A Pairwise Differencing Distribution Regression Approach for Network Models

Preliminary. Please do not cite.

Gabriela M. Szini

Amsterdam School of Economics, University of Amsterdam

Tinbergen Institute

Abstract

This paper proposes a novel estimation method for distribution regressions in a network setting, considering the effects of covariates on the entire outcome distribution rather than just on the mean. I adopt a semiparametric approach, taking into account two-way unit-specific effects that are treated as fixed parameters to be estimated. Thus, I extend the standard distribution regression approach to a network setting by estimating multiple binary choice models with two-way fixed effects for different thresholds of the distribution. Instead of using bias-correction methods to address the incidental parameter problem, as previously proposed in the literature, I propose to employ a conditional maximum-likelihood approach (Charbonneau (2017), Jochmans (2018)) that differentiates out the unit-specific effects. This method yields consistent point estimates that converge at a parametric rate and remain asymptotically unbiased in the tails of the outcome distribution, where the underlying network can be seen as sparse. Monte Carlo simulations validate these findings for both single cut-offs and the overall outcome distribution. The empirical application focuses on gravity equations for bilateral trade, demonstrating the effectiveness of the proposed approach in cases where the outcome variable is bounded below at zero.

[*]I would like to thank my advisors Frank Kleibergen and Artūras Juodis for all comments and suggestions. Moreover, I grateful for comments from Bo Honoré as well as Martin Weidner, Áureo de Paula, Stephane Bonhomme, Jonas Meier, Michal Kolesár, Mikkel Plagborg-Møller, Mark Watson, Laura Liu, Timo Schenk, Sebastian Roelsgaard, Luther Yap, Andrea Titton and seminar participants at Princeton University.

Email address: `g.m.m.szini@uva.nl.` (Gabriela M. Szini)

1. Introduction

The vast majority of studies, especially for network models, propose estimates for the effects of covariates on the mean of an outcome variable. However, in many cases, the effects on the entire distribution of the outcomes are also an object of interest. For instance, in international trade applications, one might be interested not only in the mean effects of tariffs on the level of exports from one country to another but also in how this effect may vary for different quantiles of the distribution of trade flows. In a more straightforward cross-sectional case, the estimation of such varying effects can be obtained via the distribution regression approach as initially proposed by Foresi and Peracchi (1995).

Motivated by the current abundance of network datasets and the estimation of international trade flows (which naturally constitutes a network setting), this paper provides an estimation method with valid uniform confidence bands for the distributional effects in a network setting. The contributions of this paper are threfold: (i) I propose an estimation method that is free of the incidental parameter problem, being more robust in the tails of the distribution of the outcomes; (ii) I show that uniform confidence bands can be obtained through a valid bootstrap procedure (*more details in next versions of this paper*); and (iii) I illustrate the method with an application to the estimation of gravity models for international trade flows.

A broad range of economic relationships can be modeled through a network perspective, in particular through bilateral ties of agents (for instance, in a model of risk sharing in Fafchamps and Gubert (2007), and in a model of the diffusion of microfinance in Banerjee et al. (2013)). As defined by Graham (2020), in dyadic models, the outcomes reflect pairwise interactions among sampled units. Therefore, I follow the current prominent literature on the econometrics of networks which naturally gravitated towards such models. A key aspect of dyadic regressions is the inclusion of observed dyad-level characteristics and unobserved unit-specific effects for each unit in the dyad (both senders and receivers in a directed framework). Throughout this paper, I treat the unit-specific effects as fixed parameters to be estimated, such that their distribution, conditional on the covariates, is left unrestricted. For this reason, and because the estimated effects can vary with the level of the outcome, the considered model is semiparametric.

The distribution regression (DR) approach of Foresi and Peracchi (1995) boils down to estimating the conditional distribution of the outcome of interest with a sequence of binary response estimators. The binary response is an indicator function of the outcome passing some threshold (for instance, the corresponding quantiles). Due to the non-linearity of the model, the inclusion of the two-way fixed effects to accommodate the dyadic structure leads to the incidental parameter problem (Neyman and Scott (1948)) when jointly estimating all the parameters.

To deal with the incidental parameter problem, I propose to extend the conditional maximum-likelihood approach of Charbonneau (2017) to estimate single binary choice models to multiple (possibly a continuum of) binary choice models for the thresholds. Note that the estimator of Charbonneau (2017) was initially proposed for a directed network formation model; however, since the structure of those is that of a dyadic discrete choice model, it is also suitable for the DR framework considered in this paper. The approach mentioned above relies on conditioning on sufficient statistics that, when the underlying distribution of the outcomes is logistic, differences out the fixed effects from the likelihood (a generalization of the conditional likelihood approach of Rasch (1960) for panel data binary choice models with fixed effects). As shown by Jochmans (2018), the proposed estimator for each threshold is consistent and converges asymptotically to a normal distribution centered around the true parameter value at a parametric rate.

To my knowledge, the only paper that proposes an estimator for DR in a network setting is Chernozhukov et al. (2020). Even though the model I propose and the approach of estimating the DR coefficients through a sequence of binary choice models are similar to theirs, the key difference relies on the estimation method employed for each threshold. They propose to deal with the incidental parameter problem by analytically bias-correcting the estimates, as initially shown in Fernández-Val and Weidner (2016) for a standard panel data model with two-way fixed effects and large N and T, and later formalized in Dzemski (2019) for the network's case. As pointed out by Dzemski (2019), in the context of a network formation model, an essential assumption for consistency of the bias-corrected estimator is that the underlying network is dense. In the

DR setting, this translates to the conditional probability of the outcomes being smaller than a given threshold to be bounded away from zero or one (Chernozhukov et al. (2020)). That is, the estimates are not guaranteed to be consistent and have valid inference in the extremum quantiles of the conditional distribution of the outcomes. On the other hand, as shown in Jochmans (2018), the conditional maximum likelihood approach proposed in this paper allows for a higher degree of sparsity in the underlying network, being more robust in the extremum quantiles of the conditional distribution. These results are confirmed in Monte Carlo exercises for both the estimates of a given threshold of the distribution (which is essentially a network formation model, as seen in the next Sections), and for the estimates of the entire distribution.

I consider an empirical application to gravity equations for bilateral trade between countries. As mentioned in Chernozhukov et al. (2020), the DR approach is well-suited for this application, among others, where the outcome variable is bounded below at zero, indicating the presence of a heavy upper tail in the distribution. I show that the estimated coefficients of the distribution regression have a clear relation to marginal effects of the quantile function, providing a further interpretation of the estimates. Moreover, in this particular application, joint confidence bands on the estimates allow for testing whether the elasticities of gravity models of trade are heterogeneous, which has been extensively discussed in the international trade literature.

Despite the abundance of network datasets, there is still a substantial lack of understanding of features of the estimation of models reflecting this structure where outputs contain many zeros (not only for the gravity equations application specifically). To illustrate the problem, consider the following two-way fixed effects model with a possible selection bias:

$$\begin{aligned}
y_{1,ij} &= y_{2,ij}(x'_{1,ij}\beta_{1,0} + \alpha_i + \gamma_j + u_{ij}) \\
y_{2,ij} &= \mathbb{1}\left(y^*_{2,ij} > 0\right) \\
y^*_{2,ij} &= x'_{2,ij}\beta^*_{2,0} + \xi^*_i + \zeta^*_j + \eta^*_{ij}, \\
&\quad (i = 1, \dots N; j = 1, \dots N, i \neq j)
\end{aligned}$$

where pairs ij first decide whether to form a link, in which occasion $y_{2,ij} = 1$ and then a non-zero outcome $y_{1,ij}$ is observed, generating outcomes $y_{i,ij}$ with potentially many zeros. $\alpha_i, \gamma_j, \xi_i^*$ and ζ_j^* are individual fixed effects. In both equations, the unobservable individual-specific effects might arbitrarily depend on the observable explanatory variables. Therefore, they considered nuisance parameters to be estimated in a semi-parametric model. The errors in the equations (u_{ij} and η_{ij}) might be correlated, in which case sample selectivity should be addressed. In the gravity models case, the equations above are obtained after a log-linearization of the original model, which has a multiplicative form.

Currently, there are two strands in the literature of gravity equations on how to take into account the zeros: (i) modeling through a sample selection model (Helpman et al. (2008)); or (ii) consider the model in its multiplicative form through a Poisson pseudo maximum likelihood (PPML) estimator (Silva and Tenreyro (2006)). However, due to the introduction of the two-way fixed effects and the non-linearities in both models, both estimations suffer from the incidental parameter problem. While bias-correction methods have been proposed for both the PPML (Fernández-Val and Weidner (2016)) and the first stage of the sample selection model (Dzemski (2019) and Yan et al. (2019)), the estimates are consistent and valid inference is available only under dense networks, as mentioned before. It is noteworthy that the method of Charbonneau (2017) can be employed for the first stage of the sample selection, providing asymptotically unbiased estimates even when the network is sparse. However, this method does not provide estimates of the fixed effects, hence, the estimation of the second stage is infeasible.

Therefore, the DR approach proposed in this paper fills this gap in the literature, displaying advantages compared to the previously available methods. Namely, (i) it allows for zero outcomes by relying on modeling the conditional distribution of the outcomes directly and avoiding strong assumptions on how the zeros are generated, and (ii) it allows for the presence of many zeros being suitable for sparse networks.

Plan of the paper. Section 2 outlines the main model to be estimated; Section 3 provides the estimation method; Section 4 shows the asymptotic properties of the proposed estimator; Section 5 provides the Monte Carlo simulation results for both a single threshold of the distribution and for the entire distribution; Section 6 outlines the application for gravity models of international trade; and Section 7 concludes.

2. A Distribution Regression Model for Networks

This section introduces a model for the DR approach that considers a network structure. Following the literature and as initially proposed by Foresi and Peracchi (1995), I propose a model directly for the conditional distribution of the outcomes. A general dyadic setting is considered to accommodate the network structure, which is assumed to take place mainly bilaterally. Therefore, the conditional distribution is a function of dyad-specific characteristics and fixed effects for each unit in the observed pair of nodes.

Let $\{(y_{ij}, \boldsymbol{x}_{ij}) : (i,j) \in \mathcal{D}\}$ be the observed data set, where y_{ij} is a scalar outcome variable that can be discrete, continuous or mixed for a dyad ij, and $\boldsymbol{x}_{ij}$ is a vector of covariates. I set that there is a specific region of interest $\mathcal{Y}$ for the outcome of interest and that the vector of covariates has support $\mathcal{X} \subseteq \mathbb{R}^{d_x}$. The set $\mathcal{D}$ contains the indices of the pairs (i,j) that are observed in a directed network without self-links, i.e., $\mathcal{D} = \{(i,j) : i = 1, \ldots, N, j = 1, \ldots, N\} \setminus \{(i,i) : i = 1, \ldots, N\}$. $n = |\mathcal{D}| = N(N-1)$ gives the total number of observed units. Moreover, the set of nodes in the network is given by $\mathcal{N} = \{1, 2, \ldots, N\}$.[1]

The individual fixed effects for units i and j are taken into account through vectors of unspecified dimensions $\boldsymbol{\nu}_i$ and $\boldsymbol{\omega}_j$ that contain unobserved random variables or effects that might be arbitrarily related to the covariates $\boldsymbol{x}_{ij}$. Therefore, they can be seen as nuisance parameters. The conditional distribution of y_{ij} given $(\boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j)$ is given by:

$$F_{y_{ij}}\left(y \mid x_{ij}, v_i, w_j\right) = \Lambda\left(\boldsymbol{x}_{ij}'\boldsymbol{\beta}(y)_0 + \alpha\left(\boldsymbol{\nu}_i, y\right) + \gamma\left(\boldsymbol{\omega}_j, y\right)\right), \quad y \in \mathcal{Y}, \quad (i,j) \in \mathcal{D}, \tag{1}$$

where $\Lambda(\cdot)$ is a known link function that we assume to be the logistic distribution throughout this paper. As shown in the Appendix A, modeling the conditional distribution by this link function

[1]We consider that all the nodes are senders and receivers, but the method in this paper also allows for cases where the nodes that are senders differs from the nodes that are receivers, i.e., $i = 1, \ldots, I$ and $j = 1, \ldots J$, with $I \neq J$.

is equivalent to reparametrize the problem in terms of log-odds ratios as initially proposed by Foresi and Peracchi (1995). Moreover, $\boldsymbol{\beta}(y)_0$ is an unknown parameter vector of interest that varies with the levels of y; and $\alpha(\boldsymbol{\nu}_i, y)$ and $\gamma(\boldsymbol{\omega}_j, y)$ are unspecified measurable functions that can be seen as the unobserved individual effects at a given level of y. This model is naturally semiparametric, not only because the parameters are allowed to vary with the output levels but also because it does not restrict how the individual unobserved effects correlate with the covariates.

A key feature of models of dyadic interaction is the introduction of the two-way fixed effects. Given the double indices nature of the model, it is reasonable to assume that it exhibits a two-way error component structure captured by both individuals' nuisance terms. This structure incorporates essential aspects of networks since it allows for dependence across dyads. For instance, the outcome determined from the pairwise interaction between units i and j can be correlated with the outcome resulting from the interaction between i and k due to the fixed effect for unit i. Note that one drawback of dyadic models is that, in general, the strategic dimension of how outcomes are determined is ignored. However, they can replicate important stylized features of network models (Dzemski (2019)), thus, having become widely used when modeling networks. Finally, by allowing the fixed effects for senders and receivers to be different, together with the fact that y_{ij} need not be equal to y_{ji}, this model allows for directed networks.

Finally, the conditional distribution $F_{y_{ij}}(y \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j)$ can be written as:

$$
\begin{aligned}
F_{y_{ij}}\left(y \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right) &= \mathbb{E}[1\{y_{ij} \leq y\} \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j] \\
&= \Pr[\tilde{y}_{ij} = 1 \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j] \\
&= \Lambda\left(\boldsymbol{x}_{ij}'\boldsymbol{\beta}(y)_0 + \alpha\left(\boldsymbol{\nu}_i, y\right) + \gamma\left(\boldsymbol{\omega}_j, y\right)\right). \qquad (2)
\end{aligned}
$$

Therefore, by constructing a collection of binary variables $\tilde{y}_{ij} = 1\{y_{ij} \leq y\}$, for all pairs $(i, j) \in \mathcal{D}$ and all points in the region of interest $\mathcal{Y}$, $y \in \mathcal{Y}$, we see that the parameters of the DR model can be estimated by a continuum of binary (logistic) regressions with two-way fixed effects.

As highlighted by Arellano and Hahn (2007) in a standard non-linear panel data regression with one-way fixed effects and dimensions $i = 1, \ldots N$ and $t = 1, \ldots T$, if T is fixed and $N \to \infty$, there will be an estimation error in the estimates of the fixed effects. This follows from the fact that only a finite number T of observations is available to estimate each fixed effect. Since, in general, the fixed effects can be correlated with the exogenous regressors in an arbitrary way (and its distribution is left unspecified), the estimation error contaminates the estimates of the other parameters as well, as they are not informationally orthogonal. For large enough T, this bias should be small. However, even under $T \to \infty$ and $N \to \infty$, the fixed effects estimator will be asymptotically biased, leading to incorrect inference over the parameters and the average partial effects.

Despite the use of cross-sectional data in this paper, the dyadic structure of the model proposed in this paper provides a square pseudo-panel data setting. Therefore, the conclusions above for non-linear standard panel data setting carry over for estimating each binary logistic regression for each level of y as described in Equation 2. Hence, even when considering asymptotics under sequences of both dimensions $i = 1, ...N, j = 1, \ldots N$ going to infinity, the estimates obtained via maximum-likelihood are asymptotically biased. In fact, the introduction of a second fixed effect aggravates the resulting bias.

Following the results shown by Fernández-Val and Weidner (2016), I demonstrate how the asymptotic biases arise in this framework. For the sake of simplicity, I denote by β_y the slope parameter $\beta(y)$ which explicitly depends on the level of the outcome y. After concentrating out the nuisance parameters in the maximization of the likelihood for the slope parameter β_y in Equation 2, and using an asymptotic expansion for smooth likelihoods under appropriate regularity conditions[2]:

$$\bar{\beta}_y = \beta_{y,0} + \frac{\bar{B}_\infty}{N-1} + \frac{\bar{D}_\infty}{N-1} + o_P((N-1)^{-1}), \tag{3}$$

[2]For more details, see Appendix B, and for the complete derivation see Fernández-Val and Weidner (2016)

for some constants $\bar{B}_\infty$ and $\bar{D}_\infty$, and where $\beta_{y,0}$ is the true value of the parameter. By the properties of the maximum likelihood estimator, under regularity conditions:

$$\sqrt{N(N-1)}\left(\hat{\beta}_y - \bar{\beta}_y\right) \xrightarrow{d} N(0, \bar{V}_\infty), \tag{4}$$

for some variance $\bar{V}_\infty$. From Equation 3, as $N \to \infty$, $\hat{\beta}_y \xrightarrow{p} \beta_{y,0}$. Hence, the estimate of $\beta_{y,0}$ is consistent. However, from the equation above, the estimate converges to a distribution that is not centered at zero, which leads to incorrect asymptotic confidence intervals. This demonstrates the incidental parameters problem. In this context, this asymptotic bias arises as the order of the bias is higher than the inverse of the sample size due to the smaller rate of convergence of the fixed effects.

3. Estimation method

As outlined in the previous Section, the main challenge in the estimation of the continuum of binary regressions given by Equation 2 is that, even for a single binary regression, the incidental parameter problem (Neyman and Scott (1948)) stems from the presence of the two-way fixed effects. To circumvent this problem, I propose to estimate the parameters of the model $\theta(y) := (\beta(y), \alpha_1(y), \dots \alpha_I(y), \gamma_1(y), \dots \gamma_J(y))$ for each threshold point (for a given level y), independently, with the conditional maximum-likelihood method suggested by Charbonneau (2017) (for directed networks) and concurrently by Graham (2017) (for undirected networks). The core of this approach is to extend the conditional maximum likelihood method for standard panel data models with one fixed effect in Rasch (1960) and Arellano and Honoré (2001) to models with two-way fixed effects, relying on the existence of a sufficient statistics for the fixed effects when the link function follows a logistic distribution.

Even though this approach was initially proposed for network formation models, the model I consider for a single cutoff point resembles that of a network formation model. This follows because it is a discrete choice model that considers fixed effects for each node and dyad characteristics. For the sake of simplicity, and without loss of generality, I denote $\alpha_{i,y} = \alpha(\boldsymbol{\nu}_i, y)$ and

$\gamma_{j,y} = \gamma(\boldsymbol{\omega}_j, y)$.

From the conditional distribution $F_{y_{ij}}$ and the constructed binary variables $\tilde{y}_{ij}$:

$$\tilde{y}_{ij} = 1\{\boldsymbol{x}'_{ij}\boldsymbol{\beta}_{y,0} + \alpha_{i,y} + \gamma_{j,y} + \varepsilon_{ij} \geq 0\} \quad i = 1, \ldots, I, j = 1, \ldots, J$$

where $\alpha_{i,y}$ and $\gamma_{j,y}$ are fixed effects and we assume that ε_{ij} follows a logistic distribution. Therefore:

$$\mathbb{E}[1\{y_{ij} \leq y\} \mid \boldsymbol{x}_{ij}, \alpha_{i,y}, \gamma_{j,y}] = \Pr[\tilde{y}_{ij} = 1 \mid \boldsymbol{x}_{ij}, \alpha_{i,y}, \gamma_{j,y}] \tag{5}$$

$$= \frac{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}_{y,0} + \alpha_{i,y} + \gamma_{j,y})}{1 + \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}_{y,0} + \alpha_{i,y} + \gamma_{j,y})} \tag{6}$$

Under this model, it is possible to show that the sums across each dimension of the pseudo panel, $\sum_{j=1}^{N} \tilde{y}_{ij}$ and $\sum_{i=1}^{N} \tilde{y}_{ij}$, are sufficient statistics for $\alpha_{i,y}$ and $\gamma_{j,y}$. While this statement is previously proved for the standard panel case with one fixed effect, Charbonneau (2017) only implicitly provides this result. I provide a proof for this statement in Appendix C.

Even though one could propose a conditional maximum likelihood estimator based on the sufficient statistics, the maximization might be intractable. Fortunately, Charbonneau (2017) shows that it is possible to difference out the two-way fixed effects by further conditioning the above probability on the set of events $\{\tilde{y}_{ij} + \tilde{y}_{ik} = 1, \tilde{y}_{lj} + \tilde{y}_{lk} = 1, \tilde{y}_{ij} + \tilde{y}_{lk} = 1\}$ for different indices of senders and receivers $\{i, l; j, k\}$, such that:

$$\Pr[\tilde{y}_{ij} = 1 \mid \boldsymbol{x}_{ij}, \alpha_{i,y}, \gamma_{j,y}, \tilde{y}_{ij} + \tilde{y}_{ik} = 1, \tilde{y}_{lj} + \tilde{y}_{lk} = 1, \tilde{y}_{ij} + \tilde{y}_{lk} = 1]$$
$$= \frac{\exp(((\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik}) - (\boldsymbol{x}_{lj} - \boldsymbol{x}_{lk}))'\boldsymbol{\beta}_{y,0})}{1 + \exp(((\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik}) - (\boldsymbol{x}_{lj} - \boldsymbol{x}_{lk}))'\boldsymbol{\beta}_{y,0})}, \tag{7}$$

which also no longer depends on the fixed effects. This result is obtained by applying the same trick for the logit estimation in a static standard panel model with a single fixed effect.

The last expression is then applied to all quadruples of observations that satisfy the conditions

Figure 1: Two informative tetrads

$\{\tilde{y}_{ij} + \tilde{y}_{ik} = 1, \tilde{y}_{lj} + \tilde{y}_{lk} = 1, \tilde{y}_{ij} + \tilde{y}_{lk} = 1\}$. Hence, the function to maximize is given by:

$$\sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \sum_{l,k \in Z_{ij}} \log \left(\frac{\exp(((\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik}) - (\boldsymbol{x}_{lj} - \boldsymbol{x}_{lk}))'\boldsymbol{\beta}_y)}{1 + \exp(((\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik}) - (\boldsymbol{x}_{lj} - \boldsymbol{x}_{lk}))'\boldsymbol{\beta}_y)} \right), \tag{8}$$

where Z_{ij} is the set of all potential nodes k and l that satistifies the conditions $\{\tilde{y}_{ij} + \tilde{y}_{ik} = 1, \tilde{y}_{lj} + \tilde{y}_{lk} = 1, \tilde{y}_{ij} + \tilde{y}_{lk} = 1\}$ for the pair ij. In the next Section, I show a simple pairwise differences transformation of the outcomes $\tilde{y}_{ij}$ and the covariates $\boldsymbol{x}_{ij}$ followed by a logit estimation leads to the implementation of this estimator.

A loose intuition for the identification of the common parameters is analogous to that of Graham (2017) for the undirected case. The heterogeneity parameters (fixed effects) account for the in-degree and out-degree distributions of the network (the quantity of one's for a given node when the node is a sender or a receiver). Therefore, the precise location of the ones (or links) is driven by the variation provided by the covariates and the common parameters ($\boldsymbol{x}'_{ij}\boldsymbol{\beta}$). Thus, conditioning on the set $\{\tilde{y}_{ij} + \tilde{y}_{ik} = 1, \tilde{y}_{lj} + \tilde{y}_{lk} = 1, \tilde{y}_{ij} + \tilde{y}_{lk} = 1\}$ provides the ground for an estimator that is based on the relative probability of different types of subgraphs configurations with identical degree sequences, giving the necessary variation to identify the common parameters.

For instance, the above Subgraphs 1 and 2 in Figure 1 provide the same contribution of the unobserved heterogeneity to the likelihood, such that the conditional frequency to which each is observed depends only on the variation given by the covariates associated with each. In other words, in conditioning on the degree sequences of tetrads (since they are the same in both

(a) Subgraph 3

(b) Subgraph 4

Figure 2: Two non-informative tetrads

subgraphs), the only variation is the location of the links. This intuition aligns with the fact that the sums across each dimension are sufficient statistics for the fixed effects. At the same time, the conditioning events guarantee that for a node there is variation in the outcomes such that the common parameters can be identified. This feature cannot be seen in Figure 2, where the Subgraphs are informative to the likelihood, and the outcomes for all nodes do not present variation in a given Subgraph (i.e., a node is always sending a link or never sending a link in each Subgraph).

4. Asymptotic properties

4.1. Asymptotic properties for the regression at a single cutoff point

Throughout this section I treat the sequence of individual effects $\{\alpha_i, \gamma_j\}_N$ as fixed, since we always condition on them. Moreover, I consider asymptotic approximations where both dimensions of the pseudo-panel tends to infinity at the same rate.

Following Jochmans (2018), to derive the asymptotic properties of the proposed estimator, we define the following random variables, by fixing a quadruple of distinct nodes $\{i, l; j, k\}$ from $\mathcal{N}$:

$$z(\sigma\{i, l; j, k\}) = \frac{(\tilde{y}_{ij} - \tilde{y}_{ik}) - (\tilde{y}_{lj} - \tilde{y}_{lk})}{2}$$

$$r(\sigma\{i, l; j, k\}) = (\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik}) - (\boldsymbol{x}_{lj} - \boldsymbol{x}_{lk}),$$

where we introduced a function $\sigma(\cdot)$ that maps a quadruple to the index set $\mathcal{M} = \{1, 2, \ldots, M\}$,

M denoting the number of distinct quadruples from $\mathcal{N}$, i.e, $M = \binom{N}{2}\binom{N-2}{2} = \frac{N(N-1)(N-2)(N-3)}{4}$.[3] Each distinct quadruple of nodes $\{i, l; j, k\}$ corresponds then to a unique $\sigma\{i, l; j, k\} \in \mathcal{M}$. In the remainder of this section I will use the shortcut notation r_σ and x_σ.

We further note that the transformed dependent variable can take values from the set $\{-1, -1/2, 0, 1/2, 1\}$, and that the event that $z \in \{-1, 1\}$ corresponds to the condition $\{\tilde{y}_{ij}+\tilde{y}_{ik} = 1, \tilde{y}_{lj} + \tilde{y}_{lk} = 1, \tilde{y}_{ij} + \tilde{y}_{lk} = 1\}$. Therefore, by collecting $\boldsymbol{x} = (\boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \boldsymbol{x}_{lj}, \boldsymbol{x}_{lk})$,the results in the previous section leads to the following lemma:

Lemma 1 (Sufficiency).

$$\Pr[z_\sigma = 1 \mid \boldsymbol{x}, z_\sigma \in \{-1, 1\}] = \frac{\exp(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_{y,0})}{1 + \exp(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_{y,0})}$$

As before, conditional on $\boldsymbol{x}$ and on $z_\sigma \in \{-1, 1\}$, the distribution is logistic and does not depend on fixed effects. The conditional log-likelihood of the quadruple is:

$$1\{z_\sigma = 1\}\log\Lambda(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_{y,0}) + 1\{z_\sigma = -1\}\log(1 - \Lambda(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_{y,0})),$$

which form the basis of the construction of the quasi conditional maximum likelihood estimator for $\boldsymbol{\beta}$. Hence, we estimate the model by maximizing the empirical counterpart of this conditional log-likelihood for all distinct quadruples in $\mathcal{M}$. The estimator can be written as:

$$\hat{\boldsymbol{\beta}}_y = \underset{\boldsymbol{\beta}_y \in B}{\arg\max}\, L_n(\boldsymbol{\beta}_y)$$

[3]Note that the number of quadruples reflect the fact that the senders are permutation invariant, and the receivers as well.

where B is the parameter space searched over, and

$$L_n(\boldsymbol{\beta}_y) = \sum_{\sigma \in \mathcal{M}} 1\{z_\sigma = 1\}\log\Lambda(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_y) + 1\{z_\sigma = -1\}\log(1 - \Lambda(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_y))$$

It is clear at this point that the objective function is the same as the standard logit log-likelihood function applied to all quadruples that satisfy $z_\sigma \in \{-1, 1\}$. We denote the number of quadruples satisfying it by $M^* = \sum_{\sigma \in \mathcal{M}} 1\{z_\sigma \in \{-1, 1\}\}$.

The following set of (weak) assumptions are needed to establish consistency of the estimator:

Assumption 1 (Sampling). The N nodes in $\mathcal{N}$ are sampled independently.

Assumption 2 (Parameter Space). $\boldsymbol{\beta}_{y,0}$ is interior to B, a compact subset of $\mathcal{R}^{dim\boldsymbol{\beta}_y}$.

Assumption 3 (Moments). For all $(i, j) \in \mathcal{D}$, $\mathbb{E}(||x_{ij}||^2) < C$, where C is a finite constant.

Define the expected fraction of quadruples in the data that contribute to the log-likelihood as:

$$p_n = \frac{\mathbb{E}(M^*)}{M} = \frac{\sum_{\sigma \in \mathcal{M}} 1\{z_\sigma \in \{-1, 1\}\}}{M}.$$

Assumption 4 (Identification). $Np_n \to \infty$ as $N \to \infty$ and the matrix

$$\lim_{N \to \infty} (Mp_n)^{-1} \sum_{\sigma \in \mathcal{M}} \mathbb{E}(\boldsymbol{r}_\sigma \boldsymbol{r}'_\sigma f(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_{y,0}) 1\{z_\sigma \in \{-1, 1\}\})$$

has maximal rank.

Assumption 1 allows for dependence of the covariates across dyads that have nodes in common, a key feature in network models. That is, the network dependence of the data arises from the fact that not only the same fixed effects appear across different pairs, but also, the covariates of a dyad might be correlated with those of a different dyad with one node in common. Assumption 2 is standard for establishing consistency in non-linear models. Assumption 4 allows for the expected fraction of informative quadruples to shrink as N grows, allowing for sparse networks. However, p_n should not shrink faster than N^{-1}, implying that the accumulation of informative quadruples should not cease as the sample grows. These assumptions are the equivalent of general regularity conditions for non-linear models.

Theorem 1 (Consistency). Let Assumptions 1-4 hold. Then $\hat{\boldsymbol{\beta}}_y \xrightarrow{p} \boldsymbol{\beta}_{y,0}$ as $N \to \infty$.

Proof. Follows from Jochmans (2018), a more detailed proof is available in Appendix D.

Despite the fact that the empirical counterpart of the conditional log-likelihood has the form of a standard logit model, the conventional standard errors are not valid for $\hat{\boldsymbol{\beta}}_y$. The reasons are that: (i) the estimator is based on a quasi-likelihood, hence, the information equality does not hold; and (ii) the score vector involves sums over quadruples of nodes, such that each node appears in different summands, leading to dependences over such summands that need to be taken into account. To derive the asymptotic distribution of the estimator, we first need to streghten the moment requirements:

Assumption 5 (Moments). For all $(i, j) \in \mathcal{D}$, $\mathbb{E}(||x_{ij}||^6) < C$, where C is a finite constant.

Then, we introduce each summand of the score as:

$$s(\sigma, \boldsymbol{\beta}_y) = \boldsymbol{r}_\sigma \{1\{z_\sigma = 1\}(1 - \Lambda(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_y)) + 1\{z_\sigma = -1\}\Lambda(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_y)\}.$$

Hence, the score vector is:

$$S_n(\boldsymbol{\beta}_y) = \sum_i \sum_{i<l} \sum_{j \neq i,l} \sum_{j<k, k \neq i,l} s(\sigma\{i, l; j, k\}, \boldsymbol{\beta}_y)$$

The main result to characterize the distribution of the estimator is that $\Upsilon_n(\boldsymbol{\beta}_{y,0})^{-1/2} S_n(\boldsymbol{\beta}_{y,0}) \xrightarrow{d} N(0, I)$, where:

$$\Upsilon_n(\boldsymbol{\beta}_y) = \sum_{i=1}^{N} \sum_{j \neq i} v_{ij}(\boldsymbol{\beta}_y) v_{ij}(\boldsymbol{\beta}_y)'$$

$$v_{ij}(\boldsymbol{\beta}_y) = \sum_{l \neq i,j} \sum_{k \neq i,j,l} s(\sigma\{i,l;j,k\}, \boldsymbol{\beta}_y).$$

This result, combined with the Hessian that is given by:

$$H_n(\boldsymbol{\beta}_y) = -\sum_{\sigma \in \mathcal{M}} \boldsymbol{r}_\sigma \boldsymbol{r}'_\sigma f(\boldsymbol{r}'_\sigma \boldsymbol{\beta}_y 1\{z_\sigma \in \{-1,1\}\}),$$

where f is defined as the logistic density function. And, finally, defining:

$$\hat{\Omega} = H_n(\hat{\boldsymbol{\beta}}_y)^{-1} \Upsilon_n(\hat{\boldsymbol{\beta}}_y) H_n(\hat{\boldsymbol{\beta}}_y)^{-1}$$

We have:

Theorem 2 (Asymptotic distribution). Let Assumptions 1-5 hold. Then $||\hat{\boldsymbol{\beta}}_y - \boldsymbol{\beta}_{y,0}||= O_p(1/\sqrt{N(N-1)p_n})$ and

$$\hat{\Omega}^{-1/2}(\hat{\boldsymbol{\beta}}_y - \boldsymbol{\beta}_{y,0}) \xrightarrow{d} N(0, I)$$

as $N \to \infty$.

Proof. Follows from Jochmans (2018), a more detailed proof is available in Appendix D.

The proof for this theorems follows from the fact that we can write the score vector in the form of a U-statistics. Then, it is possible to define a Hajek projection for such statistics, such that the score evaluated at the true parameter value is asymptotically equivalent to it (conditional on covariates). Finally, by using Hoeffding decomposition one can define its asymptotic variance, and by arguments of conditional independence, it is possible to derive its limiting distribution. The main argument is that, following traditional dyadic models, the probability of $\tilde{y}_{ij} = 1$ for a given dyad i, j is conditionally independent of the probability for the remaining dyads, conditioning on the node (fixed effects) and dyad (covariates) characteristics. One drawback is that transitivity across the probabilities is not taken into account by this model. It rules out interdependent link preferences, where individuals' preferences over a link may vary with the presence or absence of links elsewhere in the network. However, it is shown by Dzemski (2019) that such a dyadic structure can recover the transitivity observed in some datasets.

Imporantly, Theorem 2 shows that, pointwise, the estimator converges at a parametric rate. A natural next step is to demonstrate the uniform covergence across all thresholds of the distribution, since up to now it was shown the asymptotic properties for a single cutoff point; and to provide simultaneous confidence bands as in Chernozhukov et al. (2020), which can be estimated with a valid bootstrap procedure. However, these two steps will be provided in the next versions of this paper.

This paper is not the first in the literature to propose an estimation method for distribution regression in a network framework. Chernozhukov et al. (2020) proposes to estimate the parameters of the model $\theta(y) := (\beta(y), \alpha_1(y), \dots \alpha_I(y), \gamma_1(y), \dots \gamma_J(y))$ also separately for each threshold. The key difference to my approach is that they employ a conditional maximum likelihood method with analytical bias corrections initially proposed by Fernández-Val and Weidner (2016) and later applied to the context of a network by Yan et al. (2019) and Dzemski (2019). However, as mentioned before, in the context of a network formation model, such a method requires that the underlying network is dense, meaning that, in the DR context, the conditional probabilities of the events $\{y_{ij} \leq y\}$ are bounded away from zero and one. Therefore, in the extreme quantiles of the distribution, such an approach might be susceptible to the incidental parameter problem. As Assumption 4 indicates, the method proposed in this paper allows for the expected fraction of quadruples (and therefore, the probability of forming a link in a network formation model) to shrink to zero as N grows.[4]

Furthermore, while analytical bias corrections allow to cover a broader class of models, it does not completely eliminate the asymptotic bias. In comparison, the pairwise difference eliminates it entirely by differencing out the nuisance parameters; and allows for the presence of many zeros or ones, which are observed in the constructed binary variable $\tilde{y}$ for extreme quantiles of the conditional distribution.

[4]Note that in the usual definition of a sparse network, the probability of forming links should also not converge to one. By conjecture, Assumption 4 also allows for this notion of sparsity.

4.2. Uniform convergence over the continuum of thresholds

To be provided in next versions of this paper.

4.3. Simultaneous confidence bands

To be provided in next versions of this paper.

5. Monte Carlo simulations

5.1. Monte Carlo simulations for a single threshold

In this section, I propose a Monte Carlo simulation exercise for a single threshold y, which boils down to a network formation model. The aim is to compare the performance of the bias correction methods [5] to that of the Charbonneau (2017) approach under different levels of sparsity of the network.

I follow a standard data generating process for directed networks (and dyad settings), similar to those in Jochmans (2018), Dzemski (2019) and Yan et al. (2019). I generate the single regressor as:

$$x_{ij} = - \mid u_i - u_j \mid,$$

where $u_i = \nu_i - \frac{1}{2}$ for $\nu_i \sim Beta(2, 2)$, and the true parameter value β_0 is set to one. The fixed effects are a deterministic function of the sample size:

$$\alpha_i = -\frac{N-i}{N-1} C_n, \quad \gamma_i = \alpha_i$$

where N is the number of nodes in the network and the constant C_n usually depends on it, specifically, the larger the value of C_n, the sparser the generated network. In the above mentioned papers, $C_n \in \{0, log(log(N)), log(N)\}$. Moreover, note that the source of the dependence across dyads comes from both the covariates structure and the inclusion of the fixed effects.

[5] As initially proposed by Fernández-Val and Weidner (2016) for standard panel models with large N and T, and latter adapted to the networks framework by Yan et al. (2019) and Dzemski (2019)

I deviate from the standard specifications for C_n such that it is possible to have a better comparison between the two methods. Namely, I vary the constant C_n for different sample sizes indicated by the number of nodes N such that the number of informative quadruples for the Charbonneau (2017) estimator remains constant. The results for 1000 simulations can be seen in Table 1.

	$C_n = 7.75$ $N = 50$	$C_n = 10.9$ $N = 70$	$C_n = 14.1$ $N = 90$
Mean bias (PD)	-0.0434	-0.0304	-0.0645
Mean bias (BC)	-0.0922	-0.1085	-0.1148
Median bias (PD)	0.0486	0.0227	0.0021
Median bias (BC)	-0.0364	-0.0403	-0.0575
RMSE (PD)	2.1177	2.0462	2.0953
RMSE (BC)	1.7963	1.8264	1.8040
Mean bias (PD winsorized)	-0.0135	-0.0013	-0.0355
Mean bias (BC winsorized)	-0.0742	-0.0816	-0.0973
Median bias (PD winsorized)	0.0486	0.0227	0.0021
Median bias (BC winsorized)	-0.0364	-0.0403	-0.0575
RMSE (PD winsorized)	1.6743	1.6691	1.6660
RMSE (BC winsorized)	1.5083	1.5074	1.5165
Size t-test (PD)	0.0232	0.0228	0.0207
Avr. perc. quadruples	0.000162	4.1e-5	1.5e-5
Avr. perc. links	0.011558	0.005906	0.003547
Avr. quadruples contributing	223.5959	227.1113	224.7812

Table 1: Based on DGP by Jochmans (2018). Symmetric case with 10000 simulations.

The simulation exercise shows that when increasing both N and C_n, while the number of quadruples contributing to the likelihood function of the Pairwise Differencing (PD) estimator remains reasonably constant, the sparsity in the network increases (as it is reflected by the average percentage links). Note that, as expected from before, the mean bias of the PD estimates are generally smaller in magnitude than that of the Bias Corrected (BC) estimator. Moreover, the Median Bias displays a curious behavior: for the PD estimator, it shrinks as the sparsity grows, while for the BC increases - the latter shows that the worsening of the performance of the BC estimates are not only due to outliers. However, in the setting for $N = 50$ the median biases of both estimators are similar in magnitude, but as the network becomes sparser, the median biases for the BC estimator worsen while for the PD estimator improves considerably.

To further reduce the effect of outliers, I winsorize the top and bottom 5% of the estimates. Analyzing the winsorized result, it is even clearer that the PD estimates reduces the bias by a larger magnitude than the BC estimates throughout all the settings for the simulations. Finally, the RMSE of the PD estimator is generally bigger in magnitude, as expected, since the PD estimator is not efficient. All in all, this simulation exercise shows that the PD estimator performs better compared to the BC as the network becomes sparser, as long as the number of informative quadruples remains reasonable. Thus, the same conclusion should carry over to the DR estimators based on both methods, especially in the extreme quantiles of the distribution.

5.2. Monte Carlo simulations for the entire distribution

To analyze the finite sample properties of both DR estimators, I follow the same Monte Carlo simulations setting as Chernozhukov et al. (2020), which is calibrated to the empirical application in the following Section for gravity models of international trade. More specifically, I set the outcome to be generated by a censored logistic process

$$y_{ij}^s = \max\left\{x_{ij}'\hat{\beta} + \widehat{\alpha}_i + \widehat{\gamma}_j + \widehat{\sigma}\Lambda^{-1}\left(u_{ij}^s\right)/\sigma_L, 0\right\}, \quad (i,j) \in \mathcal{D}$$

where $\mathcal{D} = \{(i,j) : 1 \leq i,j \leq 157, i \neq j\}$, x_{ij} is the value of the covariates for the observational unit (i,j) in the trade data set, y_{ij}^s is the level of exports from country i to j, $\sigma_L = \pi/\sqrt{3}$, the standard deviation of the logistic distribution, and $\left(\widehat{\beta}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_I, \widehat{\gamma}_1, \ldots, \widehat{\gamma}_J, \widehat{\sigma}\right)$ are Tobit fixed effect estimates of the parameters in the trade data set with lower censoring point at zero. Moreover, I set the errors to be independently drawn from a uniform distribution $\mathcal{U}(0,1)$. For simplicity, in this simulation exercise, I consider only one covariate, the log of the distance between countries.

Importantly, it can be shown that the conditional distribution of the dependent variable y_{ij}^s is equivalent to a DR model as defined before, where:

$$\beta(y) = \sigma_L\left(e_1 y - \widehat{\beta}\right)/\widehat{\sigma}, \quad \alpha_i(y) = -\sigma_L\widehat{\alpha}_i/\widehat{\sigma}, \quad \text{and} \quad \gamma_j(y) = -\sigma_L\widehat{\gamma}_j/\widehat{\sigma}$$

with e_1 the unit vector of dimension d_x with a one in the first component. The results are based on 250 simulations for now due to computational limitations (*it is to be expanded in next versions of this paper*).

Figure 3 shows the absolute bias, absolute median bias, and the RMSE obtained with a naive fixed effects logit estimation (Uncorrected Logit UL), the Bias Correction (BC) method of Chernozhukov et al. (2020) and with the proposed Pairwise Differences (PD) estimator in this paper. Both BC and PD reduce the bias significantly in finite samples compared to the UL estimates. However, as expected, in the extreme quantiles, the BC method does not seem to fully correct for the bias (both mean and median), while the PD biases do not increase relative to the other quantiles. Moreover, as expected, the RMSE for the UL is the biggest due to the high magnitude of the biases. Also, since the PD estimator is not efficient, its RMSE is larger than that of the BC.

Figure 4 displays the percentage of informative quadruples to the likelihood of the PD estimator, the number of informative quadruples, and the zoomed-in number of informative quadruples (for scaling reasons). Naturally, as the quantiles increase, the percentage and number of quadruples decrease significantly. However, at the 99% quantile, there are still about 10000 informative quadruples, which renders the PD estimation robust at the tail of the outcome distribution.

6. Application to gravity models of international trade

I consider the estimation of gravity equations for bilateral trade between countries, using the same data as Helpman et al. (2008) and Chernozhukov et al. (2020). It contains information on bilateral trase flows and covariates for 157 countries in 1986 (Congo is excluded due to the perfect prediction problem, i.e., lack of variation in the dependent variable). Both i and j index countries as exporters and importers.

The outcome y_{ij} is the volume of trade in thousands of constant 2000 US dollars from country i to country j, and the covariates $P(x_{ij}) = x_{ij}$ include determinants of bilateral trade flows such as the logarithm of the distance in kilometers between country i 's capital and country j 's capital

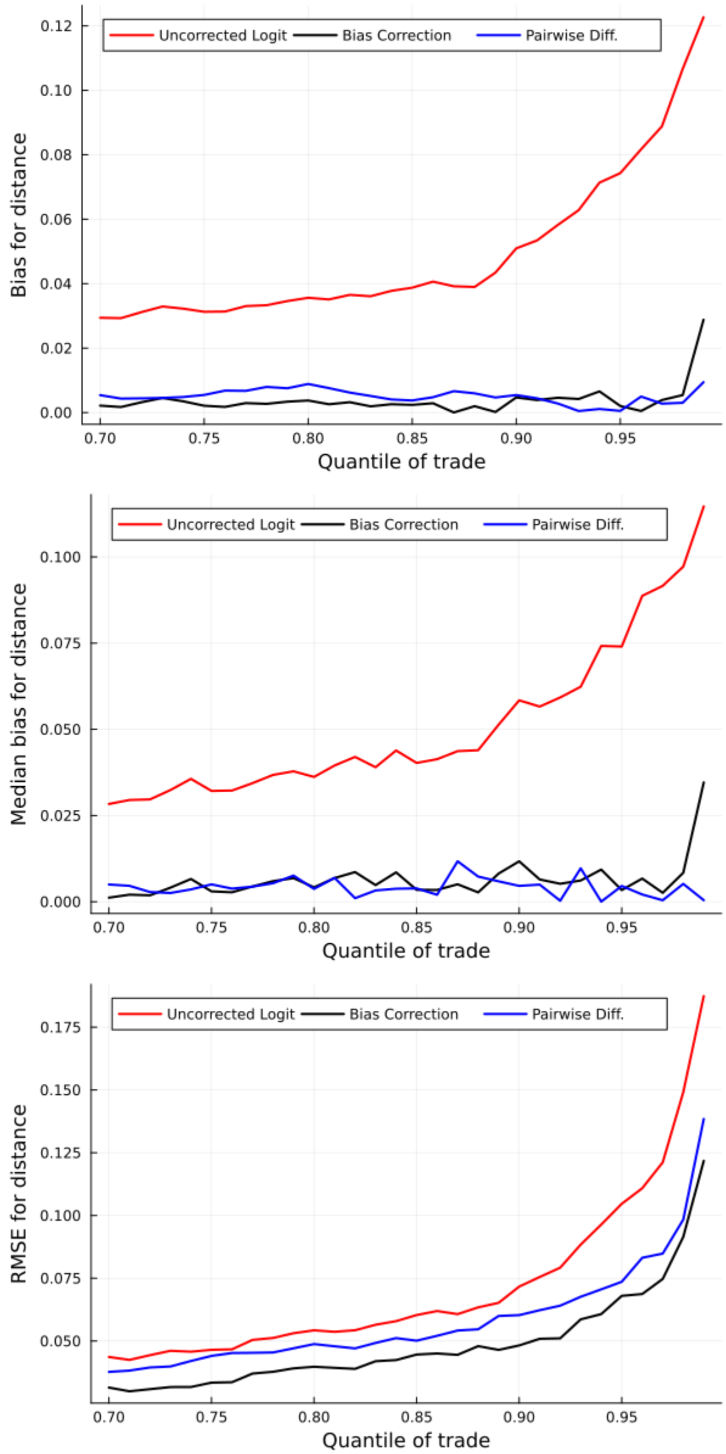

Figure 3: Simulation results for the DR coefficients of log distance.

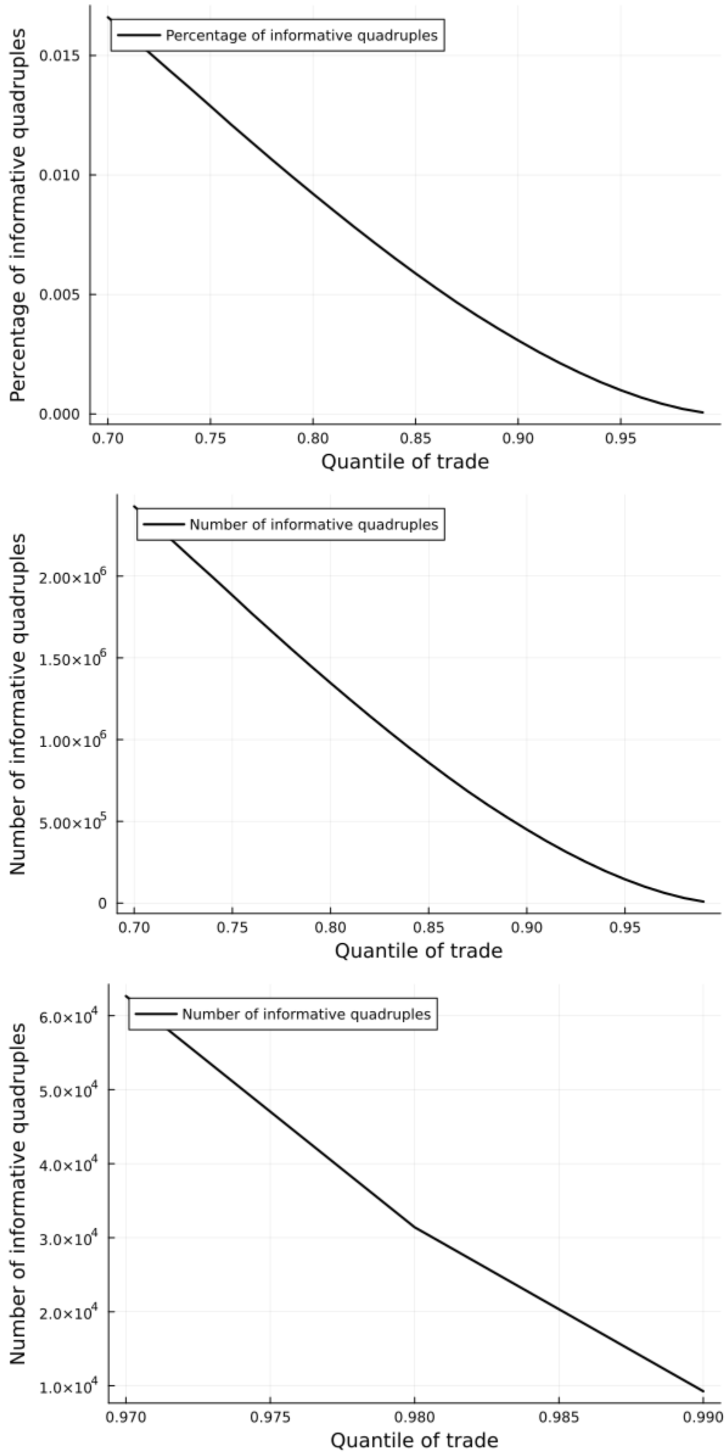

Figure 4: Simulation results for the DR regression, informative quadruples.

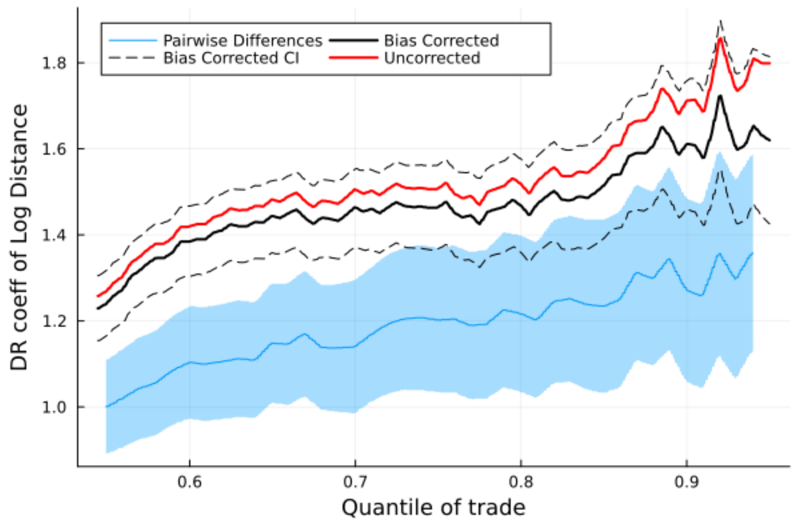

Figure 5: Estimates and 95% pointwise confidence intervals for the DR-coefficients of log distance.

and indicators for common colonial ties, currency union, regional free trade area (FTA), border, legal system, language, and religion.

There are no trade flows for 55% of the country pairs in the dataset. The volume of trade variable exhibits much larger standard deviation than the mean. Since this variable is bounded below at zero, this indicates the presence of a very heavy upper tail in the distribution. This feature also makes quantile methods specially well-suited for this application on robustness grounds.

Figure 5 shows estimates and 95% pointwise confidence intervals for the DR coefficients of log distance plotted against the quantile indexes of the volume of trade, obtained by Chernozhukov et al. (2020) using bias corrected (BC) fixed effects estimates, the method proposed in this paper; obtained when using the pairwise differencing of outcomes (PD), and the uncorrected FE logit estimation. The x-axis starts at .54, the maximum quantile index corresponding to zero volume of trade. The region of interest $\mathcal{Y}$ corresponds to the interval between zero and the 0.95-quantile of the volume of trade. Note that the sign of the effect in terms of volume of trade, y_{ij} , is the opposite to the sign of the DR coefficient. Figure 6 shows the analogous estimates for the DR coefficients of legal.

As in Chernozhukov et al. (2020), note that the difference between the uncorrected and the bias corrected estimates is of the same order as the magnitude of the width of the confidence

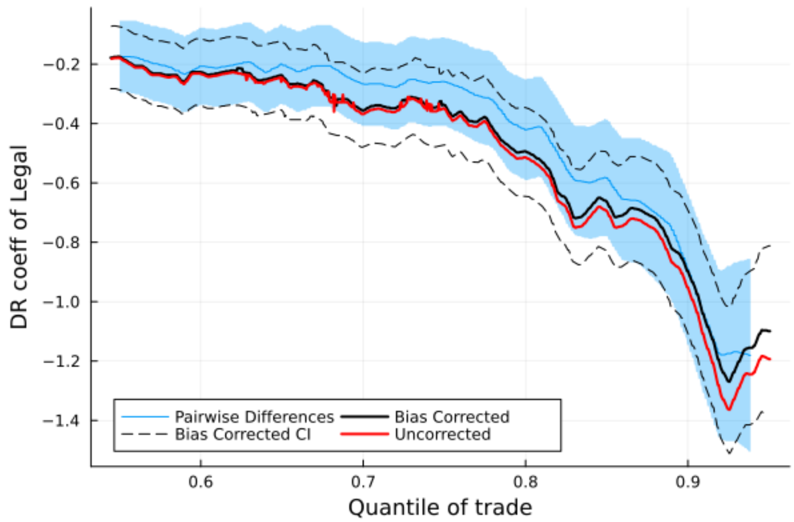

Figure 6: Estimates and 95% pointwise confidence intervals for the DR-coefficients of legal.

intervals[6] for the log distance. Moreover, the largest estimated bias when comparing the two and also when comparing with the pairwise difference estimator lies on the upper quantiles of trade, where the constructed binary variables have less variation. However, our estimates suggest an even higher bias in magnitude across the entire distribution. Interestingly, when compared to the BC estimates, the difference between the two estimates (PD and BC) seems to be constant across the distribution. Importantly, the PD estimates suggests that the effect of distance across the distribution is significant, but of smaller magnitude. A similar conclusion is drawn in general for the coefficients for legal. An exception is at the upper quantiles, where the PD estimates suggest a smalled bias relative to the BC estimates. Finally, as expected the confidence intervals around the proposed estimator are wider compared to that of the BC estimator, since this estimator is not as efficient as standard MLE.

Even though it would be desirable to compute counterfactual effects for the estimated distribution function (or, quantile functions), a drawback of this approach is that such an estimation is infeasible. This occurs since there are no available estimates for the fixed effects, or average marginal effects in general, as opposed to the bias correction method of Chernozhukov et al.

[6]For now, the confidence intervals for the PD estimator are obtained with the pointwise estimates of the standard errors.

(2020). However, there is a further interpretation to the estimated coefficients in terms of the derivatives of the conditional quantiles under certain conditions.

The conditional distribution of y_{ij} given the covariates and the unobserved effects can be represented by either the conditional distribution function or the conditional quantile function. While these equivalent representations correspond to two alternative approaches to estimation, there are relevant links between DR and quantile regression (QR) estimates, as shown by Koenker et al. (2013). In particular, following results from Chernozhukov et al. (2020) for our framework, it is possible to show that the common parameters of the model are related to the derivatives of the conditional quantiles under certain conditions.

When y_{ij} is continuous, the model given by Equation 1 has the representation as an implicit nonseparable model by the probability integral transform:

$$\Lambda\left(\boldsymbol{x}_{ij}'\boldsymbol{\beta}\left(y_{ij}\right)+\alpha\left(\boldsymbol{\nu}_i, y_{ij}\right)+\gamma\left(\boldsymbol{\omega}_j, y_{ij}\right)\right)=u_{ij}, \quad u_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j \sim U(0,1). \tag{9}$$

where, as commonly seen in DR or QR approaches, u_{ij} represents the unobserved ranking of the observation y_{ij} in the conditional distribution. Let $Q\left(u|\boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right)$ be the u-quantile of y_{ij} conditional on $(\boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j)$. This quantile function can be defined as:

$$Q\left(u \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right)=\inf\left\{y \in \mathcal{Y}: F_{y_{ij}}\left(y \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right) \geq u\right\} \wedge \sup\{y \in \mathcal{Y}\}. \tag{10}$$

It can be shown that if (i) $F_{y_{ij}}\left(y \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right)$ is strictly increasing in the support of y_{ij}; (ii) $\partial\Lambda(z)/\partial z>0$ for all y in the support of y_{ij}; and (iii) $Q\left(u \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right)$ is differentiable, then the DR coefficients are proportional to (minus) derivatives of the conditional quantile function, and ratios of the DR coefficients correspond to ratios of derivatives:

$$\left.\frac{\beta_\ell(y)}{\beta_k(y)}\right|_{y=Q\left(u|\boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right)}=\frac{\partial_{x_{ij}^\ell} Q\left(u \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right)}{\partial_{x_{ij}^k} Q\left(u \mid \boldsymbol{x}_{ij}, \boldsymbol{\nu}_i, \boldsymbol{\omega}_j\right)}, \quad \ell, k=1, \ldots, d_x \tag{11}$$

Therefore, based on the figures above for the estimated coefficients, it is possible to infer that the marginal effects of distance on the quantile functions is larger in magnitude than that of the legal variable.

Finally, there is a consolidated debate in the international trade literature regarding the homogeneity of trade elasticities, which ultimately affects welfare gains from trade (Arkolakis et al. (2012), Melitz and Redding (2015), Chen and Novy (2022)). The method presented in this paper provides a straightforward way (provided the confidence bands for the estimates) to test the heterogeneity of trade elasticities across different quantiles of the distribution of trade, which proves to be of importance in the literature.

7. Conclusion

This paper introduces a novel method for estimating distribution regressions in a network setting. To accommodate the network structure, It employs a semiparametric approach, treating two-way unit-specific effects as fixed parameters, and addresses the incidental parameter problem using a conditional maximum-likelihood approach initially proposed for network formation models (Charbonneau (2017), Jochmans (2018)). The proposed method provides consistent estimates and robust inference pointwise, particularly for extremum quantiles of the distribution. This approach fills a gap in the econometrics of network model literature by accommodating zero outcomes and sparse networks without relying on strong assumptions regarding how the zero outcomes are generated. Moreover, the empirical application demonstrates that this method is of practical relevance, allowing, for instance, to test whether the elasticities of gravity models of international trade are heterogeneous across thresholds. A current drawback of the proposed method is that estimates of counterfactual distributions are infeasible. This is because the estimates of fixed effects are unavailable, and the average (marginal) effects of network formation models remain set-identified when the network is sparse - which is the case of the underlying network in the extreme quantiles of the distribution of outcomes. Therefore, future research would involve obtaining estimates for bounds on the partially identified average effects of network models.

References

Arellano, M. and J. Hahn (2007): "Understanding bias in nonlinear panel models: Some recent developments," *Econometric Society Monographs*, 43, 381.

Arellano, M. and B. Honoré (2001): "Panel data models: some recent developments," in *Handbook of econometrics*, Elsevier, vol. 5, 3229–3296.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012): "New trade models, same old gains?" *American Economic Review*, 102, 94–130.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "The diffusion of microfinance," *Science*, 341, 1236498.

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

Chen, N. and D. Novy (2022): "Gravity and heterogeneous trade cost elasticities," *The Economic Journal*, 132, 1349–1377.

Chernozhukov, V., I. Fernandez-Val, and M. Weidner (2020): "Network and panel quantile effects via distribution regression," *Journal of Econometrics*.

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

Fafchamps, M. and F. Gubert (2007): "The formation of risk sharing networks," *Journal of development Economics*, 83, 326–350.

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

Foresi, S. and F. Peracchi (1995): "The conditional distribution of excess returns: An empirical analysis," *Journal of the American Statistical Association*, 90, 451–466.

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

——— (2020): "Dyadic regression," in *The Econometric Analysis of Network Data*, Elsevier, 23–40.

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

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

Koenker, R., S. Leorato, and F. Peracchi (2013): "Distributional vs. quantile regression," .

Melitz, M. J. and S. J. Redding (2015): "New trade models, new welfare implications," *American Economic Review*, 105, 1105–1146.

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

Peracchi, F. (2002): "On estimating conditional quantiles and distribution functions," *Computational statistics & data analysis*, 38, 433–447.

Rasch, G. (1960): *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.*, Nielsen & Lydiche.

Silva, J. S. and S. Tenreyro (2006): "The log of gravity," *The Review of Economics and statistics*, 88, 641–658.

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

Appendix A. An introduction to Distribution Regression (DR)

In this section I present an introduction to the DR approach, following the initial proposal by Foresi and Peracchi (1995), and further discussed in Peracchi (2002) and Koenker et al. (2013). Consider the problem of estimating the conditional distribution of a random variable Y given a vector of X covariates in a standard cross-sectional case. Note that the interest is not in merely a few quantiles but in the entire conditional distribution, $F(y|x)$.

It is proposed to select J distinct values $-\infty < y_1 < \cdots < y_J < \infty$ in the range of interest of Y (which is related to the quantiles of the distribution of Y), and estimate J functions $F_1(x), \ldots F_J(x)$, with $F_j(x) = F(y_j|x), j = 1, \ldots J$. It is argued that by suitably choosing J and their position, one can get a reasonably accurate description of $F(y|x)$.

If the conditional distribution of Y is continuous with support on the entire real line, then at any point x in the support of X, the sequence of conditional distribution functions must satisfy:

$$0 < F_j(x) < 1, \quad j = 1, \ldots, J, \tag{A.1}$$

$$0 < F_1(x) < \cdots < F_J(x) < 1. \tag{A.2}$$

To impose the condition given by Equation A.1, it is suggested to not model $F_j(x)$ directly, but rather to estimate the log-odds $\eta_j(x) = \ln[F_j(x)/(1 - F_j(x))]$. Then, given this estimate of the $\eta_j(x)$, one can estimate the conditional distribution at the threshold j by:

$$\hat{F}_j(x) = \frac{\exp \hat{\eta}_j(x)}{1 + \exp \hat{\eta}_j(x)}. \tag{A.3}$$

Let $\mathcal{H}$ be the class of functions of x that are possible candidates for the log-odds ratio. Since the random variable $1\{Y \leq y_j\}$ has a Bernoulli distribution with parameter $F_j(x)$, by the definition of the cumulative conditional distribution, we can define the best Kullback-Leibler approximation

$\eta_j^*(x)$ to $\eta_j(x)$ in the class of functions $\mathcal{H}$ as the minimizer of $\mathcal{K}(\eta, \eta_j) = l(\eta_j) - l(\eta)$, with:

$$\begin{aligned} l(\eta) &= \mathbb{E}[1\{Y \leq y_j\}\eta(X) - \ln(1 + \exp \eta(X))] \\ &= \mathbb{E}[F_j(X)\eta(X) - \ln(1 + \exp \eta(X))]. \end{aligned} \tag{A.4}$$

The first expectation is taken with respect to the joint distribution of (X, Y), and the second with respect to the marginal distribution of X. Therefore, the function η_j^* maximizes $l(\eta)$ over the class $\mathcal{H}$. If $\eta_j \in \mathcal{H}$, then $\eta_j^* = \eta_j$. Importantly, if X is a scalar random variable, and $\mathcal{H}$ is the class of functions linear in x, then the best Kullback-Leibler approximation to $\eta_j(x)$ is of a linear form $\eta_j^*(x) = \gamma_j + x\delta_j$, where (γ_j, δ_j) are such that the approximation error

$$F_j(X) - \frac{\exp n_j^*(X)}{1 + \exp n_j^*(X)}$$

has mean zero and is uncorrelated with X. Therefore, η_j^* can be estimated by maximizing the sample log-likelihood:

$$L(\eta) = n^{-1} \sum_{i=1}^{n} [1\{Y_i \leq y_j\}\eta(X_i) - \ln(1 + \exp \eta(X_i))].$$

over the linear functions in the class $\mathcal{H}$. Clearly, this is obtained by fitting J separate logistic regressions, one for each binary random variable $1\{Y_i \leq y_j\}, \quad j = 1, \ldots J$. Alternative specifications for the class of functions $\mathcal{H}$, for instance, non-linear specifications, entails alternative estimation methods. One caveat of this approach is that while it satisfies the condition given by Equation A.1, by modeling the log-odds ratio, it does not guarantee the monotonicity condition given by Equation A.2.

Appendix B. The incidental parameter problem

As highlighted by Arellano and Hahn (2007) in a standard panel data regression with one way fixed effects and dimensions $i = 1, \ldots N$ and $t = 1, \ldots T$, if T is fixed and $N \to \infty$, there will be an estimation error in the estimates of the fixed effects, as only a finite number T of observations are available to estimate each fixed effect. As we allow for the fixed effects to be correlated with the exogenous regressors (and its distribution is left unspecified), this estimation error contaminates the estimates of the other parameters as well, as they are not informationally orthogonal. For large enough T, this bias should be small. However, even under $T \to \infty$ and $N \to \infty$, the fixed effects estimator will be asymptotically biased, leading to incorrect inference over the parameters and the average partial effects.

The same argument holds for the present framework of a dyadic regression with two-way fixed effects. In our panel data model, we have two dimensions: $i = 1, ..N, j = 1, \ldots N$. However, both the dimensions grow at rate N. I will consider asymptotic results such that $N \to \infty$.

Note as well that for each new country in the dataset, the number of observations is increased by $2(N-1)$. Moreover, for each fixed effect in Equation 2 there are $(N-1)$ observations available for their estimation. I will now use results shown by Fernández-Val and Weidner (2016) to demonstrate how the incidental parameter problem arises in this framework, delivering consistent but asymptotic biased estimators, keeping in mind that as $N \to \infty$ both dimensions i and j go to infinity and also the number of observations $N(N-1)$ go to infinity.

Given the dataset of $N(N-1)$ observations $\left\{ \left(\tilde{y}_{ij}, x'_{ij} \right)' : 1 \leq i \leq N, 1 \leq j \leq N, i \neq j \right\}$ with $\tilde{y}_{ij} = \mathbb{1} \left(\tilde{y}^*_{ij} > 0 \right)$, we have that $\tilde{y}_{ij}$ is generated by the process:

$$\tilde{y}_{ij} \mid x_{ij}, \alpha_y, \gamma_y, \beta_y \sim f_Y \left(\cdot \mid x_{ij}, \alpha_y, \gamma_y, \beta_y \right)$$

where: $\alpha_y = \left(\alpha_{1,y}, \ldots \alpha_{N,y} \right), \gamma_y = \left(\gamma_{1,y}, \ldots \gamma_{N,y} \right), f_Y$ is a known probability function and $\alpha_{i,y}, \gamma_{j,y}$ are the unobserved fixed effects. Note here that this approach is semi-parametric in the sense that is does not specify the distribution of the fixed effects or their relationship with the explanatory variables.

We can further model the conditional distribution of $y_{2,ij}$ using a single-index specification with fixed effects, since it is a binary response model:

$$f_Y\left(\tilde{y}_{ij} \mid x_{ij}, \alpha_y, \gamma_y, \beta_y\right) = F\left(x'_{ij}\beta_y + \alpha_{i,y} + \gamma_{j,y}\right)^{\tilde{y}_{ij}} \times \left[1 - F\left(x'_{ij}\beta_y + \alpha_{i,y} + \gamma_{j,y}\right)\right]^{1-\tilde{y}_{ij}},$$

where, clearly $\tilde{y}_{ij} \in \{0, 1\}$ and F is a cumulative distribution function, defined to be a standard logistic.

I can then collect all the fixed effects to be estimated in the vector $\omega_{NN,y} = (\alpha_{1,y}, \ldots \alpha_{N,y}, \gamma_{1,y}, \ldots \gamma_{N,y})'$, which can be seen as a nuisance parameter vector. Then, the true values of the parameters, denoted by $\beta_{y,0}$ and $\omega_{NN,y,0}$ are the solution to the population conditional maximum likelihood maximization:

$$\max_{(\beta_y, \omega_{NN,y}) \in \mathbb{R}^{\dim \beta_y + \dim \omega_{NN,y}}} \mathbb{E}_\omega \left[\mathcal{L}\left(\beta_y, \omega_{NN,y}\right)\right]$$

with

$$\mathcal{L}\left(\beta_y, \omega_{NN,y}\right) = (N(N-1))^{-1} \left\{ \sum_{i=1}^{N} \sum_{j \neq i} \log f_Y\left(\tilde{y}_{ij} \mid x_{ij}, \alpha_y, \gamma_y, \beta_y\right) - b\left(\iota'_{NN}\omega_{NN,y}\right)^2/2 \right\}$$

where $\mathbb{E}_\omega$ denotes the expectation with respect to the distribution of the data conditional on the unobserved effects and strictly exogenous variables, $b > 0$ is an arbitrary constant, $\iota_{NN} = (1'_N, -1'_N)'$ and 1_N denotes a vector of ones of dimension N.

The second term of $\mathcal{L}$ relates to a penalty that imposes a normalization to identify the fixed effects in models with two-way fixed effects that enter in the log-likelihood function as $\alpha_{i,y} + \gamma_{j,y}$. To be more specific, in this case, adding a constant to all $\alpha_{i,y}$ and subtracting the same constant from all $\gamma_{j,y}$ would not change $\alpha_{i,y} + \gamma_{j,y}$. Thus, without this normalization, the parameters $\alpha_{i,y}$ and $\gamma_{j,y}$ are not identifiable.

To estimate the parameters, we solve the sample analogue of the following equation:

$$\max_{(\beta_y, \omega_{NN,y}) \in \mathbb{R}^{\dim \beta_y + \dim \omega_{NN,y}}} \mathcal{L}\left(\beta_y, \omega_{NN,y}\right)$$

In order to analyze the statistical properties of β_y, we first concentrate out the nuisance parameters $\omega_{NN,y}$, such that for given β_y, the optimal $\hat{\omega}_{NN,y}(\beta_y)$ is:

$$\hat{\omega}_{NN,y}(\beta_y) = \text{argmax}_{\omega_{NN,y} \in \mathbb{R}^{\dim \omega_{NN,y}}} \mathcal{L}(\beta_y, \omega_{NN,y})$$

Thus, the fixed effects estimator of β_y and $\omega_{NN,y}$ are, by plugging in the previous expression for $\hat{\omega}_{NN,y}(\beta_y)$:

$$\hat{\beta}_y = \text{argmax}_{\beta_y \in \mathbb{R}^{\dim \beta_y}} \mathcal{L}(\beta_y, \hat{\omega}_{NN,y}(\beta_y)) \tag{B.1}$$

$$\hat{\omega}_{NN,y}(\beta_y) = \hat{\omega}_{NN,y}\left(\hat{\beta}_y\right) \tag{B.2}$$

The source of the problem is that the dimension of the nuisance parameters $\omega_{NN,y}$ increases with the sample size under asymptotic approximations where $N \to \infty$. To further describe the incidental parameter problem, denote:

$$\bar{\beta}_y = \text{argmax}_{\beta_y \in \mathbb{R}^{\dim \beta_y}} \mathbb{E}_\omega \left[\mathcal{L}(\beta_y, \hat{\omega}_{NN,y}(\beta_y))\right]$$

Using an asymptotic expansion for smooth likelihoods under appropriate regularity conditions, provided by Fernández-Val and Weidner (2016), we have that:

$$\bar{\beta}_y = \beta_{y,0} + \frac{\bar{B}_\infty}{(N-1)} + \frac{\bar{D}_\infty}{(N-1)} + o_P\left((N-1)^{-1}\right) \tag{B.3}$$

For some constants $\bar{B}_\infty$ and $\bar{D}_\infty$. The derivation for this expression can be found in the Appendix of Fernández-Val and Weidner (2016). As explained by the authors, the expansion is obtained by first taking a firstorder Taylor expansion of the Equation B.1 around the true value $\beta_{y,0}$, as it is usually done to obtain the asymptotic properties of such estimator. Then, one should additionally take a second-order Taylor expansion of the obtained term $\frac{\partial \mathcal{L}(\beta_{y,0}, \hat{\omega}_{NN})}{\partial_{\beta_y}}$ around the true values of the nuisance terms. Intuitively, this second step demonstrates how the estimates of the fixed effects affect the estimates of the structural parameter β_y. To obtain the exact form of the expressions $\bar{B}_\infty$ and $\bar{D}_\infty$ a quite involved derivation is needed. However, this is not the focus of our study, since we show later that there are other possibilities to correct for the asymptotic bias generated by these terms other than deriving the biases themselves.

Moreover, by the properties of the maximum likelihood estimator we have that, under regularity conditions:

$$\sqrt{N(N-1)}\left(\hat{\beta}_y - \bar{\beta}_y\right) \xrightarrow{d} N\left(0, \bar{V}_{B\infty}\right)$$

For some $\bar{V}_{B\infty}$. By substituting the expression for $\beta_{y,0}$ obtained in Equation B.3, we obtain that, by Slutsky's theorem:

$$\begin{aligned}
&\sqrt{N(N-1)}\left(\hat{\beta}_y - \beta_{y,0}\right) \\
&= \sqrt{N(N-1)}\left(\hat{\beta}_y - \bar{\beta}_y\right) \\
&+ \sqrt{N(N-1)}\left(\frac{\bar{B}_\infty}{(N-1)} + \frac{\bar{D}_\infty}{(N-1)} + o_P\left((N-1)^{-1}\right)\right) \\
&\xrightarrow{d} N\left(\bar{B}_\infty + \bar{D}_\infty, \bar{V}_{B\infty}\right)
\end{aligned}$$

We can see from Equation B.3 that, as $N \to \infty, \hat{\beta}_y \xrightarrow{p} \beta_{y,0}$ ($\beta_{y,0}$ being the true value of the parameter), thus, the estimates of $\beta_{y,0}$ are consistent. However, from the equation above we see that the estimates converge to a distribution that is not centered at zero, which leads to incorrect asymptotic confidence intervals. This demonstrates the incidental parameters problem, that boils down to an asymptotic bias in the estimates of $\beta_{y,0}$. This asymptotic bias arises as the order of the bias is higher than the inverse of the sample size because of the small rate of convergence of the fixed effects.

Appendix C. Sufficient statistics

In this section, I provide a proof that $\sum_{j=1}^{N} \tilde{y}_{ij}$ and $\sum_{i=1}^{N} \tilde{y}_{ij}$ are indeed the sufficient statistics for $\boldsymbol{\alpha}_{i,y}$ and $\boldsymbol{\gamma}_{j,y}$. In the following, for the sake of simplification of notation, I omit the subscript y that denotes the threshold of the outcome variable. Denoting by $\tilde{\boldsymbol{Y}}$ the vector of all observations $(\tilde{y}_{11}, \dots \tilde{y}_{IJ})$; $\boldsymbol{r}$ the vector of sums of rows $(r_1, \dots, r_I)$ where $r_i = \sum_{j=1}^{J} \tilde{y}_{ij}$; $\boldsymbol{c}$ the vector of sums of columsn $(c_1, \ldots, c_J)$ where $c_j = \sum_{i=1}^{I} \tilde{y}_{ij}$; and $\boldsymbol{x}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ the vectors of covariates and fixed effects, we have that:

$$Pr[\tilde{\boldsymbol{Y}} \mid \boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}] = \frac{Pr[\tilde{\boldsymbol{Y}} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}]}{Pr[\boldsymbol{r}, \boldsymbol{c} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}]},$$

with $Pr[\boldsymbol{r}, \boldsymbol{c} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}] = \sum_{\bar{\tilde{\boldsymbol{Y}}} \in Q} Pr[\tilde{\boldsymbol{Y}} = \bar{\tilde{\boldsymbol{Y}}} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}]$, where Q is the set of all possible combinarions of $\tilde{y}_{ij}$ in $\tilde{\boldsymbol{Y}}$ such that the sum of the rows is given by $\boldsymbol{r}$ and the sum of columns by $\boldsymbol{c}$.

Following the proposed model for the constructed binary variables $\tilde{y}_{ij}$:

$$\tilde{y}_{ij} = 1\left\{\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \alpha_i + \gamma_j + \varepsilon_{ij} \geq 0\right\} \quad i = 1, \ldots, I, j = 1, \ldots, J$$

we have that:

$$Pr[\tilde{y}_{ij} \mid \boldsymbol{x}_{ij}, \alpha_i, \gamma_j] = \frac{\exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \alpha_i + \gamma_j)^{\tilde{y}_{ij}}}{1 + \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \alpha_i + \gamma_j)}.$$

Therefore, the joint probability of all the outcomes, conditional on $\sum_{j=1}^{N} y_{ij}$ and $\sum_{i=1}^{N} y_{ij}$ is:

$$\begin{aligned}
Pr[\tilde{\boldsymbol{Y}} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}] &= \prod_{i \neq j} Pr[\tilde{y}_{ij} \mid \boldsymbol{x}_{ij}, \alpha_i, \gamma_j; \boldsymbol{\beta}] \\
&= \frac{\exp(\sum_{i \neq j} \tilde{y}_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \sum_{i \neq j} \tilde{y}_{ij}(\alpha_i + \gamma_j))}{\prod_{i \neq j}[1 + \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \alpha_i + \gamma_j)]} \\
&= \frac{\exp(\sum_{i \neq j} \tilde{y}_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}) \exp(\sum_{i=1}^{I} \alpha_i \sum_{j=1}^{J} \tilde{y}_{ij}) \exp(\sum_{j=1}^{J} \gamma_j \sum_{i=1}^{I} \tilde{y}_{ij})}{\prod_{i \neq j}[1 + \exp(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \alpha_i + \gamma_j)]}
\end{aligned}$$

And analogously for $Pr[\bar{\tilde{\boldsymbol{Y}}} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}]$. Then, we can write:

$$\frac{Pr[\tilde{\boldsymbol{Y}} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}]}{\sum_{\bar{\tilde{Y}} \in Q} Pr[\bar{\tilde{\boldsymbol{Y}}} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}]} = \frac{\exp(\sum_{i \neq j} \tilde{y}_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}) \exp(\sum_{i=1}^{I} \alpha_i \sum_{j=1}^{J} \tilde{y}_{ij}) \exp(\sum_{j=1}^{J} \gamma_j \sum_{i=1}^{I} \tilde{y}_{ij})}{\sum_{\bar{\tilde{Y}} \in Q} \exp(\sum_{i \neq j} \bar{\tilde{y}}_{ij}\boldsymbol{x}_{ij}'\boldsymbol{\beta}) \exp(\sum_{i=1}^{I} \alpha_i \sum_{j=1}^{J} \bar{\tilde{y}}_{ij}) \exp(\sum_{j=1}^{J} \gamma_j \sum_{i=1}^{I} \bar{\tilde{y}}_{ij})}$$

Finally, independently of which set in Q we consider, we have that, by the construction of the set, $\sum_{i=1}^{I} \bar{\tilde{y}}_{ij} = \sum_{i=1}^{I} \tilde{y}_{ij}$ and $\sum_{j=1}^{J} \bar{\tilde{y}}_{ij} = \sum_{j=1}^{J} \tilde{y}_{ij}$, such that:

$$\frac{Pr[\tilde{\boldsymbol{Y}} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}]}{\sum_{\bar{\tilde{Y}} \in Q} Pr[\bar{\tilde{\boldsymbol{Y}}} \mid \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{x}; \boldsymbol{\beta}]} = \frac{\exp(\sum_{i \neq j} \tilde{y}_{ij} \boldsymbol{x}'_{ij} \boldsymbol{\beta})}{\sum_{\bar{\tilde{Y}} \in Q} \exp(\sum_{i \neq j} \bar{\tilde{y}}_{ij} \boldsymbol{x}'_{ij} \boldsymbol{\beta})}$$

which does not depend on the fixed effects, rendering the result.