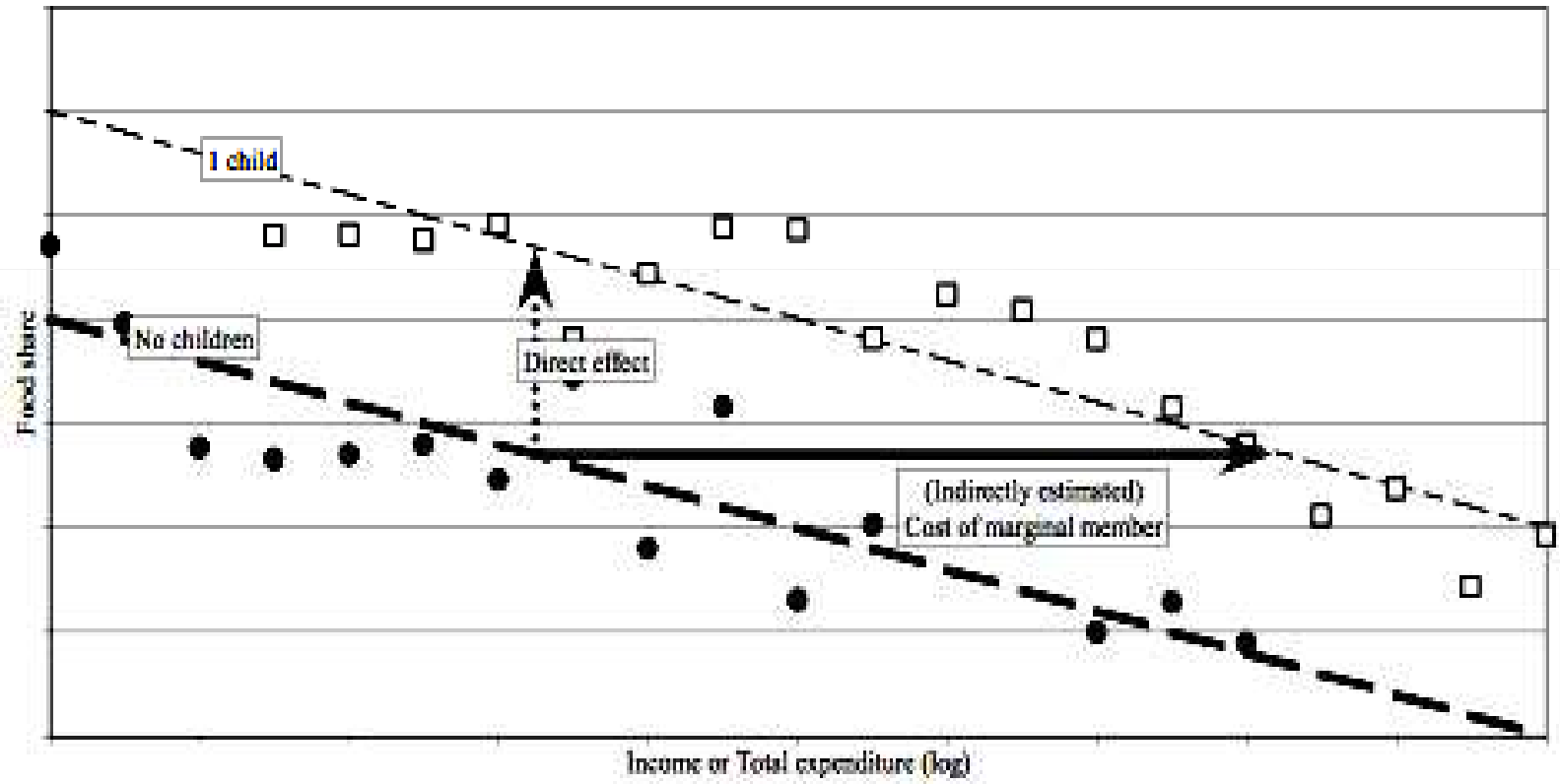
THEORETICAL METHODS TO COMPUTE COST OF A CHILD

- Subsistence cost vs. cost of a child in an intact family

- Dividing up existing expenditures of intact family
  - Ernst Engel, 1857, "food"
  - Espenshade, 1984, "food at home"
  - Rothbart, 1943, "adult goods"

ENGEL'S METHOD

Figure 2. Food share and child costs in Engel's method



Source: De Santis, Gustavo and Mauro Maltagliati, Child-cost estimates: the great leap forward, Department of Statistics, November 2001

DATA ISSUES

- Need data for detailed consumption expenditures for families with and without children
- Need complete income data for families
- Bureau of Labor Statistics publishes a survey of families in the U.S., called CES or sometimes CEX data.

MANY ECONOMIC ESTIMATES ARE BASED ON SURVEY DATA

- The **Consumer Expenditure Survey (CE)** provides information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics. The survey data are collected for the Bureau of Labor Statistics by the U.S. Census Bureau.

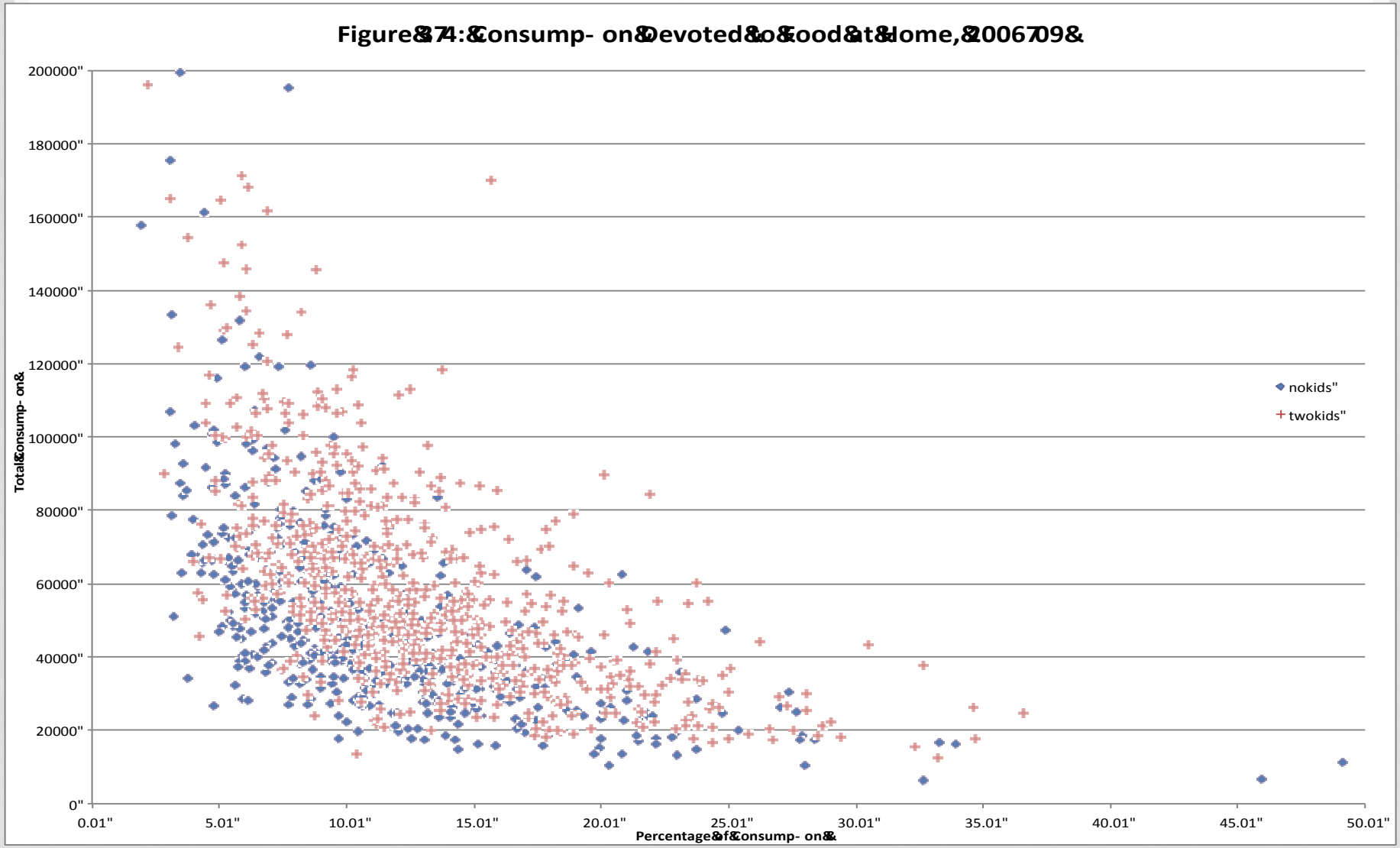
- Important because it is the only Federal survey to provide information on the complete range of consumers' expenditures and incomes, as well as the characteristics of those consumers.

- It is used by economic policymakers examining the impact of policy changes on economic groups, by businesses and academic researchers studying consumers' spending habits and trends, by other Federal agencies, and, perhaps most importantly, to regularly revise the Consumer Price Index market basket of goods and services and their relative importance.

Source: BLS, website

PROCESS OF COMPUTING COST OF CHILDREN

CONSUMPTION DEVOTED TO FOOD AT HOME, 2006-09



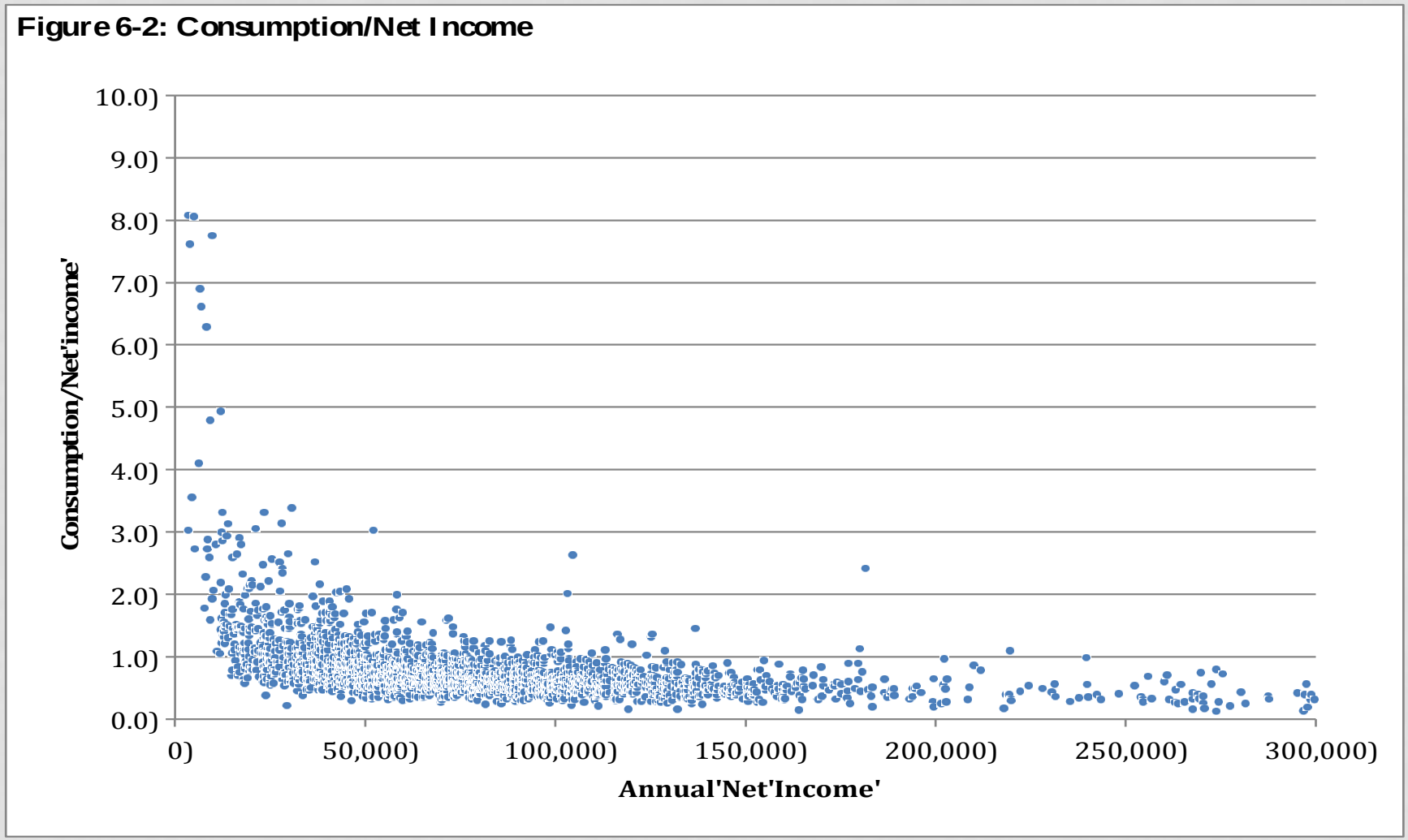
THREE DATA SETS

- Full data set (has some basic restrictions)
- Manual restrictions (we often restrict variables in a univariate manner, e.g. values > 0)
- Mahalanobis selected data

FULL DATA SET

Table 2-1: Sample Restrictions		
	Deletions	**Remaining Sample Size**
Total Number of Consumer Units		43,850
Sample Restriction		
Full Year	28,721	15,129
Income Not Imputed	6,977	8,152
Family Income Greater Than 0	30	8,122
Married	4,027	4,095
Under Age 55 If No Children	1,314	2,781
All Children Age 24 or Younger	199	2,582
No Non-Family Members living with Family	193	2,389
Not missing Data on Location	9	2,380

CONSUMPTION/INCOME



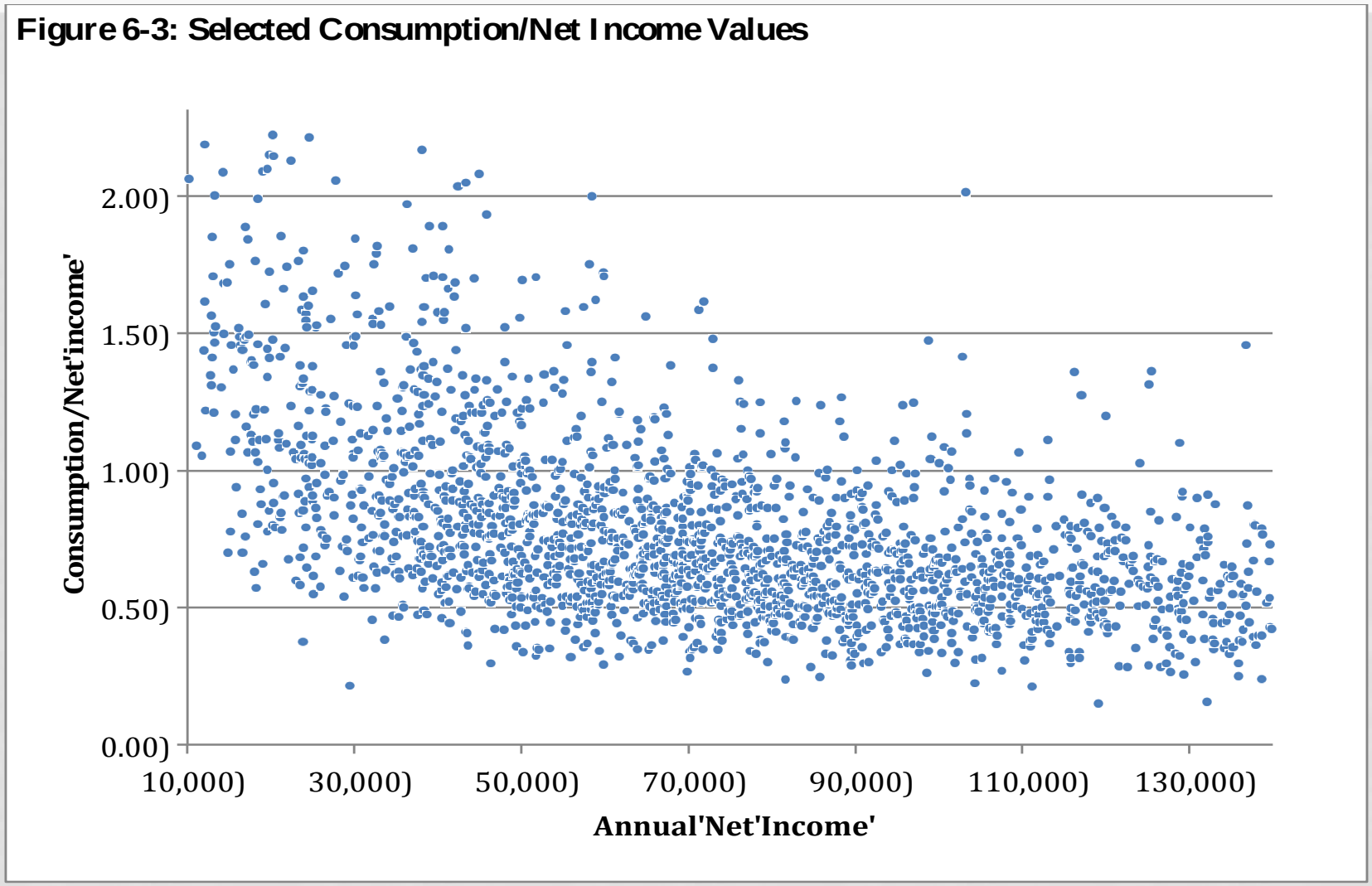
CONSUMPTION/INCOME

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.1397  0.5755  0.7518  0.9847  1.0130 65.3900 
```

MANUAL SELECTION

- Fraction of consumption/net income
  - >= 0
  - <= 2.315
- Annual net income (after tax)
  - >= 10,000
  - <= 140,000

MANUALLY SELECTED DISTRIBUTION



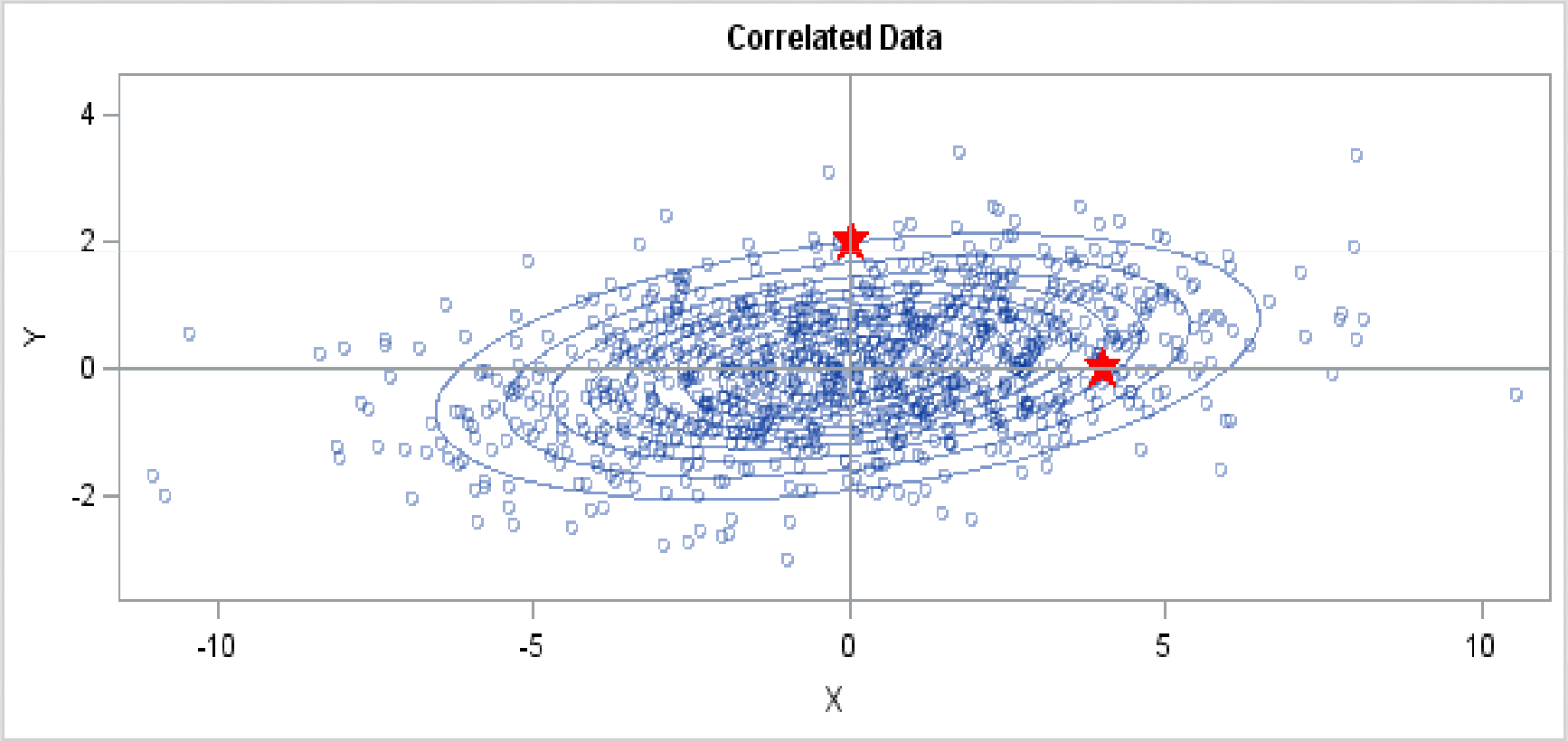
MAHALANOBIS DISTANCE

- Mahalanobis, 1936 grouping individuals

$$D_M(x) = \sqrt{(x-\mu)^T S^{-1}(x-\mu)}$$

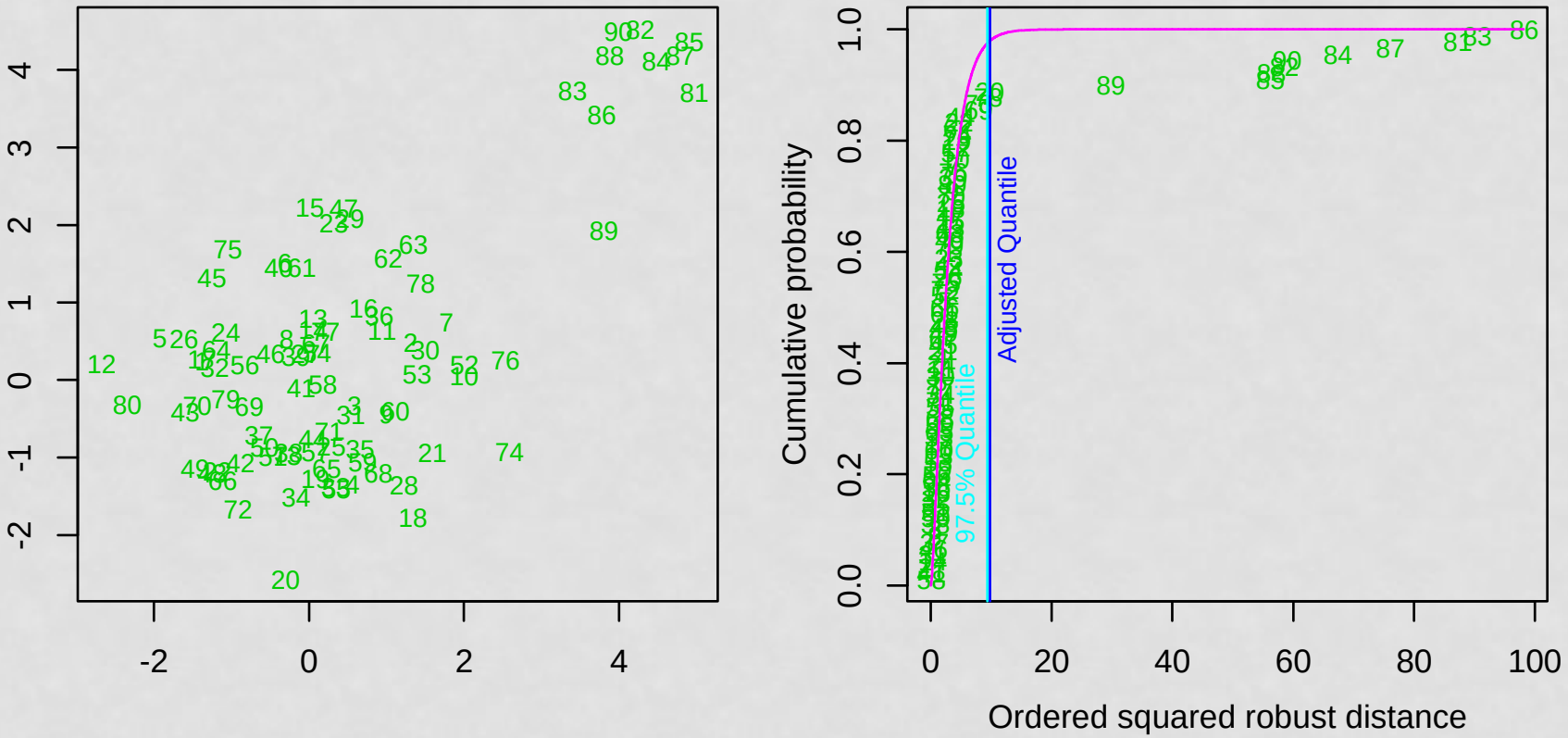
- Filzmoser, Garrett, Reimann, 2005 Geochemistry

MAHALANOBIS DISTANCE

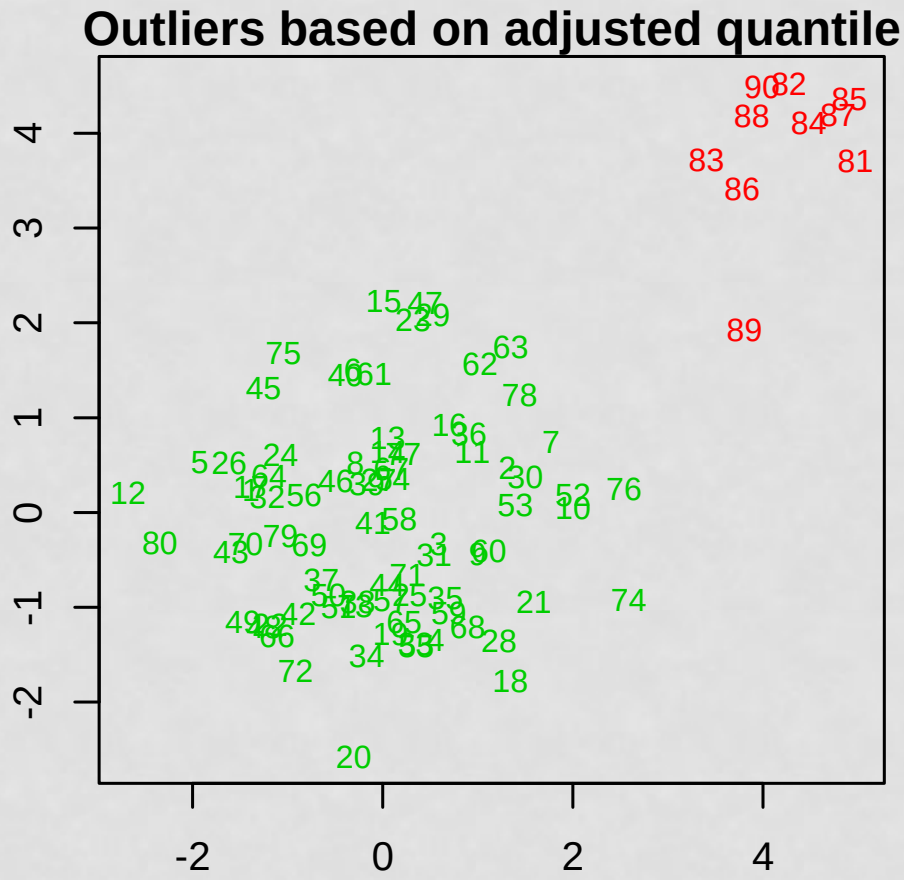


Source: Wilkin, Rick, SAS Blog, February 15, 2012

PICKING OUTLIERS



SELECTING "NORMAL" DATA



ESTIMATING THE ENGEL CURVES

To derive nonlinear Engel curves for calculating expenditures on children, the following equation for food as a share of total consumption is estimated:

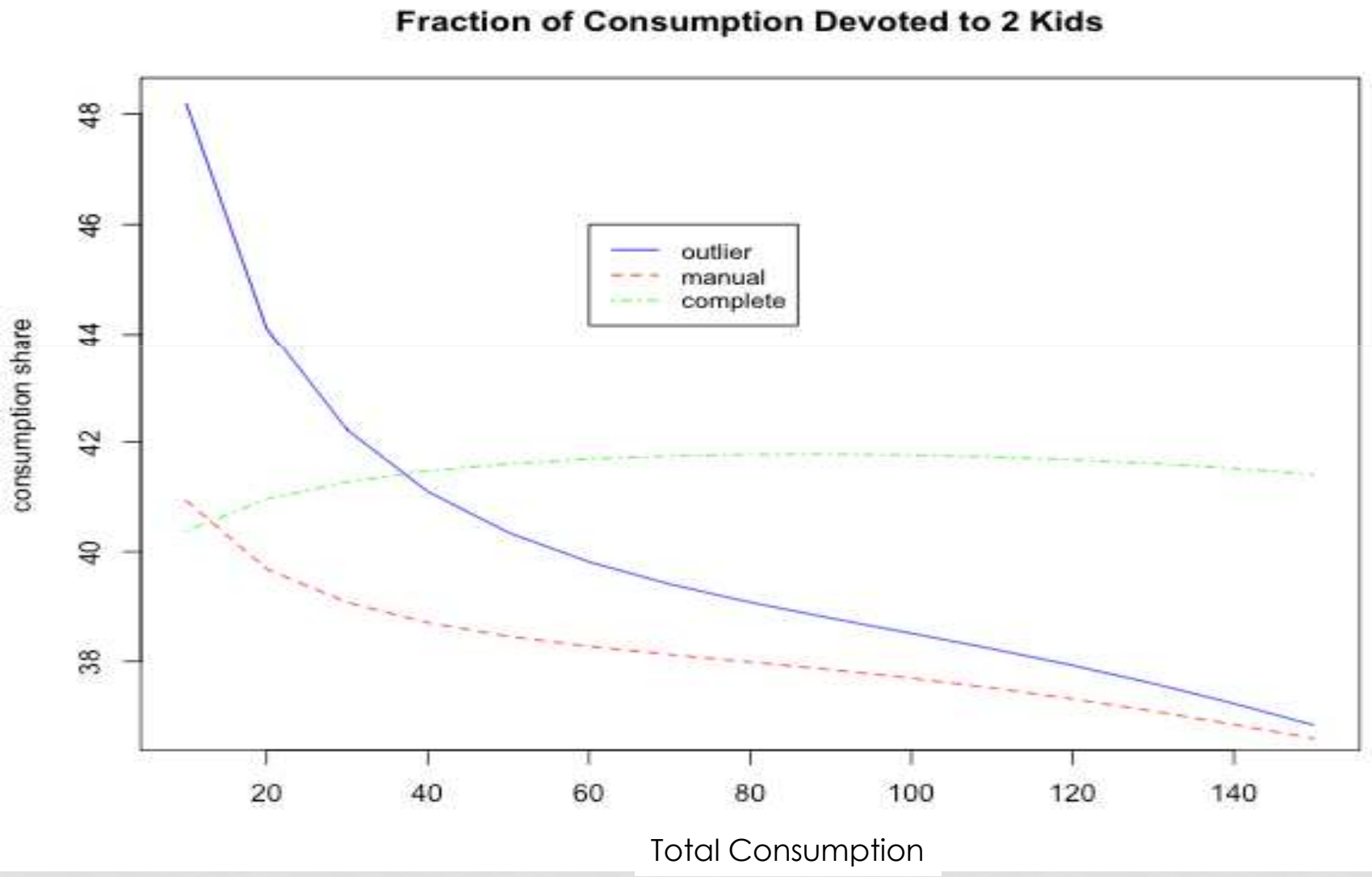
$$\ln\left(F/(1-F)\right)=\delta\ln(S)+\beta\ln(S)^2+\alpha(K)+\gamma(X)$$

- ln (*F*/1-*F*), is the log of the ratio of the food budget share to one minus the food budget share.
- the log of total spending, $\delta ln(S)$, and its square, $\beta ln(S)^2$
- the number of children in the family, $\alpha(K)$
- a set of characteristics of the adults in the family, $\gamma(X)$

OUTLIER SELECTION

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -0.9567259  0.0833146 -11.483  < 2e-16 ***
family09$lexp      -0.5805907  0.1001828  -5.795 7.92e-09 ***
family09$lexp2     -0.0305982  0.0305155  -1.003 0.316123
family09$kid1       0.1387079  0.0226647   6.120 1.13e-09 ***
family09$kid2       0.2370661  0.0216315  10.959  < 2e-16 ***
family09$kid3       0.3106914  0.0295497  10.514  < 2e-16 ***
family09$kid4       0.3316829  0.0507897   6.531 8.32e-11 ***
family09$kid5       0.5045945  0.0963713   5.236 1.82e-07 ***
family09$black     -0.1645684  0.0328817  -5.005 6.09e-07 ***
family09$northcen  -0.0962964  0.0265310  -3.630 0.000291 ***
family09$west      -0.0748085  0.0272227  -2.748 0.006050 **
family09$south     -0.0480304  0.0249428  -1.926 0.054295 .
family09$husb_nohs  0.0079879  0.0344845   0.232 0.816844
family09$husb_hsplus -0.0481151  0.0210192  -2.289 0.022179 *
family09$wife_nohs -0.0002407  0.0362315  -0.007 0.994699
family09$wife_hsplus -0.0246009  0.0216010  -1.139 0.254893
family09$wife_weeks -0.0006159  0.0004945  -1.245 0.213104
family09$wife_full -0.0241015  0.0222580  -1.083 0.279019
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.366 on 1971 degrees of freedom
Multiple R-squared: 0.4549, Adjusted R-squared: 0.4502
F-statistic: 96.74 on 17 and 1971 DF,  p-value: < 2.2e-16
```

FRACTION OF CONSUMPTION COST OF CHILDREN



CONSUMPTION TO INCOME EQUATION

$$C / NI = \alpha_0 + \alpha_1 (NI) + \alpha_2 (NI)^2$$

COMPUTING CONSUMPTION (COMPLETE DATA)

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.998e+00  9.578e-02  20.857  < 2e-16 ***
family09$inc_ataxnew   -1.488e-02  1.351e-03 -11.017  < 2e-16 ***
family09$inc_atax2new   2.651e-05  3.461e-06   7.658 2.72e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 272.7 on 2372 degrees of freedom
Multiple R-squared: 0.05532,    Adjusted R-squared: 0.05452
F-statistic: 69.45 on 2 and 2372 DF,  p-value: < 2.2e-16
```

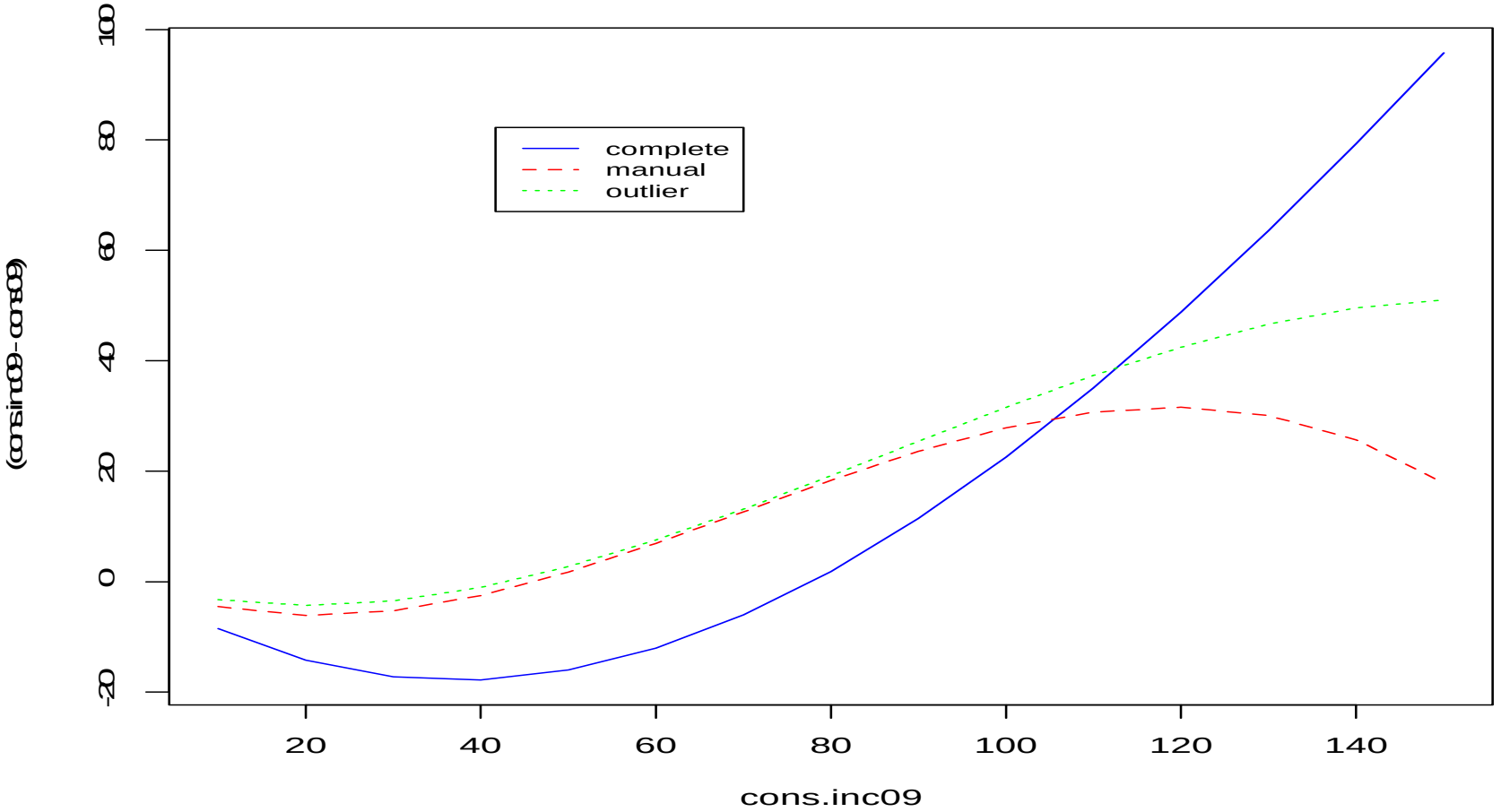
COMPUTING CONSUMPTION (MANUAL SELECTION)

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.612e+00  3.331e-02   48.41   <2e-16 ***
family09$inc_ataxnew   -1.698e-02  9.550e-04  -17.78   <2e-16 ***
family09$inc_atax2new   8.066e-05  6.681e-06   12.07   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 39.56 on 2017 degrees of freedom
Multiple R-squared: 0.3003, Adjusted R-squared: 0.2996
F-statistic: 432.9 on 2 and 2017 DF,  p-value: < 2.2e-16
```

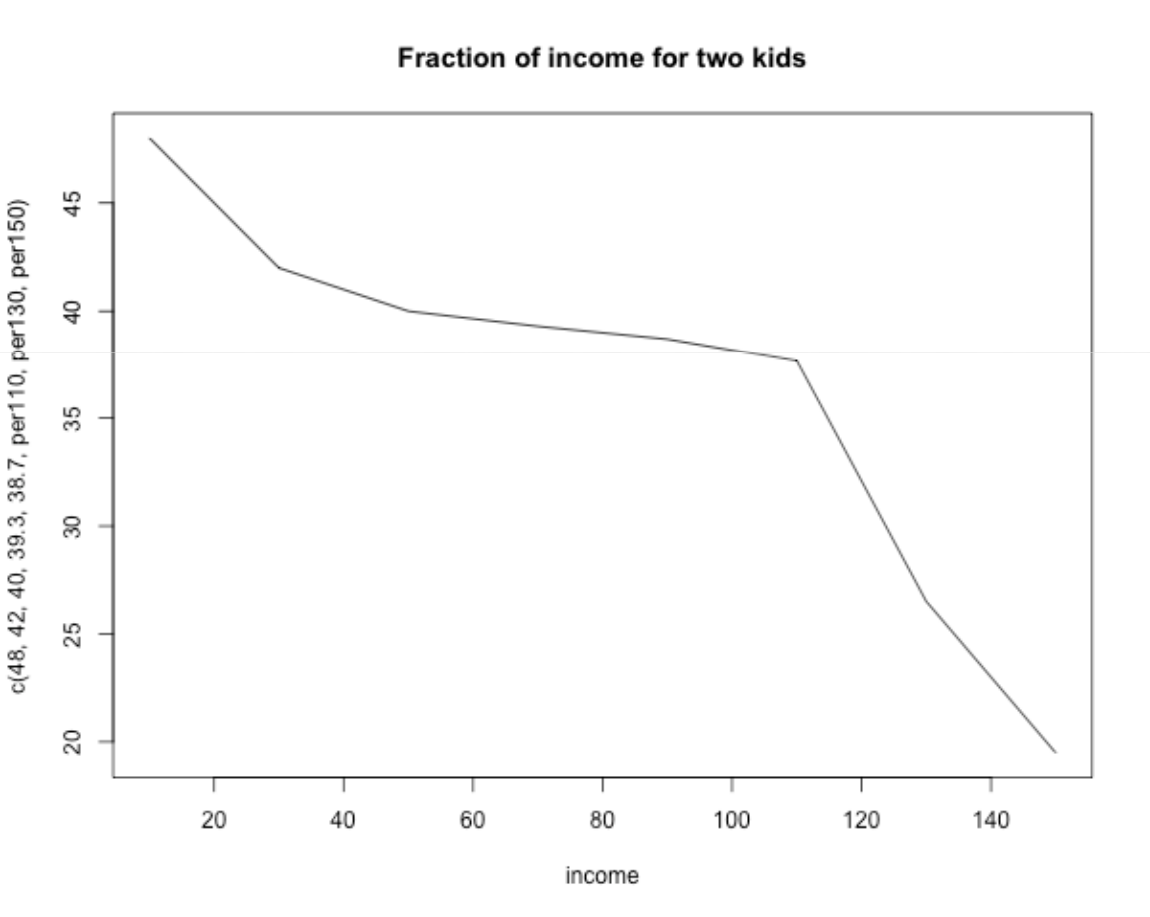
COMPUTING CONSUMPTION (OUTLIER SELECTION)

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.445e+00  2.675e-02   54.03   <2e-16 ***
familyst$inc_ataxnew   -1.235e-02  6.725e-04  -18.37   <2e-16 ***
familyst$inc_atax2new   4.745e-05  4.099e-06   11.57   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 35.42 on 1986 degrees of freedom
Multiple R-squared: 0.3201, Adjusted R-squared: 0.3194
F-statistic: 467.5 on 2 and 1986 DF,  p-value: < 2.2e-16
```

SAVING BY INCOME



FINAL ESTIMATE



CONCLUSIONS

- Survey data may have mistaken entries, misunderstanding of questions or odd survey participants
- Using manual or mechanical outlier techniques may improve the stability of the estimates
- Cannot completely solve data problems where the data is questionable. Can only solve "outliers"

PERMANENT INCOME HYPOTHESIS

- $C = f(Y_{permanent})$
- Here we have $C = f(Y_{current})$
- If $Y_{current} < Y_{permanent}$ then dissaving
- If $Y_{current} > Y_{permanent}$ then saving
- So we expect C/NI > 1 for some incomes (especially low incomes), and
- C/NI substantially less than one for some incomes (especially high incomes)
- As child support payments are cut at

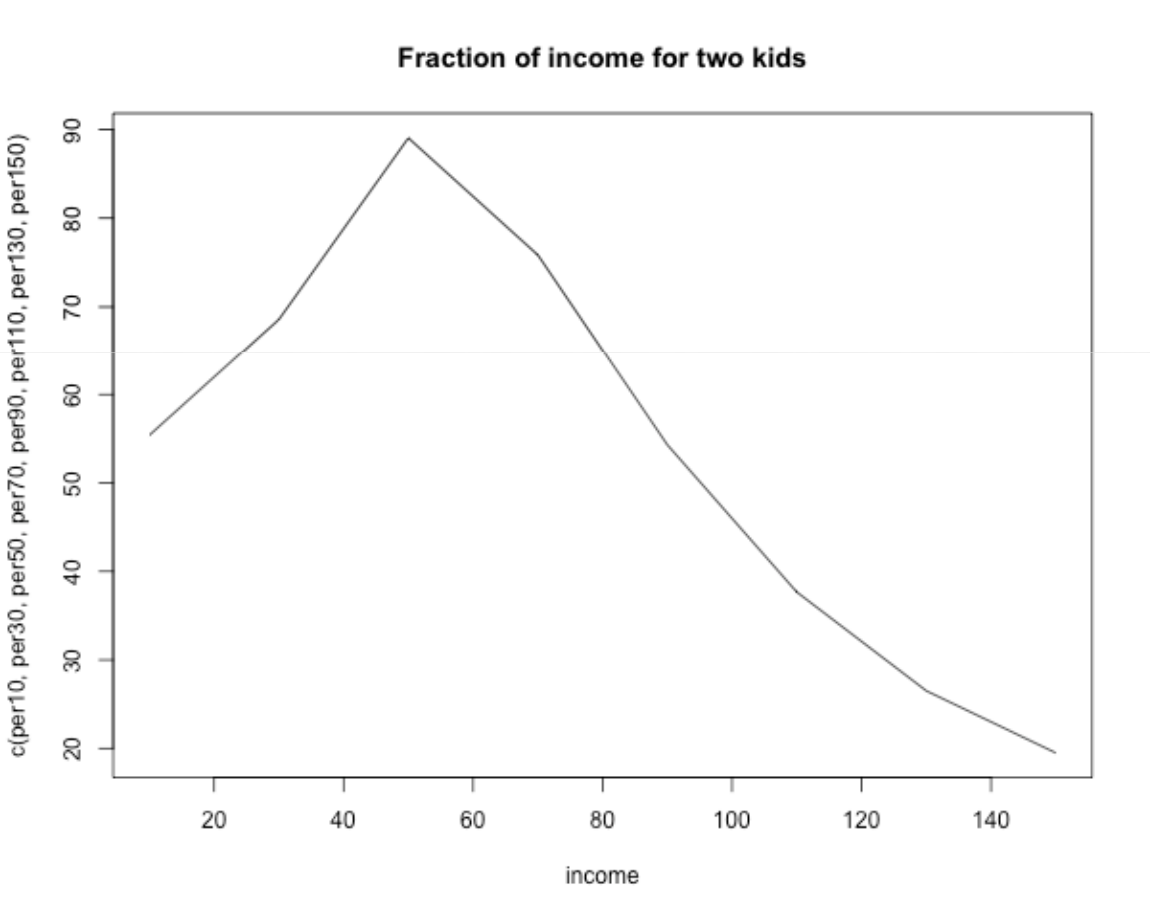
MANUAL SELECTION

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -0.8670455  0.0818010 -10.599  < 2e-16 ***
family09$lexp       -0.6706993  0.0973875  -6.887 7.61e-12 ***
family09$lexp2      -0.0074343  0.0295810  -0.251  0.80159
family09$kid1        0.1310781  0.0232612   5.635 2.00e-08 ***
family09$kid2        0.2297338  0.0223259  10.290  < 2e-16 ***
family09$kid3        0.3209391  0.0300994  10.663  < 2e-16 ***
family09$kid4        0.3910295  0.0491322   7.959 2.88e-15 ***
family09$kid5        0.5276924  0.0999813   5.278 1.45e-07 ***
family09$black      -0.1691434  0.0342939  -4.932 8.80e-07 ***
family09$northcen   -0.0885583  0.0275405  -3.216  0.00132 **
family09$west       -0.0649408  0.0280350  -2.316  0.02064 *
family09$south      -0.0288574  0.0257615  -1.120  0.26277
family09$husb_nohs   0.0198613  0.0346342   0.573  0.56640
family09$husb_hsplus -0.0526840  0.0213207  -2.471  0.01356 *
family09$wife_nohs   0.0039566  0.0364835   0.108  0.91365
family09$wife_hsplus -0.0307723  0.0220373  -1.396  0.16276
family09$wife_weeks  -0.0005598  0.0005033  -1.112  0.26611
family09$wife_full   -0.0328805  0.0229110  -1.435  0.15140
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3805 on 2001 degrees of freedom
Multiple R-squared: 0.4655, Adjusted R-squared: 0.461
F-statistic: 102.5 on 17 and 2001 DF,  p-value: < 2.2e-16
```

COMPLETE DATA

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          -0.8304178  0.0658959 -12.602  < 2e-16 ***
family09$lexp        -0.7176383  0.0704983 -10.180  < 2e-16 ***
family09$lexp2        0.0079715  0.0193737   0.411 0.680771
family09$kid1         0.1376135  0.0215335   6.391 1.98e-10 ***
family09$kid2         0.2379756  0.0204653  11.628  < 2e-16 ***
family09$kid3         0.3254450  0.0280842  11.588  < 2e-16 ***
family09$kid4         0.4087590  0.0450503   9.073  < 2e-16 ***
family09$kid5         0.4770289  0.0911290   5.235 1.80e-07 ***
family09$black       -0.1904258  0.0320391  -5.944 3.20e-09 ***
family09$northcen    -0.0999651  0.0250900  -3.984 6.97e-05 ***
family09$west        -0.0831931  0.0250794  -3.317 0.000923 ***
family09$south       -0.0427946  0.0234088  -1.828 0.067654 .
family09$husb_nohs    0.0097169  0.0332359   0.292 0.770035
family09$husb_hsplus -0.0491905  0.0203718  -2.415 0.015827 *
family09$wife_nohs    0.0156882  0.0350253   0.448 0.654259
family09$wife_hsplus -0.0231640  0.0208485  -1.111 0.266653
family09$wife_weeks  -0.0004978  0.0004662  -1.068 0.285657
family09$wife_full   -0.0326094  0.0212109  -1.537 0.124332
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3797 on 2356 degrees of freedom
Multiple R-squared: 0.5096, Adjusted R-squared: 0.5061
F-statistic:   144 on 17 and 2356 DF,  p-value: < 2.2e-16
```

UNADJUSTED


Fraction of income for two kids
c(per10, per30, per50, per70, per90, per110, per130, per150)
20
30
40
50
60
70
80
90
20
40
60
80
100
120
140
income