Risk and Rationality: Uncovering Heterogeneity in Probability Distortion

Supplementary Materials

February 1, 2010

1 Estimation of the Finite Mixture Model

As it is generally the case in finite mixture models, direct maximization of the log likelihood function

$$\ln L\left(\Psi; ce, \mathcal{G}\right) = \sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c \, f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right)$$

may encounter several problems, even if it is in principle feasible (for a general treatise see for example McLachlan and Peel (2000)). First, the highly non-linear form of the log likelihood causes the optimization algorithm to be rather slow or even incapable of finding the maximum. Second, the likelihood of a finite mixture model is often multimodal and therefore we have no guaranty that a standard optimization routine will converge towards the global maximum rather than to one of the local maxima.

However, if individual group-membership were observable and indicated by $t_{ic} \in \{0,1\}$ the individual contribution to the likelihood function would be given by

$$\tilde{\ell}\left(\Psi_i; ce_i, \mathcal{G}, t_i\right) = \prod_{c=1}^{C} \left[\pi_c \, f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right)\right]^{t_{ic}}$$

By using the above formulation and taking logarithms, the complete-data log likelihood function

$$\ln \tilde{L}\left(\Psi; ce, \mathcal{G}, t\right) = \sum_{i=1}^{N} \sum_{c=1}^{C} t_{ic} \left[\ln \pi_c + \ln f\left(ce_i, \mathcal{G}; \theta_c, \xi_i\right)\right]$$

would follow directly. As relative group sizes sum up to one, their maximum likelihood estimates, $\hat{\pi}_c = 1/N \sum_{i=1}^{N} t_{ic}$, would be given analytically by the relative number of individuals in the respective group. Furthermore, the maximum likelihood estimates of the group-specific parameters could be obtained separately in each group by numerically maximizing the corresponding joint density function which would simplify the optimization problem considerably.

The EM algorithm proceeds iteratively in two steps, E and M, while it treats the unobservable t_{ic} as missing data. In the E-step of the $(k+1)$-th iteration the expectation of the complete-data log likelihood $\tilde{L}$, given the actual fit of the data $\Psi^{(k)}$, is computed. This yields, according to Bayes' law, the posterior probabilities of individual group-membership

$$\tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right) = \frac{\pi_c^{(k)} f\left(ce_i, \mathcal{G}; \theta_c^{(k)}, \xi_i^{(k)}\right)}{\sum_{m=1}^{C} \pi_m^{(k)} f\left(ce_i, \mathcal{G}; \theta_m^{(k)}, \xi_i^{(k)}\right)}$$

which replace the unknown indicators of individual group-membership, t_{ic}. Given $\tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right)$, the complete-data log likelihood, $\tilde{L}$, is maximized in the following M-step which yields the updates of the model parameters,

$$\pi_c^{(k+1)} = \frac{1}{N}\sum_{i=1}^{N} \tau_{ic}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right),$$

and

$$\left(\theta_1^{(k+1)}, \ldots, \theta_C^{(k+1)}, \xi_1^{(k+1)}, \ldots, \xi_N^{(k+1)}\right) = \\ \underset{\theta_1,\ldots,\theta_C,\xi_1,\ldots,\xi_N}{\arg\max} \sum_{i=1}^{N}\sum_{m=1}^{C} \tau_{im}\left(ce_i, \mathcal{G}; \Psi_i^{(k)}\right) \ln f\left(ce_i, \mathcal{G}; \theta_m^{(k)}, \xi_i^{(k)}\right).$$

As Dempster, Laird, and Rubin (1977) show, the likelihood never decreases from one iteration to the next, i.e. $L\left(\Psi^{(k+1)}; ce, \mathcal{G}\right) \geq L\left(\Psi^{(k)}; ce, \mathcal{G}\right)$, which makes the EM algorithm converge monotonically towards the nearest maximum of the likelihood function regardless whether this maximum is global or just local. In the Zurich 2003 data set, we therefore needed to apply a stochastic extension, the Simulated Annealing Expectation Maximization (SAEM) algorithm proposed by Celeux, Chauveau, and Diebolt (2001), in order to overcome the EM algorithm's tendency to converge towards local maxima. In each iteration, there is a non-zero probability that the SAEM algorithm leaves the current optimization path and starts over in a different region of the likelihood function which results in much higher chances of finding the global maximum. But this robustness against multimodality of the objective function comes at the cost of much higher computational demands.

As the EM algorithm is computationally highly demanding, even in its basic form, and tends to become tediously slow when close to convergence our estimation routine relies on a hybrid estimation algorithm (Render and Walker, 1984): It first uses either the EM or the SAEM algorithm and takes advantage of their robustness before it switches to the direct maximization of the log likelihood by the much faster BFGS algorithm. The estimation routine in this form turned out to be efficient and robust as it reliably converged towards the same maximum likelihood estimates regardless of the randomly chosen start values.

2 Aggregate Behavior

Table 1: Single-Component Models

	Gains			Losses		
Parameters	ZH 03	ZH 06	BJ 05	ZH 03	ZH 06	BJ 05
α/β	1.041	0.916	0.443	1.077	1.093	1.131
	(0.021)	(0.021)	(0.116)	(0.025)	(0.036)	(0.123)
γ	0.482	0.519	0.318	0.487	0.579	0.383
	(0.010)	(0.017)	(0.016)	(0.012)	(0.027)	(0.015)
δ	0.869	0.886	1.296	1.030	1.039	0.944
	(0.020)	(0.022)	(0.081)	(0.026)	(0.033)	(0.062)
$\ln L$	19,563	10,671	9,550			
Parameters	364	242	308			
Individuals	179	118	151			
Observations	8,906	4,669	4,225			

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications. ZH stands for Zurich, BJ for Beijing.

3 Classification and Demographics

Table 2: Summary Statistics for Demographic Variables

Zurich 03		**Mean**	**Std. Err.**
Individuals	179		
female		0.430	0.037
semester		3.676	0.159
highincome		0.162	0.028
Zurich 06		**Mean**	**Std. Err.**
Individuals	118		
female		0.441	0.046
semester		3.551	0.240
highincome		0.051	0.020
Beijing 05		**Mean**	**Std. Err.**
Individuals	151		
female		0.483	0.041
semester		2.238	0.133
highincome		0.146	0.029

highincome: equals one if disposable income
per month above is CHF 1,500 and CHN 1,000, respectively.
Thresholds chosen by distributional considerations and
relative students' hourly wages.

Table 3: Classification of Behavior with $C = 3$, Pooled: Men

	Gains				Losses		
	EUT	*CPT-I*	*CPT-II*		*EUT*	*CPT-I*	*CPT-II*
π	0.182	0.333	0.485				
	(0.014)	(0.022)	(0.025)				
α	0.981	0.925	0.988	β	1.018	1.280	1.066
	(0.015)	(0.035)	(0.028)		(0.028)	(0.099)	(0.082)
γ	0.963	0.260	0.505	γ	0.970	0.285	0.543
	(0.044)	(0.117)	(0.108)		(0.040)	(0.124)	(0.117)
δ	0.908	0.896	0.993	δ	1.078	0.963	0.956
	(0.016)	(0.052)	(0.046)		(0.023)	(0.033)	(0.026)
$\ln L$			24,114				
Parameters			512				
Individuals			246				
Observations			9,874				

Standard errors (in parentheses) are based on the bootstrap method with 2,000 replications.
Parameters include estimates of ξ_i for domain- and individual-specific error variances.

Table 4: Classification of Behavior with $C = 3$, Pooled: Women

	Gains				**Losses**		
	EUT	*CPT-I*	*CPT-II*		*EUT*	*CPT-I*	*CPT-II*
π	0.240	0.369	0.391				
	(0.038)	(0.028)	(0.031)				
α	0.936	0.967	0.914	β	1.159	1.186	1.296
	(0.032)	(0.049)	(0.045)		(0.069)	(0.081)	(0.088)
γ	0.780	0.317	0.281	γ	0.714	0.327	0.312
	(0.092)	(0.049)	(0.031)		(0.102)	(0.043)	(0.027)
δ	0.925	1.153	0.686	δ	0.960	0.748	1.264
	(0.045)	(0.254)	(0.216)		(0.069)	(0.305)	(0.236)
$\ln L$			18,213				
Parameters			424				
Individuals			202				
Observations			7,926				

Standard errors (in parentheses) are based on the bootstrap method with 2,000 replications.
Parameters include estimates of ξ_i for domain- and individual-specific error variances.

Table 5: Effects of Socio-Economic Variables on Parameters

	Gains			Losses		
Regressors	ZH 03	ZH 06	BJ 05	ZH 03	ZH 06	BJ 05
α/β						
constant	1.101**	0.935**	0.538**	1.075**	1.049**	1.553**
	(0.051)	(0.039)	(0.189)	(0.061)	(0.047)	(0.373)
female	-0.008	-0.041	-0.424	0.103	0.136	-0.347
	(0.042)	(0.044)	(0.325)	(0.069)	(0.069)	(0.351)
semester	-0.016	0.002	0.096	-0.009	-0.006	-0.095
	(0.012)	(0.006)	(0.091)	(0.013)	(0.008)	(0.106)
highincome	-0.024	-0.049	-0.436	0.078	0.064	-0.450
	(0.059)	(0.112)	(0.251)	(0.085)	(0.126)	(0.387)
γ						
constant	0.434**	0.562**	0.374**	0.472**	0.746**	0.454**
	(0.037)	(0.057)	(0.025)	(0.037)	(0.063)	(0.035)
female	**-0.143****	**-0.186****	**-0.113****	**-0.149****	**-0.324****	**-0.112****
	(0.022)	(0.057)	(0.031)	(0.026)	(0.054)	(0.036)
semester	0.031**	0.023	0.001	0.019	0.011*	0.001
	(0.012)	(0.010)	(0.009)	(0.011)	(0.005)	(0.015)
highincome	0.204**	-0.110	-0.007	0.002	-0.051	-0.046
	(0.079)	(0.098)	(0.034)	(0.071)	(0.070)	(0.033)
δ						
constant	0.848**	0.945**	1.295**	1.008**	0.990**	0.754**
	(0.051)	(0.042)	(0.125)	(0.068)	(0.047)	(0.176)
female	-0.147**	-0.134**	0.195	0.091	0.021	0.186
	(0.041)	(0.045)	(0.227)	(0.074)	(0.065)	(0.172)
semester	0.021	-0.001	-0.062	-0.001	0.008	0.038
	(0.013)	(0.006)	(0.063)	(0.014)	(0.006)	(0.053)
highincome	-0.072	-0.064	0.214	-0.059	-0.016	0.227
	(0.060)	(0.123)	(0.185)	(0.084)	(0.156)	(0.238)
$\ln L$	19,755	10,816	9,601			
Parameters	382	260	326			
Observations	8,906	4,669	4,225			

Standard errors (in parentheses) are based on the bootstrap method with 4,000 replications.
** Significant at 1%-level; * significant at 5%-level. ZH stands for Zurich, BJ for Beijing.

References

Celeux, G., D. Chauveau, and J. Diebolt (2001): "On Stochastic Versions of the EM Algorithm," *Biometrica*, 88(1), 281–286.

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

McLachlan, G., and D. Peel (2000): *Finite Mixture Models*. Wiley Series in Probabilities and Statistics.

Render, R. A., and H. F. Walker (1984): "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review*, 26(2), 195–239.