Conditional likelihood inference for binary and ordinal responses: review and developments

Francesco Bartolucci

Department of Economics
University of Perugia (IT)

joint work with

Claudia Pigini and Francesco Valentini

Polytechnic University of Marche (IT)

January 19^{th}, 2024

Workshop in honor of Franco Peracchi

Outline

1. Preliminaries
   - Static logit model for binary and ordinal panel data
   - Conditional maximum likelihood (CML) estimation
   - Quadratic exponential model
2. Main extensions of CML in the context of panel data
   - Estimation of the dynamic logit model
   - Testing for true state dependence
   - Testing for time-invariant unobserved heterogeneity
3. Extension to network data
   - Model assumptions
   - MCMC-CML estimation procedure
4. Main conclusions and further developments

Preliminaries

Static logit model model for binary panel data

- For a sequence of binary responses y_{it} and corresponding vectors of covariates $\boldsymbol{x}_{it}$, the *static logit (SL) model* (see Hsiao, 2022, for a review) assumes that:

$$\begin{aligned} y_{it} &= \mathrm{I}\{y_{it}^* > 0\} \\ y_{it}^* &= \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, \dots, n, \ t = 1, \dots, T \end{aligned}$$

  - $\mathrm{I}\{\cdot\}$: indicator function
  - α_i: individual-specific intercept (*unobserved heterogeneity*), considered as a fixed parameter
  - $\boldsymbol{\beta}$: regression coefficients for the covariates
  - ε_{it}: idiosyncratic error term having standard logistic distribution

- *Incidental parameters problem* (Neyman and Scott, 1948): with fixed T, the maximum likelihood (ML) estimator of $\boldsymbol{\beta}$ is *not consistent*

Conditional maximum likelihood (CML) estimation

- The joint probability of $\boldsymbol{y}_i = (y_{i1}, \dots, y_{iT})'$ is

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i) = \prod_{t=1}^{T} \frac{\exp[y_{it}(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})]}{1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})} = \frac{\exp(\alpha_i y_{i+}) \exp\left[\left(\sum_t y_{it}\boldsymbol{x}_{it}\right)' \boldsymbol{\beta}\right]}{\prod_t [1 + \exp(\alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta})]},$$

  - $\boldsymbol{X}_i$: matrix of all covariates for unit i
  - $y_{i+} = \sum_t y_{it}$ (total score): *sufficient statistic* for the intercept α_i

- The *conditional probability* of $\boldsymbol{y}_i$, given y_{i+}, does not depend on α_i:

$$p(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i+}) = \frac{\exp\left[\left(\sum_t y_{it}\boldsymbol{x}_{it}\right)' \boldsymbol{\beta}\right]}{\sum_{\boldsymbol{z}: z_+ = y_{i+}} \exp\left[\left(\sum_t z_t \boldsymbol{x}_{it}\right)' \boldsymbol{\beta}\right]} = p(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i+})$$

  - $\sum_{\boldsymbol{z}: z_+ = y_{i+}}$: sum over all vectors of binary variables $\boldsymbol{z} = (z_1, \dots, z_T)'$ such that $z_+ = y_{i+}$, where $z_+ = \sum_t z_t$

- The corresponding *conditional log-likelihood* has expression

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{I}(0 < y_{i+} < T) \log p(\boldsymbol{y}_i | \boldsymbol{X}_i, y_{i+})$$

  - $\mathbf{I}(0 < y_{i+} < T)$ takes into account that observations whose total score is 0 or T do not contribute to the likelihood

- $\ell(\boldsymbol{\beta})$ is maximized with respect to $\boldsymbol{\beta}$ by the Newton-Raphson (NR) algorithm obtaining the *CML estimator* $\hat{\boldsymbol{\beta}}$, which is consistent for fixed T as $n \to \infty$ and with asymptotic normal distribution

- Standard errors for $\hat{\boldsymbol{\beta}}$ may be obtained from the *information matrix* that is of simple computation

Case of ordinal variables

- With *ordinal responses* having $J > 2$ categories (from 0 to $J-1$), y_{it} is tied to y_{it}^* through the *observation rule*

$$y_{it} = \sum_{j=0}^{J-1} j \cdot \mathrm{I}\{\tau_j \leq y_{it}^* < \tau_{j+1}\}$$

  - $\tau_0 < \tau_1 < \cdots < \tau_{J-1} < \tau_J$: *thresholds* ($\tau_0 = -\infty$, $\tau_J = \infty$)

- A *proportional odds regression model* (McCullagh, 1980) based on global-logits results:

$$\log \frac{p(y_{it} \geq j | \alpha_{it}, \mathbf{x}_{it})}{p(y_{it} < j | \alpha_{it}, \mathbf{x}_{it})} = \alpha_i + \mathbf{x}_{it}' \boldsymbol{\beta} + \tau_j^*, \quad j = 1, \dots, J-1$$

  - τ_j^*: intercepts related to the thresholds τ_j

- CML estimation is applied after the outcomes are *dichotomized* in all the possible ways (Baetschmann et al., 2015):

$$y_{it}^{(j)} = \mathbf{I}\{y_{it} \geq j\}, \quad j = 1, \ldots, J-1$$

- The *conditional log-likelihood* is modified as:

$$\ell(\boldsymbol{\beta}) = \sum_{j=1}^{J-1} \sum_{i=1}^{n} \log p(\mathbf{y}_i^{(j)} | \mathbf{X}_i, y_{i+}^{(j)})$$

  - $\mathbf{y}_i^{(j)} = (y_{i1}^{(j)}, \ldots, y_{iT}^{(j)})$: vector of dichotomized outcomes at level j

- This *dichotomization method* can be generally used in many contexts, so as to adapt models for binary data to ordinal data

Quadratic exponential model

- An extension of the SL model allowing for serial dependence is the *Quadratic Exponential (QE)* model defined in Bartolucci and Nigro (2010) and based on the family of distributions for multivariate binary data proposed by Cox (1972)
- The QE model directly *formulates* the conditional distribution of $\boldsymbol{y}_i$:

$$p(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp\left[y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\eta}_1 + y_{iT}\left(\phi + \boldsymbol{x}'_{iT}\boldsymbol{\eta}_2\right) + y_{i*}\psi\right]}{\sum_{\boldsymbol{z}} \exp[z_+\delta_i + \sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\eta}_1 + z_T\left(\phi + \boldsymbol{x}'_{iT}\boldsymbol{\eta}_2\right) + z_{i*}\psi]}$$

  - δ_i: individual specific intercept
  - $\boldsymbol{\eta}_1$: parameters for the covariates (ϕ and $\boldsymbol{\eta}_2$ are not of main interest)
  - ψ: parameter for the serial dependence
  - $y_{i*} = \sum_t y_{i,t-1}y_{it}$
  - $\sum_{\boldsymbol{z}}$: sum over all possible binary vectors $\boldsymbol{z} = (z_1, \dots, z_T)$

- The QE model *allows for state dependence and unobserved heterogeneity* other that the effect of covariates

- An important distinction is between:

  - *true state dependence* (Heckman, 1981): due the effect that experiencing a particular event in the present has on the probability of the same event occurring in the future (e.g., direct effect of being employed this year has on the probability of being employed next year)

  - *spurious state dependence*: corresponding to the individual time-invariant unobserved heterogeneity due to unobservable factors (e.g., motivational factors)

- Under the QE model, each y_{i+} (total score) is a *sufficient statistic* for the incidental parameter δ_i (main advantage with respect to other extensions of the SL model for state dependence)

Main extensions of CML in the context of panel data

1. Estimation of the dynamic logit model

- The *dynamic logit (DL) model* assumes:

$$
\begin{aligned}
y_{it} &= \mathbf{I}\{y_{it}^* > 0\} \\
y_{it}^* &= \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma + \varepsilon_{it}, \quad i = 1, \ldots, n, \; t = 1, \ldots, T
\end{aligned}
$$

  - γ: regression coefficient for the lag-response (*true state dependence*)

- Differently from the SL model ($\gamma = 0$), the total score y_{i+} *is not a sufficient statistic* for the incidental parameter α_i

- CML estimation is viable:
  - in *absence of covariates* with $T = 3$ (Chamberlain, 1980)
  - *with covariates* on basis of a weighted conditional log-likelihood (Honoré and Kyriazidou, 2000); the (HK) estimator is consistent only under certain conditions on the covariates (time dummies cannot be included), but the rate of convergence is slower than $\sqrt{n}$

Pseudo CML (PCML) estimation of the DL model

- Bartolucci and Nigro (2012) proposed an *estimation method* of the structural parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}', \gamma)'$ of the DL model based on approximating it by a modified version of the QE model

- The approximating QE model is based on a *Taylor series expansion* of the log-probability of $\boldsymbol{y}_i$ under the DL model around $\alpha_i = \bar{\alpha}_i$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$

- *Probability* of $\boldsymbol{y}_i$ under the approximating model:

$$\tilde{p}(\boldsymbol{y}_i|\alpha_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\alpha_i + \sum_t y_{it}\boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i*}\gamma - \sum_{t>1} \bar{q}_{it} y_{i,t-1}\gamma)}{\sum_{\boldsymbol{z}} \exp(z_+\alpha_i + \sum_t z_t\boldsymbol{x}_{it}'\boldsymbol{\beta} + z_{i*}\gamma - \sum_{t>1} \bar{q}_{it} y_{i,t-1}\gamma)}$$

$$\bar{q}_{it} = \frac{\exp(\bar{\alpha}_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}})}{1 + \exp(\bar{\alpha}_i + \boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}})} :$$ *probability of success* at occasion t

Features of the approximating model

- The model is *similar to the initial QE model* (Bartolucci and Nigro, 2010): the main difference is the inclusion of the correction term $-\sum_{t>1} \bar{q}_{it} y_{i,t-1}\gamma$
- The approximating model *coincides with the true DL model* when $\gamma = 0$ (absence of state dependence) and in general the two models share many common properties in terms of dependence between the response variables
- Under the approximating model, each y_{i+} is a *sufficient statistic* for the incidental parameter α_i; then, the incidental parameters may be removed by conditioning on these statistics

Two-step estimator

- The PCML estimator is based on *two steps*:
  1. obtain $\bar{\boldsymbol{\beta}}$ as the CML estimate of $\boldsymbol{\beta}$ under the SL model, by maximizing

$$\bar{\ell}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathrm{I}\{0 < y_{i+} < T\}\bar{\ell}_i(\boldsymbol{\beta}),$$

$$\bar{\ell}_i(\boldsymbol{\beta}) = \log \frac{\exp(\sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\beta})}{\sum_{\boldsymbol{z}:z_+=y_{i+}} \exp(\sum_t z_t \boldsymbol{x}'_{it}\boldsymbol{\beta})},$$

  and the corresponding $\bar{\alpha}_i$ by ML estimation, so as to *obtain the $\bar{q}_{it}$*

  2. estimate $\boldsymbol{\theta}$ by maximizing the conditional log-likelihood of the approximating model

$$\tilde{\ell}(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \mathrm{I}\{0 < y_{i+} < T\}\tilde{\ell}_i(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}),$$

$$\tilde{\ell}(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) = \log \tilde{p}_{\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}}(\boldsymbol{y}_i|\boldsymbol{X}_i, y_{i0}, y_{i+}),$$

  by a *simple NR algorithm* similar to that used for CML estimation of the initial QE model

- Asymptotic properties of the *PCML estimator* $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}', \tilde{\gamma})'$ are studied under the DL model, showing that when $\gamma_0 = 0$ the estimator is consistent; otherwise the estimator converges to a point "close" to the true parameter value

- *Finite-sample properties* are studied by simulation under different settings; main conclusions:
  - the PCML estimator has *negligible bias* when γ_0 is close to 0 and a *reduced bias* even for values of γ_0 rather different from 0
  - *confidence intervals* based on the PCML estimator usually attain the nominal coverage level even for γ_0 far from 0
  - the PCML estimator *outperforms* the HK estimator, in particular for short panels and for high values of γ

- The PCML estimator is also much *simpler to compute* than the HK estimator and can be used with $T \geq 2$ instead of $T \geq 3$ and with no limitations on the covariate structure

2. Testing for true state dependence

- Apart from estimation of the state dependence parameter under the DL model, an important issue is testing that $H_0 : \gamma = 0$ (*absence of state dependence*)
- The nonparametric test proposed by Halliday (2007) is based on the construction of *conditional probability inequalities* that depend only on the sign of the state dependence parameter γ under the DL model, *avoiding distributional assumptions* on the unobserved heterogeneity parameters
- The test *cannot be directly generalized* to $T > 2$ and there are some difficulties in the presence of individual covariates

- In principle, both (initial and modified for PCML) versions of the *QE model* (Bartolucci and Nigro, 2010, 2012) could be use to test H_0 by a t-test

- The reason why the QE models (initial and modified) may be used to test for state dependence is that they *include the SL model* as a special case and then the DL model under H_0

- The same happens for a *simplified version*, named QE1, based on the assumption

$$p_1(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\phi} + y_{i*}\psi)}{\sum_{\boldsymbol{z}} \exp(z_+\delta_i + \sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\phi} + z_{i*}\psi)}$$

- Hypothesis H_0 *may be tested* on the basis of the QE1 model:

  1. ϕ and ψ are estimated by the *CML method* based on the NR algorithm similar to that available for the initial QE model (the estimator is $\sqrt{n}$ consistent), obtaining $\hat{\phi}_1$ and $\hat{\psi}_1$
  2. the *test statistic*

     $$W_1 = \frac{\hat{\psi}_1}{\mathrm{se}(\hat{\psi}_1)}$$

     is used (it has asymptotic $N(0,1)$ distribution under H_0)

- Though the test is asymptotically unbiased and may be used even with $T > 2$ and covariates, Bartolucci et al. (2023) noted that it is *less powerful* than that proposed by Halliday (2007) in certain particular cases

A modified version of the QE model: QE2 model

- A *different version of the QE model* (Bartolucci et al., 2023), named QE2, is defined as

$$p_2(\boldsymbol{y}_i|\delta_i, \boldsymbol{X}_i, y_{i0}) = \frac{\exp(y_{i+}\delta_i + \sum_t y_{it}\boldsymbol{x}'_{it}\boldsymbol{\phi} + \tilde{y}_{i*}\psi)}{\sum_{\boldsymbol{z}} \exp(z_+\delta_i + \sum_t z_t\boldsymbol{x}'_{it}\boldsymbol{\phi} + \tilde{z}_{i*}\psi)}$$

  - $\tilde{y}_{i*} = \sum_t \mathbf{I}\{y_{it} = y_{i,t-1}\}$

- The difference between QE1 and QE2 is in how the *association between the response variables* is accounted for

- In QE2, it is based on the statistic $\tilde{y}_{i*}$ that, differently from y_{i*}, is equal to the *number of consecutive pairs of outcomes which are equal each other*, regardless if they are 0 or 1

- The different association modeling allows us to use a *larger set of information* with respect to the basic model QE1 in testing for state dependence

- *CML estimation* of the QE2 model and *testing for state dependence* is performed in a very similar way as for the QE1 model

- Once the parameters of QE2 are estimated, we use a *t-statistic* of type

$$W_2 = \frac{\hat{\psi}_2}{\mathrm{se}(\hat{\psi}_2)}$$

- Under the DL model with strictly exogenous covariates, if $H_0 : \gamma = 0$ holds, the test statistic W_2 has *asymptotic distribution* $N(0, 1)$ as $n \to \infty$

- If $\gamma \neq 0$, W_2 is expected to *diverge* to $+\infty$ or $-\infty$ according to whether γ is positive or negative

- We also proposed an extension of the test to deal with:
  - *ordered response variables* (considering all possibly dichotomizations of the response categories)
  - *predetermined covariates* (when $\boldsymbol{x}_{it}$ may depend on $y_{i,t-1}$)

- The *finite sample properties of the method are studied* by simulation, showing that:
  - the test based on QE2 model has *more power* than the test based on QE1 model
  - both tests always attain the *nominal significant level* under H_0
  - when $T = 2$ and in absence of covariates, the test of Halliday (2007) and that based on QE2 model have a *very similar behavior*
  - in the other cases the Halliday's test has an unsatisfactory behavior and, in particular with covariates, it has a *wrong significance level* under H_0

3. Testing for time-invariant unobserved heterogeneity

- An important element when modeling panel data is the treatment of *unobserved heterogeneity*, that is typically assumed to to be time invariant

- A more general version of the SL model is based on assuming that y^*_{it} follows the *linear model*

$$y^*_{it} = \alpha_{it} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T$$

  - α_{it}: unobservable unit-specific time-varying effects

- y_{it} is related to y^*_{ij} by the same *observation rules* introduced for the SL model depending on the binary or ordinal nature of the responses

Hypotheses of interest

- *Null hypothesis (H_0)*: Unit-specific unobserved heterogeneity is *constant over time*

$$\alpha_{i1} = \alpha_{i2} = \cdots = \alpha_{iT} = \alpha_i, \qquad i = 1, \ldots, n$$

- *Alternative hypothesis (H_1)*: Unit-specific unobserved heterogeneity is *time-varying*, with no *a priori* assumptions on how it evolves over time

- Bartolucci et al. (2015) proposed a test of H_0 vs H_1 based on the comparison of *standard* and *pairwise* CML estimators of $\boldsymbol{\beta}$

Standard and pairwise CML estimators

- Regarding *standard CML estimator* $\hat{\boldsymbol{\beta}}_1$:
  - under H_0, $\hat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}_0$
  - under H_1, $\hat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}_{1*} \neq \boldsymbol{\beta}_0$

- When y_{it} is *binary*, the *pairwise* conditional log-likelihood has expression

$$\ell_2(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{t=2}^{T} \log p(y_{i,t-1}, y_{it} | \mathbf{x}_{i,t-1}, \mathbf{x}_{it}, y_{i,t-1} + y_{it})$$

- Regarding the *pairwise CML estimator* $\hat{\boldsymbol{\beta}}_2$:
  - under H_0, $\hat{\boldsymbol{\beta}}_2 \xrightarrow{p} \boldsymbol{\beta}_0$
  - under H_1, $\hat{\boldsymbol{\beta}}_2 \xrightarrow{p} \boldsymbol{\beta}_{2*} \neq \boldsymbol{\beta}_0$, with $\boldsymbol{\beta}_{2*} \neq \boldsymbol{\beta}_{1*}$ (in general)

Test implementation

- We first consider the *difference*: $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2$
- *Under H_0*, both estimators are consistent: $\hat{\boldsymbol{\delta}} \xrightarrow{p} \mathbf{0}$ and $\sqrt{n}\hat{\boldsymbol{\delta}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V})$
- *Under H_1*, the two estimators diverge: $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_* \neq \mathbf{0}$ and $\sqrt{n}\hat{\boldsymbol{\delta}} \xrightarrow{d} \mathcal{N}(\boldsymbol{\delta}_*, \mathbf{V}_*)$
- The *test statistic* Hausman (1978) is then

$$U = n\hat{\boldsymbol{\delta}}'\hat{\mathbf{V}}^{-}\hat{\boldsymbol{\delta}}$$

  - $\hat{\mathbf{V}}^{-}$: generalized inverse of the estimate of $\mathbf{V}$
- *Under H_0*, $U \xrightarrow{d} \chi^2_{k^*}$
  - $k^* = \text{rank}(\mathbf{V}) \leq k$

- The test may be also used with *ordinal responses* by using the dichotomization rule
- Being a pure specification test, the test may *lack power* in some cases (Holly, 1982)
- One such case is when unobserved heterogeneity effects are *serially correlated* but the first-order autocorrelation is zero (e.g., some form of seasonality)
- To handle this case, we may generalize the test by considering *all possible pairs* of observations for the same unit (in progress)
- We analyzed the *size* and *power* of the proposed test using a set of *Monte Carlo experiments*, confirming the good properties of the proposed method

Extension to network data

Model assumptions

- *Network models* are able to represent links between agents and have application in trade and other fields (see De Paula, 2020, for a review)
- In the network, *n nodes (agents)* are observed and each pair of nodes (i, j) can form a *link* without self-ties
- A single link is represented as

$$y_{ij} = \begin{cases} 1 & \text{if } i \text{ is connected to } j \\ 0 & \text{otherwise} \end{cases}$$

  and the whole network is represented by the $n \times n$ *adjacency matrix* $\boldsymbol{Y}$ with elements y_{ij}
- *Directed* or *undirected* ($y_{ij} = y_{ji}$) networks can be considered

- The *model of the link formation* assumes that:

$$
\begin{aligned}
y_{ij} &= \mathbf{I}\{y^*_{ij} > 0\} \\
y^*_{ij} &= \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \gamma_j + \varepsilon_{ij}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T
\end{aligned}
$$

  - $\boldsymbol{x}_{ij}$: a vector of k dyad-specific covariates
  - $\boldsymbol{\beta}$: homophily parameters
  - α_i, γ_j: sender and receiver fixed-effects
  - ε_{ij}: idiosyncratic shock following a standard logistic distribution

Likelihood function

- The *likelihood* for an observed adjacency matrix $\boldsymbol{Y}$ has expression

$$L(\boldsymbol{\beta}) = p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \prod_{i=1}^{n}\prod_{j=1}^{n} p(y_{ij}|\boldsymbol{x}_{ij}, \alpha_i, \gamma_j) =$$

$$\frac{\exp\left(\sum_{i=1}^n \sum_{j=1}^n y_{ij}\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sum_{i=1}^n y_{i+}\alpha_i + \sum_{j=1}^n y_{+j}\gamma_j\right)}{\prod_{i=1}^n \prod_{j=1}^n [1 + \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \gamma_j)]}$$

  - $\boldsymbol{X}$ collects $\boldsymbol{x}_{ij}$, $i, j = 1, \ldots, n$, $i \neq j$
  - $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)'$, $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)'$
  - $y_{i+} = \sum_{j=1}^n y_{ij}$: number of outgoing links from i
  - $y_{+j} = \sum_{i=1}^n y_{ij}$: number of incoming links to j

- As usual the ML estimation is affected by the *incidental parameters problem*

Conditional Likelihood function

- *Sufficient statistics* for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$:
  - $\boldsymbol{y}_+ = (y_{1+}, \dots, y_{n+})'$: outdegree sequence of the nodes
  - $\boldsymbol{y}_{(+)} = (y_{+1}, \dots, y_{+n})'$: indegree sequence of the nodes

- The *conditional likelihood* has expression

$$p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{y}_+, \boldsymbol{y}_{(+)}) = \frac{p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\alpha}, \boldsymbol{\gamma})}{p(\boldsymbol{y}_+, \boldsymbol{y}_{(+)}|\boldsymbol{X}, \boldsymbol{\alpha}, \boldsymbol{\gamma})} = \frac{\exp\left(\sum_{i=1}^n \sum_{j=1}^n y_{ij} \boldsymbol{x}'_{ij} \boldsymbol{\beta}\right)}{\sum_{\boldsymbol{Z}} \exp\left(\sum_{i=1}^n \sum_{j=1}^n z_{ij} \boldsymbol{x}'_{ij} \boldsymbol{\beta}\right)}$$

  - $\sum_{\boldsymbol{Z}}$: sum over all possible adjacency matrices $\boldsymbol{Z}$ with the same degree sequences as $\boldsymbol{Y}$; this sum makes the conditional likelihood *computationally intractable*

- The *available solution* (Charbonneau, 2017) consists in:
  - scanning all possible 2x2 *submatrices* in $\boldsymbol{Y}$ and picking those having a two different values (0,1) in each row and column
  - building a *composite conditional likelihood* referred to all these submatrices $\ell_C(\boldsymbol{\beta})$
  - *maximizing* $\ell_C(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ by a NR algorithm
- The procedure may be seen as an *extension* for networks of the method proposed by Bartolucci et al. (2015)

MCMC-CML estimation procedure

- Bartolucci et al. (2024) proposed a *Markov Chain Monte Carlo (MCMC) approximation* (Geyer, 1991) of the conditional likelihood function, getting rid of the intractable normalizing constant

- Starting from a fixed parameter vector $\bar{\boldsymbol{\beta}}$, chosen as the ML estimate of $\boldsymbol{\beta}$, the *proposed method* consists in:
  - *sampling $n \times n$ binary matrices* having the same values of the sufficient statistics as $\boldsymbol{Y}$: this is performed by a Metropolis algorithm (Metropolis et al., 1953) based on switching the values of certain randomly selected rectangles (Diaconis and Gangolli, 1995) and hexagons (Rao et al., 1996) in $\boldsymbol{Y}$
  - *maximizing a likelihood ratio* between the conditional likelihood at $\boldsymbol{\beta}$ and that at $\bar{\boldsymbol{\beta}}$: this is performed by a simple NR algorithm

- The estimator is *consistent* and has *asymptotically normal* distribution under standard assumptions and given that the number of MCMC iterations goes to ∞

- *Good finite sample properties* are established by simulation in comparison to the ML and the composite likelihood estimator, and also in comparison to the bias corrected estimator of Fernández-Val and Weidner (2016)

- Overall the proposal overcomes computational intractability of CML, although there is a *computational burden* of the Metropolis Algorithm

Main conclusions and further developments

- *Main features the CML method*:
  - + straightforward solution to the *incidental parameters problem* with repeated observations for the same units (panel, network)
  - + the estimator is of *simple implementation* (`cquad` package in `R` and `Stata`, Bartolucci and Pigini, 2017; Bartolucci et al., 2020)
  - – the method crucially depends on the *assumption of logistic distribution* for the error terms
  - – as for any other fixed-effects method, regression coefficients of the *covariates constant within the same unit* are not estimable
- Nowadays a competitive alternative is represented by *bias corrected estimators* (Fernández-Val, 2009; Fernández-Val and Weidner, 2016), which are more flexible in terms of assumptions, although of more complex implementation and less efficient for specific models (Bartolucci et al., 2016; Valentini et al., 2023, for a recent review)
- There is still *room for application* of CML models for complex data structures (e.g., dynamic network models for repeated adjacency matrices across time)

References

References I

Baetschmann, G., Staub, K. E., and Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society: Series A*, 178:685–703.

Bartolucci, F., Bellio, R., Salvan, A., and Sartori, N. (2016). Modified profile likelihood for fixed-effects panel data models. *Econometric Reviews*, 35:1271–1289.

Bartolucci, F., Belotti, F., and Peracchi, F. (2015). Testing for time-invariant unobserved heterogeneity in generalized linear models for panel data. *Journal of Econometrics*, 184:111–123.

Bartolucci, F. and Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator. *Econometrica*, 78:719–733.

Bartolucci, F. and Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data. *Journal of Econometrics*, 170:102–116.

Bartolucci, F. and Pigini, C. (2017). cquad: An R and Stata package for conditional maximum likelihood estimation of dynamic binary panel data models. *Journal of Statistical Software*, 78(i07).

Bartolucci, F., Pigini, C., and Valentini, F. (2020). Cquadr: Stata module to estimate quadratic exponential models running the cquad R package.

References II

Bartolucci, F., Pigini, C., and Valentini, F. (2023). Testing for state dependence in the fixed-effects ordered logit model. *Economics Letters*, 222:110964.

Bartolucci, F., Pigini, C., and Valentini, F. (2024). MCMC conditional maximum likelihood for the two-way fixed-effects logit. Forthcoming *Econometric Reviews*.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47:225–238.

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal*, 20:S1–S13.

Cox, D. (1972). The analysis of multivariate binary data. *Applied Statistics*, 21:113–120.

De Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics*, 12:775–799.

Diaconis, P. and Gangolli, A. (1995). Rectangular arrays with fixed margins. In *Discrete Probability and Algorithms*, pages 15–41. Springer.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150:71–85.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics*, 192:291–312.

References III

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In *Proceedings of the 23rd Symposium on the Interface*, pages 156–163. Interface Foundation of North America.

Halliday, T. J. (2007). Testing for state dependence with time-variant transition probabilities. *Econometric Reviews*, 26:685–703.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 46:1251–1271.

Heckman, J. J. (1981). Heterogeneity and state dependence. In *Studies in labor markets*, pages 91–140. University of Chicago Press.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68:839–874.

Hsiao, C. (2022). *Analysis of panel data*. Cambridge university press, 4th edition.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B*, 42:109–127.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092.

References IV

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32.

Rao, A. R., Jana, R., and Bandyopadhyay, S. (1996). A Markov chain Monte Carlo method for generating random (0, 1)-matrices with given marginals. *Sankhyā, Series A*, 58:225–242.

Valentini, F., Pigini, C., and Bartolucci, F. (2023). Advances in maximum likelihood estimation of fixed-effects binary panel data models. In *Trends and Challenges in Categorical Data Analysis: Statistical Modelling and Interpretation*, pages 275–315. Springer.