Shall We Mixed Logit?

Estimation stability and prediction reliability of error component mixed logit models

Shusaku NAKAI

Ryuichi KITAMURA

Kyoto University

Toshiyuki YAMAMOTO

Nagoya University

Outline

- Introduction
- Error component MXL models
  - Identification issue
  - Variability of parameter estimates
  - Estimation of choice probabilities
- Usefulness of MNL models
- Conclusions and future research

Introduction

MXL models

- considered the most promising discrete choice model
- widespread applications in recent years

However

- properties of parameter estimates are not well understood

Objective

- Estimation stability and prediction reliability of error component MXL models are examined with simulated data

Error component MXL models

- Examined is a trinomial MXL model

2 explanatory variables

Standard iid Gumbel

$$
\begin{cases}
u_{1n} = \beta_1 X_{11n} + \beta_2 X_{21n} + \mu_{1n} + \varepsilon_{1n} \\
u_{2n} = \beta_1 X_{12n} + \beta_2 X_{22n} + \mu_{2n} + \varepsilon_{2n} \\
u_{3n} = \beta_1 X_{13n} + \beta_2 X_{23n} + \mu_{2n} + \varepsilon_{3n}
\end{cases}
$$

2 error components

$$\mu_{1n} \sim N(0, s_1^{\,2})$$

$$\mu_{2n} \sim N(0, s_2^{\,2})$$

$$
\Sigma = \begin{pmatrix}
s_1^{\,2} + \frac{\pi^2}{6} & 0 & 0 \\
 & s_2^{\,2} + \frac{\pi^2}{6} & s_2^{\,2} \\
 & & s_2^{\,2} + \frac{\pi^2}{6}
\end{pmatrix}
$$

Simulated discrete choice data

Generated by a probit model

$$\begin{cases} u_{1n} = \beta_1 X_{11n} + \beta_2 X_{21n} + \xi_{1n} \\ u_{2n} = \beta_1 X_{12n} + \beta_2 X_{22n} + \xi_{2n} \\ u_{3n} = \beta_1 X_{13n} + \beta_2 X_{23n} + \xi_{3n} \end{cases}$$

$$\Sigma_\xi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}$$

$$\begin{cases} \beta_1 = 1.0 \\ \beta_2 = 0.5 \end{cases}$$

$$X_{jin} \sim N(0,1)$$

ρ = 0.00, 0.10, 0.30, 0.50, 0.70, 0.90, 0.95, 0.99
Each data set contains 1,000 cases
25 data sets are generated for each value of ρ

Identification issue

For trinomial probit models, Dansie (1985) suggests

$$\boxed{\mathbf{\Sigma}_A = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & 1 & \sigma_{23} \\ 0 & \sigma_{23} & 1 \end{bmatrix}} \quad \mathbf{\Sigma}_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \sigma_{23}' \\ 0 & \sigma_{23}' & 1 \end{bmatrix} \quad \mathbf{\Sigma}_C = \begin{bmatrix} \sigma_{11}'' & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- 3 matrices are equivalent, and produce the same likelihood value, thus Σ_A is not estimable
- Model estimation would not be able to indicate which is most likely

Identification issue (cont.)

For GEV models, Börsch-Supan (1990) and Munizaga et al. (2000) estimated in the case of 4 alternatives

Nested logit model

HEV model

$$\boldsymbol{\Sigma}_A = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & 1 & \sigma_{23} \\ 0 & \sigma_{23} & 1 \end{bmatrix} \quad \boldsymbol{\Sigma}_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \sigma_{23}' \\ 0 & \sigma_{23}' & 1 \end{bmatrix} \quad \boldsymbol{\Sigma}_C = \begin{bmatrix} \sigma_{11}'' & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- and found that nested logit models have some capacity to accommodate heteroscedasticity

Identification issue (cont.)

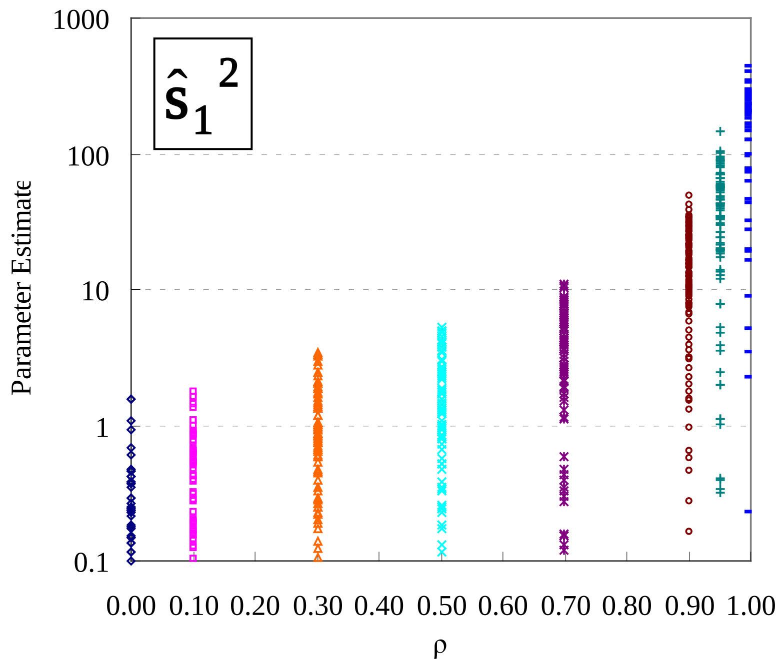
- In this study, data sets are simulated by Σ_B

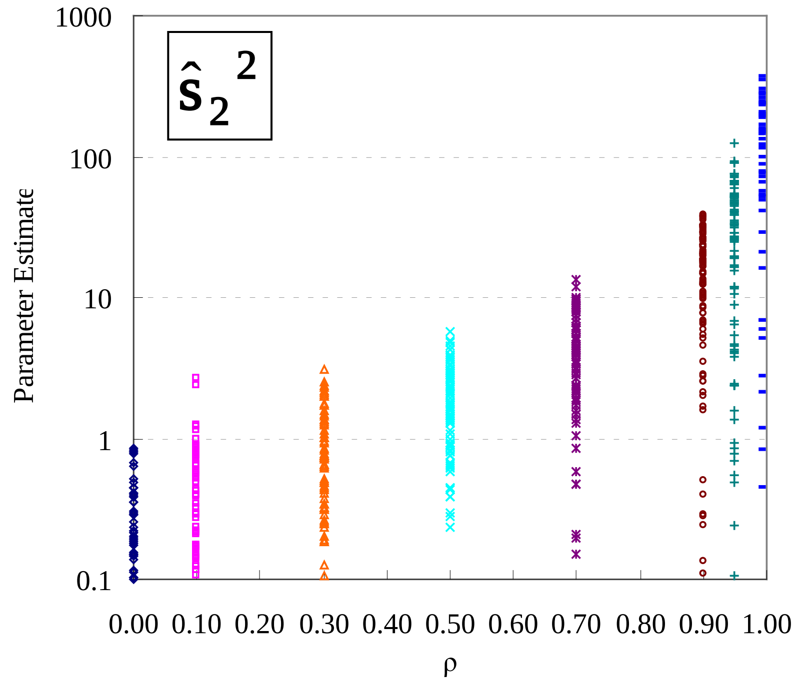
$$\boldsymbol{\Sigma}_A = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & 1 & \sigma_{23} \\ 0 & \sigma_{23} & 1 \end{bmatrix} \quad \boldsymbol{\Sigma}_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \sigma_{23}' \\ 0 & \sigma_{23}' & 1 \end{bmatrix} \quad \boldsymbol{\Sigma}_C = \begin{bmatrix} \sigma_{11}'' & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\Sigma = \begin{pmatrix} s_1{}^2 + \frac{\pi^2}{6} & 0 & 0 \\ & s_2{}^2 + \frac{\pi^2}{6} & s_2{}^2 \\ & & s_2{}^2 + \frac{\pi^2}{6} \end{pmatrix}$$

- Error component MXL model examined in this study is consistent with Σ_A

Identification issue (cont.)





- Standard deviation becomes extremely large, implying covariance structure is unidentified
- MXL model is subject to the same identification problem of probit model (consistent with Walker et al. (2007))

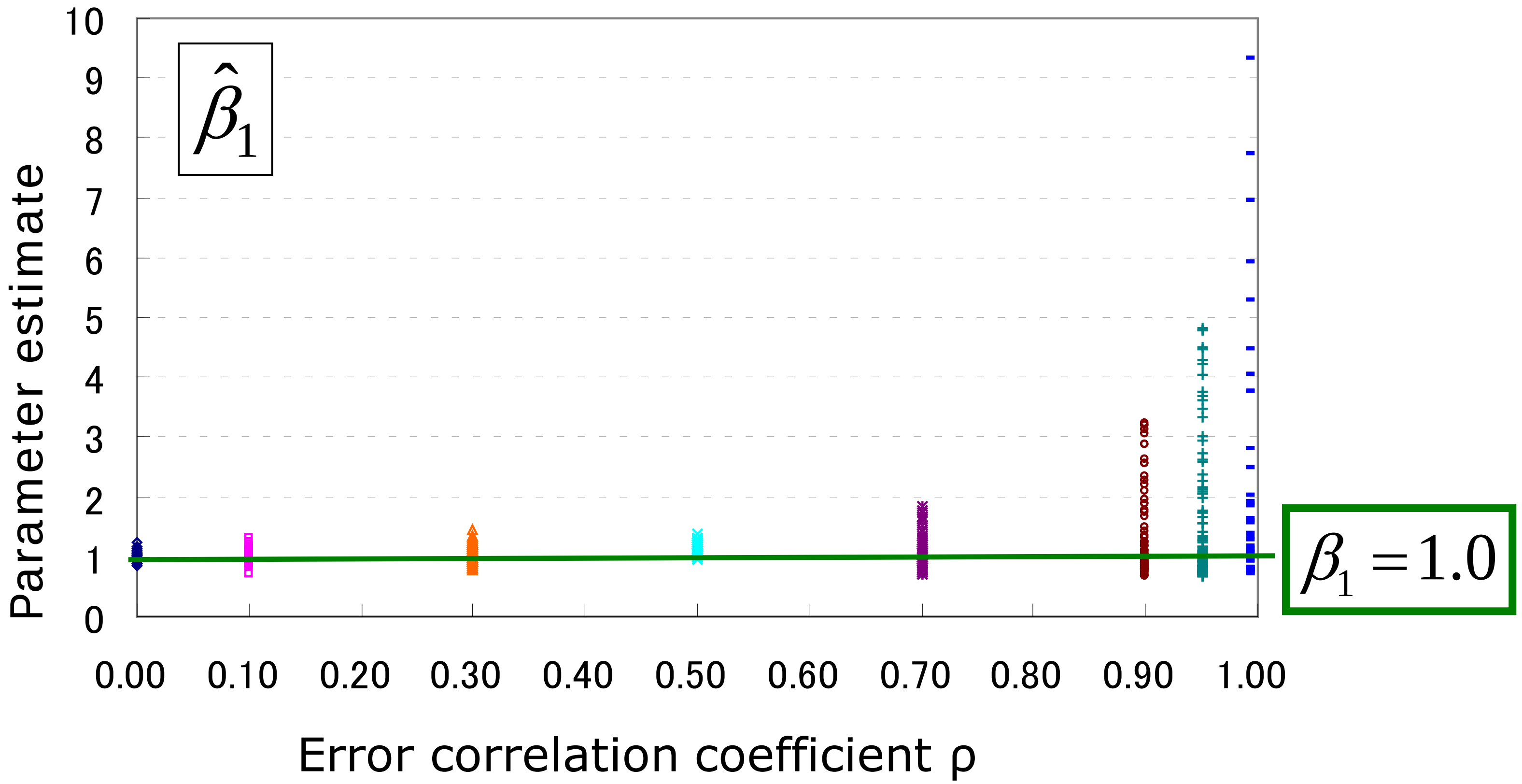
Identification issue (cont.)

- Hereafter, we constrain $\boxed{s^2 = s_1^{\,2} = s_2^{\,2}}$

$$\boxed{\mathbf{\Sigma}_A = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & 1 & \sigma_{23} \\ 0 & \sigma_{23} & 1 \end{bmatrix}} \quad \boxed{\mathbf{\Sigma}_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \sigma_{23}' \\ 0 & \sigma_{23}' & 1 \end{bmatrix}} \quad \mathbf{\Sigma}_C = \begin{bmatrix} \sigma_{11}'' & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boxed{\Sigma = \begin{pmatrix} s_1^{\,2} + \dfrac{\pi^2}{6} & 0 & 0 \\ & s_2^{\,2} + \dfrac{\pi^2}{6} & s_2^{\,2} \\ & & s_2^{\,2} + \dfrac{\pi^2}{6} \end{pmatrix}} \Rightarrow \boxed{\begin{aligned} &\left(s^2 + \frac{\pi^2}{6}\right)\begin{pmatrix} 1 & 0 & 0 \\ & 1 & \rho \\ & & 1 \end{pmatrix} \\ &\rho = \frac{s^2}{s^2 + \dfrac{\pi^2}{6}} \end{aligned}}$$

Variability of parameter estimates



- Parameter estimates are quite instable especially for the case with higher ρ

Variability of parameter estimates (cont.)

- This instability is caused by the dependence of coefficient estimates on error variance

Probit model

$$\Sigma_{\xi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}$$

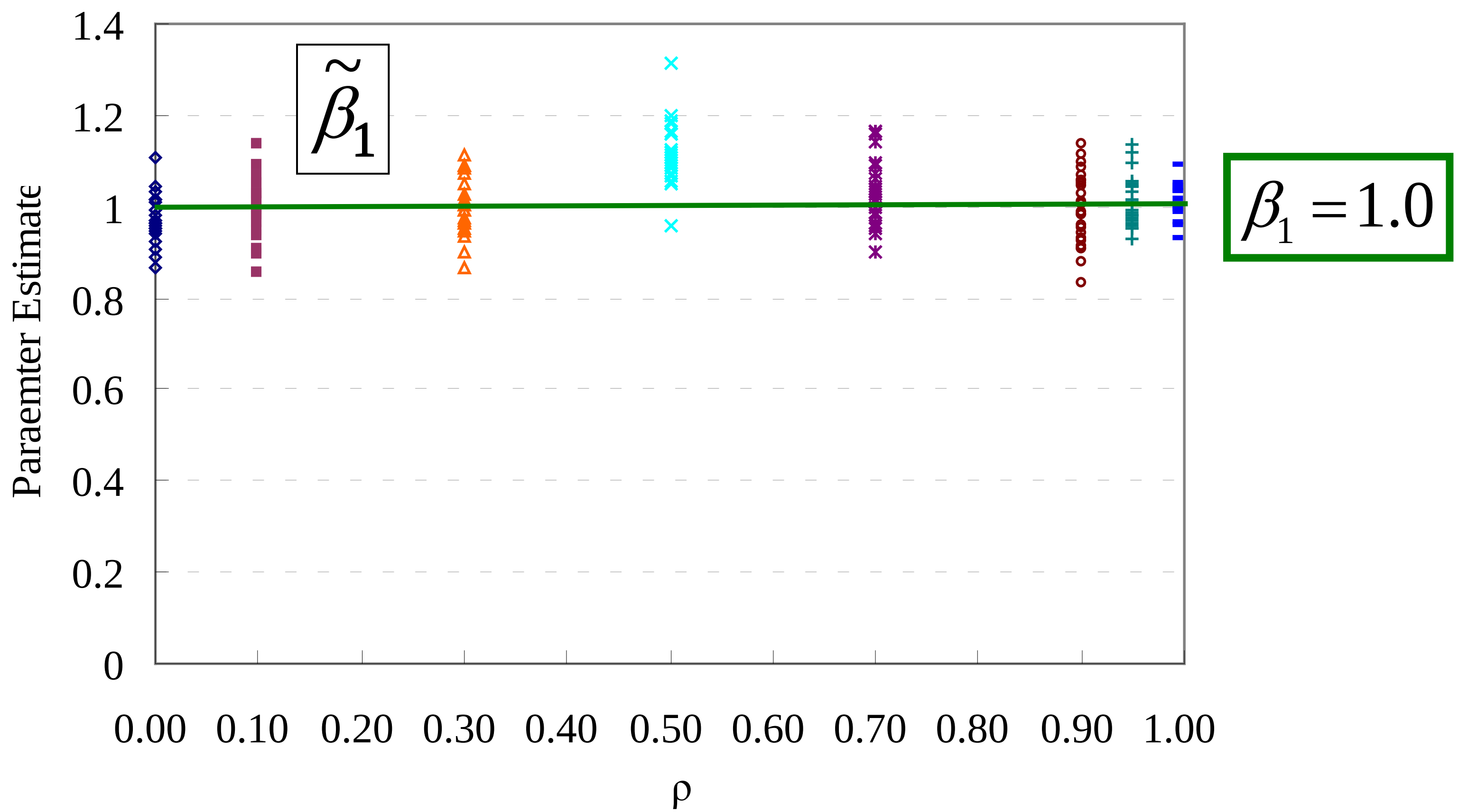
Error component MXL model

$$\Sigma = \left(s^2 + \frac{\pi^2}{6} \right) \begin{pmatrix} 1 & 0 & 0 \\ & 1 & \rho \\ & & 1 \end{pmatrix} \qquad \rho = \frac{s^2}{s^2 + \frac{\pi^2}{6}}$$

- Error variance is not standardized in MXL model
- Needs for normalization of parameter estimates

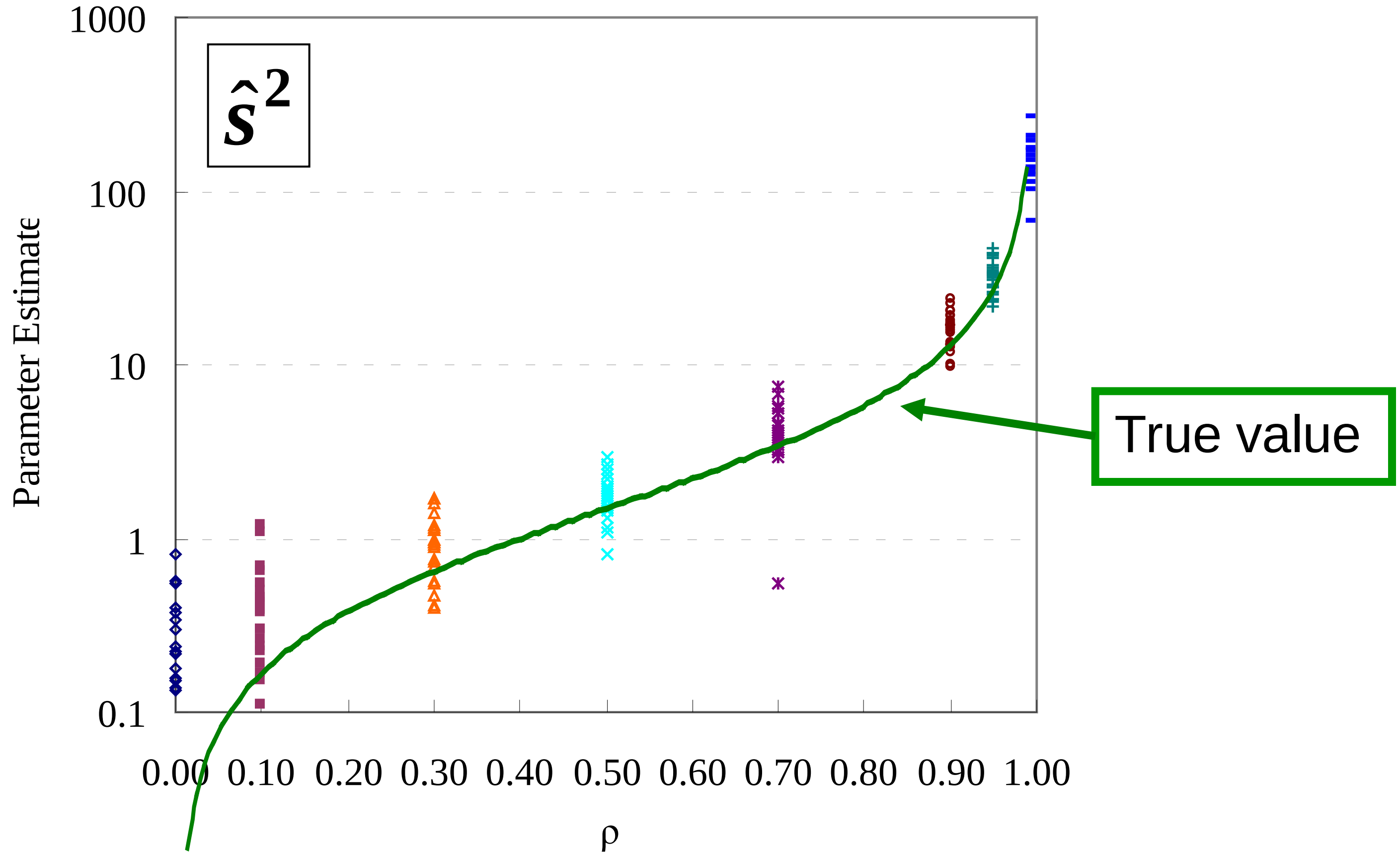
$$\tilde{\beta}_j = \frac{1}{\sqrt{\hat{s}^2 + \frac{\pi^2}{6}}} \hat{\beta}_j$$

Variability of parameter estimates (cont.)



- After normalization, utility coefficients are unbiased and stable

Variability of parameter estimates (cont.)

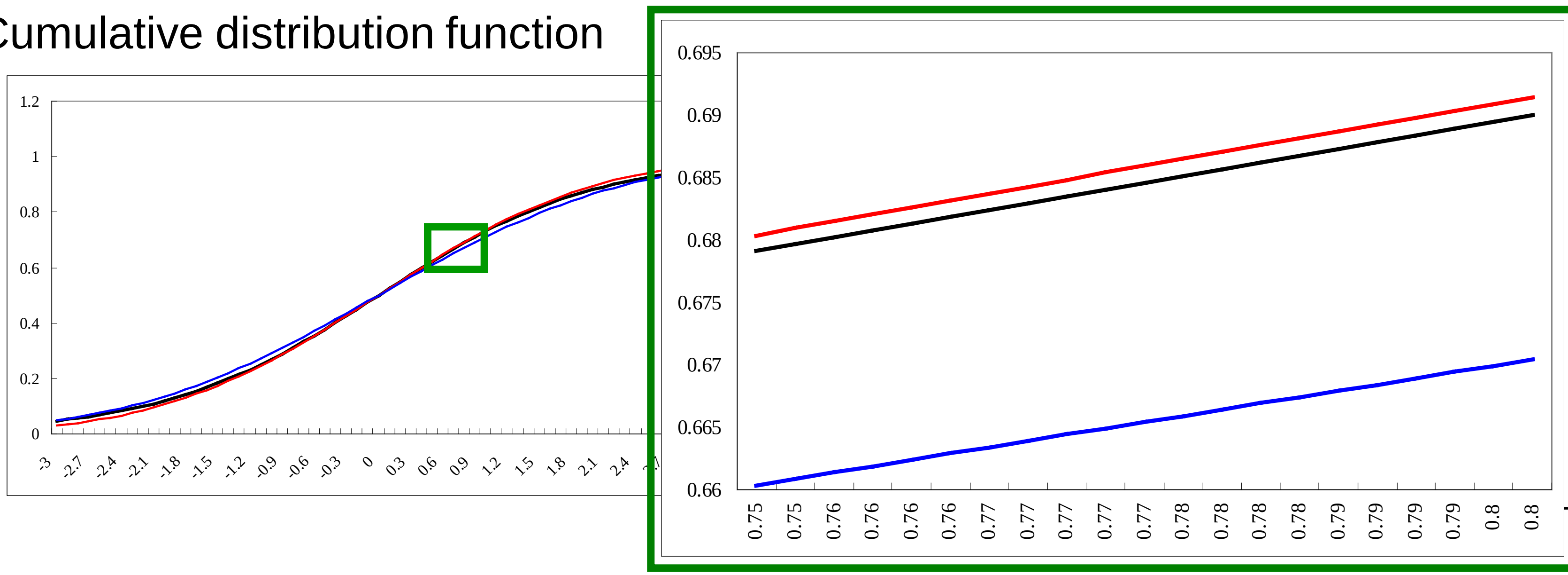


- Estimated variances of the error components tend to be biased upward

Variability of parameter estimates (cont.)

- Biases in estimated variances might be related to the difference in shape of Normal and Gumbel distribution
- Amemiya (1981) suggests in binary case $\boxed{N(0, 1.6^2)}$ rather than $\boxed{N(0, \pi^2/3)}$ fits better to $\boxed{L(0, \pi^2/3)}$ though the latter has equal variance to $L(0, \pi^2/3)$ $(1.6 < \pi/3^{0.5} \approx 1.8)$

Cumulative distribution function



Estimation of choice probabilities

- Choice probabilities are calculated by

$$\hat{P}_n(i) = \iint \frac{\exp\left(\hat{\beta}_1 X_{1in} + \hat{\beta}_2 X_{2in} + \hat{s}\eta_i\right)}{\sum_j \exp\left(\hat{\beta}_1 X_{1jn} + \hat{\beta}_2 X_{2jn} + \hat{s}\eta_j\right)} df(\eta_1) df(\eta_2) \qquad \eta_i, \eta_j = \begin{cases} \eta_1 & if\ i \text{ or } j = 1 \\ \eta_2 & if\ i \text{ or } j = 2 \text{ or } 3 \end{cases}$$

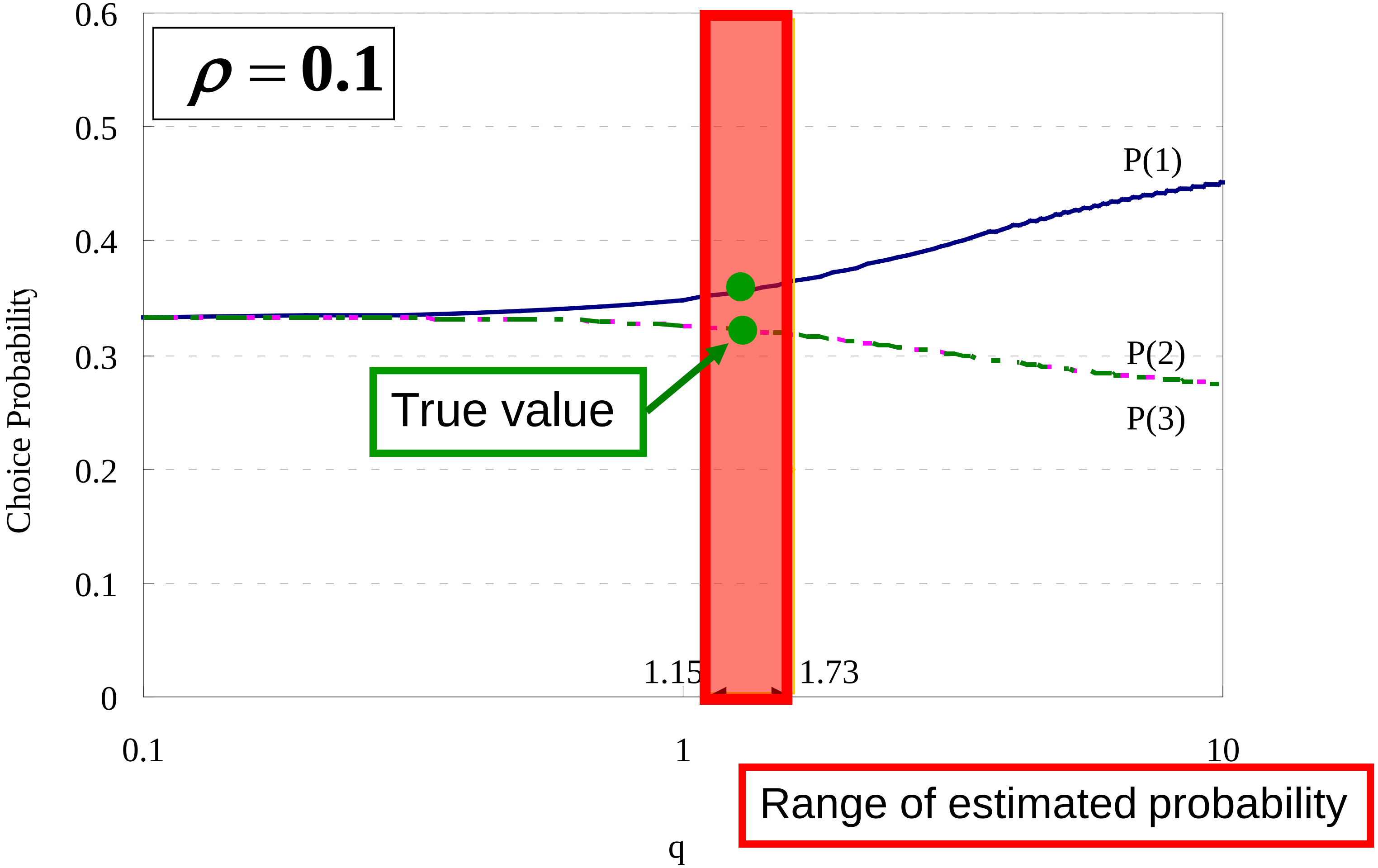
for the case $X_{11} = X_{21} = X_{12} = X_{22} = X_{13} = X_{23} = 1.0$

- The effects of biased estimate of s is examined by introducing q, and calculate

$$P(i \mid q) = \iint \frac{\exp q\left(\beta_1 X_{1i} + \beta_2 X_{2i} + s\eta_i \big/ \sqrt{\pi^2/6}\right)}{\sum_j \exp q\left(\beta_1 X_{1j} + \beta_2 X_{2j} + s\eta_j \big/ \sqrt{\pi^2/6}\right)} df(\eta_1) df(\eta_2)$$

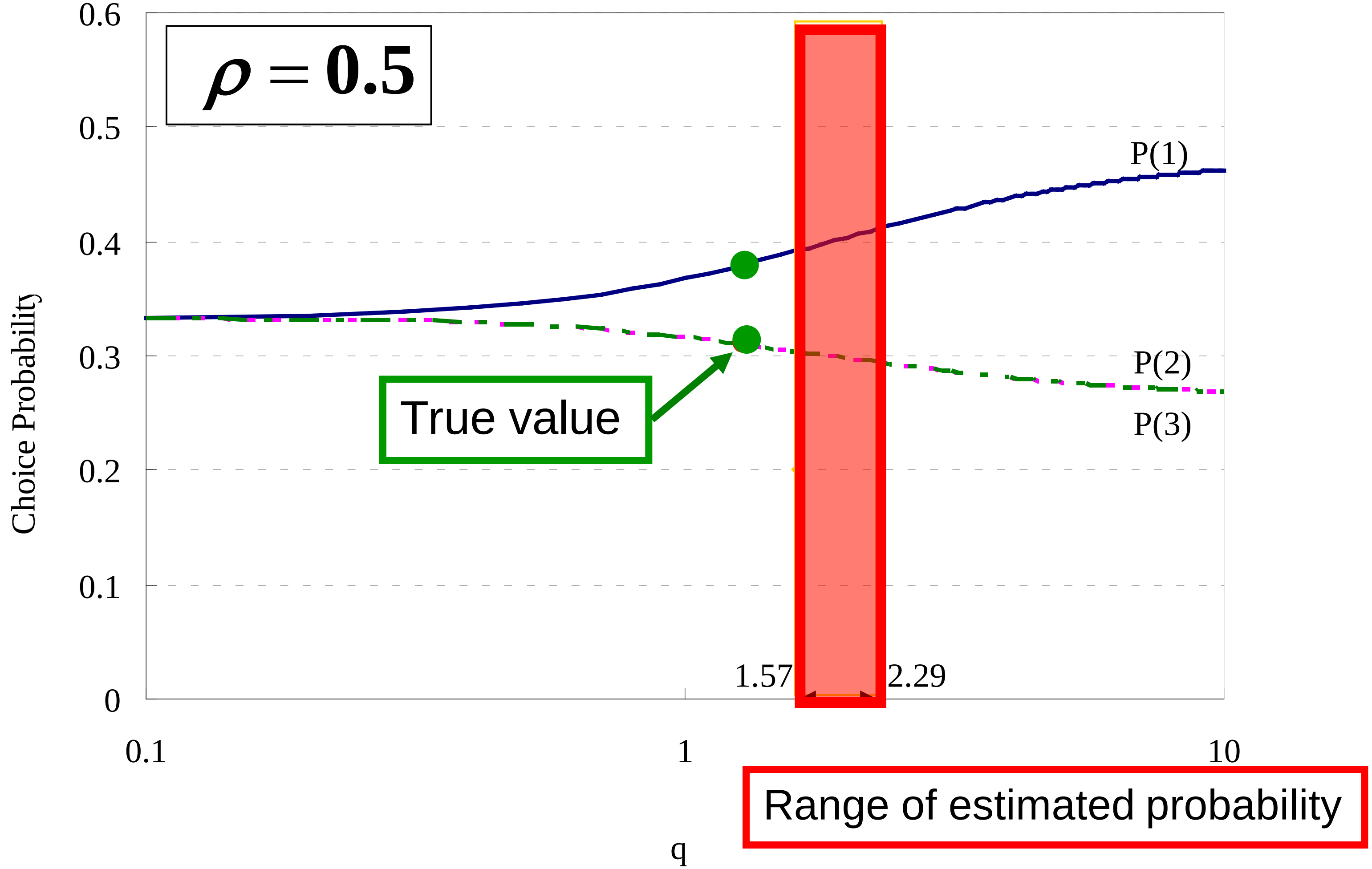
- True probability is obtained when q ≈ 1.29

Estimation of choice probabilities (cont.)



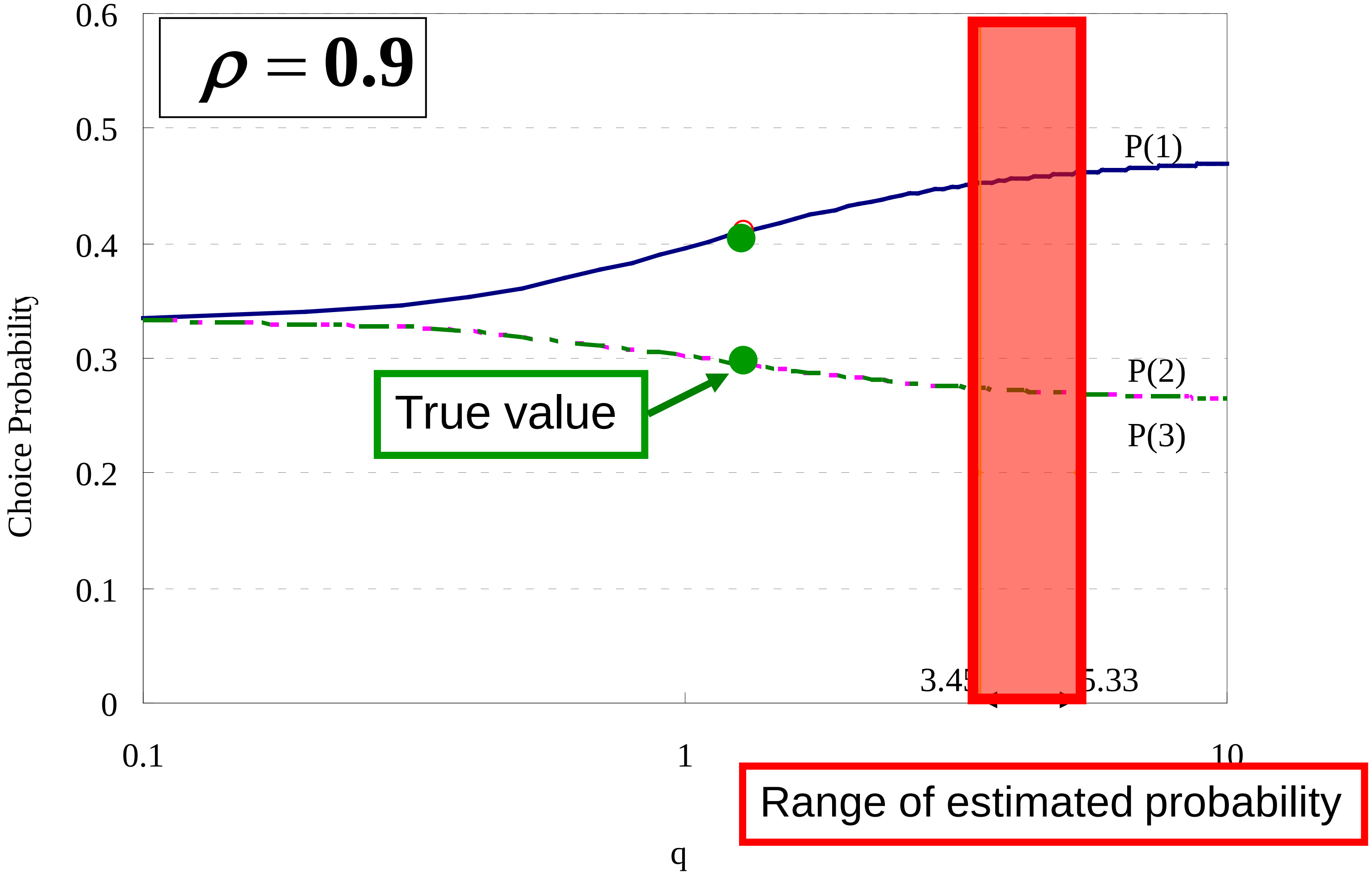
- True probabilities are contained in the range of the estimated probability

Estimation of choice probabilities (cont.)



- True probabilities are NOT contained in the range of the estimated probability

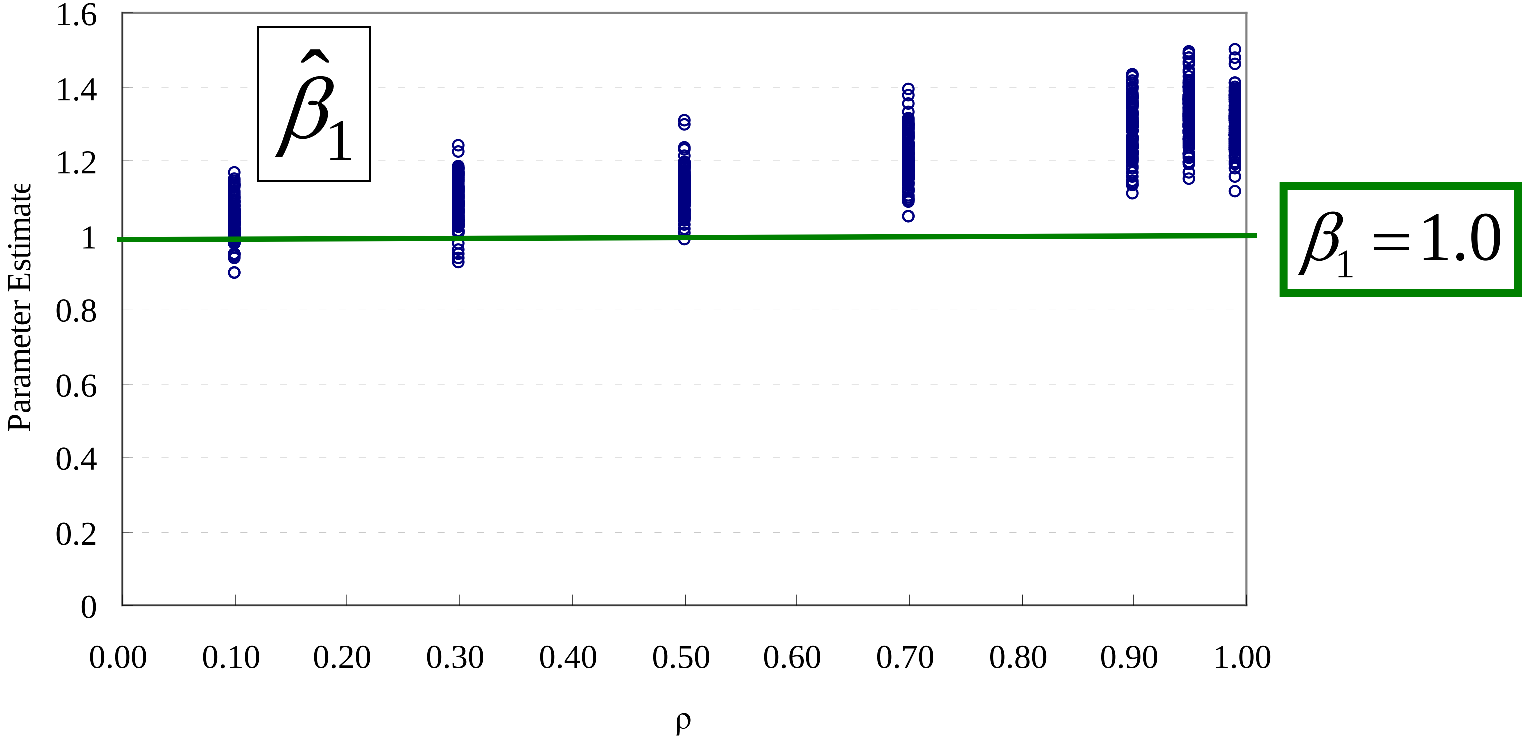
Estimation of choice probabilities (cont.)



- True probabilities are NOT contained in the range of the estimated probability

Usefulness of MNL models

- MNL models are estimated using the same data sets



- Utility coefficient estimates are biased upward, but up to about 30%, smaller than MXL model

Conclusions and future research

For the error component MXL model

1. Variance structure cannot be uniquely identified through model estimation
2. Parameter estimates are quite instable especially for the case with a high error correlation
3. After proper normalization, utility coefficients are unbiased and stable
4. Estimated variances of the error components tend to be biased upwards
5. Choice probabilities are biased unless the error correlation is very small

MNL model can produce relatively unbiased utility coefficient estimates

Conclusions and future research (cont.)

- One would adopt MXL model in search of covariance specification

-> The model is incapable of identifying the true structure, and parameter estimates are instable

- One may opt to develop adequately specified MNL through careful selection of explanatory variables, utility formulation or definition of alternatives (consistent with suggestion by Pinjari & Bhat (2006))

- Needs for further research on properties of parameter estimates of MXL model with taste heterogeneity as well as error components