MCMC Conditional Maximum Likelihood for the two-way fixed-effects logit

Bartolucci, Francesco and Pigini, Claudia and Valentini, Francesco

6 October 2021

MCMC Conditional Maximum Likelihood for the two-way fixed-effects logit

Francesco Bartolucci
Università di Perugia (IT)
francesco.bartolucci@unipg.it

Claudia Pigini
Università Politecnica
delle Marche and MoFiR (IT)
c.pigini@univpm.it

Francesco Valentini
Università Politecnica
delle Marche (IT)
f.valentini@staff.univpm.it

Abstract

We propose a Markov chain Monte Carlo Conditional Maximum Likelihood (MCMC-CML) estimator for two-way fixed-effects logit models for dyadic data. The proposed MCMC approach, based on a Metropolis algorithm, allows us to overcome the computational issues of evaluating the probability of the outcome conditional on nodes in and out degrees, which are sufficient statistics for the incidental parameters. Under mild regularity conditions, the MCMC-CML estimator converges to the exact CML one and is asymptotically normal. Moreover, it is more efficient than the existing pairwise CML estimator. We study the finite sample properties of the proposed approach by means of a simulation study and three empirical applications, where we also show that the MCMC-CML estimator can be applied to binary logit models for panel data with both subject and time fixed effects. Results confirm the expected theoretical advantage of the proposed approach, especially with small and sparse networks or with rare events in panel data.

Directed network, Fixed effects, Link formation, Metropolis algorithm, Panel data

JEL Classification: C23, C25, C63

1 Introduction

Network models have received increasing attention in the economic literature, as these models can effectively represent links and interactions between economic agents that emerge in several fields of application such as trade (Helpman et al., 2008), risk sharing (Fafchamps and Gubert, 2007; Attanasio et al., 2012), knowledge spillovers (Zacchia, 2020), financial contagion (Acemoglu et al., 2015), and diffusion of microfinance (Banerjee et al., 2013). Such an interest has spurred the development of related statistical and econometric methods, especially for the analysis on network formation (see De Paula, 2020, for a recent and extensive review).

The model typically applied is based on the directed Erdős-Rényi random graph (Erdős and Rényi, 1960) where the probability of link formation is a function of node-specific parameters, capturing the network agents heterogeneity, and observable covariates that vary with each pair of nodes. The parameters associated with these variables represent homophily, which is the attitude of agents to create links with other agents that have similar characteristics. This model can be seen as an extension of the popular p_1 model proposed by Holland and Leinhardt (1981) to include covariates. Consistent Maximum Likelihood (ML) estimation of all the model parameters is however hampered by the incidental parameters problem (Neyman and Scott, 1948; Lancaster, 2000), which arises in fixed-effects models as the number of parameters for the node heterogeneity increases with the number of nodes. The asymptotic properties of the ML estimator for a model of link formation in the directed case are established by Yan et al. (2019).

Different strategies have been put forward to overcome the incidental parameter problem so as to obtain either a consistent or bias reduced estimate of the model parameters. With regards to the latter, Yan et al. (2019) and Dzemski (2019) consider an analytical bias correction for the ML estimator that builds on Fernández-Val and Weidner (2016), who studied the incidental parameters problem in nonlinear panel data models with both individual and time fixed effects. Similarly, for the undirected case, Graham (2017) adapts the iterated bias correction technique developed by Hahn and Newey (2004) for nonlinear panel data models with individual heterogeneity parameters. Despite their ease of implementation, the estimation approaches based on bias corrections are viable only with dense graph sequences (Graham, 2017), a condition required for the estimation of the node-specific parameters. However most social and economic networks are sparse, meaning that only a small fraction of all possible links are observed. Also, bias corrected estimators have been shown to exhibit poor finite sample properties in such cases (Jochmans, 2018).

An alternative approach to bias-corrections is based on conditioning the link formation logit probability on suitable sufficient statistics to eliminate the incidental parameters representing node heterogeneity. In this vein is the strategy adopted by Graham (2017) for undirected networks and by Charbonneau (2017) and Jochmans (2018) for the directed case. In particular, the conditional estimator of the homophily parameters proposed by Charbonneau (2017) maximizes the pairwise or quasi-likelihood (P-CML, hereafter), which does not depend on the node-specific parameters. Doing so allows them to avoid the specification of the full conditional likelihood, which would arise as a generalization of the approaches put forward by Rasch (1961) and Chamberlain (1980) specifically to binary panel data models. Differently from the quasi-likelihood method, the Conditional ML (CML) approach uses the number of incoming and outgoing links of all nodes as sufficient statistics for the heterogeneity parameters, but the resulting likelihood is computationally intractable. The bulk of the problem rests on the enumeration of binary tables with fixed marginal sums, a challenging task that has been debated in the literature (Wang and Zhang, 1998; Pérez-Salvador et al., 2002). Nevertheless, the strategy of conditioning on sufficient statistics is preferable because applicable to sparse as well as dense network, as it does not require to estimate the incidental parameters.

In this paper we overcome the computational issue of evaluating the full conditional likelihood of a logit model for network formation with homophily parameters and node heterogeneity. Once the incidental parameters are eliminated by conditioning on the nodes out and in degree, the estimator for the homophily parameters is then obtained by maximizing the resulting likelihood which is approximated by Markov chain Monte Carlo (MCMC). MCMC-ML methods have been proposed by Geyer (1991, 1992); Geyer and Thompson (1992); Geyer (1994); Huffer and Wu (1998), while a review of MCMC methods for conditional inference is provided by Caffo and Booth (2003). We show that under mild regularity conditions the MCMC-CML estimator converges to the full CML estimator and has an asymptotic normal distribution.

The standard inferential framework for the conditional logit model (Andersen, 1970) does not apply to the pairwise likelihood method considered by Charbonneau (2017) and Jochmans (2018). In particular, in deriving consistency and limit distribution of the resulting estimator, Jochmans (2018) shows that uniform convergence of the objective function rests on the assumption that the accumulation of the informative quadruples, corresponding to any pair of senders forming one out of two possible links with any two receivers, does not cease as the sample size grows. This rate condition becomes relevant with sparse networks, whereas it is not at all required within the standard CML framework. Furthermore, since the pairwise likelihood is a misspecified density function giving rise to unbiased estimating equations, the resulting estimator is consistent but generally less efficient than the full ML one (Varin, 2008; Varin et al., 2011). The practitioner dealing with socio-economic applications based on network models could actually benefit from the use of the standard CML estimator, as the datasets in this field typically depict small and sparse networks.

We investigate the finite-sample properties of the MCMC-CML estimator by means of a simulation study based the same design presented by Charbonneau (2017). Results show that the MCMC-CML estimator proves to be effective in overcoming the incidental parameters problem, as its small sample bias is negligible and comparable to that of the P-CML estimator. Moreover, the MCMC-CML brings an efficiency gain with respect to the pairwise CML, which is reflected by a sensible difference in the estimators standard deviations.

We also consider two empirical applications on network data. The first concerns a model for the formation of trade relationships between countries, based on data provided by Helpman et al. (2008); the same exercise has been considered by Charbonneau (2017) and Jochmans (2018), as an application of the P-CML, and by Chen et al. (2021) in the context of interactive fixed effects. The second application is based on the sparse network of US attorneys provided by (Lazega et al., 2001), where links are formed whenever professional consultations occur. It emerges that the proposed MCMC-CML estimator is able to reproduce the results obtained by the alternative methods, with the advantage of offering more precise estimates than those obtained by P-CML, especially when the network is small and sparse.

It is worth noticing that the proposed approach can also be applied to the estimation of panel data logit models with both individual and time fixed effects that have been considered in the recent stream of literature regarding panel factor models for binary data (Fernández-Val and Weidner, 2016; Boneva and Linton, 2017; Jochmans and Otsu, 2019; Ando et al., 2021; Chen et al., 2021). In particular, under the assumption of additive fixed effects, the proposed estimator is the conditional counterpart of the analytical and jackknife bias corrections by Fernández-Val and Weidner (2016). We show that the MCMC-CML approach can be applied to a binary panel logit model with both subject and time fixed effects by means of an empirical application based on banking crises data provided by Laeven and Valencia (2018) to identify early warning signs of financial distress.

The remainder of the paper is organized as follows: Section 2 briefly describes the logit model for network formation and establishes notation; Section 3 illustrates the CML estimation and its computational issues; Section 4 describes the proposed MCMC-CML approach; Section 5 contains the simulation study and Section 6 the empirical applications; Finally Section 7 concludes.

2 Background

In this section we briefly describe the logit model for network formation, that has also been considered by Charbonneau (2017) and Jochmans (2018). Assume we observe a random sample of n nodes, describing agents such as individuals, firms, banks, or countries, and consider the possibility that a pair of nodes (i, j) forms a link (edge), with $i, j = 1, \dots, n$ and $i \neq j$. For each pair define $y_{ij} = 1$ if node i is connected (linked) to node j and $y_{ij} = 0$ otherwise. We assume that the connections are directed, allowing for the possibility that $y_{ij} \neq y_{ji}$; yet the application of the proposed approach to the analysis of an undirected network, that is $y_{ij} = y_{ji}$, is straightforward. The set of connections can be represented by the adjacency matrix $\boldsymbol{Y}$, defined as a $n \times n$ matrix such that each entry $\boldsymbol{Y}(i, j) = y_{ij}$. We exclude the possibility of self-ties, so that we characterize the elements on the main diagonal of $\boldsymbol{Y}$ as missing values, which also allows us to avoid abuse of notation when specifying the sample likelihood function.

Following Jochmans (2018), the link formation decision for a given dyad takes the threshold-crossing form $y_{ij} = \mathbf{I}(s_{ij}) > 0$, where $I(\cdot)$ is the indicator function and s_{ij} is the surplus generated by the connection between the nodes i and j. The surplus is defined as

$$s_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \gamma_j + \varepsilon_{ij}, \tag{1}$$

where α_i and γ_j denote node-specific unobserved traits, $\boldsymbol{x}_{ij}$ denotes a set of k dyad-specific covariates associated with the vector of the homophily parameters $\boldsymbol{\beta}$, and ε_{ij} is an idiosyncratic component. Under the assumption that ε_{ij} follows a standard logistic distribution, we can define the probability of link formation for the dyad i, j as

$$p(y_{ij} = 1 | \boldsymbol{x}_{ij}, \alpha_i, \gamma_j) = \frac{\exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \gamma_j\right)}{1 + \exp\left(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha_i + \gamma_j\right)}, \tag{2}$$

where the node heterogeneity components α_i and γ_j are treated as fixed and enter in fact the conditioning set. Such model is coherent with the Erdős-Rényi random graph and represents an extension of the p_1 model by Holland and Leinhardt (1981), in that the probability of link formation is a function of the node heterogeneity and observable dyad-specific covariates. Notice that the model described by Equation (2) can also be used to specify a logit panel data model with two-way fixed effects (Fernández-Val and Weidner, 2016), by letting i denote the subject and j the time occasion, for $i = 1, \dots, N$ and $j = 1, \dots, T$, with $\boldsymbol{x}_{ij}$ collecting a set of time-varying covariates. The panel data counterpart of the adjacency matrix has dimension $N \times T$ and is not restricted to be a square matrix.

According to the model reported in (2), the likelihood function for the sample is

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

$$\frac{\exp\left[\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sum_{i=1}^{n}(y_{i+}\alpha_i) + \sum_{j=1}^{n}(y_{*j}\gamma_j)\right]}{\prod_{i=1}^{n}\prod_{j=1}^{n}[1+\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \alpha_i + \gamma_j)]}, \tag{3}$$

where $\boldsymbol{Y}$ and $\boldsymbol{X}$ collect y_{ij} and $\boldsymbol{x}_{ij}$, $\forall i, j$, with $i \neq j$, respectively, and where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)'$ and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)'$ are vectors collecting the node heterogeneity parameters. Moreover, $y_{i+} = \sum_{j=1}^{n} y_{ij}$ denotes the number of outgoing links from node i, which is the sum of the dependent variable for row i of the adjacency matrix. Similarly, $y_{+j} = \sum_{i=1}^{n} y_{ij}$ is the number of incoming links to node j, which is the sum of y_{ij} for column j of the $\boldsymbol{Y}$ matrix. In the panel setting, the quantities y_{i+} and y_{+j} would be referred to as the row and the column *total scores*, respectively.

Other than Charbonneau (2017) and Jochmans (2018), who also consider a conditional approach, the same model is studied by Dzemski (2019) to develop the analytical bias correction for the ML estimator that, however, is only viable when the network is dense. The above model is instead different from the one considered by Graham (2017), which describes an undirected network.

3 Conditional Maximum Likelihood

In the following we describe the conditional inference approach to the estimation of the two-way fixed-effects logit model for network formation described by (2). We show that the conditional likelihood does not depend on the incidental parameters after conditioning on the nodes outdegrees and indegrees. Furthermore, we pin down the conditions under which the related CML estimator is consistent and asymptotically normal.

For ease of notation, let $\boldsymbol{y}_+ = (y_{1+}, \dots, y_{n+})'$ be the vector collecting all the nodes outdegrees (or row total scores) and $\boldsymbol{y}_{(+)} = (y_{+1}, \dots, y_{+n})'$ be the one containing all nodes indegrees (or column total scores). Furthermore, let $\boldsymbol{Z}$ indicate a generic adjacency matrix of dimension $n \times n$ and $\sum_{\boldsymbol{Z}}$ denote the sum across all the possible sample configurations such that $\boldsymbol{Z}$ has the same row and column totals as $\boldsymbol{Y}$, that is $\boldsymbol{Z} : \boldsymbol{z}_+ = \boldsymbol{y}_+, \boldsymbol{z}_{(+)} = \boldsymbol{y}_{(+)}$.

Lemma 1 (Sufficiency)**.** *For the model described in Equations* (2) *and* (3)*,* $\boldsymbol{y}_+$ *and* $\boldsymbol{y}_{(+)}$ *are sufficient statistics for the incidental parameters, so that the conditional probability* $p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{y}_+, \boldsymbol{y}_{(+)})$ *does not depend on* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\gamma}$*, that is*

$$p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{y}_+, \boldsymbol{y}_{(+)}) = \frac{\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}\right)}{\sum_{\boldsymbol{Z}}\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{n} z_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}\right)}. \tag{4}$$

Proof. Define the joint probability of observing $\boldsymbol{Y}$, $\boldsymbol{y}_+$ and $\boldsymbol{y}_{(+)}$ as $p(\boldsymbol{Y}\cap\boldsymbol{y}_+\cap\boldsymbol{y}_{(+)}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma})$. From the Bayes rule, we have that

$$p(\boldsymbol{Y}\cap\boldsymbol{y}_+\cap\boldsymbol{y}_{(+)}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma}) = p(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{y}_+,\boldsymbol{y}_{(+)})p(\boldsymbol{y}_+\cap\boldsymbol{y}_{(+)}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma})$$

and similarly

$$p(\boldsymbol{Y}\cap\boldsymbol{y}_+\cap\boldsymbol{y}_{(+)}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma}) = p(\boldsymbol{y}_+\cap\boldsymbol{y}_{(+)}|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma})p(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma}).$$

In the second expression, $p(\boldsymbol{y}_+\cap\boldsymbol{y}_{(+)}|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma}) = 1$ by definition. Then rearranging the the first expression gives

$$p(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{y}_+,\boldsymbol{y}_{(+)}) = \frac{p(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma})}{p(\boldsymbol{y}_+\cap\boldsymbol{y}_{(+)}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma})}. \tag{5}$$

Under model (2), the probability of observing the row and column total scores $\boldsymbol{y}_+$ and $\boldsymbol{y}_{(+)}$ is given by the sum of the probabilities of observing all the possible sample configurations for which the adjacency matrix $\boldsymbol{Z}$ has the same total scores as $\boldsymbol{Y}$, that is $\boldsymbol{Z} : \boldsymbol{z}_+ = \boldsymbol{y}_+, \boldsymbol{z}_{(+)} = \boldsymbol{y}_{(+)}$. Therefore we have

$$p(\boldsymbol{y}_+\cap\boldsymbol{y}_{(+)}|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma}) = \frac{\sum_{\boldsymbol{Z}} \exp\left[\sum_{i=1}^n\sum_{j=1}^n z_{ij}\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sum_{i=1}^n(z_{i+}\alpha_i) + \sum_{j=1}^n(z_{(+)j}\gamma_j)\right]}{\prod_{i=1}^n\prod_{j=1}^n[1+\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\alpha_i+\gamma_j)]}.$$

By substituting (3) and the above expression in Equation (5), we obtain

$$p(\boldsymbol{Y}\,|\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{y}_+,\boldsymbol{y}_{(+)}) =$$

$$\frac{\exp\left[\sum_{i=1}^n\sum_{j=1}^n y_{ij}\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sum_{i=1}^n(y_{i+}\alpha_i) + \sum_{j=1}^n(y_{(+)j}\gamma_j)\right]}{\sum_{\boldsymbol{Z}} \exp\left[\sum_{i=1}^n\sum_{j=1}^n z_{ij}\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sum_{i=1}^n(z_{i+}\alpha_i) + \sum_{t=1}^n(z_{(+)j}\gamma_j)\right]} = p(\boldsymbol{Y}\,|\boldsymbol{X},\boldsymbol{y}_+,\boldsymbol{y}_{(+)})$$

which is Equation (4), as $\boldsymbol{z}_+ = \boldsymbol{y}_+$ and $\boldsymbol{z}_{(+)} = \boldsymbol{y}_{(+)}$ are constant across the sample configurations $\boldsymbol{Z}$, and therefore no longer depends on the incidental parameters. □

The score vector and information matrix can be obtained from the conditional likelihood in Equation (4). In this regard, it is useful to write the corresponding conditional log-likelihood $\ell(\boldsymbol{\beta}) = \log p(\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{y}_+,\boldsymbol{y}_{(+)})$ in matrix notation as

$$\ell(\boldsymbol{\beta}) = \boldsymbol{u}(\boldsymbol{y},\boldsymbol{X})\boldsymbol{\beta} - \log\sum_{\boldsymbol{Z}}\exp\left[\boldsymbol{u}(\boldsymbol{z},\boldsymbol{X})\boldsymbol{\beta}\right], \tag{6}$$

where $\boldsymbol{y}$ collects the $m = n(n-1)$ observations contained in $\text{vec}(\boldsymbol{Y})$, excluding the missing values on the main diagonal; similarly $\boldsymbol{z} = \text{vec}(\boldsymbol{Z})$, $\boldsymbol{X}$ denotes the related $m \times k$ matrix of covariates, and $\boldsymbol{u}(\boldsymbol{y}, \boldsymbol{X}) = \sum_{i=1}^{n} \sum_{j=1}^{n} y_{ij} \boldsymbol{x}'_{ij}$. Using the standard theory about the regular exponential family, we have the following expression for the score of $\ell(\boldsymbol{\beta})$

$$\boldsymbol{s}(\boldsymbol{\beta}) = \nabla_{\beta} \ell(\boldsymbol{\beta}) = \boldsymbol{u}(\boldsymbol{y}, \boldsymbol{X}) - \mathrm{E}_{\boldsymbol{\beta}} \left[\boldsymbol{u}(\boldsymbol{y}, \boldsymbol{X}) | \boldsymbol{X}, \boldsymbol{y}_{+}, \boldsymbol{y}_{(+)}) \right],$$

where

$$\mathrm{E}_{\boldsymbol{\beta}} \left[\boldsymbol{u}(\boldsymbol{y}, \boldsymbol{X}) | \boldsymbol{X}, \boldsymbol{y}_{+}, \boldsymbol{y}_{(+)} \right] = \sum_{\boldsymbol{Z}} p(\boldsymbol{Z} | \boldsymbol{X}, \boldsymbol{y}_{+}, \boldsymbol{y}_{(+)}) \boldsymbol{u}(\boldsymbol{z}, \boldsymbol{X}).$$

As for the information matrix, we have

$$\boldsymbol{J}(\boldsymbol{\beta}) = -\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}} \ell(\boldsymbol{\beta}) = \boldsymbol{V}_{\boldsymbol{\beta}} \left[\boldsymbol{u}(\boldsymbol{y}, \boldsymbol{X}) | \boldsymbol{X}, \boldsymbol{y}_{+}, \boldsymbol{y}_{(+)}) \right],$$

where

$$\boldsymbol{V}_{\boldsymbol{\beta}} \left[\boldsymbol{u}(\boldsymbol{y}, \boldsymbol{X}) | \boldsymbol{X}, \boldsymbol{y}_{+}, \boldsymbol{y}_{(+)}) \right] = \sum_{\boldsymbol{Z}} p(\boldsymbol{Z} | \boldsymbol{X}, \boldsymbol{y}_{+}, \boldsymbol{y}_{(+)}) \boldsymbol{d}(\boldsymbol{z}, \boldsymbol{X})' \boldsymbol{d}(\boldsymbol{z}, \boldsymbol{X}),$$

with $\boldsymbol{d}(\boldsymbol{z}, \boldsymbol{X}) = \boldsymbol{u}(\boldsymbol{z}, \boldsymbol{X}) - \mathrm{E}_{\boldsymbol{\beta}} \left[\boldsymbol{u}(\boldsymbol{y}, \boldsymbol{X}) | \boldsymbol{X}, \boldsymbol{y}_{+}, \boldsymbol{y}_{(+)} \right]$. Matrix $\boldsymbol{J}(\boldsymbol{\beta})$ is negative semi-definite as $\boldsymbol{V}_{\boldsymbol{\beta}} \left[\boldsymbol{u}(\boldsymbol{y}, \boldsymbol{X}) | \boldsymbol{X}, \boldsymbol{y}_{+}, \boldsymbol{y}_{(+)}) \right]$ is a variance-covariance matrix, so that $\ell(\boldsymbol{\beta})$ is concave. By simple algebra, it is also possible to show that $\boldsymbol{J}(\boldsymbol{\beta})$ is negative definite when $\boldsymbol{X}'\boldsymbol{X}$ is of full rank. Together with standard regularity conditions, non-singularity of $\boldsymbol{X}'\boldsymbol{X}$ ensures identification of $\boldsymbol{\beta}$ (cf. Lemma 2.2 of Newey and McFadden, 1994). As a consequence, the possibility of including node-constant covariates in the model specification is ruled out.

Assuming that the evaluation of all the necessary components is feasible, the CML estimator of $\boldsymbol{\beta}$, denoted $\hat{\boldsymbol{\beta}}$, could be obtained by maximizing (6). Under the conditions stated above, $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically normal, as per Theorems 2.7 and 3.1 by Newey and McFadden (1994). Finally, standard errors could be obtained as the square root of the diagonal elements of $\hat{\boldsymbol{J}}^{-1}$, a consistent estimate of $\boldsymbol{J}(\boldsymbol{\beta})^{-1}$.

Unfortunately, computing the sum at the denominator of (4) is often not feasible even with samples of moderate size, as the number of matrices $\boldsymbol{Z}$ such that $\boldsymbol{z}_{+} = \boldsymbol{y}_{+}, \boldsymbol{z}_{(+)} = \boldsymbol{y}_{(+)}$, for any configuration of $\boldsymbol{y}_{+}$ and $\boldsymbol{y}_{(+)}$, increases dramatically with n. The problem is equivalent to the enumeration of binary tables with fixed marginal sums, a challenging task that has been debated in the literature (Wang and Zhang, 1998; Pérez-Salvador et al., 2002). One solution is to adopt techniques that allow us to sample zero-one contingency tables, which are widely exploited in many scientific fields. Among other works, (Chen et al., 2005) report applications concerning ecology and psychometrics, making use of importance sampling and MCMC algorithms. As described in the following section, we rely on the latter to approximate the conditional likelihood and obtain the MCMC-CML estimator.

4 MCMC Conditional Maximum Likelihood

In this section, we first provide a brief and general overview of MCMC-ML and recall its asymptotic properties, following Geyer (1991) and Huffer and Wu (1998), and then illustrate the proposed approach.

4.1 Monte Carlo Maximum Likelihood

Consider the probability density function $f_{\boldsymbol{\theta}}$, with respect to a measure μ. Assume that $f_{\boldsymbol{\theta}}$ is know up to a normalizing constant and can be written as

$$f_{\boldsymbol{\theta}}(x) = \frac{1}{z(\boldsymbol{\theta})} h_{\boldsymbol{\theta}}(x),$$

where $h_{\boldsymbol{\theta}}$ is a known function of $\boldsymbol{\theta}$ but the normalizing constant is not, except for its definition as a the integral

$$z(\boldsymbol{\theta}) = \int h_{\boldsymbol{\theta}}(x) d\mu(x), \tag{7}$$

which is analytically intractable.

MCMC algorithms can be used to draw samples $X_1, \ldots, X_N$ from any ψ in the parameter space, that are then used to estimate the log-likelihood ratio for an observation x

$$\ell(\boldsymbol{\theta}) = \log \frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\psi}}(x)} = \log \frac{h_{\boldsymbol{\theta}}(x)}{h_{\boldsymbol{\psi}}(x)} - \log \frac{z(\boldsymbol{\theta})}{z(\boldsymbol{\psi})}.$$

This is because the last term can be written as

$$\frac{z(\boldsymbol{\theta})}{z(\boldsymbol{\psi})} = \frac{1}{z(\boldsymbol{\psi})} \int h_{\boldsymbol{\theta}}(x) d\mu(x) = \mathrm{E}_{\boldsymbol{\psi}} \left[\frac{h_{\boldsymbol{\theta}}(X)}{h_{\boldsymbol{\psi}}(X)} \right], \tag{8}$$

of which a natural estimator is

$$\frac{1}{N} \sum_{j=1}^{N} \frac{h_{\boldsymbol{\theta}}(X_j)}{h_{\boldsymbol{\psi}}(X_j)}.$$

As a result, the log-likelihood $\ell_N(\boldsymbol{\theta})$ can be obtained as

$$\ell_N(\boldsymbol{\theta}) = \log \left[\frac{h_{\boldsymbol{\theta}}(X_{obs})}{h_{\boldsymbol{\psi}}(X_{obs})} \right] - \log \left[\frac{1}{N} \sum_{j=1}^{N} \frac{h_{\boldsymbol{\theta}}(X_j)}{h_{\boldsymbol{\psi}}(X_j)} \right], \tag{9}$$

where X_{obs} denotes the observed data. The maximum of $\ell_N(\boldsymbol{\theta})$ is denoted as $\hat{\boldsymbol{\theta}}_N$ and is the MCMC approximation of the ML estimator $\hat{\boldsymbol{\theta}}$.

Asymptotic results for the MCMC-ML estimator are given in Geyer and Thompson (1992) and Geyer (1994). Provided that samples are taken from an ergodic Markov chain, for any fixed $\boldsymbol{\theta}$ the ergodic theorem then ensures that $\ell_N(\hat{\boldsymbol{\theta}}_N)$ converges almost surely to

$\ell(\hat{\boldsymbol{\theta}})$. If $\hat{\boldsymbol{\theta}}_N$ is a maximizer of $\ell_N(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}$ the the unique maximizer of $\ell(\boldsymbol{\theta})$, then

$$\hat{\boldsymbol{\theta}}_N \xrightarrow{a.s.} \hat{\boldsymbol{\theta}}$$

follows from concavity of both $\ell_N(\boldsymbol{\theta})$ and $\ell(\boldsymbol{\theta})$ (cf. Theorem 4 by Geyer, 1994). Notice that here the MCMC-ML estimator converges almost surely to the ML estimator as the number of the Markov chain samples goes to infinity. By further assuming that $\sqrt{N}\nabla_{\boldsymbol{\theta}}\ell_N(\hat{\boldsymbol{\theta}}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{A})$ for a suitable variance-covariance matrix $\boldsymbol{A}$, and that $\boldsymbol{B} = -\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\hat{\boldsymbol{\theta}})$ is positive definite, along with other standard regularity conditions, we have that $-\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_N(\hat{\boldsymbol{\theta}}_N) \xrightarrow{p} \boldsymbol{B}$ and

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_N - \hat{\boldsymbol{\theta}}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}^{-1}\right),$$

as per Theorem 7 in Geyer (1994). Estimation of the asymptotic covariance matrix is challenging due to the presence of A, which should account for the correlation within the Markov chain. A HAC-type estimator is therefore advisable, but with the additional concern for the sensitivity of the resulting estimate to the choice of the lag bandwidth. For these reasons, Huffer and Wu (1998) propose to use $\boldsymbol{B}$ evaluated at the MCMC-ML estimate $\hat{\boldsymbol{\theta}}_N$ to approximate the asymptotic covariance matrix, motivated by its good finite-sample performance. We refer the reader to Huffer and Wu (1998) for a thorough discussion of the theoretical issues and results of extensive simulation studies.

4.2 Proposed methodology

We propose approximating the CML estimator of $\boldsymbol{\beta}$ based on the conditional probability in (4) by means of an MCMC-CML estimator. In the following we describe the three stages of our procedure: parameter initialization, drawing samples from a Markov chain, and maximization of the resulting log-likelihood.

4.2.1 Initialization

First of all, we need a value for the parameter vector that will be used to draw the samples and is a reference value in the likelihood function. With a notation similar to that adopted in the previous section, we denote this vector by $\boldsymbol{\psi}$. We set this value equal to the ML estimator of $\boldsymbol{\beta}$ based on (3), that is $\boldsymbol{\psi} = \hat{\boldsymbol{\beta}}_{ML}$. The closer $\boldsymbol{\psi}$ to the true value of the model parameters, the better the MCMC approximation of the likelihood function.

4.2.2 Metropolis algorithm

The observed data $\boldsymbol{Y}$ can be represented as a 0-1 table with fixed margins. Therefore we want to randomly generate samples $\boldsymbol{Y}_s$ with $s = 1, \ldots, N$ such that sums over rows, $\boldsymbol{y}_+$, and columns, $\boldsymbol{y}_{(+)}$, are fixed over draws.

Diaconis and Gangolli (1995) and Diaconis and Sturmfels (1998) propose a random walk MCMC that allows us to generate samples form a uniform distribution. We mimic their strategy within a Metropolis algorithm in order to obtain the distribution of the samples (0-1 tables) under the model described in (2) as follows:

1. Define a data table $\boldsymbol{Y}_{old}$ starting from the observed data $\boldsymbol{Y}$.

2. Find all the possible 2×2 sub-matrices in $\boldsymbol{Y}_{old}$ equal to $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ or $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

3. Randomly choose one of the suitable sub-matrices above and update the table according to: $\begin{bmatrix} +1 & -1 \\ -1 & +1 \end{bmatrix}$ or $\begin{bmatrix} -1 & +1 \\ +1 & -1 \end{bmatrix}$, respectively, so that row and column sums remain unchanged. The updated table is the candidate denoted by $\boldsymbol{Y}_{cand}$.

4. Accept $\boldsymbol{Y}_{cand}$ according to

$$\min \left[1; \frac{L(\boldsymbol{Y}_{cand})}{L(\boldsymbol{Y}_{old})} \right] > U(0,1),$$

   where $L(\boldsymbol{Y}) = \log p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{y}_+, \boldsymbol{y}_{(+)})$, with $p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{y}_+, \boldsymbol{y}_{(+)})$ defined by (4), and $U(0,1)$ denotes a random draw from a uniform distribution in $[0,1]$. According to the above formulation, we have that

$$\frac{L(\boldsymbol{Y}_{cand})}{L(\boldsymbol{Y}_{old})} = \exp \left[\sum_{i=1}^{n} \sum_{j=1}^{n} \left(y_{ij}^{cand} - y_{ij}^{old} \right) \boldsymbol{x}'_{ij} \boldsymbol{\psi} \right],$$

   where y_{ij}^{cand} and y_{ij}^{old} are the ij-th dyad in $\boldsymbol{Y}_{cand}$ and $\boldsymbol{Y}_{old}$, respectively.

5. Store $\boldsymbol{Y}_{cand}$ if accepted, $\boldsymbol{Y}_{old}$ otherwise.

6. If $\boldsymbol{Y}_{cand}$ is accepted, $\boldsymbol{Y}_{cand}$ will be labeled $\boldsymbol{Y}_{old}$ in the following steps.

7. Iterate steps 2 to 6 $\tilde{N}$ times.

8. In order to reduce the autocorrelation among the generated samples, we drop the first 1000 tables (burn-in) and we set a thinning factor of 100. We end up with N generated samples.

It can be shown that the Markov chain generated according to this procedure is ergodic (cf. Lemma 2.1 in Diaconis and Sturmfels, 1998).

4.2.3 The MCMC-CML estimator

Assume we have drawn the samples $\boldsymbol{Y}_s$, with $s = 1, \ldots, N$. We can build the log-likelihood as in (9) as follows. Let us define

$$h_{\boldsymbol{\beta}}(\boldsymbol{Y}) = \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}\right)$$

and

$$z_{\boldsymbol{\beta}}(\boldsymbol{Y}) = \sum_{\boldsymbol{Z}} \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{n} z_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}\right).$$

The approximated log-likelihood can therefore be written as

$$\ell_N(\boldsymbol{\beta}) = \log\left[\frac{h_{\boldsymbol{\beta}}(Y_{obs})}{h_{\boldsymbol{\psi}}(Y_{obs})}\right] - \log\left[\frac{1}{N}\sum_{s=1}^{N}\frac{h_{\boldsymbol{\beta}}(Y_s)}{h_{\boldsymbol{\psi}}(Y_s)}\right].$$

With our formulation and by making use of the matrix notation outlined in Section 3, the log of the ratio $h_{\boldsymbol{\beta}}(\boldsymbol{Y})/h_{\boldsymbol{\psi}}(\boldsymbol{Y})$ simplifies to

$$\log\frac{h_{\boldsymbol{\beta}}(\boldsymbol{Y})}{h_{\boldsymbol{\psi}}(\boldsymbol{Y})} = \log\frac{\exp\left(\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}\right)}{\exp\left(\sum_{i=1}^{n}\sum_{t=1}^{n} y_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\psi}\right)} = \boldsymbol{u}(\boldsymbol{y},\boldsymbol{X})\left(\boldsymbol{\beta}-\boldsymbol{\psi}\right).$$

The log-likelihood function can therefore be rewritten as

$$\ell_N(\boldsymbol{\beta}) = \boldsymbol{u}(\boldsymbol{y},\boldsymbol{X})'(\boldsymbol{\beta}-\boldsymbol{\psi}) - \log\left\{\frac{1}{N}\sum_{s=1}^{N}\exp\left[\boldsymbol{u}(\boldsymbol{y}_s,\boldsymbol{X})'(\boldsymbol{\beta}-\boldsymbol{\psi})\right]\right\},$$

where $\boldsymbol{y}_s = \mathrm{vec}(\boldsymbol{Y}_s)$, missing vales excluded, from the $s-$th MC sample.

Let us denote the MCMC-CML estimator $\hat{\boldsymbol{\beta}}_N$, which can be obtained by maximizing the MC conditional log-likelihood $\ell_N(\boldsymbol{\beta})$ by a Newton-Raphson algorithm using the following score and Information matrix. The score function is

$$\boldsymbol{s}_N(\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}}\ell_N(\boldsymbol{\beta}) = \boldsymbol{u}(\boldsymbol{y},\boldsymbol{X}) - \mathrm{E}_N\left[\boldsymbol{u}(\boldsymbol{y}_s,\boldsymbol{X})|\boldsymbol{X}\right],$$

with

$$\mathrm{E}_N\left[\boldsymbol{u}(\boldsymbol{y}_s|\boldsymbol{X})|\boldsymbol{X}\right] = \sum_{s=1}^{N} p(\boldsymbol{y}_s|\boldsymbol{X})\boldsymbol{u}(\boldsymbol{y}_s,\boldsymbol{X}),$$

where we avoid writing the expected value with respect to the vector $\boldsymbol{\beta}-\boldsymbol{\psi}$ to avoid abuse of notation and instead emphasize that it is taken over the N samples, and

$$p(\boldsymbol{y}_s|\boldsymbol{X}) = \frac{\exp\left[\boldsymbol{u}(\boldsymbol{y}_s,\boldsymbol{X})(\boldsymbol{\beta}-\boldsymbol{\psi})\right]}{\sum_{s=1}^{N}\exp\left[\boldsymbol{u}(\boldsymbol{y}_s,\boldsymbol{X})(\boldsymbol{\beta}-\boldsymbol{\psi})\right]}.$$

The Information matrix can therefore be written as

$$\boldsymbol{J}_N(\boldsymbol{\beta}) = -\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}}\ell_N(\boldsymbol{\beta}) = \boldsymbol{V}_N\left[\boldsymbol{u}(\boldsymbol{y}_s, \boldsymbol{X})|\boldsymbol{X}\right],$$

where

$$\boldsymbol{V}_N\left[\boldsymbol{u}(\boldsymbol{y}_s, \boldsymbol{X})|\boldsymbol{X}\right] = \sum_{s=1}^{N} p(\boldsymbol{y}_s|\boldsymbol{X})\boldsymbol{e}(\boldsymbol{y}_s, \boldsymbol{X})'\boldsymbol{e}(\boldsymbol{y}_s, \boldsymbol{X}),$$

with $\boldsymbol{e}(\boldsymbol{y}_s, \boldsymbol{X}) = \boldsymbol{u}(\boldsymbol{y}_s, \boldsymbol{X}) - \mathrm{E}_N\left[\boldsymbol{u}(\boldsymbol{y}_s, \boldsymbol{X})|\boldsymbol{X}\right]$. Akin to the CML estimator, matrix $\boldsymbol{J}_N(\boldsymbol{\beta})$ is negative semi-definite as $\boldsymbol{V}_N\left[\boldsymbol{u}(\boldsymbol{y}_s, \boldsymbol{X})|\boldsymbol{X}\right]$ is a variance-covariance matrix, so that $\ell_N(\boldsymbol{\beta})$ is concave, and again by simple algebra it is possible to show that $\boldsymbol{J}_N(\boldsymbol{\beta})$ is negative definite when $\boldsymbol{X}'\boldsymbol{X}$ is of full rank.

Together with the ergodic property of the Markov chain used to generate tables with fixed margins, the concavity of $\ell_N(\boldsymbol{\beta})$ is now enough to ensure that $\hat{\boldsymbol{\beta}}_N$ converges almost surely to the CML estimator $\hat{\boldsymbol{\beta}}$, which is in turn a consistent estimator of $\boldsymbol{\beta}$. By further assuming that $\sqrt{N}\boldsymbol{s}_N(\hat{\boldsymbol{\beta}}_N)$ converges in distribution to a normal random variable, along with other standard regularity conditions, $\hat{\boldsymbol{\beta}}_N$ is also asymptotically normally distributed and, according Huffer and Wu (1998), the asymptotic variance-covariance matrix can be well approximated by $\boldsymbol{J}(\hat{\boldsymbol{\beta}}_N)^{-1}$, where $\boldsymbol{J}(\cdot)$ is the Information matrix of the conditional likelihood.

5 Simulations

In this section we study the finite sample performance of the proposed MCMC-CML estimator. The data generating process is as follows

$$y_{ij} = \mathbf{I}\left(x_{1,ij}\beta_1 + x_{2,ij}\beta_2 + \alpha_i + \gamma_j + \varepsilon_{ij} > 0\right),$$

for $i, j = 1, \ldots, n$. Here $x_{1,ij}$ is a continuous regressor such that $x_{1,ij} = \alpha_i + \gamma_j + \eta_{ij}$, with $\eta_{ij} \sim N(0,1)$, and $x_{2,ij}$ is a binary covariate observed according to $x_{2,ij} = \mathbf{I}\{u_{ij} > 0.5\}$, where u_{ij} is a $[0,1]$ uniform random variable. Both the fixed effects α_i and γ_j are drawn from independent standard normal random variables and ε_{ij} is a standard logistic error term. Finally, the homophily parameters are set to $\beta_1 = 1$ and $\beta_2 = 2.5$.

We consider networks of $n = (25, 50)$ nodes and for each of the two scenarios we perform 1,000 Monte Carlo replications. As for the proposed MCMC-CML estimator, for each of the 1,000 Monte Carlo draws, we generate a Markov chain of 100,000 samples where the first 20,000 are excluded (burn-in) and we apply a thinning factor of 100 on the rest of the chain, ending up with $N = 800$ generated samples.

Tables 1 and 2 report the simulation results regarding the networks with $n = 25$ and $n = 50$, respectively, for the proposed MCMC-CML estimator along with the ML estimator of the two-way fixed-effects logit and the pairwise CML estimator (P-CML) proposed by Charbonneau (2017) an Jochmans (2018). For each parameter, we report the mean, median, standard deviation (sd), the interquartile range (iqr), the ratio between estimated standard errors (se/sd), and the rejection rate of a simple t-test on the coefficient for $H_0 : \beta_1 = 1$ and $H_0 : \beta_2 = 2.5$, respectively, where the nominal size of the test is set to 5% (p.05).

The reported results show that the bias of the ML estimator, due to the incidental parameters problem, is sizable for both parameters with a network of $n = 25$ nodes. On average, ML estimates are about 20% larger than the true value of the parameters and the rejection rate of the t-test is far from the nominal size. The bias, however, seems to be decreasing in n. On the contrary, P-CML and MCMC-CML exhibit a negligible bias with both networks. However, the MCMC-CML estimator seems to outperform the P-CML one, at least with $n = 25$, in terms of bias and size of the t-test, even tough both estimators here tend to overestimate standard errors. Finally, and in line with the theoretical framework outlined above, we observe a remarkable advantage in terms of efficiency of the MCMC-CML estimator with respect to the P-CML estimator. The latter exhibits a standard deviation about 20% larger and a wider interquartile range with a network of $n = 25$. Instead, the difference in the standard deviation reduces to about 10% and interquartile ranges a are substantially the same with $n = 50$. In this respect, it is worth clarifying that convergence of the two conditional estimators is based on the number of likelihood components, equal to $m = 2450$ for the MCMC-CML estimator and to the number of informative quadruples for the P-CML, that in the current design amounts to about $49,000$ on average. It therefore stands to reason that, in larger networks with an abundance of informative quadruples, the advantage in terms precision of the MCMC-CML is masked by the faster convergence rate of the P-CML. Nevertheless, these results suggest that the computational effort required by the MCMC-CML estimator is justified by a sizable efficiency gain over the P-CML one, especially with smaller networks.

Table 1: Simulation results: $n = 25$

	$\beta_1 = 1$				$\beta_2 = 2.5$		
	ML	P-CML	MCMC-CML		ML	P-CML	MCMC-CML
mean	1.206	1.042	1.001	mean	2.982	2.576	2.481
median	1.193	1.035	0.993	median	2.964	2.545	2.478
sd	0.208	0.201	0.164	sd	0.437	0.414	0.341
iqr	0.268	0.255	0.216	iqr	0.597	0.578	0.454
se/sd	0.887	1.107	1.117	se/sd	0.877	1.129	1.127
p.05	0.189	0.037	0.041	p.05	0.228	0.021	0.037

Table 2: Simulation results: $n = 50$ (Preliminary)

	$\beta_1 = 1$				$\beta_2 = 2.5$		
	ML	P-CML	MCMC-CML		ML	P-CML	MCMC-CML
mean	1.059	0.983	0.969	mean	2.668	2.491	2.459
median	1.049	0.987	0.965	median	2.660	2.486	2.479
sd	0.083	0.088	0.078	sd	0.159	0.165	0.150
iqr	0.114	0.103	0.103	iqr	0.261	0.207	0.226
se/sd	0.968	1.081	1.145	se/sd	1.037	1.167	1.218
p.05	0.109	0.036	0.074	p.05	0.164	0.018	0.000

6 Empirical examples

In this section we illustrate the proposed approach by means of three applications. We first estimate a two-way fixed-effects gravity model for the extensive margin of trade; we then turn to a sparse network of attorneys consulting professionally; finally we move away from the network setting and apply the proposed approach to the estimation of a binary panel data model for banking crises.

6.1 A binary two-way gravity model

We use data from Helpman et al. (2008) consisting of a cross section of 158 countries and we focus on the existence of trade flows between pairs of countries, which give rise to a trade network. The two-way fixed-effects logit model is used to describe the probability that an exchange between country i and j occurs, namely the trade extensive margin, where covariates and the related homophily parameters as well as node-specific heterogeneity are included.

The dependent variable is therefore a binary variable denoting whether a trade flow occurs from country i to country j. As for dyadic-specific covariates, we include a set of geographical regresses that are: *Distance*, which is the logarithm of the geographical distance (in kilometers) between the capitals of each pair of countries; *Island*, a dummy variable equal to 1 if one or both countries are islands; *Landlock*, equal to 1 if one or both do not have access to the sea; *Border*, which is a dummy indicating whether the two countries share a border. Furthermore, an additional set of variables captures the institutional and cultural similarities of each pair of countries: *Legal*, *Language*, and *Currency* are binary variables equal to one if countries i and j share the same legal origin, the same language, and the same currency (or they are in the same currency union), respectively; *Religion* is measure of cult similarities, and for details on its construction we refer the reader to Helpman et al. (2008) and Charbonneau (2017). Finally, *Colonial Ties* is a dummy variable assuming value 1 if country i colonized country j (or viceversa) and FTA is a binary variable capturing whether the two countries belong to a common trade agreement.

Table 3 collects the estimation results for four different estimators based on models including both importer and exported fixed effects: ML is the (dummy variables) Maximum Likelihood estimator, BC denotes the analytical bias-corrected ML estimator, P-CML is the pairwise CML, and MCMC-CML is the proposed estimator. The bias corrected estimator used here is one provided by Fernández-Val and Weidner (2016), which was originally proposed for panel data but also applied to the estimation gravity equations for the extensive margin of trade (Cruz-Gonzalez et al., 2017). Jochmans (2018) reports that the mean out and in degrees in the sample are approximately 50%, describing a rather dense network, which would lead us expect similar results from BC, P-CML, and MCMC-CML. As a matter of fact, the estimates obtained by the proposed MCMC-CML approach are rather similar those produced by ML, with and without a bias correction. There are small differences between these coefficients and those resulting from the P-CML estimator, for which however, with the exception of the coefficient associated with geographical distance, the 95% confidence intervals overlap. For some covariates, the related homophily parameters are more precisely estimated by MCMC-CML rather than by P-CML. Yet this evidence might be here mitigated by the fact that the likelihood components exploited by the P-CML are $2,852,337$ against the $24,806$ used to build the MCMC approximate likelihood function.

Estimation results are in line with those presented already by Charbonneau (2017). The estimates presented for the four methods considered all agree on the sign and magnitude of the homophily parameters. In particular, geographical distance, as well as the fact that one or both countries are either islands or are landlocks, have a negative effect on the probability of trading. An exception is represented by sharing a common border, which surprisingly seems to hamper the chances of an exchange. However both Charbonneau (2017) and Jochmans (2018) argue that the sign could be due to the sparsity of this variable in the sample (17% of non-zero instances) and to its strong correlation with geographical distance. Finally, all the regressors reflecting institutional and cultural similarities, as well as an history of colonial ties and trade agreements, exert a positive effect on the probability of trading.

6.2 A network of attorneys

We here consider an advice network created by 71 attorneys employed in US law firms. The data are provided by Lazega et al. (2001) and the same example below is also presented by Jochmans (2018). Here the response variable takes value one if lawyer i has consulted with lawyer j. The set of covariates associated with the network homophily parameters comprises three dummy variables describing whether the attorneys have the

Table 3: Estimation results: gravity model for the extensive margin of trade

	ML	BC	P-CML	MCMC-CML
Distance	-1.252	-1.224	-0.995	-1.246
	(0.039)	(0.039)	(0.056)	(0.040)
Island	-0.602	-0.589	-0.392	-0.623
	(0.135)	(0.135)	(0.137)	(0.153)
Landlock	-0.361	-0.354	-0.198	-0.339
	(0.189)	(0.189)	(0.199)	(0.178)
Border	-0.762	-0.745	-0.517	-0.759
	(0.169)	(0.169)	(0.236)	(0.178)
Legal	0.186	0.182	0.183	0.199
	(0.054)	(0.054)	(0.061)	(0.053)
Language	0.504	0.492	0.416	0.496
	(0.069)	(0.069)	(0.075)	(0.075)
Currency	0.898	0.881	1.057	0.890
	(0.233)	(0.233)	(0.243)	(0.236)
Religion	0.416	0.408	0.485	0.398
	(0.109)	(0.109)	(0.120)	(0.118)
Colonial ties	0.534	0.502	1.137	0.374
	(0.529)	(0.532)	(0.661)	(0.629)
FTA	3.471	3.374	3.523	3.531
	(0.553)	(0.553)	(0.575)	(0.836)

No. of countries: 158; No. of dyads: $24,806$. Standard error in parentheses. ML: dummy variable maximum likelihood estimator; BC: bias corrected estimator (Fernández-Val and Weidner, 2016); P-CML: pairwise conditional maximum likelihood estimator (Charbonneau, 2017; Jochmans, 2018); MCMC-CML: proposed estimator. No. of Markov chain samples: 500k; Burn-in: 100k; Thinning factor: 100.

Table 4: Estimation results: network model for consulting attorneys

	ML	BC	P-CML	MCMC-CML
Same status	0.958	0.920	0.944	0.930
	(0.126)	(0.126)	(0.137)	(0.119)
Same gender	0.244	0.234	0.204	0.231
	(0.125)	(0.125)	(0.131)	(0.127)
Same office	2.209	2.110	1.981	2.155
	(0.125)	(0.125)	(0.141)	(0.134)
Difference in tenure	-0.040	-0.038	-0.034	-0.037
	(0.010)	(0.010)	(0.012)	(0.011)
Difference in age	-0.016	-0.016	-0.018	-0.015
	(0.009)	(0.009)	(0.009)	(0.009)

No. of attorneys: 71; No. of dyads: 4,970. Standard error in parentheses. ML: dummy variable maximum likelihood estimator; BC: bias corrected estimator (Fernández-Val and Weidner, 2016); P-CML: pairwise conditional maximum likelihood estimator (Charbonneau, 2017; Jochmans, 2018); MCMC-CML: proposed estimator. No. of Markov chain samples: 500k; Burn-in: 100k; Thinning factor: 100.

same status (partner or associate) in the their respective law firms, have the same gender, and work in the same firm, denoted *same status*, *same gender*, and *same office*, respectively. Furthermore, the set of regressors include the *difference in tenure* and *difference in age* between the two attorneys.

Table 4 reports the results for the four estimators considered also in Section 6.1 and, all in all, results agree on a role of similarity between attorneys in forming a network of professional consultations. The network statistics reported by Jochmans (2018) show that both the mean in an out degree are about 18%, therefore depicting a considerably sparser network than the one presented in the previous section. Jochmans (2018) argues that in this setting the conditional approach should be favored over the bias correction ML estimation. Even in this sparser network, however, the results obtained by both the conditional approaches, P-CML and MCMC-CML, look quite similar to those obtained by the bias-corrected ML estimator. Rather, it is worth to note that, compared with the P-CML approach, most of the homophily parameters are more precisely estimated by MCMC-CML.

6.3 A binary panel data model for banking crises

In the following, we illustrate how the proposed and alternative approaches can be applied to the estimation of a binary panel data logit model with both subject and year fixed effects. To this aim we consider a so-called logit early warning system for banking crises, that is a binary choice model where the response variable takes value 1 if a banking crisis occurs in country i at time t and 0 in the tranquil periods. The set of covariates comprises macroeconomic variables, usually lagged, considered relevant in signaling imminent spans of financial turmoil. Accounting for both subject and time effects in this context can help identify the early warnings and enhance the model ability to perform in-sample forecasts.

We consider a dataset consisting of 33 countries over the years $1985 - 2016$. For the sake of comparing the results obtained with the proposed MCMC-CML estimator with other approaches, for which we use ready-to-use software, we selected a balanced panel. It is worth mentioning, however, that the proposed approach can be extended to the case of unbalanced panel data as well. The definition of banking crisis for a large set of countries is provided by Laeven and Valencia (2018) and identifies 73 crisis episodes in the data at hand. Such a structure of the dataset depicts a similar situation to that presented in Section 6.2. Instead of small and sparse network, here we have a panel dataset with similar and small n and T, and sparsity is somewhat represented by rarity of the crisis events. The set of macroeconomic variables, available as International Financial Statistics (International Monetary Fund) and/or World Development Indicators (World Bank), is fairly self-explanatory and comprises *Real GPD growth*, the *Log of per capita GDP*, *Inflation*, the *Real interest rate*, the ratio of *M2* (broad money) *to foreign exchange reserves*, the *Growth rate of real domestic credit*, and the *Growth of net foreign assets to GDP*. All explanatory variables are lagged by one period. A broader version of the same data has been considered by Pigini (2021) and Caggiano et al. (2016), to which we refer the reader for more details on the variable definition and descriptive statistics.

Table 5 reports the estimation results for the same estimators considered in Section 6.1. Although, to the best of our knowledge, there are currently no panel data applications available of the P-CML estimator, its possible viability for the estimation of binary panel data models with both subject and time fixed effects is implied by Charbonneau (2017). As with the previous applications, results are all coherent on the early warnings. In particular, only the real GDP growth seems to be able to signal, with a statistically significant effect, the outbreak of a banking crisis. The sign of the coefficients are in line with those found in this strand of literature, with the exception of the those associated with last two variables, which are however not statistically significant.

The results in table 5 also show a few discrepancies in the estimates obtained by ML and those by the two conditional approaches. In this respect, it is worth to mention that the ML estimator, with and without the bias correction, uses only 520 out of the $nT = 1,056$ available observations, as countries where and years when a crisis never occurs are dropped from the estimation sample. On the contrary, the P-CML estimator exploits $1,756$ informative quadruples. More importantly, though, as in the previous example it emerges here that the proposed MCMC-CML approach brings a gain in estimation precision over the P-CML method.

Table 5: Estimation results: early warning system for banking crises

	ML	BC	P-CML	MCMC-CML
Real GDP growth$_{t-1}$	-0.207	-0.170	-0.157	-0.170
	(0.045)	(0.050)	(0.069)	(0.045)
Log of per capita GDP$_{t-1}$	-0.862	-0.544	-0.552	-0.594
	(0.337)	(0.707)	(0.770)	(0.580)
Inflation$_{t-1}$	0.019	0.019	0.027	0.020
	(0.019)	(0.018)	(0.027)	(0.016)
Real interest rate$_{t-1}$	0.023	0.003	0.005	0.003
	(0.021)	(0.020)	(0.023)	(0.019)
M2 to foreign exchange reserves$_{t-1}$	0.016	0.017	0.018	0.018
	(0.010)	(0.011)	(0.016)	(0.010)
Growth of real domestic credit$_{t-1}$	-0.013	-0.031	-0.024	-0.033
	(0.030)	(0.030)	(0.039)	(0.033)
Growth of net foreign assets to GDP$_{t-1}$	1.484	2.755	2.107	2.909
	(2.976)	(2.985)	(3.911)	(2.634)

No. of countries: 33; No. of years: 32. Standard error in parentheses. ML: dummy variable maximum likelihood estimator; BC: bias corrected estimator (Fernández-Val and Weidner, 2016); P-CML: pair-wise conditional maximum likelihood estimator (Charbonneau, 2017; Jochmans, 2018); MCMC-CML: proposed estimator. No. of Markov chain samples: 500k; Burn-in: 100k; Thinning factor: 100.

7 Conclusion

In this work we overcome the computational issue that arises with the evaluation of the conditional probability for a two-way fixed-effects logit model. Logit models based on dyadic data are relevant in describing link formation in a network where the quantities of interest are the homophily parameters, that is, the propensity of subjects to form ties with similar agents, and they need to be identified once node heterogeneity has been accounted for. As illustrated by the empirical applications, these models are relevant to describe trade relationships, social networks, and can also be adapted for the estimation of binary panel data models with both individual and time permanent unobserved heterogeneity.

We have shown that, under mild regularity conditions, the MCMC-CML estimator converges to the CML and is asymptotically normal. Also, the simulation study and empirical applications here proposed confirm the well-known theoretical result, according to which the full CML is more efficient than the estimator maximizing the composite conditional log-likelihood, which is the alternative approach available in the literature.

References

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(2):283–301.

Ando, T., Bai, J., and Li, K. (2021). Bayesian and maximum likelihood analysis of large-scale panel choice models with unobserved heterogeneity. *Journal of Econometrics*.

Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G., and Meghir, C. (2012). Risk pooling, risk preferences, and social networks. *American Economic Journal: Applied Economics*, 4(2):134–67.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144).

Boneva, L. and Linton, O. (2017). A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance. *Journal of Applied Econometrics*, 32(7):1226–1243.

Caffo, B. S. and Booth, J. (2003). Monte carlo conditional inference for log-linear and logistic models: a survey of current methodology. *Statistical Methods in Medical Research*, 12(2):109–123.

Caggiano, G., Calice, P., Leonida, L., and Kapetanios, G. (2016). Comparing logit-based early warning systems: Does the duration of systemic banking crises matter? *Journal of Empirical finance*, 37:104–116.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The review of economic studies*, 47(1):225–238.

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal*, 20(3):S1–S13.

Chen, M., Fernández-Val, I., and Weidner, M. (2021). Nonlinear factor models for network and panel data. *Journal of Econometrics*, 220(2):296–324.

Chen, Y., Diaconis, P., Holmes, S. P., and Liu, J. S. (2005). Sequential monte carlo methods for statistical analysis of tables. *Journal of the American Statistical Association*, 100(469):109–120.

Cruz-Gonzalez, M., Fernández-Val, I., and Weidner, M. (2017). Bias corrections for probit and logit models with two-way fixed effects. *The Stata Journal*, 17(3):517–545.

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.

Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of statistics*, 26(1):363–397.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics*, 101(5):763–776.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60.

Fafchamps, M. and Gubert, F. (2007). Risk sharing and network formation. *American Economic Review*, 97(2):75–79.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics*, 192(1):291–312.

Geyer, C. J. (1991). Markov chain monte carlo maximum likelihood.

Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical science*, pages 473–483.

Geyer, C. J. (1994). On the convergence of monte carlo maximum likelihood calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):261–274.

Geyer, C. J. and Thompson, E. A. (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The quarterly journal of economics*, 123(2):441–487.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373):33–50.

Huffer, F. W. and Wu, H. (1998). Markov chain monte carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*, pages 509–524.

Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics*, 36(4):705–713.

Jochmans, K. and Otsu, T. (2019). Likelihood corrections for two-way models. *Annals of Economics and Statistics*, 2019(134):227–242.

Laeven, L. and Valencia, F. (2018). Systemic banking crises revisited. IMF Working Paper 18/206, International Monetary Fund.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of econometrics*, 95(2):391–413.

Lazega, E. et al. (2001). *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford University Press on Demand.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, pages 1–32.

Pérez-Salvador, B. R., de-los Cobos-Silva, S., Gutiérrez-Andrade, M. A., and Torres-Chazaro, A. (2002). A reduced formula for the precise number of (0, 1)-matrices in a (r, s). *Discrete mathematics*, 256(1-2):361–372.

Pigini, C. (2021). Penalized maximum likelihood estimation of logit-based early warning systems. *International Journal of Forecasting*, 37(3):1156–1172.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 4, pages 321–333.

Varin, C. (2008). On composite marginal likelihoods. *Asta advances in statistical analysis*, 92(1):1–28.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42.

Wang, B.-Y. and Zhang, F. (1998). On the precise number of (0, 1)-matrices in a (r, s). *Discrete mathematics*, 187(1-3):211–220.

Yan, T., Jiang, B., Fienberg, S. E., and Leng, C. (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association*, 114(526):857–868.

Zacchia, P. (2020). Knowledge spillovers through networks of scientists. *The Review of Economic Studies*, 87(4):1989–2018.