*Econometrica Supplementary Material*

# SUPPLEMENT TO "RISK AND RATIONALITY: UNCOVERING HETEROGENEITY IN PROBABILITY DISTORTION"

(*Econometrica*, Vol. 78, No. 4, July 2010, 1375–1412)

BY ADRIAN BRUHIN, HELGA FEHR-DUDA, AND THOMAS EPPER

## S1. ESTIMATION OF THE FINITE MIXTURE MODEL

AS IT IS GENERALLY THE CASE in finite mixture models, direct maximization of the log likelihood function

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

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 nonlinear 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 toward 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

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

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

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

would follow directly. As relative group sizes sum up to 1, 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,

 DOI: 10.3982/ECTA7139

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

$$\tau_{ic}\big(\text{ce}_i, \mathcal{G}; \Psi_i^{(k)}\big) = \frac{\pi_c^{(k)} f(\text{ce}_i, \mathcal{G}; \theta_c^{(k)}, \xi_i^{(k)})}{\sum_{m=1}^{C} \pi_m^{(k)} f(\text{ce}_i, \mathcal{G}; \theta_m^{(k)}, \xi_i^{(k)})},$$

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

$$\pi_c^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} \tau_{ic}\big(\text{ce}_i, \mathcal{G}; \Psi_i^{(k)}\big)$$

and

$$\begin{aligned}&\big(\theta_1^{(k+1)}, \dots, \theta_C^{(k+1)}, \xi_1^{(k+1)}, \dots, \xi_N^{(k+1)}\big)\\ &\quad = \operatorname*{arg\,max}_{\theta_1,\dots,\theta_C,\xi_1,\dots,\xi_N} \sum_{i=1}^{N} \sum_{m=1}^{C} \tau_{im}\big(\text{ce}_i, \mathcal{G}; \Psi_i^{(k)}\big) \ln f\big(\text{ce}_i, \mathcal{G}; \theta_m^{(k)}, \xi_i^{(k)}\big).\end{aligned}$$

As Dempster, Laird, and Rubin (1977) showed, the likelihood never decreases from one iteration to the next, that is, $L(\Psi^{(k+1)}; \text{ce}, \mathcal{G}) \geq L(\Psi^{(k)}; \text{ce}, \mathcal{G})$, which makes the EM algorithm converge monotonically toward the nearest maximum of the likelihood function regardless of 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), to overcome the EM algorithm's tendency to converge toward local maxima. In each iteration, there is a nonzero 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 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 Broyden–Fletcher–Goldfarb–Shanno algorithm. The estimation routine in this form turned out to be efficient and robust as it reliably converged toward the same maximum likelihood estimates regardless of the randomly chosen start values.

## S2. AGGREGATE BEHAVIOR

TABLE S.I

SINGLE-COMPONENT MODELS[a]

| | Gains | | | Losses | | |
|---|---|---|---|---|---|---|
| Parameters | ZH 03 | ZH 06 | BJ 05 | ZH 03 | ZH 06 | BJ 05 |
| $\alpha/\beta$ | 1.041 | 0.916 | 0.443 | 1.077 | 1.093 | 1.131 |
| | (0.021) | (0.021) | (0.116) | (0.025) | (0.036) | (0.123) |
| $\gamma$ | 0.482 | 0.519 | 0.318 | 0.487 | 0.579 | 0.383 |
| | (0.010) | (0.017) | (0.016) | (0.012) | (0.027) | (0.015) |
| $\delta$ | 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 | 9550 | | | |
| Parameters | 364 | 242 | 308 | | | |
| Individuals | 179 | 118 | 151 | | | |
| Observations | 8906 | 4669 | 4225 | | | |

[a] Standard errors (in parentheses) are based on the bootstrap method with 4000 replications. ZH stands for Zurich; BJ stands for Beijing.

## S3. CLASSIFICATION AND DEMOGRAPHICS

TABLE S.II

SUMMARY STATISTICS FOR DEMOGRAPHIC VARIABLES[a]

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

[a] The variable *highincome* equals 1 if disposable income per month is above 1500 swiss francs and 1000 yuan, respectively. Thresholds chosen by distributional considerations and relative students' hourly wages.

TABLE S.III

CLASSIFICATION OF BEHAVIOR WITH $C = 3$, POOLED: MEN[a]

| | Gains | | | | Losses | | |
|---|---|---|---|---|---|---|---|
| | *EUT* | *CPT-I* | *CPT-II* | | *EUT* | *CPT-I* | *CPT-II* |
| $\pi$ | 0.182 | 0.333 | 0.485 | | | | |
| | (0.014) | (0.022) | (0.025) | | | | |
| $\alpha$ | 0.981 | 0.925 | 0.988 | $\beta$ | 1.018 | 1.280 | 1.066 |
| | (0.015) | (0.035) | (0.028) | | (0.028) | (0.099) | (0.082) |
| $\gamma$ | 0.963 | 0.260 | 0.505 | $\gamma$ | 0.970 | 0.285 | 0.543 |
| | (0.044) | (0.117) | (0.108) | | (0.040) | (0.124) | (0.117) |
| $\delta$ | 0.908 | 0.896 | 0.993 | $\delta$ | 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 | | | 9874 | | | | |

[a]Standard errors (in parentheses) are based on the bootstrap method with 2000 replications. Parameters include estimates of $\xi_i$ for domain- and individual-specific error variances.

TABLE S.IV

CLASSIFICATION OF BEHAVIOR WITH $C = 3$, POOLED: WOMEN[a]

| | Gains | | | | Losses | | |
|---|---|---|---|---|---|---|---|
| | *EUT* | *CPT-I* | *CPT-II* | | *EUT* | *CPT-I* | *CPT-II* |
| $\pi$ | 0.240 | 0.369 | 0.391 | | | | |
| | (0.038) | (0.028) | (0.031) | | | | |
| $\alpha$ | 0.936 | 0.967 | 0.914 | $\beta$ | 1.159 | 1.186 | 1.296 |
| | (0.032) | (0.049) | (0.045) | | (0.069) | (0.081) | (0.088) |
| $\gamma$ | 0.780 | 0.317 | 0.281 | $\gamma$ | 0.714 | 0.327 | 0.312 |
| | (0.092) | (0.049) | (0.031) | | (0.102) | (0.043) | (0.027) |
| $\delta$ | 0.925 | 1.153 | 0.686 | $\delta$ | 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 | | | 7926 | | | | |

[a]Standard errors (in parentheses) are based on the bootstrap method with 2000 replications. Parameters include estimates of $\xi_i$ for domain- and individual-specific error variances.

TABLE S.V

EFFECTS OF SOCIO-ECONOMIC VARIABLES ON PARAMETERS[a]

| | Gains | | | Losses | | |
|---|---|---|---|---|---|---|
| Parameters | ZH 03 | ZH 06 | BJ 05 | ZH 03 | ZH 06 | BJ 05 |
| $\alpha/\beta$ | | | | | | |
| *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) |
| $\gamma$ | | | | | | |
| *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) |
| $\delta$ | | | | | | |
| *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 | 9601 | | | |
| Parameters | 382 | 260 | 326 | | | |
| Observations | 8906 | 4669 | 4225 | | | |

[a]Standard errors (in parentheses) are based on the bootstrap method with 4000 replications.

**Significant at 1% level; *significant at 5% level. ZH stands for Zurich, BJ stands for Beijing.

## REFERENCES

CELEUX, G., D. CHAUVEAU, AND J. DIEBOLT (2001): "On Stochastic Versions of the EM Algorithm," *Biometrica*, 88, 281–286. [2]

DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977): "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 39, 1–38. [2]

MCLACHLAN, G., AND D. PEEL (2000): *Finite Mixture Models*. Wiley Series in Probabilities and Statistics. New York: Wiley. [1]

RENDER, R. A., AND H. F. WALKER (1984): "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review*, 26, 195–239. [2]

*Institute for Empirical Research in Economics, University of Zurich, Bluemlisalpstrasse 10, 8006 Zurich, Switzerland; bruhin@iew.uzh.ch,*

*Chair of Economics, ETH Zurich, Weinbergstrasse 35, 8092 Zurich, Switzerland; fehr@econ.gess.ethz.ch,*

*and*

*Chair of Economics, ETH Zurich, Weinbergstrasse 35, 8092 Zurich, Switzerland; epper@econ.gess.ethz.ch.*

*Manuscript received May, 2007; final revision received February, 2010.*