Fit Income Distribution from Gini

Ben Mazzotta

Wednesday, October 15, 2014

Overview

Personal income data can be fit most accurately to a lognormal distribution of income, when calibrating to the Gini coefficient on income (Eric Kemp-Benedict 2001). The evidence for this is quantile regression of the Lorenz curve that generates the Gini coefficient, when compared to similar analysis for the gamma, Pareto and beta distributions. The lognormal distribution has a distinct advantage over polynomial distributions in that it is characterized by just two parameters, μ and σ.

```
require(data.table); require(plyr); require(ggplot2)
```

```
## Loading required package: data.table
## Loading required package: plyr
## Loading required package: ggplot2
```

```
setwd("../data")
load("current.Rdata")
```

Model

Kemp-Benedict (2001:7) shows that the Gini index depends only on the parameter σ of the lognormal distribution.

$$g = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$$

... where Φ is the cumulative standard normal distribution. If we knew the value of σ, we could write the following code.

```
g <- 2 * dnorm(sigma/sqrt(2)) - 1
```

And hence to calculate σ from a known Gini coefficient, we would invert as follows.

$$\sigma = \sqrt{2}\Phi^{-1}\left(\frac{g+1}{2}\right)$$

The corresponding R code is:

```
sigma = sqrt(2) * qnorm( (g+1)/2 )
```

Parameters

The scale and location parameters of the lognormal distribution are μ and σ. From Kemp above we can calculate σ.

We also know that average income is a function of μ and σ. So inverting, we can derive:

$$E(I) = \exp\mu + \frac{\sigma^2}{2}$$

$$\mu = \log(E(I)) - \frac{\sigma^2}{2}$$

And hence the procedure should be to calculate:

1. Sigma from Gini coefficient
2. Mu from GNI/capita and sigma
3. Quantiles from known distribution I = f(mu, sigma; X).

```
inc <- inc2010[!is.na(gini),]
setkey(inc, iso2c, country, year)

##      Step 1: calculate sigma from Gini coefficient.
##          (BDM: is this possible w/o scale parameter?)
inc[,sigma:= sqrt(2)*qnorm((gini/100+1)/2)]
```

```
##      iso2c       country year    income inccap  gini    saving savingrate
##   1:    AL       Albania 2010 1.087e+10   4493 28.96 2.011e+09      16.36
##   2:    AM       Armenia 2010 6.557e+09   3530 30.90 1.582e+09      15.44
##   3:    AO        Angola 2010 4.629e+10   4150 42.66 2.027e+10      22.62
##   4:    AR     Argentina 2010 2.368e+11   6310 44.03 1.010e+11      19.12
##   5:    AT       Austria 2010 3.313e+11  47737 29.50 9.782e+10      24.80
##  ---
## 136:    VE Venezuela, RB 2010 1.781e+11  11913 44.77 1.059e+11      29.81
## 137:    VN       Vietnam 2010 7.944e+10   1403 37.44 4.072e+10      31.24
## 138:    YE   Yemen, Rep. 2010 1.514e+10   1200 35.91 3.331e+09      18.66
## 139:    ZA  South Africa 2010 2.923e+11   6770 65.02 5.888e+10      15.68
## 140:    ZM        Zambia 2010 9.592e+09   1223 57.49 5.422e+09      30.72
##       savers  sigma
##   1:  8.5616 0.5251
##   2:  0.8191 0.5621
##   3: 15.9156 0.7963
##   4:  3.8007 0.8250
##   5: 51.5938 0.5355
##  ---
## 136: 13.5693 0.8405
## 137:  7.7408 0.6899
## 138:  1.0740 0.6596
## 139: 22.0800 1.3223
## 140: 11.7659 1.1280
```

```
summary(inc)
```

```
##     iso2c             country                year          income
##  Length:140         Length:140         Min.   :2010   Min.   :1.20e+08
##  Class :character   Class :character   1st Qu.:2010   1st Qu.:5.68e+09
##  Mode  :character   Mode  :character   Median :2010   Median :2.37e+10
##                                        Mean   :2010   Mean   :3.53e+11
##                                        3rd Qu.:2010   3rd Qu.:1.70e+11
##                                        Max.   :2010   Max.   :1.41e+13
##
##      inccap           gini          saving            savingrate
```

```
##  Min.   :  220   Min.   :24.8   Min.   :-3.10e+08   Min.   :-6.95
##  1st Qu.: 1182   1st Qu.:33.0   1st Qu.: 1.58e+09   1st Qu.:14.44
##  Median : 3980   Median :37.4   Median : 7.02e+09   Median :19.15
##  Mean   :10824   Mean   :39.4   Mean   : 1.15e+11   Mean   :20.85
##  3rd Qu.:10709   3rd Qu.:44.1   3rd Qu.: 5.54e+10   3rd Qu.:26.14
##  Max.   :91570   Max.   :65.8   Max.   : 3.65e+12   Max.   :72.12
##                                 NA's   :7           NA's   :7
##      savers          sigma
##  Min.   : 0.29   Min.   :0.447
##  1st Qu.: 5.46   1st Qu.:0.602
##  Median :11.77   Median :0.689
##  Mean   :16.68   Mean   :0.736
##  3rd Qu.:19.90   3rd Qu.:0.827
##  Max.   :63.58   Max.   :1.343
##  NA's   :15
```

```
##      Step 2: calculate mu from sigma and mean
inc[,mu:=log(inccap)-(sigma^2/2)]
```

```
##      iso2c       country year    income inccap  gini    saving savingrate
##   1:    AL       Albania 2010 1.087e+10   4493 28.96 2.011e+09      16.36
##   2:    AM       Armenia 2010 6.557e+09   3530 30.90 1.582e+09      15.44
##   3:    AO        Angola 2010 4.629e+10   4150 42.66 2.027e+10      22.62
##   4:    AR     Argentina 2010 2.368e+11   6310 44.03 1.010e+11      19.12
##   5:    AT       Austria 2010 3.313e+11  47737 29.50 9.782e+10      24.80
##  ---
## 136:    VE Venezuela, RB 2010 1.781e+11  11913 44.77 1.059e+11      29.81
## 137:    VN       Vietnam 2010 7.944e+10   1403 37.44 4.072e+10      31.24
## 138:    YE   Yemen, Rep. 2010 1.514e+10   1200 35.91 3.331e+09      18.66
## 139:    ZA  South Africa 2010 2.923e+11   6770 65.02 5.888e+10      15.68
## 140:    ZM        Zambia 2010 9.592e+09   1223 57.49 5.422e+09      30.72
##       savers  sigma     mu
##   1:  8.5616 0.5251  8.272
##   2:  0.8191 0.5621  8.011
##   3: 15.9156 0.7963  8.014
##   4:  3.8007 0.8250  8.410
##   5: 51.5938 0.5355 10.630
##  ---
## 136: 13.5693 0.8405  9.032
## 137:  7.7408 0.6899  7.009
## 138:  1.0740 0.6596  6.873
## 139: 22.0800 1.3223  7.946
## 140: 11.7659 1.1280  6.473
```

```
summary(inc)
```

```
##     iso2c             country              year         income
##  Length:140         Length:140         Min.   :2010   Min.   :1.20e+08
##  Class :character   Class :character   1st Qu.:2010   1st Qu.:5.68e+09
##  Mode  :character   Mode  :character   Median :2010   Median :2.37e+10
##                                        Mean   :2010   Mean   :3.53e+11
##                                        3rd Qu.:2010   3rd Qu.:1.70e+11
##                                        Max.   :2010   Max.   :1.41e+13
```

```
##
##      inccap           gini          saving            savingrate
##  Min.   :  220   Min.   :24.8   Min.   :-3.10e+08   Min.   :-6.95
##  1st Qu.: 1182   1st Qu.:33.0   1st Qu.: 1.58e+09   1st Qu.:14.44
##  Median : 3980   Median :37.4   Median : 7.02e+09   Median :19.15
##  Mean   :10824   Mean   :39.4   Mean   : 1.15e+11   Mean   :20.85
##  3rd Qu.:10709   3rd Qu.:44.1   3rd Qu.: 5.54e+10   3rd Qu.:26.14
##  Max.   :91570   Max.   :65.8   Max.   : 3.65e+12   Max.   :72.12
##                                 NA's   :7           NA's   :7
##      savers          sigma             mu
##  Min.   : 0.29   Min.   :0.447   Min.   : 5.21
##  1st Qu.: 5.46   1st Qu.:0.602   1st Qu.: 6.76
##  Median :11.77   Median :0.689   Median : 7.97
##  Mean   :16.68   Mean   :0.736   Mean   : 7.97
##  3rd Qu.:19.90   3rd Qu.:0.827   3rd Qu.: 8.93
##  Max.   :63.58   Max.   :1.343   Max.   :11.31
##  NA's   :15
```

```
##      Step 3: Read quantiles from mu and sigma
inc[,p20:=qlnorm(.20, mu, sigma)]
```

```
##      iso2c       country year    income inccap  gini    saving savingrate
##   1:    AL       Albania 2010 1.087e+10   4493 28.96 2.011e+09      16.36
##   2:    AM       Armenia 2010 6.557e+09   3530 30.90 1.582e+09      15.44
##   3:    AO        Angola 2010 4.629e+10   4150 42.66 2.027e+10      22.62
##   4:    AR     Argentina 2010 2.368e+11   6310 44.03 1.010e+11      19.12
##   5:    AT       Austria 2010 3.313e+11  47737 29.50 9.782e+10      24.80
##  ---
## 136:    VE Venezuela, RB 2010 1.781e+11  11913 44.77 1.059e+11      29.81
## 137:    VN       Vietnam 2010 7.944e+10   1403 37.44 4.072e+10      31.24
## 138:    YE   Yemen, Rep. 2010 1.514e+10   1200 35.91 3.331e+09      18.66
## 139:    ZA  South Africa 2010 2.923e+11   6770 65.02 5.888e+10      15.68
## 140:    ZM        Zambia 2010 9.592e+09   1223 57.49 5.422e+09      30.72
##       savers  sigma     mu     p20
##   1:  8.5616 0.5251  8.272  2516.2
##   2:  0.8191 0.5621  8.011  1878.1
##   3: 15.9156 0.7963  8.014  1546.4
##   4:  3.8007 0.8250  8.410  2242.3
##   5: 51.5938 0.5355 10.630 26354.7
##  ---
## 136: 13.5693 0.8405  9.032  4125.0
## 137:  7.7408 0.6899  7.009   618.9
## 138:  1.0740 0.6596  6.873   554.1
## 139: 22.0800 1.3223  7.946   928.2
## 140: 11.7659 1.1280  6.473   250.6
```

```
inc[,psav:=qlnorm(1-savers/100, mu, sigma)]
```

```
##      iso2c       country year    income inccap  gini    saving savingrate
##   1:    AL       Albania 2010 1.087e+10   4493 28.96 2.011e+09      16.36
##   2:    AM       Armenia 2010 6.557e+09   3530 30.90 1.582e+09      15.44
##   3:    AO        Angola 2010 4.629e+10   4150 42.66 2.027e+10      22.62
##   4:    AR     Argentina 2010 2.368e+11   6310 44.03 1.010e+11      19.12
```

```
##   5:    AT       Austria 2010 3.313e+11  47737 29.50 9.782e+10      24.80
##  ---
## 136:    VE Venezuela, RB 2010 1.781e+11  11913 44.77 1.059e+11      29.81
## 137:    VN       Vietnam 2010 7.944e+10   1403 37.44 4.072e+10      31.24
## 138:    YE   Yemen, Rep. 2010 1.514e+10   1200 35.91 3.331e+09      18.66
## 139:    ZA  South Africa 2010 2.923e+11   6770 65.02 5.888e+10      15.68
## 140:    ZM        Zambia 2010 9.592e+09   1223 57.49 5.422e+09      30.72
##       savers  sigma     mu     p20  psav
##   1:  8.5616 0.5251  8.272  2516.2  8030
##   2:  0.8191 0.5621  8.011  1878.1 11617
##   3: 15.9156 0.7963  8.014  1546.4  6691
##   4:  3.8007 0.8250  8.410  2242.3 19407
##   5: 51.5938 0.5355 10.630 26354.7 40485
##  ---
## 136: 13.5693 0.8405  9.032  4125.0 21092
## 137:  7.7408 0.6899  7.009   618.9  2952
## 138:  1.0740 0.6596  6.873   554.1  4400
## 139: 22.0800 1.3223  7.946   928.2  7813
## 140: 11.7659 1.1280  6.473   250.6  2470
```

```
options(digits=4)
summary(inc)
```

```
##     iso2c             country                year          income
##  Length:140         Length:140         Min.   :2010   Min.   :1.20e+08
##  Class :character   Class :character   1st Qu.:2010   1st Qu.:5.68e+09
##  Mode  :character   Mode  :character   Median :2010   Median :2.37e+10
##                                        Mean   :2010   Mean   :3.53e+11
##                                        3rd Qu.:2010   3rd Qu.:1.70e+11
##                                        Max.   :2010   Max.   :1.41e+13
##
##      inccap           gini          saving            savingrate
##  Min.   :  220   Min.   :24.8   Min.   :-3.10e+08   Min.   :-6.95
##  1st Qu.: 1182   1st Qu.:33.0   1st Qu.: 1.58e+09   1st Qu.:14.44
##  Median : 3980   Median :37.4   Median : 7.02e+09   Median :19.15
##  Mean   :10824   Mean   :39.4   Mean   : 1.15e+11   Mean   :20.85
##  3rd Qu.:10709   3rd Qu.:44.1   3rd Qu.: 5.54e+10   3rd Qu.:26.14
##  Max.   :91570   Max.   :65.8   Max.   : 3.65e+12   Max.   :72.12
##                                 NA's   :7           NA's   :7
##      savers          sigma             mu             p20
##  Min.   : 0.29   Min.   :0.447   Min.   : 5.21   Min.   :  106
##  1st Qu.: 5.46   1st Qu.:0.602   1st Qu.: 6.76   1st Qu.:  456
##  Median :11.77   Median :0.689   Median : 7.97   Median : 1425
##  Mean   :16.68   Mean   :0.736   Mean   : 7.97   Mean   : 5273
##  3rd Qu.:19.90   3rd Qu.:0.827   3rd Qu.: 8.93   3rd Qu.: 4174
##  Max.   :63.58   Max.   :1.343   Max.   :11.31   Max.   :54134
##  NA's   :15
##       psav
##  Min.   :  541
##  1st Qu.: 2846
##  Median : 7813
##  Mean   :12685
##  3rd Qu.:18151
##  Max.   :56142
```

```
##  NA's   :15
```

```
summary(1-inc$savers/100)
```

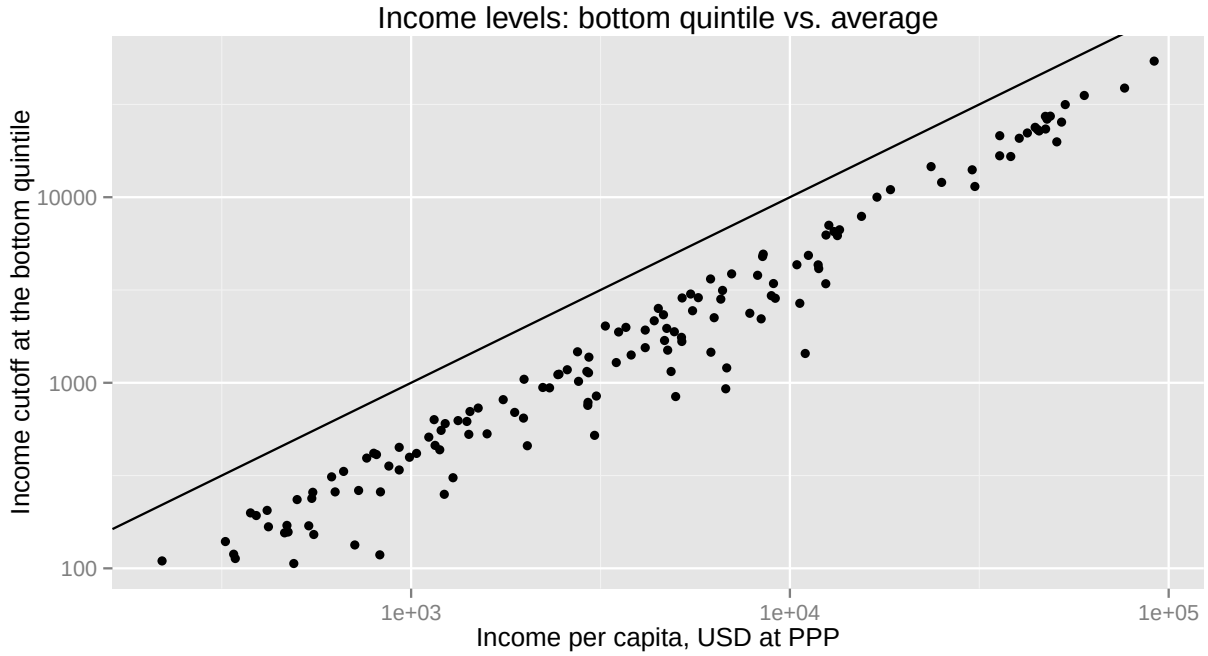
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.364   0.801   0.882   0.833   0.945   0.997      15
```

So the object `inc` in memory has accurate parameters for the lognormal distribution of per capita income in memory.

It also has the 20th percentile of per capita income, and the nth percentile of per capita income, where n corresponds to the share of the population that do not save at a formal financial institution.

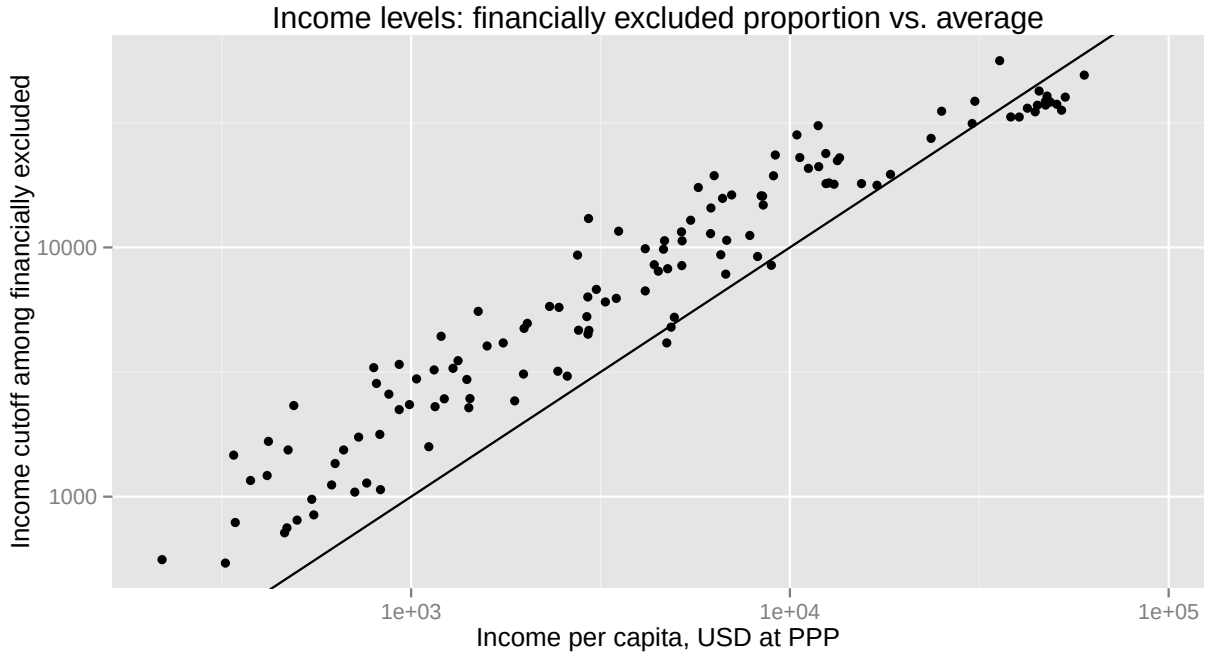
Are these estimates plausible?

```
## Plot p20 vs income per capita.
##      Has to be all below 45 degree line
qplot(inccap, p20, data=inc, geom="point", log="xy")+geom_abline(intercept=0,slope=1)+
  labs(list(title="Income levels: bottom quintile vs. average", x="Income per capita, USD a
```




```
## Plot psavers vs income per capita.
##      Rich countries below the line, poor countries above
qplot(inccap, psav, data=inc, geom="point", log="xy")+geom_abline(intercept=0,slope=1)+
  labs(list(title="Income levels: financially excluded proportion vs. average", x="Income p
```

```
## Warning: Removed 15 rows containing missing values (geom_point).
```



Simulate distributions

Now let's take the distribution parameters mu and sigma for each country. Using a large-enough population, simulate the share of income accruing to the bottom quintile of the distribution.

```
# inc.siminput <- inc[,.SD, .SDcols=c("mu","sigma")]

## Make a list of simulated income data
##      Each of length 10^6, using the observed values of mu and sigma
sim <- mapply(rlnorm, rep(1e6, nrow(inc)), inc$mu, inc$sigma)
##      It's very large.
dim(sim)
```

```
## [1] 1000000     140
```

```
object.size(sim)/1e6
```

```
## 1120.0002 bytes
```

```
##      It took several seconds to compute the summary statistics.
# summary(sim)
sim <- data.table(sim)

sim.sums <- colSums(sim)

str(inc$p20)
```

```
##  num [1:140] 2516 1878 1546 2242 26355 ...
```

```r
head(inc$p20)
```

```
## [1]  2516  1878  1546  2242 26355 25399
```

```r
# ##    How to use data.table to get fast sums of subsets?
# sim[V1<235,sum(V1)]
# ## You can run this code to get the sum of bottom quintile values in a single column
# myvar <- "V1"
# sim[get(myvar)<234.4945, sum(.SD), .SDcols=myvar]
# # Can I rewrite so that only number 1 is an input?
# sim[get(paste("V",1,sep=""))<inc$p20[1], sum(.SD), .SDcols=paste("V",1,sep="")]

## Write a function to do this.
calc.p20sum <- function(x) sim[get(paste("V",x,sep=""))<inc$p20[x], sum(.SD), .SDcols=paste

# Does it still work? Yes. :)
calc.p20sum(1)
```

```
## [1] 386124694
```

```r
sim.sump20 <- sapply(c(1:length(inc$p20)), calc.p20sum)
summary(sim.sump20)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 1.28e+07 6.21e+07 1.85e+08 7.83e+08 5.89e+08 8.43e+09
```

```r
sim.p20share <- sim.sump20/sim.sums
sim.p20share *100
```

```
##    V1    V2    V3    V4    V5    V6    V7    V8    V9   V10   V11   V12
## 8.599 8.004 5.059 4.780 8.412 7.176 7.426 7.417 7.672 7.366 5.704 6.899
##   V13   V14   V15   V16   V17   V18   V19   V20   V21   V22   V23   V24
## 7.361 4.883 4.295 3.137 5.969 2.060 9.143 7.237 4.694 2.608 5.611 7.496
##   V25   V26   V27   V28   V29   V30   V31   V32   V33   V34   V35   V36
## 4.946 3.477 5.511 5.219 2.896 3.920 9.307 8.093 5.686 9.179 4.246 4.111
##   V37   V38   V39   V40   V41   V42   V43   V44   V45   V46   V47   V48
## 7.573 8.077 6.700 7.245 8.954 5.028 1.973 7.843 5.181 6.140 5.296 5.056
##   V49   V50   V51   V52   V53   V54   V55   V56   V57   V58   V59   V60
## 4.148 7.255 6.971 3.233 6.757 2.754 7.294 2.204 8.529 6.746 7.700 5.051
##   V61   V62   V63   V64   V65   V66   V67   V68   V69   V70   V71   V72
## 7.177 8.396 6.068 9.412 6.763 2.920 7.252 7.686 4.036 7.035 7.516 1.599
##   V73   V74   V75   V76   V77   V78   V79   V80   V81   V82   V83   V84
## 8.703 6.585 6.562 6.090 2.936 7.391 6.732 5.439 7.922 8.433 5.536 4.902
##   V85   V86   V87   V88   V89   V90   V91   V92   V93   V94   V95   V96
## 7.437 6.504 5.535 6.654 6.308 4.323 4.054 4.337 4.429 1.943 7.951 5.005
##   V97   V98   V99  V100  V101  V102  V103  V104  V105  V106  V107  V108
## 4.447 8.639 9.204 7.461 3.315 4.516 4.993 8.404 7.447 6.514 3.472 9.006
##  V109  V110  V111  V112  V113  V114  V115  V116  V117  V118  V119  V120
## 8.408 6.155 3.490 1.453 6.820 9.247 9.849 9.172 6.815 5.622 7.196 5.028
##  V121  V122  V123  V124  V125  V126  V127  V128  V129  V130  V131  V132
## 6.687 3.359 4.945 4.367 5.831 7.845 8.159 6.694 5.804 6.181 9.870 4.904
##  V133  V134  V135  V136  V137  V138  V139  V140
## 5.415 4.958 7.146 4.620 6.276 6.664 1.533 2.444
```

```
## And to do the same thing for the proportion saving:

## This function returns the gross income beneath the nth percentile of income,
##      Where n is the percentage of those that do not save.
calc.finexsum <- function(x) sim[get(paste("V",x,sep=""))<inc$psav[x], sum(.SD), .SDcols=pa

# Does it still work? Yes. :)
calc.finexsum(1)
```

```
## [1] 3.595e+09
```

```
# calc.finexsum(17)

# inc$psav

## Running the same command as above threw an error
##      Not all the psav observations are defined.
##      Need to substitute something when undefined.....
sim.sumfinex <- sapply(which(!is.na(inc$psav)), calc.finexsum)
summary(sim.sumfinex)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 1.96e+08 8.28e+08 2.67e+09 4.35e+09 7.11e+09 2.30e+10
```

```
sim.finexshare <- rep(NA, nrow(inc))
sim.finexshare[which(!is.na(inc$psav))] <- sim.sumfinex / sim.sums[which(!is.na(inc$psav))]

# options(digits=3)
sim.finexshare*100
```

```
##   [1] 80.06 96.72 57.91 82.79 28.21 17.73 93.85 82.74 64.83 33.67 75.00
##  [12] 84.65 89.04 74.56 52.98 59.68    NA 41.01 84.11 24.26 91.11 80.51
##  [23] 80.37    NA    NA 57.17 70.10 37.45 60.84 47.07 45.63 24.10 86.11
##  [34] 25.77 54.60 56.52 48.59 97.14 39.32    NA 25.51    NA    NA 28.69
##  [45] 71.82 29.22 93.77 57.54    NA 92.31 58.30 60.40    NA 61.49 70.94
##  [56] 39.77 66.18 64.37 26.85 45.35 71.75 85.62 55.88    NA 64.11 29.57
##  [67] 77.94 26.80 43.02 95.83 96.47 47.47 83.52 57.92 46.50 64.75 64.22
##  [78] 58.58 67.60 65.62 89.12 90.17 92.68 72.69 86.14 52.47 77.71 43.57
##  [89]    NA 69.78 72.36 31.05 53.17    NA 95.54 46.71 74.31 23.60    NA
## [100] 75.41 56.19 69.77 59.62 95.02 62.35 82.39 62.73 80.52 90.58 70.49
## [111] 48.09    NA 88.00 20.30 54.27 44.02 66.01 84.96    NA 63.05 83.64
## [122] 47.62 75.02 82.24 29.26 98.57    NA 83.96 84.08 68.55 87.70 56.67
## [133] 21.98 78.04 96.22 60.30 76.79 94.87 29.16 52.53
```

```
gini.dt <- inc
gini.dt[,bottom20share <- sim.p20share]
```

```
##      V1      V2      V3      V4      V5      V6      V7      V8      V9
## 0.08599 0.08004 0.05059 0.04780 0.08412 0.07176 0.07426 0.07417 0.07672
##     V10     V11     V12     V13     V14     V15     V16     V17     V18
## 0.07366 0.05704 0.06899 0.07361 0.04883 0.04295 0.03137 0.05969 0.02060
```

```
##      V19     V20     V21     V22     V23     V24     V25     V26     V27
## 0.09143 0.07237 0.04694 0.02608 0.05611 0.07496 0.04946 0.03477 0.05511
##      V28     V29     V30     V31     V32     V33     V34     V35     V36
## 0.05219 0.02896 0.03920 0.09307 0.08093 0.05686 0.09179 0.04246 0.04111
##      V37     V38     V39     V40     V41     V42     V43     V44     V45
## 0.07573 0.08077 0.06700 0.07245 0.08954 0.05028 0.01973 0.07843 0.05181
##      V46     V47     V48     V49     V50     V51     V52     V53     V54
## 0.06140 0.05296 0.05056 0.04148 0.07255 0.06971 0.03233 0.06757 0.02754
##      V55     V56     V57     V58     V59     V60     V61     V62     V63
## 0.07294 0.02204 0.08529 0.06746 0.07700 0.05051 0.07177 0.08396 0.06068
##      V64     V65     V66     V67     V68     V69     V70     V71     V72
## 0.09412 0.06763 0.02920 0.07252 0.07686 0.04036 0.07035 0.07516 0.01599
##      V73     V74     V75     V76     V77     V78     V79     V80     V81
## 0.08703 0.06585 0.06562 0.06090 0.02936 0.07391 0.06732 0.05439 0.07922
##      V82     V83     V84     V85     V86     V87     V88     V89     V90
## 0.08433 0.05536 0.04902 0.07437 0.06504 0.05535 0.06654 0.06308 0.04323
##      V91     V92     V93     V94     V95     V96     V97     V98     V99
## 0.04054 0.04337 0.04429 0.01943 0.07951 0.05005 0.04447 0.08639 0.09204
##     V100    V101    V102    V103    V104    V105    V106    V107    V108
## 0.07461 0.03315 0.04516 0.04993 0.08404 0.07447 0.06514 0.03472 0.09006
##     V109    V110    V111    V112    V113    V114    V115    V116    V117
## 0.08408 0.06155 0.03490 0.01453 0.06820 0.09247 0.09849 0.09172 0.06815
##     V118    V119    V120    V121    V122    V123    V124    V125    V126
## 0.05622 0.07196 0.05028 0.06687 0.03359 0.04945 0.04367 0.05831 0.07845
##     V127    V128    V129    V130    V131    V132    V133    V134    V135
## 0.08159 0.06694 0.05804 0.06181 0.09870 0.04904 0.05415 0.04958 0.07146
##     V136    V137    V138    V139    V140
## 0.04620 0.06276 0.06664 0.01533 0.02444
```

```
gini.dt[,finexshare <- sim.finexshare]
```

```
##   [1] 0.8006 0.9672 0.5791 0.8279 0.2821 0.1773 0.9385 0.8274 0.6483 0.3367
##  [11] 0.7500 0.8465 0.8904 0.7456 0.5298 0.5968     NA 0.4101 0.8411 0.2426
##  [21] 0.9111 0.8051 0.8037     NA     NA 0.5717 0.7010 0.3745 0.6084 0.4707
##  [31] 0.4563 0.2410 0.8611 0.2577 0.5460 0.5652 0.4859 0.9714 0.3932     NA
##  [41] 0.2551     NA     NA 0.2869 0.7182 0.2922 0.9377 0.5754     NA 0.9231
##  [51] 0.5830 0.6040     NA 0.6149 0.7094 0.3977 0.6618 0.6437 0.2685 0.4535
##  [61] 0.7175 0.8562 0.5588     NA 0.6411 0.2957 0.7794 0.2680 0.4302 0.9583
##  [71] 0.9647 0.4747 0.8352 0.5792 0.4650 0.6475 0.6422 0.5858 0.6760 0.6562
##  [81] 0.8912 0.9017 0.9268 0.7269 0.8614 0.5247 0.7771 0.4357     NA 0.6978
##  [91] 0.7236 0.3105 0.5317     NA 0.9554 0.4671 0.7431 0.2360     NA 0.7541
## [101] 0.5619 0.6977 0.5962 0.9502 0.6235 0.8239 0.6273 0.8052 0.9058 0.7049
## [111] 0.4809     NA 0.8800 0.2030 0.5427 0.4402 0.6601 0.8496     NA 0.6305
## [121] 0.8364 0.4762 0.7502 0.8224 0.2926 0.9857     NA 0.8396 0.8408 0.6855
## [131] 0.8770 0.5667 0.2198 0.7804 0.9622 0.6030 0.7679 0.9487 0.2916 0.5253
```

```
save(gini.dt, file="income share of financially excluded.Rda")
rm(sim)
save.image("../data/working.Rdata")
```

The lognormal distribution

Key features of the lognormal distribution

$$X \sim lornorm(\mu, \sigma)$$

$$E(X) = exp(\mu + \frac{\sigma^2}{2}) = e^{\mu + \frac{\sigma^2}{2}}$$

$$X_{median} = e^{\mu}$$

And hence we can calculate from average income per capita:

$$\mu = ln(inccap) - \frac{\sigma^2}{2}$$

Data

This is a dataset that uses sloppy averaging from World Development Indicators for proof of concept (world Bank 2014).

Var	Data	Notes
iso2c	2-letter abbreviation	Provided by the Bank, supposedly ISO format.
country	Country name	
year	Year	Grossly inaccurate!
income	National income	Current US$
incap	Per capita income	Current US$
gini	Gini index	A pure number on the (0,1 interval)
saving	National saving	Current US$
savrate	Saving rate	Percent of GNI

This is a summary of the most recent data available.

```
## Classes 'data.table' and 'data.frame':   213 obs. of  9 variables:
##  $ iso2c     : chr  "AD" "AE" "AF" "AG" ...
##  $ country   : chr  "Andorra" "United Arab Emirates" "Afghanistan" "Antigua and Barbuda"
##  $ year      : num  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
##  $ income    : num  2.54e+09 2.13e+11 6.29e+09 9.60e+08 1.09e+10 ...
##  $ inccap    : num  38413 35600 590 12553 4493 ...
##  $ gini      : num  NaN NaN NaN NaN 29 ...
##  $ saving    : num  NaN NaN -2.98e+09 2.48e+08 2.01e+09 ...
##  $ savingrate: num  NaN NaN -16.5 22.3 16.4 ...
##  $ savers    : num  NaN 19.16 2.82 NaN 8.56 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
##     iso2c              country                year          income
##  Length:213         Length:213         Min.   :2010   Min.   :3.74e+07
##  Class :character   Class :character   1st Qu.:2010   1st Qu.:3.49e+09
##  Mode  :character   Mode  :character   Median :2010   Median :1.56e+10
```

```
##                                         Mean   :2010   Mean   :2.69e+11
##                                         3rd Qu.:2010   3rd Qu.:1.10e+11
##                                         Max.   :2010   Max.   :1.41e+13
##                                                        NA's   :18
##      inccap            gini           saving             savingrate
##  Min.   :   220   Min.   :24.8   Min.   :-2.98e+09   Min.   :-16.5
##  1st Qu.:  1513   1st Qu.:33.0   1st Qu.: 1.16e+09   1st Qu.: 13.9
##  Median :  5183   Median :37.4   Median : 6.70e+09   Median : 19.1
##  Mean   : 14880   Mean   :39.4   Mean   : 9.82e+10   Mean   : 21.6
##  3rd Qu.: 16483   3rd Qu.:44.1   3rd Qu.: 4.97e+10   3rd Qu.: 28.4
##  Max.   :161983   Max.   :65.8   Max.   : 3.65e+12   Max.   : 72.1
##  NA's   :14       NA's   :73     NA's   :43          NA's   :43
##      savers
##  Min.   : 0.12
##  1st Qu.: 5.59
##  Median :13.32
##  Mean   :18.32
##  3rd Qu.:24.17
##  Max.   :63.58
##  NA's   :66
```

□

Bibliography

Eric Kemp-Benedict. 2001. Income Distribution and Poverty Data: Methods for Using Available Data in Global Analysis. Polestar Technical Note No. 4.

World Bank. 2014. World Development Indicators.

Ben Mazzotta is a postdoc investigating the Cost of Cash at the Institute for Business in the Global Context (IBGC). Contact: benjamin.mazzotta@tufts.edu +1.617.627.4561 160 Packard Ave, Medford MA 02155.