Fitting Zenga's new model to real income data

L. Pasquazzi, Ma. Zenga

September 13, 2010

Our Purpose

Compare different methods for fitting Zenga's new model to empirical income distributions:

- Matching model characteristics with empirical counterparts
  - method of moments
  - D'Addario's invariants method
- Minimum distance methods
  - maximum likelihood
  - minimization of measures for the goodness of fit
- minimum distance methods with restrictions to preserve characteristics of the empirical distribution in the fitted model

Goodness of fit measures (I)

- We consider measures for the goodness of fit based on a histogram representation of the empirical income distribution.
- The endpoints of the 25 histogram-classes we used are quantiles of the observed incomes, which are choosen to give an accurate description of the empirical distribution in both tails. We also added the empirical mean as an endpoint, since its value is needed to compute many descriptive indexes.
- The goodness of fit measures are different kinds of averages of the relative deviations

$$\rho_j = \frac{n_j - \widehat{n}_j}{\widehat{n}_j} \quad \text{or} \quad \rho'_j = \frac{n_j - \widehat{n}_j}{n_j}$$

between observed (n_j) and model $(\widehat{n}_j)$ frequencies.

Goodness of fit measures (II)

- Thus, we have

$$
\begin{aligned}
A_1 &= \frac{1}{n}\sum_{j=1}^{s}\frac{|n_j-\widehat{n}_j|}{\widehat{n}_j}\widehat{n}_j = \frac{1}{n}\sum_{j=1}^{s}\frac{|n_j-\widehat{n}_j|}{n_j}n_j = \frac{1}{n}\sum_{j=1}^{s}|n_j-\widehat{n}_j| \\
A_2 &= \left\{\frac{1}{n}\sum_{j=1}^{s}\left|\frac{n_j-\widehat{n}_j}{\widehat{n}_j}\right|^2\widehat{n}_j\right\}^{1/2} = \left\{\frac{1}{n}\sum_{j=1}^{s}\frac{|n_j-\widehat{n}_j|^2}{\widehat{n}_j}\right\}^{1/2} \\
A_2' &= \left\{\frac{1}{n}\sum_{j=1}^{s}\left|\frac{n_j-\widehat{n}_j}{n_j}\right|^2 n_j\right\}^{1/2} = \left\{\frac{1}{n}\sum_{j=1}^{s}\frac{|n_j-\widehat{n}_j|^2}{n_j}\right\}^{1/2}.
\end{aligned}
$$

Method of Moments

- According to this method the parameters of the fitted model are obtained equating emprical and model moments.
- In Zenga's model, the r-th order moment is finite if and only if $r < \alpha + 1$. In this case it is given by

$$\mu_r = \frac{\mu^r}{2r-1} \frac{1}{B(\alpha;\theta)} \sum_{i=1}^{2r-1} B(\alpha - r + i; \theta).$$

  If $\theta > 1$, this formula simplifies to

$$\mu_r = \frac{\mu^r}{2r-1} \frac{1}{B(\alpha;\theta)} \left[B(\alpha - r + 1; \theta - 1) - B(\alpha + r; \theta - 1) \right].$$

- Since the empirical third moment is always finite, the method of moments can only yield values of $\alpha > 2$.

Method of Moments: fitted models

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A'_2
CH 00	3628	6764	2.2639	3.3526	0.18814	0.23478	0.31359
CH 01	3729	6706	2.4945	2.4638	0.14328	0.17825	0.17144
CH 02	3709	6915	2.8680	3.1782	0.10322	0.13110	0.12862
CH 03	3454	6786	4.8198	5.7336	0.15884	0.24503	0.19554
CH 04	3249	6560	3.9073	4.4095	0.09777	0.18888	0.14259
CH 05	3071	6784	4.0210	4.8195	0.13758	0.17274	0.17007
I 06 eq	7762	19121	2.1939	3.6394	0.24793	0.29701	0.41231
I 06	7762	31919	2.4447	4.0653	0.12317	0.16592	0.20295
I CE 06	1574	21543	2.1880	5.9061	0.65516	0.73107	1.22580
I CE 06	1574	35596	2.3365	4.2729	0.21665	0.28156	0.41298
I NW 06	1987	21572	2.7120	3.0018	0.15727	0.19938	0.23182
I NW 06	1987	34693	2.5007	3.6521	0.13266	0.18971	0.24103
I S 06	1618	13666	2.6086	3.6902	0.16516	0.23363	0.29830
I S 06	1618	24550	2.4047	4.2064	0.15111	0.20619	0.25150
I 06 ind	13401	18503	2.2678	4.6497	0.25900	0.32080	0.29740
US 00	5199184	56453	3.7480	11.0614	0.27891	0.68294	0.33490
US 08	2899458	82460	3.9922	10.7071	0.26766	0.59313	0.32365
UK 99/00	7086	446	2.8227	5.6448	0.13566	0.19394	0.24924

D'Addario's Method of Invariants

- Matching model characteristics - which are functions of the parameters - with counterparts in the empirical distribution.
- D'Addario proposed to use those characteristics that are actually employed to describe income distributions like the arithmetic mean, the median, the Gini inequality ratio, the Pietra index, etc.
- We consider the following characteristics
  a) Mean - Median - Pietra index
  b) Mean - Median - Zenga index

D'Addario's Method of Invariants

invariants	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
Italian individual income 2006 (13401 observations)						
μ, med, P	18503	1.9697	2.8967	0.0731	0.1827	0.1168
μ, med, Z	18503	1.9694	2.8962	0.0731	0.1826	0.1167
Italian household income 2006 (7762 observations)						
μ, med, P	31919	3.3995	5.1556	0.06632	0.07682	0.07911
μ, med, Z	31919	3.3959	5.1500	0.06637	0.07682	0.07911
Swiss household income 2000 (3628 observations)						
μ, med, P	6764	2.8514	3.3605	0.06574	0.11195	0.11924
μ, med, Z	6764	2.8507	3.3596	0.06576	0.11196	0.11928
Swiss household income 2005 (3071 observations)						
μ, med, P	6784	1.8587	1.9291	0.21440	0.26323	0.27810
μ, med, Z	6784	1.8390	1.9112	0.21603	0.26568	0.28253
US household income 2000 (5199184 observations)						
μ, med, P	56453	1.3836	2.8887	0.0735	0.1050	0.1133
μ, med, Z	56453	1.3755	2.8712	0.0739	0.1055	0.1165
US household income 2008 (2899458 observations)						
μ, med, P	82460	1.4456	2.7929	0.07062	0.08628	0.08804
μ, med, Z	82460	1.4374	2.7767	0.07104	0.08726	0.08932
UK equivalent gross income 99/00 (7086 observations)						
μ, med, P	446	2.9344	5.3059	0.11272	0.17232	0.18890
μ, med, Z	446	2.9335	5.3046	0.11275	0.17237	0.18899

Minimum A_1 method and Maximum Likelihood

method	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
Italian individual income 2006 (13401 observations)						
μ, med, P	18503	1.9697	2.8967	0.0731	0.1827	0.1168
μ, med, Z	18503	1.9694	2.8962	0.0731	0.1826	0.1167
min A_1	18652	2.0349	2.9558	0.06980	0.20330	0.11750
ML	18573	1.6916	2.4545	0.0914	0.1376	0.1217
Swiss household income 2000 (3628 observations)						
μ, med, P	6764	2.8514	3.3605	0.06574	0.11195	0.11924
μ, med, Z	6764	2.8507	3.3596	0.06576	0.11196	0.11928
min A_1	6754	3.1849	3.7692	0.06103	0.11609	0.11609
ML	6724	2.9740	3.4348	0.0654	0.1113	0.1125
Swiss household income 2005 (3071 observations)						
μ, med, P	6784	1.8587	1.9291	0.21440	0.26323	0.27810
μ, med, Z	6784	1.8390	1.9112	0.21603	0.26568	0.28253
min A_1	7155	2.8340	3.4984	0.10044	0.15147	0.18499
ML	6910	3.0855	3.6550	0.1099	0.1427	0.1526
US household income 2000 (5199184 observations)						
μ, med, P	56453	1.3836	2.8887	0.0735	0.1050	0.1133
μ, med, Z	56453	1.3755	2.8712	0.0739	0.1055	0.1165
min A_1	59371	1.4497	3.3310	0.04140	0.09400	0.11280
ML	57894	1.4772	3.2883	0.0505	0.0911	0.1036
US household income 2008 (2899458 observations)						
μ, med, P	82460	1.4456	2.7929	0.07062	0.08628	0.08804
μ, med, Z	82460	1.4374	2.7767	0.07104	0.08726	0.08932
min A_1	85925	1.5266	3.2534	0.03958	0.06912	0.07550
ML	84449	1.5625	3.2262	0.0458	0.0648	0.0676
UK equivalent gross income 99/00 (7086 observations)						
μ, med, P	446	2.9344	5.3059	0.11272	0.17232	0.18890
μ, med, Z	446	2.9335	5.3046	0.11275	0.17237	0.18899
min A_1	446	3.4541	6.1752	0.10221	0.15182	0.15508
ML	434	4.2329	7.7547	0.1102	0.1349	0.1414

Minimum A_1 and Maximum Likelihood

method	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1
	Italian individual income 2006 13401 observations				**Italian household income 2006** 7762 observations			
μ, med, P	18503	1.9697	2.8967	0.0731	31919	3.3995	5.1556	0.0663
μ, med, Z	18503	1.9694	2.8962	0.0731	31919	3.3959	5.1500	0.0664
min A_1	18652	2.0349	2.9558	0.0698	31301	4.2575	6.3816	0.0608
ML	18573	1.6916	2.4545	0.0914	31475	3.3967	4.9924	0.0644
	Swiss household income 2000 3628 observations				**Swiss household income 2005** 3071 observations			
μ, med, P	6764	2.8514	3.3605	0.0657	6784	1.8587	1.9291	0.2144
μ, med, Z	6764	2.8507	3.3596	0.0658	6784	1.8390	1.9112	0.2160
min A_1	6754	3.1849	3.7692	0.0610	7155	2.8340	3.4984	0.1004
ML	6724	2.9740	3.4348	0.0654	6910	3.0855	3.6550	0.1099
	US household income 2000 5199184 observations				**US household income 2008** 2899458 observations			
μ, med, P	56453	1.3836	2.8887	0.0735	82460	1.4456	2.7929	0.0706
μ, med, Z	56453	1.3755	2.8712	0.0739	82460	1.4374	2.7767	0.0710
min A_1	59371	1.4497	3.3310	0.0414	85925	1.5266	3.2534	0.0396
ML	57894	1.4772	3.2883	0.0505	84449	1.5625	3.2262	0.0458

Minimum A_1 preserving empirical mean

method	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1
	Italian individual income 2006				**Italian household income 2006**			
	13401 observations				7762 observations			
μ, med, P	18503	1.9697	2.8967	0.0731	31919	3.3995	5.1556	0.0663
μ, med, Z	18503	1.9694	2.8962	0.0731	31919	3.3959	5.1500	0.0664
min A_1	18652	2.0349	2.9558	0.0698	31301	4.2575	6.3816	0.0608
ML	18573	1.6916	2.4545	0.0914	31475	3.3967	4.9924	0.0644
min $A_1 \vert \mu$	18503	2.0015	2.9215	0.0715	31919	3.8584	5.8840	0.0629
	Swiss household income 2000				**Swiss household income 2005**			
	3628 observations				3071 observations			
μ, med, P	6764	2.8514	3.3605	0.0657	6784	1.8587	1.9291	0.2144
μ, med, Z	6764	2.8507	3.3596	0.0658	6784	1.8390	1.9112	0.2160
min A_1	6754	3.1849	3.7692	0.0610	7155	2.8340	3.4984	0.1004
ML	6724	2.9740	3.4348	0.0654	6910	3.0855	3.6550	0.1099
min $A_1 \vert \mu$	6764	3.1451	3.7117	0.0612	6784	3.0513	3.5553	0.1145
	US household income 2000				**US household income 2008**			
	5199184 observations				2899458 observations			
μ, med, P	56453	1.3836	2.8887	0.0735	82460	1.4456	2.7929	0.0706
μ, med, Z	56453	1.3755	2.8712	0.0739	82460	1.4374	2.7767	0.0710
min A_1	59371	1.4497	3.3310	0.0414	85925	1.5266	3.2534	0.0396
ML	57894	1.4772	3.2883	0.0505	84449	1.5625	3.2262	0.0458
min $A_1 \vert \mu$	56453	1.4140	3.1376	0.0609	82460	1.5375	3.1801	0.0521

Minimum A_1 methods

- Lower values of the A_1 index for the goodness of fit may be achieved by minimizing A_1 with respect to the model parameters.
- Restrictions on the parameter space to preserve some characteristics (mean, median, an inequality index) of the empirical distribution in the fitted model.

Minimum A_1 with two restrictions

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
I 06							
med	7762	31919	3.8348	5.7891	0.06301	0.08836	0.08667
$F(\mu)$	7762	31919	3.7855	5.5456	0.06396	0.08991	0.08543
P	7762	31919	3.8696	5.9613	0.06325	0.09131	0.09143
A	7762	31919	3.8683	5.9595	0.06326	0.09125	0.09140
CH 05							
med	3071	6784	2.7099	2.6892	0.18106	0.21209	0.20118
$F(\mu)$	3071	6784	5.3561	4.4032	0.29720	0.62936	0.35435
P	3071	6784	3.1243	3.5665	0.11974	0.15132	0.15119
A	3071	6784	3.0936	3.5399	0.11848	0.15102	0.15105
US 08							
med	2899458	82460	1.5263	2.9493	0.06717	0.08066	0.08021
$F(\mu)$	2899458	82460	1.9966	3.5227	0.11461	0.20097	0.15979
P	2899458	82460	1.6763	3.3661	0.05993	0.07822	0.07578
A	2899458	82460	1.6713	3.3586	0.05965	0.07753	0.07528
UK 99/00							
med	7086	446	3.4837	6.2750	0.10280	0.14971	0.15467
$F(\mu)$	7086	446	3.5395	6.4789	0.10432	0.14702	0.15497
P	7086	446	3.4700	6.4187	0.10644	0.14879	0.16017
A	7086	446	3.4699	6.4186	0.10644	0.14880	0.16017

The first restriction is always put on the mean.

Minimum A_1 with restriction on the mean and unrestricted

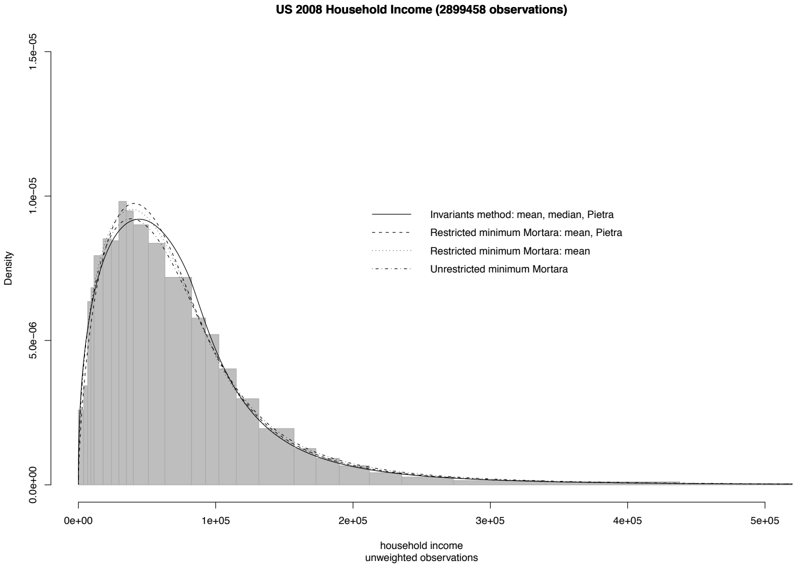
Restriction on mean

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
I 06	7762	31919	3.8584	5.8840	0.06293	0.08975	0.08889
I 06 ind	13401	18503	2.0015	2.9215	0.07150	0.19150	0.11680
I NW 06	1987	34693	3.6027	4.8350	0.08780	0.12168	0.13009
I S 06	1618	24550	3.8224	5.9336	0.11234	0.21542	0.14959
CH 00	3628	6764	3.1451	3.7117	0.06116	0.11502	0.11540
CH 05	3071	6784	3.0513	3.5553	0.11453	0.14937	0.15098
UK 99/00	7086	446	3.4562	6.1753	0.10225	0.15147	0.15492
US 00	5199184	56453	1.4140	3.1376	0.06090	0.09760	0.11390
US 08	2899458	82460	1.5375	3.1801	0.05213	0.07265	0.07462

Unrestricted

dataset	n	$\widehat{\mu}$	$\widehat{\alpha}$	$\widehat{\theta}$	A_1	A_2	A_2'
I 06	7762	31301	4.2575	6.3816	0.06083	0.10927	0.09841
I 06 ind u	13401	18652	2.0349	2.9558	0.06980	0.20330	0.11750
I NW 06	1987	33924	3.9262	5.1371	0.08633	0.12302	0.12476
I S 06	1618	24207	3.8902	6.0044	0.11115	0.22039	0.15230
CH 00	3628	6754	3.1849	3.7692	0.06103	0.11609	0.11609
CH 05	3071	7155	2.8340	3.4984	0.10044	0.15147	0.18499
UK 99/00	7086	446	3.4541	6.1752	0.10221	0.15182	0.15508
US 00	5199184	59371	1.4497	3.3310	0.04140	0.09400	0.11280
US 08	2899458	85925	1.5266	3.2534	0.03958	0.06912	0.07550

US 2008 Household Income



Maximum Likelihood Estimation

- Fit Zenga's model by maximizing the *likelihood function*

$$L(\mu, \alpha, \theta) \propto \prod_{j=1}^{k} \widehat{n}_j^{\,n_j},$$

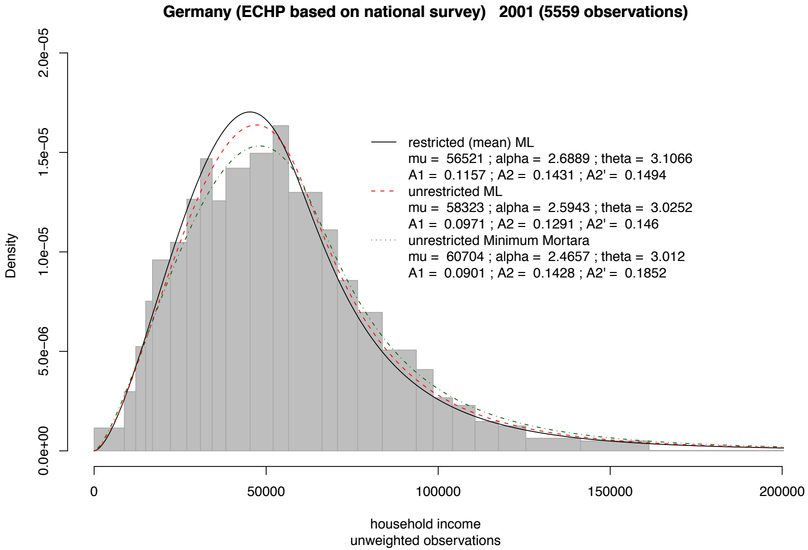
where $\widehat{n}_j = \widehat{n}_j(\mu, \alpha, \theta)$.

- Compare restricted maximum (to preserve empirical mean) with unrestricted maximum.
- Assess goodness of fit by A_1, A_2 and A_2'.

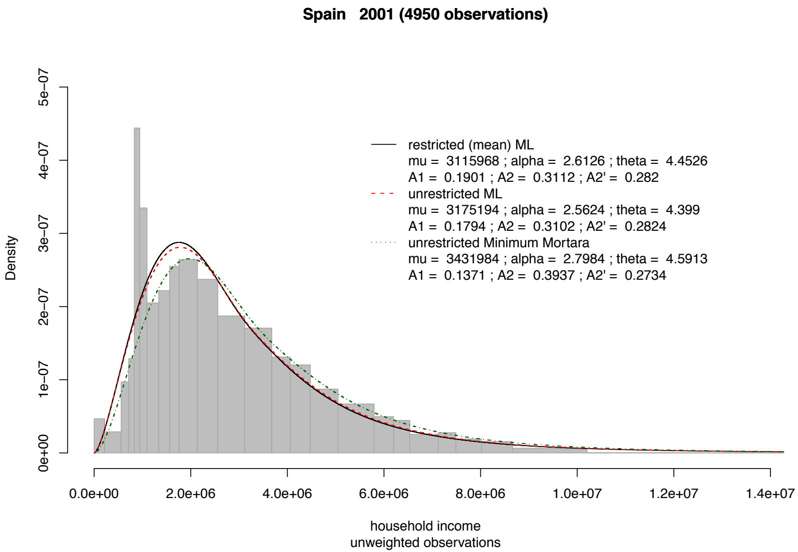
Maximum Likelihood - EHCP 2001 data

country	n	μ	α	θ	A_1	A_2	A_2'	χ^2
				restricted maximum likelihood				
D	5559	56521	2,689	3,107	0,116	0,143	0,149	113,8
F	5268	175603	2,772	3,958	0,105	0,173	0,164	157,3
UK	4779	21006	2,440	3,886	0,170	0,193	0,194	177,8
I	5525	40183	2,566	3,859	0,132	0,203	0,181	226,8
E	4950	3115968	2,613	4,453	0,190	0,311	0,282	479,4
Sw	5085	249974	2,439	2,935	0,196	0,230	0,232	269,9
				unrestricted maximum likelihood				
D	5559	58323	2,5943	3,0252	0,0971	0,1291	0,146	92,7
F	5268	180310	2,6971	3,8904	0,0962	0,1646	0,1680	142,7
UK	4779	21848	2,3063	3,6962	0,1479	0,1832	0,1945	160,4
I	5525	41603	2,4704	3,7626	0,1257	0,1909	0,1849	201,3
E	4950	3175194	2,5624	4,3990	0,1794	0,3102	0,2824	476,3
Sw	5085	263358	2,2083	2,6560	0,1657	0,2074	0,2307	218,7

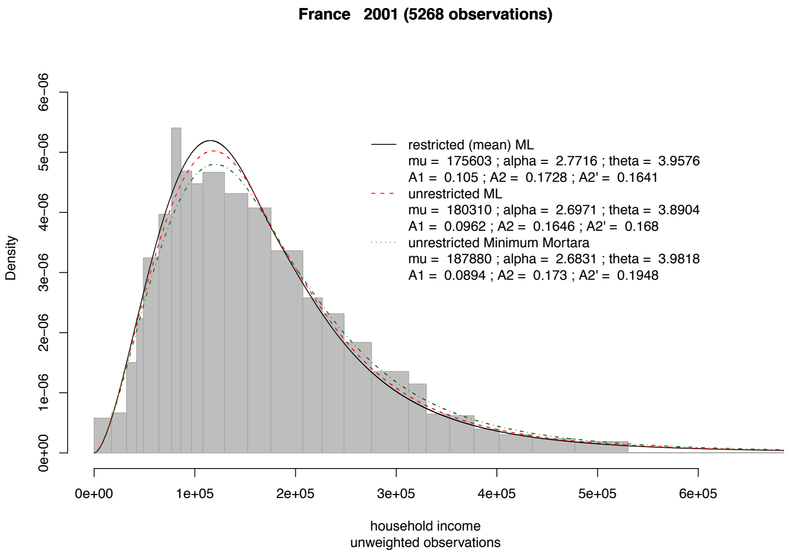
Maximum Likelihood - EHCP 2001 data



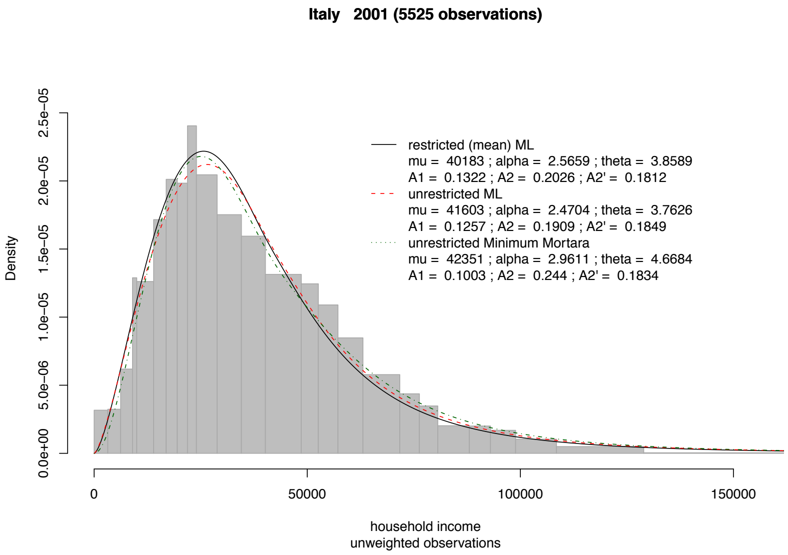
Maximum Likelihood - EHCP 2001 data



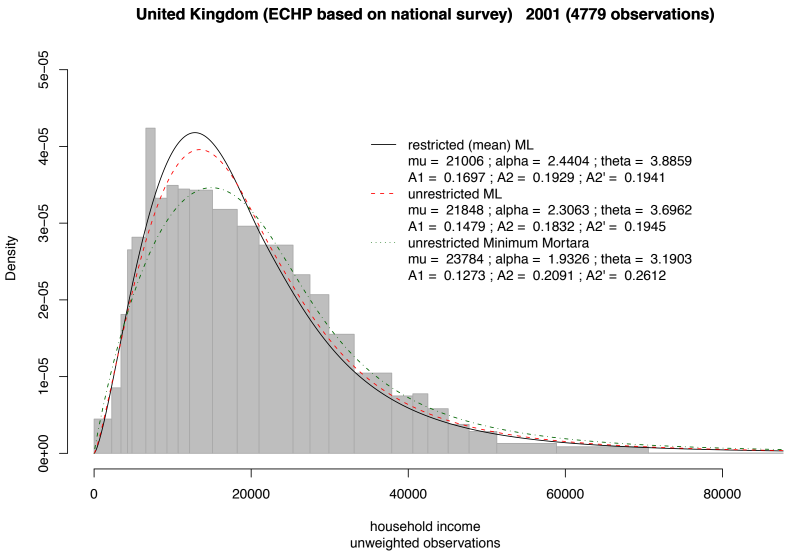
Maximum Likelihood - EHCP 2001 data



Maximum Likelihood - EHCP 2001 data



Maximum Likelihood - EHCP 2001 data



Conclusions

- Overall, the new model provides a good fit to empirical income distributions
- In the fitted model, we usually get $2 < \alpha < 4$ and $2 < \theta < 6$.
- We may expect to achieve values of Mortara's index between $0,04$ and $0,12$.
- We strongly recommend not to use the method of moments. D'Addario's method provides much lower values of Mortara's A_1 index, when we match mean, median and a inequality measure.
- Lower values of the A_1-index may be achieved, by minimizing its value with respect to the model parameters. However, this may lead to a considerable difference between empirical and model mean.
- Minimizing Mortara's index restricting the model mean to match the empirical mean seems to provide a good compromise between D'Addadario's method and minimization of Mortara's A_1-index.