UNIVERSITÄT BIELEFELD

Fakultät für Wirtschaftswissenschaften

Die Verwendung des MaCML Ansatzes für die Schätzung von multinomialen Probit Modellen auf der Basis von Paneldaten

Using the MaCML approach for the estimation of multinomial probit models for panel data

Dietmar Bauer

Bielefeld University

Jahrestagung des Ausschusses für Ökonometrie, Verein für Socialpolitik,

Berlin, 21.-23.3. 2024

Introduction

Multinomial Probit Models (MNP)

MNPs are used for discrete choice modeling:

- Deciders pick one out of a finite and exhaustive number of distinct alternatives.
- No ordering necessary: categorical data.
- Alternatives could be ordered: for example answer to survey questions (leads to ordered probit/logit).
- Typical setting: conjoint analysis.

Example data set: Helveston et al. (2015)

Purchase decisions for automobiles within a conjoint study.

SECTION 3

Suppose these 3 vehicles below were the only vehicles available for purchase, which would you choose?

Each option will look like this:

Attribute*	Option 1	Option 2	Option 3
Vehicle Type	Conventional 300 mile range on 1 tank	Plug-In Hybrid & 300 mile range on 1 tank (first 40 miles electric)	Electric 75 mile range on full charge
Brand	German	American	Japanese
Purchase Price	$18,000	$32,000	$24,000
Fast Charging Capability	--	Not Available	Available
Operating Cost (Equivalent Gasoline Fuel Efficiency)	19 cents per mile (20 MPG equivalent)	12 cents per mile (30 MPG equivalent)	6 cents per mile (60 MPG equivalent)
0 to 60 mph Acceleration Time**	8.5 seconds (Medium-Slow)	8.5 seconds (Medium-Slow)	7 seconds (Medium-Fast)
	○	○	○

*To view an attribute description, click on:
**The average acceleration for cars in the U.S. is 0 to 60 mph in 7.4 seconds

Source: Helveston et al. (2015).

Each respondent sees 15 of these images and takes 15 decisions (balanced panel structure).

Other example: mode choice

- For each trip a mode is chosen.
- Data typically collected in household surveys for one week (collecting information for each trip for each member of household).
- rolling panel design: each household is kept in the panel for three consecutive years.
- In Germany: Mopilitätspanel (MOP), collected by KIT Karlsruhe.
- Unbalanced panel data set: respondents have different number of trips.

Interpretation of MNP as random utility model (RUM)

- Alternatives are characterized via regressors: $X_{i,t} \in \mathbb{R}^K$.
- Some regressors may characterize deciders: $S_{i,t} \in \mathbb{R}^d$.
- Each decider assigns a utility to each alternative:

$$U_{i,t,j} = V_{i,t,j} + e_{i,t,j}$$

- systematic part of the utility $V_{i,t,j}$ depends on the characteristics, often linear:

$$V_{i,t,j} = \beta_{i,t}' X_{i,t,j}$$

- preferences are encoded using the parameter vectors $\beta_{i,t}$.
- Error terms and regressors are assumed to be independent of each other.
- Probit models are obtained for jointly normally distributed error term $e_{i,t,:} \in \mathbb{R}^J \sim \mathcal{N}(0, \Sigma)$.
- Choice $y_{i,t}$: utility maximizing alternative.

Remark: Multinomial Logit Models (MNL) use a different distribution for the error term.

Interpretation of MNP as random utility model (RUM)

For the mode choice example:

- $X_{i,t,j}$: i.A. travel time [in min] and costs [in Euro].
- $S_{i,t}$ characterizing deciders: age, income, ...
- systematic part of the utility $V_{i,t,j}$:

$$V_{i,t,j} = \beta_p \text{price}_{i,t,j} + \beta_t \text{traveltime}_{i,t,j} + \ldots$$
$$= \beta_p(\text{price}_{i,t,j} + \frac{\beta_t}{\beta_p}\text{traveltime}_{i,t,j}) + \ldots$$

- interpretation: tradeoffs: identical systematic part of utility for one minute more travel time, when price drops by $\frac{\beta_t}{\beta_p}$ (VoT)
- The VoT may depend on income, for example.

Probability structure

- MNP provides choice probabilities conditional on regressors: $\mathbb{P}(y_{i,t} = j | X_{i,t,:}, S_{i,t}; \beta, \Sigma), j = 1, ..., J$.

Different choices for one decider are not necessarily (conditionally) independent:

- $e_{i,t,:}$ and $e_{i,s,:}$ can be autocorrelated: temporal dependence.
- $\beta_{i,t}$ can be modelled explicitly as $\beta(S_{i,t})$
- ... or implicitly using random effects: assume that β_i is chosen from some underlying distribution for each decider once: mixing.

Random Effects Models

MNP with random effects: Mixed MNP (MMNP).

$$\beta_i \sim F(.;\theta)$$

- *parametric*: multivariate Gaussian, often independent over parameters (diagonal variance matrix).
- extension: *latent class* models using, e.g., Gaussian mixtures for the mixing.
- *non-parametric* models, estimated using, e.g., penalization, see e.g. Heiss, Hetzenecker, Osterhaus (2021) or Train (2016).

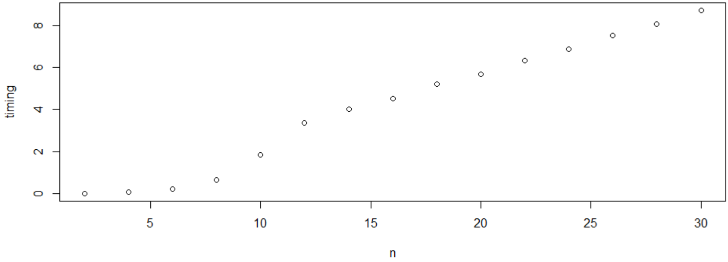
Identification: holds, if usual restrictions to identify level and scale are used and if regressors are rich enough. *Details:*

REMARK: One does not have to believe in the RUM interpretation of the MNP. One can see it merely as one way to model the conditional choice probabilities.

Estimation

- Since choice probabilities are directly given, use of (conditional) MLE is straightforward.
- If choices are not independent (temporal autocorrelation and/or random coefficients), they must be dealt with jointly.
- Likelihood involves evaluating Gaussian CDF for large dimensions (Helveston et al.: 30; mode choice: typically dozens).
- This is numerically costly.

Evaluation of CDF at 100 points for different dimensions (in sec.): R fct. `pmvnorm`



Why then not use MNL?

- This is a possibility.
- MNL models suffer from IIA (independence of irrelevant alternatives) property, due to independence of errors across alternatives.
- Blue-bus red-bus paradoxon: Adding an additional alternative that is identical to one already there changes the choice probabilities of all alternatives.
- This can be alleviated using many different model classes (nested, cross-nested, ...), that are often hard to interpret.
- Heteroskedastic errors are harder to include, temporal autocorrelation hard to justify (extreme value distributions are not closed under linear transformations).
- MNP offers the advantage of being able to introduce correlations between the errors across alternatives.

Bringing down numerical costs

If there is no temporal autocorrelation of the errors, but only random effects:

1. Maximum simulated likelihood: mixing over random effects is an expectation, which can be approximated using simulation: sample means as estimators of expectations.
2. analytic approximation methods
3. using composite marginal likelihood (CML)

This talk: combining the last two methods: Maximum approximate Composite Marginal Likelihood (MaCML).

REMARKS:

1. This is a methodological talk. Data examples are just used for demonstration purposes and not seen as finished models.
2. There are not too many people in Germany working in the area. Hence the talk may feel somewhat like a lecture. Apologies, if this is too boring or inappropriate.

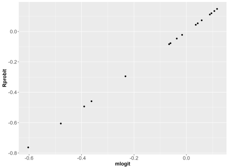
MaCML

Comparison

Estimation of (M)MNP model for Helveston et al. (2015) US data set:

- $N = 384$ deciders, 3 alternatives, $T = 15$ choice occasions
- 5760 choices on total
- Estimation using `mlogit` (MSL with $R = 200$) without panel structure versus `Rprobit` without random coefficients and `Rprobit` with random coefficient for `opCost`.

	without panel		with panel	
	mlogit	MLE	FP	AP
opCost/price	1.66	1.66	1.84	1.84
accelTime/price	1.80	1.80	1.78	1.78
timing [sec]	630	9	309	32
in sample log-like	-4633.0	-4632.8	-4612.2	-4612.2



Origin of ideas

- Varin and Vidoni (2005) introduce *composite marginal likelihood* (see also Varin (2008), Varin, Reid, Firth (2011)).

Main idea

If evaluating the full likelihood is too costly, but some of the marginals are easier to obtain, a collection of marginals may contain enough information for consistent estimation.

- Oftentimes the accuracy loss is not huge (see later).
- Concept aims at numerical speed in situation of large datasets.
- CML is a general principle, not limited to discrete choice models.
- Introduced to MMNP model estimation by Bhat (2011).
- Motivated by speed advantages compared to MSL approach in some simulation setting (Bhat, Sidharthan, 2011).

Definition

- Evaluating the Gaussian CDF shows strongly non-linear costs for increasing dimensions.
- No issue for small sizes that are typically involved in pairs of choices: for example: evaluating a 30-dim. CDF is roughly as costly as approx. 133 times evaluating 4- dim CDFs.
- Change the criterion function from log-likelihood to pairwise log-CML:

$$
\begin{aligned}
ll(\beta, \Sigma; y, X, S) &= \sum_{i=1}^{N} lcml(\beta, \Sigma; y_i, X_i, S_i), \\
lcml(\beta, \Sigma; y_i, X_i, S_i) &= \sum_{1 \le a < b \le T_i} w_{t_a, t_b, i} \log \mathbb{P}(y_{i,t_a}, y_{i,t_b} | X_{i,t_a}, X_{i,t_b}, S_i; \beta, \Sigma) \\
&\quad + \sum_{1 \le a \le T_i} v_{t_a, i} \log \mathbb{P}(y_{i,t_a} | X_{i,t_a} S_i; \beta, \Sigma).
\end{aligned}
$$

$w_{t_a,t_b,i} \ge 0$, $v_{t_a,i} \ge 0$... power weights.

Definition (II)

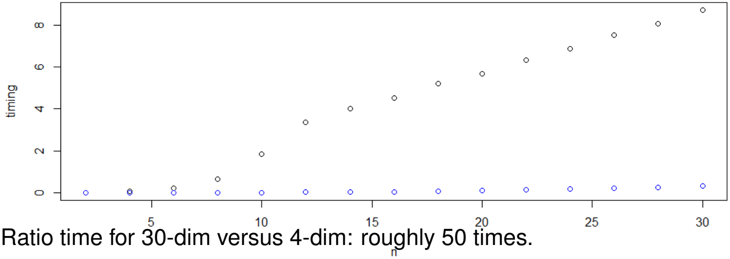
- Power weights can be used to tune the numerically load and have impact on accuracy.
- $w_{t_a,t_b,i} = 0$ or $v_{t_a,i} = 0$ implies: Do not need to calculate this term.
- Often used:
  - full pairwise $w_{t_a,t_b,i} = 1$, $v_{t_a,i} = 0$ (FP).
  - adjacent pairwise: $w_{t_a,t_b,i} = \mathbb{I}(b = a + 1)$, $v_{t_a,i} = 0$ (AP).
  - adjacent pairwise plus marginals: $v_{t_a,i} = 1$ (APU)
- While speedup is clear, impact on accuracy is less clear a-priori.

Asymptotics

- CML is a weighted sum of log-likelihoods for reduced data set.
- For iid draws over $i = 1, ..., N$ all appropriately scaled log-likelihoods (corresponding to one pair of choices or one choice) converge to limiting function, which has a minimum at the true parameter vector.
- Combining them leads to consistent estimates for $N \to \infty$, if the parameters are identified from at least one pair of choices with positive weight. example
- Asymptotic normality (with $N \to \infty$) with sandwich covariance matrix follows as usual under appropriate assumptions.

Impact of approximation

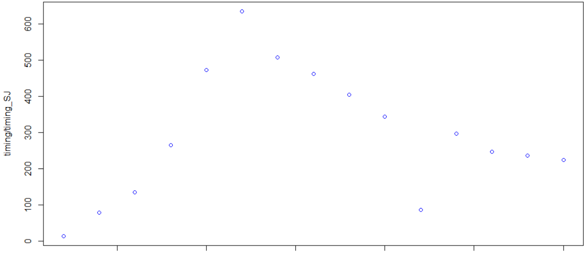
- Different approximation concepts: Solow-Joe (Joe, 1995), Mendel-Elston (1974), TVBS (Bhat, 2018).
- **Common feature:** analytic approximation, not numerical. No 'bandwidth' parameter that trades off accuracy and numerical speed.
- Two interpretations:
  1. Approximation changes the criterion function
  2. Approximation changes the mapping from parameters to conditional choice probabilities (Batram and Bauer, 2019).



Ratio time for 30-dim versus 4-dim: roughly 50 times.

Impact of approximation

- Different approximation concepts: Solow-Joe (Joe, 1995), Mendel-Elston (1974), TVBS (Bhat, 2018).
- **Common feature:** analytic approximation, not numerical. No 'bandwidth' parameter that trades off accuracy and numerical speed.
- Two interpretations:
  1. Approximation changes the criterion function
  2. Approximation changes the mapping from parameters to conditional choice probabilities (Batram and Bauer, 2019).



Ratio time for 30-dim versus 4-dim: roughly 50 times.

Impact of the power weights

Balanced panel case

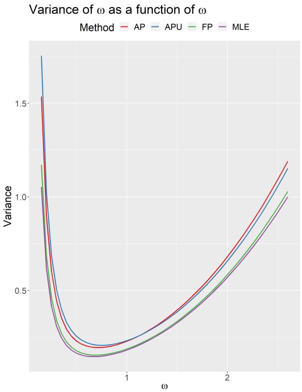
We use a simple model that allows to separate a number of effects:

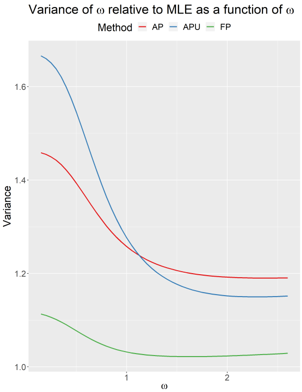
- MMNP with two alternative varying regressors iid uniformly $[-1, 1]$ distributed. [Leaving out the ASCs identifies the level.]
- $\beta = [\tilde{\beta}_1, 1]$: The second coefficient is fixed to 1 [This fixes the scale.].
- $\tilde{\beta}_1 \sim \mathcal{N}(\beta_1, \omega^2)$ is drawn randomly from a normal distribution with expectation β_1 and variance ω^2 (random effect).
- The error $e_{:,i,t} \in \mathbb{R}^J$ is normally distributed with expectation zero and variance $\sigma^2 I_J$.
- Each individual chooses from $J = 2$ alternatives in T consecutive choice occasions (balanced case).
- The choice probabilities for the choices y_i of the i-th decider depend on X_i and the true parameters $\theta_\circ = [\beta_1, \omega, \sigma]$ and are given as $p(y_i|X_i; \theta_\circ)$.

3. Impact of the power weights

$\beta_1 = 1, \omega \in (0.15, 2.6), \sigma = 1, T = 3$

- When estimating ω the situation of small ω is particularly critical.
- APU comes close to 70% efficiency loss, AP 45% while FP looses only 15%.

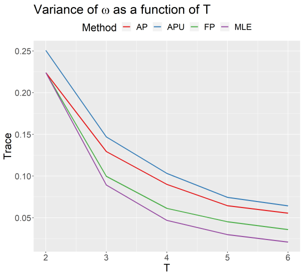


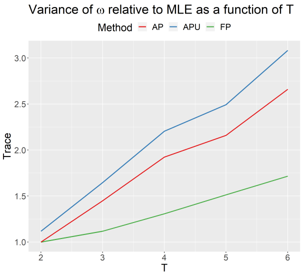


The influence of the number of choice occasions T

$\beta_1 = 0.5, \omega = 0.1, \sigma = 0.5$

- As expected the variance decreases as the number of choice situations increases.
- At the same time for the estimation of ω relative to MLE the other variants get worse the larger T.





Details on Bhat & Sidharthan (2011) system

Impact of pow. w.: unbalanced case

Motivation

- Consider two different groups of deciders for mode choice:
  - immobile persons with 2 trips within the observation week
  - agile persons with 28 trips.
- If decisions are independent, pairwise choice probabilities equal product of single choices.
- For log CML this adds two terms for each pair of choices.
- For the immobile persons each of the two choices occurs once in the CML.
- For the agile persons each trip adds 27 terms in the FP CML.
- It is a-priori unlikely, that this leads to the smallest variance (across all weighting schemes $w_{a,b,i}$), as the observations for the immobile persons are dominated by the mobile persons' trips.

Power weights depending on the total number of choices may help.

3. 1. Impact of pow. w.: unbalanced case

Assumption (DGP)

The data set (y_n, X_n), $n = 1, ..., N$ is generated by the following mechanism:

1. A number T_n of choice occasions are drawn from a discrete random distribution supported in $\{2, 3, ..., \mathbf{S}\}$.
2. For each decider facing T_n choice occasions, a matrix $X_n = [X_{n,1}, ..., X_{n,T_n}] \in \mathbb{R}^{JR \times T_n}$ of regressors is chosen iid over deciders such that for each pair of choice occasions (a, b), the distribution of the matrix $[X_{n,a}, X_{n,b}]$ is identical. Furthermore, $\|X_{n,a}\| \leq M$ (uniform norm bound).
3. For given T_n and X_n, the vector of choices $y_n = [y_{n,1}, ..., y_{n,T_n}]' \in \mathcal{J}^{T_n}, \mathcal{J} = \{1, ..., J\}$ is chosen according to the mixed MNP model corresponding to parameter vector $\theta_0 \in \mathbb{R}^p$ for appropriate integer p.

For conjoint studies these assumptions are realistic, also for surveys.

3. 1. Impact of pow. w.: unbalanced case

Theorem (Asymptotic Variance (Büscher and Bauer, 2024))

Let the data be generated according to Assumption DGP with parameter vector θ_0 and let $\hat{\theta}$ be the CML estimator maximising the weighted CML function using the weights $w_{n,a,b} = w_{T_n} \mathring{w}_{a,b}$, where $C_s(\mathring{W}) = \sum_{1 \leq a < b \leq s} \mathring{w}_{a,b} > 0, s = 2, ..., \mathbf{S}$. Further, let $w_s \geq 0$ denote group-specific weights according to the number of observations $T_n = s,\ s \in \{2, \ldots, \mathbf{S}\}$ of decider n, such that $\sum_{s=2}^{\mathbf{S}} f_s w_s C_s(\mathring{W}) = 1$. Then the following hold:

1. $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_\theta(W_s))$, where the asymptotic variance-covariance matrix $V_\theta(W_s)$ for a given vector $W_s = (w_2, \ldots, w_{\mathbf{S}})'$ of group-specific weights has the form

$$V_\theta(W_s) = H_0^{-1} \left(\sum_{s=2}^{\mathbf{S}} f_s w_s^2 V_s \right) H_0^{-1},$$

with $g_n(\mathring{W}) = \partial lcml(\beta, \Sigma; y_i, X_i, S_i, \mathring{W})$

$$V_s = \mathbb{E}\ g_n(\mathring{W}) g_n(\mathring{W})', \quad H_0 = \mathbb{E}\ \partial_\theta^2 \log \mathbb{P}(y_{n,1}, y_{n,2}, X_n; \theta_0).$$

3. 1. Impact of pow. w.: unbalanced case

Theorem ('Optimal' Group-Specific Weights (Büscher and Bauer, 2024))

Let $l : \mathbb{R}^{p\times p} \to \mathbb{R}$ be a linear mapping of the form $l(V) = \mathrm{tr}(VA)$, with $A \in \mathbb{R}^{p\times p}$, $A \neq 0$ symmetric and positive semidefinite, $\hat{\theta}$ be the CML estimator maximising the weighted CML function, and let the data be generated subject to Assumption DGP.
Let $\mathring{w}_{a,b}$ be the initial weights, where $\sum_{s=2}^{\mathbf{S}} f_s w_s C_s(\mathring{W}) = 1$.
Then $l(V_\theta(W_s))$ is minimised over W_s by

$$w_s^* = \left(\sum_{s=2}^{\mathbf{S}} f_s C_s(\mathring{W})^2 / v_s(\mathring{W}) \right)^{-1} C_s(\mathring{W})/v_s(\mathring{W}) \propto C_s(\mathring{W})/v_s(\mathring{W}),$$

with $v_s(\mathring{W}) = l\left(H_0^{-1} V_s(\mathring{W}) H_0^{-1}\right)$.

- Reminiscent of GLS estimator: optimal weighting is proportional to inverse of variance.
- Using estimates we can define two-step 'FGLS' estimators.

Model diagnostics

What to diagnose?

Some objectives for diagnosing the appropriateness of the models are:

- unobserved heterogeneity in groups of deciders?
- structural breaks: panel waves in successive years?
- temporal auto-correlation in the error terms?

Diagnostics in the likelihood setting

- enlarge the model by including terms that describe the deviation from the restricted model
- use LR or LM tests based on likelihood optimization
- calculate different test statistics for different deviations

Thus: typical each test statistic requires additional calculations (at least evaluating the typically non-linear gradient of the unrestricted model).

Example: pooling two groups of individuals in linear panel models

$$y_{i,t} = x_{i,t}'\beta + x_{i,t}'\gamma\mathbb{I}(g(i)=2) + u_{i,t}, \quad g(i) = \begin{cases} 1 & , \quad 1 \le i \le N/2, \\ 2 & , \quad N/2 < i \le N \end{cases}$$

- Restricted model: $\gamma = 0$. Estimate OLS: $\hat{\beta} = (X'X)^{-1}X'y$ with residuals $\hat{u}_{i,t} = y_{i,t} - x_{i,t}'\hat{\beta}$.
- Score for the unrestricted model (for known variance $\sigma^2 = 1$) evaluated at the estimate in the restricted model:

$$\begin{pmatrix} X'\hat{u} \\ X_2'\hat{u}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ X_2'X_2\gamma + X_2'u_2 - X_2'X_2(X_1'X_1 + X_2'X_2)^{-1}(X_1'u_1 + X_2'u_2) \end{pmatrix}.$$

- Expectation of score: $\mathbb{E}\begin{pmatrix} 0 \\ X_2'X_2\gamma \end{pmatrix}$

Example: pooling two groups of individuals in linear panel models (II)

- Score under restricted model $\gamma = 0$ (for known variance $\sigma^2 = 1$):
$\begin{pmatrix} X'\hat{u} \\ X_2'\hat{u}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ X_2'u_2 - X_2'X_2(X_1'X_1 + X_2'X_2)^{-1}(X_1'u_1 + X_2'u_2) \end{pmatrix}.$
- Assume $X_1 = X_2 \Rightarrow X_2'X_2(X_1'X_1 + X_2'X_2)^{-1} = X_2'X_2(2X_2'X_2)^{-1} = \frac{1}{2} \Rightarrow$

$$\frac{1}{2}(X_2'u_2 - X_1'u_1) = \sum_{t=1}^{T}(\sum_{i=1}^{N/2} x_{i,t}'u_{i,t} - \sum_{i=N/2+1}^{N} x_{i,t}'u_{i,t}).$$

- Under the usual panel assumptions of $u_{i,t}$ iid (over individuals and over time) we then get an iid sample (over t) of score elements of differences between the first and the second group.

Calculation of LM-type-test for structural break in time

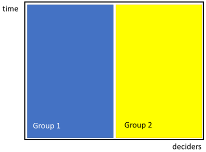
$$\left\{ \partial \left[\sum_{1 \le t_a < t_b \le T_0} w_{a,b} \log \hat{P}_{i,a,b} - \sum_{T_0+1 \le t_a < t_b \le T} w_{a,b} \log \hat{P}_{i,a,b} \right], i = 1, ..., N \right\}$$

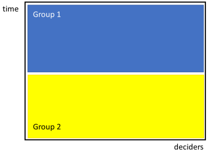
- under the null constitutes a zero mean iid sample
- tests for zero expectation are immediate
- we can test single entries with t-tests or more entries using F-tests
- for the calculation all that is needed are $\partial \log \hat{P}_{i,a,b}$, which are calculated in gradient type methods anyway

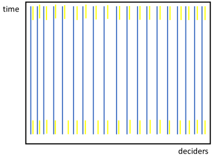
Via different groupings this provides diagnostics that are not available in the likelihood setting.

Application for CML

- Key to the LM calculation: exchangeability of pairs of choices
- For MMNP models this is often the case: same questions asked to all persons; same characteristics for mode choice for same trips on consecutive days.
- Then we can use different groupings:
  - two groups of individuals
  - two groups of time points: early in the sample and late to detect structural breaks
  - two groups of temporal distance: close by versus far apart (start of sample, end of sample).
- This leads to different LM test statistics all calculated based on the score for individual deciders for different pairs of decisions.
- Minimal extra numerical load.







Summary

5. Summary

- MaCML reduces the numerical load compared to MLE for the estimation of MMNP models from panel data.
- The loss in accuracy has been found to be small in some simulations/calculations.
- It appears to be strongest for large T situations.
- Compared to MLE the CML approach allows for simple diagnostic tools that provide insights into the suitability of pooling, temporal stability and the existence of temporal autocorrelation.
- Thus MaCML is an option primarily for large data sets.
- Using analytical approximations involves changes to the model, that need be taken into account.
- Temporal correlation of the error term can be detected using the LM-type tests, but it is currently not clear, how to include it in the model (identification issues with random ASCs; potentially large parameter set).

6. Literature

- Batram, M. and Bauer, D. (2019). On consistency of the MaCML approach to discrete choice modelling, *Journal of Choice Modelling* **30** 1-16.
- Bhat, C. R. (2011). The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models, *Transportation Research Part B* **45**(7): 923-939
- Bhat, C. R. (2018). New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function, *Transportation Research Part B: Methodological* **109**, 238-256
- Bhat, C. R., Sidharthan, R. (2011). A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B: Methodological*, **45(7)**, 940-953
- Büscher, S., Bauer, D. (2024). Weighting strategies for pairwise composite marginal likelihood estimation in case of unbalanced panels and unaccounted autoregressive structure of the errors. *Transportation Research Part B: Methodological* **181**, 102890.
- Fox, J. (2017) A Note on Nonparametric Identification of Distributions of Random Coefficients in Multinomial Choice Models. Technical report, National Bureau of Economic Research, Cambridge, MA.
- Fox, J., il Kim, K., Ryan, St., and Bajari, P. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, **166**(2):204-212.
- Grün, B. and Leisch, F. (2008). Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. *Journal of Classification*, **25**(2):225-247.
- Heiss, F., Hetzenecker, S., Osterhaus, M. (2022). Nonparametric estimation of the random coefficients model: An elastic net approach. *Journal of Econometrics*, **229**(2), 299-321.
- Helveston, J. P., Liu, Y., Feit, E. M., Fuchs, E., Klampfl, E., Michalek, J. J. (2015). Will subsidies drive electric vehicle adoption? Measuring consumer preferences in the US and China. *Transportation Research Part A: Policy and Practice*, **73**, 96-112
- Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations, *Journal of the American Statistical Association* **90**(431): 957-964
- Mendell, N. and Elston, R. (1974). Multifactorial qualitative traits: Genetic analysis and prediction of recurrence risks, *Biometrics* **30**, 41-57
- Train, K. (2009) Discrete choice methods with simulation. Cambridge university press.
- Train, K. (2016) Mixed logit with a flexible mixing distribution. *Journal of Choice Modelling*, 19:40-53.
- Varin, C. (2008). On composite marginal likelihoods, *Advances in Statistical Analysis* **92**, 1-28.
- Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods, *Statistica Sinica* **21**, 5-42.
- Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection, *Biometrika* **92**(3): 519-528.

6. Literature

Thank you for your attention!

Questions:
Dietmar.Bauer@uni-bielefeld.de

UNIVERSITÄT KLAGENFURT

UNIVERSITÄT BIELEFELD
Fakultät für Wirtschaftswissenschaften

4th Klagenfurt-Bielefeld Summer School
Modern Topics in Time Series Analysis
Klagenfurt, September 16–20, 2024

Target audience and conditions of participation

The summer school's primary target audience are PhD students. However, subject to capacity the school also welcomes post-doctoral researchers (from universities, research institutes or (central) banks) and professionals with a strong mathematical and statistical background seeking to extend their knowledge base in time series analysis. A good preliminary knowledge of the main work-horses in time series analysis, e.g., estimation and specification of vector autoregressive models, is expected.

The summer school is limited to 40 participants. The fees for the summer school amount to **€550.00 (PhD students)** and **€700.00 (others)** - including all lectures and materials, lunches, snacks, coffee breaks and the social event.

For additional information and details on the application process see http://www.aau.at/econ/mtsa2024.

Program

The summer school provides a PhD-level introduction to a range of modern topics in time series analysis taught (each one on a half-day in two 90min lectures) by a selection of top researchers in Europe:

Topics (in alphabetical order)	Lecturer	Affiliation
Bayesian Methods for (Macro and Financial) Time Series Analysis	Gregor Kastner	University of Klagenfurt
Dynamic Factor Models	Manfred Deistler	Vienna University of Technology
Econometric Modelling of Climate Change	Eric Hillebrand	Aarhus University
Functional Time Series Analysis	Siegfried Hörmann	Graz University of Technology
Hidden Markov Models	Timo Adam	Bielefeld University
Nonlinear Cointegration	James Duffy	Oxford University
Seasonal and Trend Modelling for Forecasting	Harry Haupt	University of Passau
Introduction and Overview	Dietmar Bauer Martin Wagner	Bielefeld University University of Klagenfurt

Important Dates	
May 15, 2024	Deadline for Application
June 1, 2024	Notification of Acceptance
June 15, 2024	Registration

We gratefully acknowledge support by the **University of Klagenfurt Research Council** and **VEREIN** ZUR FÖRDERUNG DER **WIRTSCHAFTSWISSENSCHAFTEN** AN DER ALPEN-ADRIA-UNIVERSITÄT KLAGENFURT

Problems with identification

- Binary choice, $U_{i,t,0} = 0$.
- $U_{i,t,1} = \alpha_i + e_{i,t}$, $e_{i,t} = \rho e_{i-1,t} + \varepsilon_{i,t}$ (Gaussian AR(1) process, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma^2)$ iid)
- $\alpha_i \sim \mathcal{N}(0, \sigma_a^2)$ iid over deciders standard normally distributed.
- Hence $U_{i,t,1} \sim \mathcal{N}(0, V)$ where

$$V = \frac{\sigma^2}{1-\rho^2} + \sigma_a^2$$

- Two consecutive choices have an error term for the utilities to alternative 1:

$$\begin{pmatrix} \frac{\sigma^2}{1-\rho^2} + \sigma_a^2 & \frac{\rho\sigma^2}{1-\rho^2} + \sigma_a^2 \\ \frac{\rho\sigma^2}{1-\rho^2} + \sigma_a^2 & \frac{\sigma^2}{1-\rho^2} + \sigma_a^2 \end{pmatrix}$$

- Identify three parameters at max (scale normalization may add one restriction) from 1 or 3 equations.

Identification

- Level: only utility differences are identified, hence the level needs to be specified. $U_{i,t,1} = 0$ works.
- Scale: $cU_{i,t,j}$ leads to the same choices for $c > 0$. Often the variance is fixed to identify scale.

Mixing:

- Parametric mixing for MNL models is identified for different mixing distributions (Train, 2009) and also for non-parametric mixing (Train, 2016).
- Fox (2017) in the general case and Fox et al. (2012) for MNL models provide assumptions such that the mixing distribution is (non-parametrically) identified from cross-sectional data: at least one regressor needs to be supported in $\mathbb{R}$ where the coefficient is supported in subset of $\mathbb{R}_+$ (as would be typical for costs).
- Grün and Leisch (2008) show that in order to identify S latent classes in mixed MNL models the regressors need to take on at least $2S - 1$ different values.

System in Bhat and Sidhartan (2011)

$J = 5, K = 5, T \in \{2, 3, 4\}$

$$\beta = \begin{pmatrix} 1.5 \\ -1 \\ 2 \\ 1 \\ -2 \end{pmatrix}, \Omega = \begin{pmatrix} 1.0 & -0.5 & 0.25 & 0.75 & 0 \\ -0.5 & 1.0 & 0.25 & -0.5 & 0 \\ 0.25 & 0.25 & 1 & 0.33 & 0 \\ 0.75 & -0.5 & 0.33 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \Sigma = 0.5 I_5.$$

	Trace of Var for β			Trace of Var for Ω		
	$T = 2$	$T = 3$	$T = 4$	$T = 2$	$T = 3$	$T = 4$
MLE	0.91	0.47	0.31	3.24	1.66	0.94
FP	0.91	0.52	0.31	3.24	1.79	1.09
AP	0.91	0.62		3.24	2.15	
APU	0.99	0.65		3.52	2.28	
FP /MLE	1.00	1.10	1.01	1.00	1.08	1.16
AP /MLE	1.00	1.32		1.00	1.30	
APU /MLE	1.08	1.40		1.08	1.37	