

Functional Differencing in Networks*

Stéphane Bonhomme[†]

Kevin Dano[‡]

July 17, 2023

Abstract

Economic interactions often occur in networks where heterogeneous agents (such as workers or firms) sort and produce. However, most existing estimation approaches either require the network to be dense, which is at odds with many empirical networks, or they require restricting the form of heterogeneity and the network formation process. We show how the functional differencing approach introduced by [Bonhomme \(2012\)](#) in the context of panel data, can be applied in network settings to derive moment restrictions on model parameters and average effects. Those restrictions are valid irrespective of the form of heterogeneity, and they hold in both dense and sparse networks. We illustrate the analysis with linear and nonlinear models of matched employer-employee data, in the spirit of the model introduced by [Abowd, Kramarz, and Margolis \(1999\)](#).

KEYWORDS: Econometric models of networks, matching, sorting, heterogeneity, functional differencing.

*This paper has been prepared for the special issue of the *Revue Économique* in honor of Jean-Marc Robin. We thank the editor, an anonymous reviewer, Jaime Arellano-Bover, Jesús Carro, Martin Weidner, and Tom Zohar for comments.

[†]University of Chicago.

[‡]University of California Berkeley.

1 Introduction

Network data is increasingly prevalent in applied economics. In this paper we focus on models where agents (e.g., workers and firms) sort and interact on a network. In such settings, accounting for unobserved heterogeneity is empirically key. However, existing approaches to estimation in the presence of flexible heterogeneity are imperfect.

A first approach consists in treating the heterogeneity as “fixed effects” parameters to be estimated. Bias reduction methods, initially developed for single-agent panel data (Hahn and Newey, 2004, Dhaene and Jochmans, 2015, Fernández-Val and Weidner, 2016), have been recently extended to networks (e.g., Graham, 2017, Hughes, 2022). The fixed-effects approach is appealing since it does not require modeling the distribution of heterogeneity and how it correlates with conditioning variables. Additionally, in models where agents interact on an exogenous network, the fixed-effects approach does not require specifying a model of network formation.

However, the performance of bias reduction methods in fixed-effects models hinges crucially on the network being sufficiently dense. This requirement is at odds with the nature of several empirical networks. For instance, in applications to wage determination in the presence of worker and firm heterogeneity (Abowd, Kramarz, and Margolis, 1999), a dense network approximation is typically inappropriate, and fixed-effects estimates suffer from a “limited mobility bias” that may be substantial (Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler, 2023). More generally, sparsity is a feature of many empirical networks (Graham, 2020).

A second approach consists in postulating a “random effects” model for the unobserved heterogeneity. For example, Bonhomme, Lamadon, and Manresa (2019) and Lentz, Piyapromdee, and Robin (2022) propose and estimate random-effects models of worker heterogeneity in the presence of firm heterogeneity to account for sorting and complementarity on the labor market. Studying a different setting, Bonhomme (2021) develops random effects models of agent heterogeneity in team production networks. Random-effects methods enjoy theoretical guarantees in sparse networks, under the assumption that the model is correctly specified.

However, modeling the full distribution of heterogeneity given conditioning variables can be challenging. In models of wage determination in the presence of worker and firm

heterogeneity, this requires modeling the heterogeneity conditional on the entire network of employment relationships and job transitions, as proposed, for example, by [Woodcock \(2008\)](#) and [Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler \(2023\)](#). More generally, the random-effects approach effectively requires modeling the network formation process, which can be a difficult task due to dimensionality and equilibrium multiplicity challenges.

In this paper our aim is to achieve the best of these two approaches, in the sense that we seek estimators that are fully robust to the form of heterogeneity, and that behave well in denser and sparser networks. Such estimators currently exist only in very special cases. Notably, [Andrews, Gill, Schank, and Upward \(2008\)](#) and [Kline, Saggio, and Sølvsten \(2020\)](#) propose exact bias corrections for fixed-effects estimators of variance components in linear regressions on networks, while [Graham \(2017\)](#) proposes a “tetrad logit” estimator in a logistic model of network formation. These strategies mimic panel data methods that are consistent in fixed-length panels, such as conditional logit estimators ([Rasch, 1960](#), [Andersen, 1970](#)) or estimators of variance components ([Arellano and Bonhomme, 2012](#)).

In panel data, the functional differencing approach ([Bonhomme, 2012](#)) provides a general methodology to find moment restrictions on parameters that are robust to any distribution of heterogeneity and correlation with conditioning variables, and hold in fixed-length panels. Recent applications of the approach include the derivation of new moment restrictions in binary and discrete choice models, both static and dynamic ([Honoré and Weidner, 2020](#), [Honoré, Muris, and Weidner, 2021](#), [Dano, 2023](#)). In certain models, no exact moment restrictions exist. However, [Dhaene and Weidner \(2023\)](#) show how to regularize the functional differencing moments to provide restrictions that are satisfied up to a vanishing approximation error as the length of the panel tends to infinity.

The starting point of this paper is the observation that the scope of functional differencing is not limited to panel data, and the approach can be applied to any setting with a parametric conditional distribution involving latent variables. Our main goal is to apply the functional differencing approach to derive moment restrictions on parameters in some network settings. Specifically, we consider linear and binary choice logit models on networks, including a novel “AKM logit model” that provides a counterpart to the AKM estimator of [Abowd, Kramarz, and Margolis \(1999\)](#) for binary outcomes. In those models, we characterize the available moment restrictions on parameters.

In addition, we study average effects that depend on the joint distribution of unobserved heterogeneity and observed covariates. In panel binary choice models, average effects have been studied by various authors (see, e.g., [Chernozhukov, Fernández-Val, Hahn, and Newey, 2013](#), [Davezies, D’Haultfoeuille, and Laage, 2021](#), [Dobronyi, Gu, and Kim, 2021](#), [Aguirregabiria and Carro, 2021](#), and [Pakel and Weidner, 2021](#)). The functional differencing approach can be applied to average effects, as initially shown in the working paper version of [Bonhomme \(2012\)](#). This approach applies to general panel data models where the outcome distribution is parametrically specified and the distribution of heterogeneity and covariates is unrestricted. Here we use it to derive moment conditions on average effects in network settings.

Lastly, as in its panel data applications, in the settings that we consider in this paper the functional differencing approach delivers moment restrictions on parameters, yet it does not guarantee identification or consistent estimation of the parameters given those restrictions. Although, in several examples that we study, identification can be verified directly and analog estimators can be constructed, applying the approach to other models will generally require careful analysis of statistical properties. We only briefly touch on estimation at the end of the paper, and leave a deeper analysis of identification and estimation to future work. At the same time, we see the functional differencing approach as a promising building block for researchers to discover novel moment restrictions and estimators in network settings in the future.

The outline of the paper is as follows. In [Section 2](#) we present a class of models with multi-sided heterogeneity in networks. In [Section 3](#) we describe the functional differencing approach. In [Sections 4, 5, 6, and 7](#) we illustrate the approach with various examples. Finally, in [Section 8](#) we briefly sketch how to construct estimators based on functional differencing moment restrictions.

2 Models with multi-sided heterogeneity in networks

In this section we introduce and describe a framework for heterogeneous agents interacting on a network. The subsequent sections show how to derive moment restrictions on parameters and average effects in this framework.

2.1 Description and examples

The model consists of two layers: a model of the network, and a model of agents' outcomes on the network.

The *network* is represented by a graph, or more generally a hypergraph, featuring nodes and edges. Nodes in the network correspond to economic agents, and edges represent their links or collaborations. The network can be static or dynamic, and we will see that, under the assumption that the network is exogenous, a specification of the network formation model is not needed.

A key feature of the framework is the presence of agent-specific *types*, which govern sorting patterns and affect outcomes, yet are latent to the econometrician.

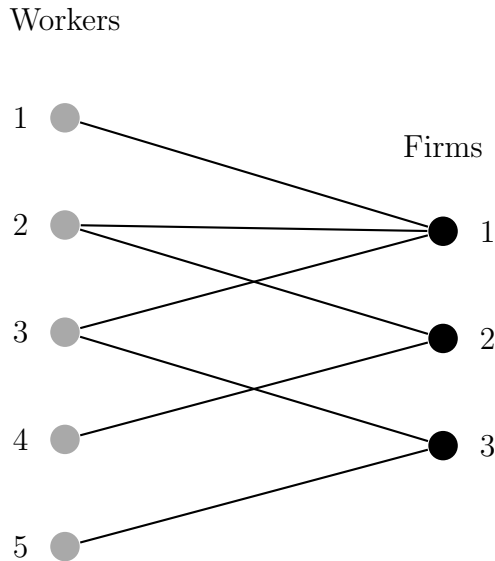
Throughout the paper, we will refer to economic agents as workers and firms as a leading example. In settings with workers and firms, we model the network as bipartite, links are employment relationships, and both workers and firms are heterogeneous. We show an example in Figure 1. Employment, job mobility and wages all depend on the workers' and firms' latent types. Prominent models of this kind were proposed in [Becker \(1973\)](#), [Shimer and Smith \(2000\)](#), [Postel-Vinay and Robin \(2002\)](#), and [Lentz, Piyapromdee, and Robin \(2022\)](#), among many others.

Outcomes in the network depend on the agents' latent types, and on covariates that are observed by the econometrician. Outcomes also depend on idiosyncratic errors, or "shocks".

A key assumption that we will maintain in this paper is that the network is *exogenous*, in the sense that network links are independent of the shocks conditional on agents' types and covariates. In models with workers and firms, [Bonhomme, Lamadon, and Manresa \(2019\)](#) show that this assumption is satisfied in the models of [Shimer and Smith \(2000\)](#) and [Lentz, Piyapromdee, and Robin \(2022\)](#) but that it fails in the model with sequential bargaining of [Postel-Vinay and Robin \(2002\)](#), for example.

Our main focus will be on estimating parameters governing the model of outcomes while conditioning on the network. This approach only restricts the network insofar as exogeneity with respect to idiosyncratic shocks is required. However, it allows for general forms of matching and sorting patterns. Beyond the example of workers and firms, this setup is relevant to other bipartite network settings appearing in the economics of

Figure 1: A worker-firm network



Notes: Nodes are workers (numbered from 1 to 5) and firms (numbered from 1 to 3). Edges are lines linking workers to firms, representing employment relationships. Workers 1,4,5 stay in a single firm, while workers 2 and 3 move between two firms.

education, empirical finance, and trade (Bonhomme, 2020). Non-bipartite settings, such as models of team production, are also encompassed by this framework (Ahmadpoor and Jones, 2019, Bonhomme, 2021).

In applications, the network formation model may also be of interest. Since the framework accounts for unobserved heterogeneity, it can be used to study network formation models with additive or non-additive heterogeneity and independent link-specific shocks (Graham, 2017, Bickel and Chen, 2009). However, since our approach relies on a parametric likelihood for the distribution of outcomes, it is not well-suited to study the determinants of link formation in models with strategic interactions (De Paula, Richards-Shubik, and Tamer, 2018, Gualdani, 2021, Sheng, 2020).

2.2 Probabilistic framework

Let Y denote a vector of outcomes, one for every link (or edge) in the network. Let A denote the vector of latent types, one for every agent (or node). Finally, let X denote a matrix indicating which agents are linked together. In models with covariates, which may

be agent-specific or link-specific, we include those in X .

We postulate a parametric model for outcomes conditional on the latent types and the network links (and possibly covariates),

$$(Y | X = x, A = a) \sim f_{\theta}(\cdot | x, a), \quad (1)$$

where f_{θ} is a parametric distribution indexed by a finite-dimensional parameter vector θ .

We leave the joint distribution of latent types and network links (and possibly covariates) fully unrestricted. Formally,

$$(A, X) \sim \pi, \quad (2)$$

where π is an unknown (i.e., nonparametric) distribution.

The combination of a parametric outcome distribution and a nonparametric distribution of heterogeneity and covariates is common in the panel data literature. Indeed, (1)-(2) nests the standard panel data case, where X simply indicates which observations correspond to the same worker. Nevertheless, the current framework is not limited to panel data.

A key assumption implied by the specification of a likelihood conditional on X (and A) is that the network is assumed exogenous. In panel data models, this corresponds to the assumption of strict exogeneity of covariates. This assumption is commonly relaxed in linear panel models, for example using the sequential moment restrictions introduced by [Arellano and Bond \(1991\)](#). A subsequent literature allows for sequentially exogenous network links in linear models (e.g., [Kuersteiner and Prucha, 2020](#)). Allowing for sequential exogeneity of covariates in nonlinear panel data models is still a frontier research area (see [Bonhomme, Dano, and Graham, 2023](#)).

In a setting with workers and firms, a leading example of (1) is a model of wage determination. Following the pioneering approach of [Abowd, Kramarz, and Margolis \(1999\)](#), a linear specification for log-wages is

$$Y_{it} = \alpha_i + \psi_{j(i,t)} + \varepsilon_{it}, \quad (3)$$

where $i \in \{1, \dots, N\}$ are workers, $t \in \{1, \dots, T\}$ are time periods, $j(i, t) \in \{1, \dots, J\}$ denotes the firm where i is employed at t , ε_{it} is an idiosyncratic shock, and we are abstracting from

exogenous covariates for simplicity. In this model, network exogeneity is often referred to as “exogenous mobility”, meaning that job transitions (represented by the firm indicators $j(i, t)$) are assumed independent of the ε_{it} ’s conditional on the α_i ’s and the ψ_j ’s.

In order to allow for complementarity patterns, one may be interested in a nonlinear extension of (3), such as a Constant Elasticity of Substitution (CES) specification in logarithms,

$$Y_{it} = \ln \lambda + \frac{1}{\gamma} \ln \left(\rho \alpha_i^\gamma + (1 - \rho) \psi_{j(i,t)}^\gamma \right) + \varepsilon_{it}. \quad (4)$$

Nonlinear log wage specifications have been proposed and estimated by [Bonhomme, Lamadon, and Manresa \(2019\)](#) and [Lentz, Piyapromdee, and Robin \(2022\)](#). However, a difference between their approaches and the one we propose here is that those authors restrict the distribution of heterogeneity and how it relates to employment relationships and job transitions. That is, in both papers, π in (2) is restricted, while it is fully unrestricted in the current framework. In addition, both [Bonhomme, Lamadon, and Manresa \(2019\)](#) and [Lentz, Piyapromdee, and Robin \(2022\)](#) rely on large firms to recover firm heterogeneity, whereas our approach will yield valid moment restrictions irrespective of the degree of sparsity of the network.

In models (3) and (4) one can form an $NT \times (N + J)$ matrix X , by having each row denoting an observation (or edge) (i, t) , and having both sets of agents (or nodes) i and j in the columns of X . In this case, one can equivalently write (3) and (4) as

$$(3') \quad Y = XA + \varepsilon, \quad \text{and} \quad (4') \quad Y = g(X, A, \lambda, \rho, \gamma) + \varepsilon,$$

where the vector A contains the α ’s and ψ ’s, and the function g takes a known (CES) form. Since we assume that the model is parametric, we specify the distribution of ε conditional on X and A , for example as a normal distribution with zero mean and diagonal covariance matrix with variance σ^2 . In this case, $\theta = \sigma^2$ in the linear model, and $\theta = (\lambda, \rho, \gamma, \sigma^2)$ in the CES model.

The framework in (1)-(2) also nests certain models of network formation. As an example, consider the model of link formation in [Graham \(2017\)](#). Undirected binary links $Y_{ij} \in \{0, 1\}$ between agents i and j are determined based on covariates X_{ij} and

agent-specific types A_i and A_j as

$$Y_{ij} = \mathbf{1} \{X'_{ij}\theta + A_i + A_j + \varepsilon_{ij} \geq 0\}, \quad (5)$$

where the ε_{ij} are standard logistic, independent of A and X , and i.i.d. across pairs of agents $(i, j) \in \{1, \dots, N\}^2$, $i \neq j$. [Graham \(2017\)](#) leaves the joint distribution of X and A unrestricted, and his model is thus a special case of the framework in (1)-(2).

3 Functional differencing: A general presentation

In the framework (1)-(2) we ask two questions. First, how can one derive moment restrictions on θ ? Second, how can one find moment restrictions on quantities depending on θ and (A, X) , such as average effects? Both questions can be answered using the functional differencing approach, and we address them in turn.

3.1 Parameter θ

To answer the first question, the following proposition shows how to characterize moment restrictions on θ that are valid irrespective of the heterogeneity distribution.

Proposition 1 *Let $\phi_\theta(y, x)$ be a function of outcomes and network links (and/or covariates). Suppose that $\mathbb{E}[\phi_\theta(Y, X) | A, X]$ is bounded. The following three statements are equivalent:*

(i) *For any joint distribution of (A, X) we have*

$$\mathbb{E}[\phi_\theta(Y, X)] = 0.$$

(ii) *Almost surely in A, X ,*

$$\mathbb{E}[\phi_\theta(Y, X) | A, X] = 0.$$

(iii) *For almost all a and x ,*

$$\int \phi_\theta(y, x) f_\theta(y | x, a) dy = 0. \quad (6)$$

It is immediate that (ii) implies (i), and that (ii) and (iii) are equivalent. The implication from (i) to (ii) comes from the fact that we require the moment restriction to hold

irrespective of the distribution of A and X . A formal argument is provided in Appendix A. Note also that (ii) implies moment restrictions conditional on X , $\mathbb{E}[\phi_\theta(Y, X) | X] = 0$. Proposition 1 implies that, to look for moment restrictions on θ , it is necessary and sufficient to find solutions to the linear functional equation in (6).

Proposition 1 is the main insight of the functional differencing approach (Bonhomme, 2012). Indeed, finding a ϕ satisfying (6) amounts to finding an element in the null space of the conditional expectation operator associated with the parametric conditional model of Y given (A, X) . To this end, Bonhomme (2012) proposed numerical projection methods, while Honoré and Weidner (2020) relied on symbolic computing to find analytical ϕ functions in dynamic discrete choice settings. In some models, however, one can show that no non-trivial solution ϕ exists, which implies the absence of informative moment equality restrictions on θ . To deal with such cases, Dhaene and Weidner (2023) propose an approximate functional differencing approach. In dynamic panel data logit models, Dobronyi, Gu, and Kim (2021) derive the identified set on the parameters, which combines moment equality and inequality restrictions.

While most applications of the functional differencing approach so far are confined to panel data settings without interactions between agents, Proposition 1 makes clear that the scope of the approach is not limited to those settings. A conventional panel data model consists of a collection of individual-specific submodels indexed by the individual heterogeneity A_i , the vector of individual covariates $(X'_{i1}, \dots, X'_{iT})'$, and the parameter θ that is common across individuals. In contrast, in network settings all units may potentially be related, and both the network matrix X and the vector of heterogeneity A may affect all outcomes in the network. As Proposition 1 illustrates, this difference between panel data and network settings is immaterial from the point of view of the applicability of functional differencing. In the next sections we will provide examples to illustrate the usefulness of the functional differencing approach when applied to networks.

3.2 Average effects

We now turn our attention to linear functionals of the distribution of A and X , and answer the second question. Using functional differencing, we show how to obtain, for

given θ , moment restrictions on an average effect of the form

$$\mu = \mathbb{E}[m_\theta(A, X)],$$

where $m_\theta(\cdot)$ is known given θ , and the expectation is taken with respect to the joint distribution π of (A, X) . As an example, in the CES model of wage determination (4), one may be interested in estimating average marginal effects of worker or firm heterogeneity on log wages, while accounting for the presence of complementarity between worker and firm effects. The following proposition provides a counterpart to Proposition 1 for such target parameters.

Proposition 2 *Let $\psi_\theta(y, x)$ be a function of outcomes and network links (and/or covariates). Suppose that $\mathbb{E}[\psi_\theta(Y, X) | A, X] - m_\theta(A, X)$ is bounded. The following three statements are equivalent:*

(i) *For any joint distribution of (A, X) we have*

$$\mathbb{E}[\psi_\theta(Y, X)] = \mathbb{E}[m_\theta(A, X)].$$

(ii) *Almost surely in A, X ,*

$$\mathbb{E}[\psi_\theta(Y, X) | A, X] = m_\theta(A, X).$$

(iii) *For almost all a and x ,*

$$\int \psi_\theta(y, x) f_\theta(y | x, a) dy = m_\theta(a, x). \tag{7}$$

The equivalence between the three parts is again easy to see (see Appendix A), and the usefulness of the result comes from the fact that (7) is a linear functional equation. In this case as well, the linear operator in (7) is known given θ . Proposition 2 shows that the functional differencing approach initially applied in Bonhomme (2012) to obtain restrictions on model parameters in panel data models can be applied to derive restrictions on average effects in other models with latent variables, including network settings.

Finding a moment representation for μ amounts to solving a linear functional system, which is a Fredholm integral equation of the first kind. Numerical and analytical methods can be used to construct ψ functions that satisfy (7). In dynamic panel logit models,

Aguirregabiria and Carro (2021), Dobronyi, Gu, and Kim (2021) and Dano (2023) show that the computation of average marginal effects is analytically straightforward. However, in models with continuous outcomes the inverse problem in (7) is generally ill-posed (Carrasco, Florens, and Renault, 2007, Engl, Hanke, and Neubauer, 1996). Given a solution ψ to the functional system, using it for estimation in a finite sample thus typically requires regularization. In this paper we focus on finding moment functions ϕ and ψ , and leave a detailed study of estimators and their properties to future work. See Section 8 for further discussion.

4 Linear network models

In this section and the next three we illustrate Propositions 1 and 2 through various examples. Consider first a linear model with an X matrix that consists of two parts, $X = (X_1, X_2)$, where X_1 is a matrix of network links and X_2 is a matrix of covariates. An example is the AKM model of Abowd, Kramarz, and Margolis (1999), given by an augmented version of (3) that includes covariates. We specify

$$Y = X_1 A + X_2 \beta + \varepsilon, \quad (\varepsilon | X, A) \sim iid\mathcal{N}(0, \sigma^2 I_n), \quad (8)$$

where n is the number of observations, I_n denotes the $n \times n$ identity matrix, and we denote $\theta = (\beta', \sigma^2)'$.

In this model we will focus on the parameters β and σ^2 , and on quadratic forms $\mu = \mathbb{E}[A'QA]$ for a symmetric matrix Q . Variance components, which can be written as quadratic forms, are of interest for, e.g., decomposing the variance of log wages into components reflecting worker heterogeneity, firm heterogeneity, and sorting patterns between heterogeneous workers and firms (Abowd, Kramarz, and Margolis, 1999, Card, Heining, and Kline, 2013, Song, Price, Guvenen, Bloom, and Von Wachter, 2019).

4.1 Parameters β and σ^2

In this subsection we derive moment restrictions on β and σ^2 . For this purpose we rely on Proposition 1. We start by noting that (6) can be equivalently written as

$$\int \phi_\theta(y, x) \exp\left(-\frac{1}{2\sigma^2}(y - x_1a - x_2\beta)'(y - x_1a - x_2\beta)\right) dy = 0.$$

It is useful to introduce the Moore-Penrose pseudo-inverse x_1^\dagger of x_1 , and to write the two orthogonal projectors associated with x_1 as $x_1x_1^\dagger = u_1u_1'$ and $I_n - x_1x_1^\dagger = u_2u_2'$, where $u = (u_1, u_2)$ is orthogonal. Let $v = y - x_2\beta$, $v_1 = u_1'v$, and $v_2 = u_2'v$.

We then note that (6) can equivalently be written as

$$\begin{aligned} \int \phi_\theta(y, x) \exp\left(-\frac{1}{2\sigma^2}(y - x_1a - x_2\beta)'x_1x_1^\dagger(y - x_1a - x_2\beta)\right) \\ \times \exp\left(-\frac{1}{2\sigma^2}(y - x_2\beta)'[I_n - x_1x_1^\dagger](y - x_2\beta)\right) dy = 0. \end{aligned}$$

Since $u = (u_1, u_2)$ is orthogonal, we equivalently obtain

$$\begin{aligned} \int \left[\int \phi_\theta(x_2\beta + u_1v_1 + u_2v_2, x) \exp\left(-\frac{1}{2\sigma^2}v_2'v_2\right) dv_2 \right] \\ \times \exp\left(-\frac{1}{2\sigma^2}(v_1 - u_1'x_1a)'(v_1 - u_1'x_1a)\right) dv_1 = 0. \end{aligned}$$

In Proposition 1 we are looking for restrictions holding for all real vectors a . Since the rows of $u_1'x_1$ are linearly independent, we equivalently search for the following equation being satisfied for all real vectors b :¹

$$\begin{aligned} \int \left[\int \phi_\theta(x_2\beta + u_1v_1 + u_2v_2, x) \exp\left(-\frac{1}{2\sigma^2}v_2'v_2\right) dv_2 \right] \\ \times \exp\left(-\frac{1}{2\sigma^2}(v_1 - b)'(v_1 - b)\right) dv_1 = 0. \end{aligned}$$

This convolution equation has the unique solution:

$$\int \phi_\theta(x_2\beta + u_1v_1 + u_2v_2, x) \exp\left(-\frac{1}{2\sigma^2}v_2'v_2\right) dv_2 = 0.$$

We have thus shown the following.

¹ b has the same dimension as v_1 .

Proposition 3 *In model (8), the following two statements are equivalent:*

(i) $\int \phi_\theta(y, x) f_\theta(y | x, a) dy = 0.$

(ii) $\phi_\theta(y, x) = \varphi_\theta(u'_1(y - x_2\beta), u'_2(y - x_2\beta), x)$, where the function φ_θ is such that

$$\int \varphi_\theta(v_1, v_2, x) \exp\left(-\frac{1}{2\sigma^2} v'_2 v_2\right) dv_2 = 0.$$

By Proposition 1, Proposition 3 characterizes all the available moment restrictions on the parameter $\theta = (\beta', \sigma^2)$ in model (8). Now, there are many possible choices for φ_θ , leading to many choices of moment functions in this model.

As a first example, let us take

$$\varphi_\theta(v_1, v_2, x) = u_2 v_2.$$

We obtain

$$\begin{aligned} \phi_\theta(y, x) &= \varphi_\theta(u'_1(y - x_2\beta), u'_2(y - x_2\beta), x) \\ &= u_2 u'_2(y - x_2\beta) \\ &= (I_n - x_1 x_1^\dagger)(y - x_2\beta). \end{aligned}$$

This implies the following conditional moment restrictions on β :

$$\mathbb{E} \left[(I_n - X_1 X_1^\dagger)(Y - X_2\beta) \mid X_1, X_2 \right] = 0. \tag{9}$$

In a panel data setting, Chamberlain (1992) shows that the efficiency bound for β based on the quasi-differencing restrictions (9) coincides with the bound based on a semiparametric model where ε has mean zero but is otherwise not restricted. In our setup, (9) remains valid in general non-Gaussian linear network regression models where $\mathbb{E}[\varepsilon \mid X] = 0$. Moreover, the Gaussian assumption on ε provides additional restrictions that are fully characterized by Proposition 3. Some of those additional restrictions arise from the variance matrix of outcomes.

As a second example, let $n_2 = \dim v_2 = \text{Trace}(I_n - x_1 x_1^\dagger)$, and take

$$\varphi_\theta(v_1, v_2, x) = v'_2 v_2 - n_2 \sigma^2.$$

We obtain

$$\phi_\theta(y, x) = (y - x_2\beta)'[I_n - x_1x_1^\dagger](y - x_2\beta) - n_2\sigma^2.$$

This implies the following conditional moment restrictions on β and σ^2 :

$$\mathbb{E} \left[(Y - X_2\beta)'[I_n - X_1X_1^\dagger](Y - X_2\beta) - n_2\sigma^2 \mid X_1, X_2 \right] = 0. \quad (10)$$

Note that (10) remains valid under non-Gaussianity, provided $\mathbb{E}[\varepsilon\varepsilon' \mid X] = \sigma^2 I_n$. Restrictions akin to (10) were considered in [Arellano and Bonhomme \(2012\)](#) in a panel data setting. In a network context, [Andrews, Gill, Schank, and Upward \(2008\)](#) derived an unconditional version of (10), and applied it to the decomposition of the variance of log wages.²

4.2 Quadratic forms

In this subsection we derive moment restrictions on a quadratic form $\mu = \mathbb{E}[m_\theta(A, X)]$, where $m_\theta(a, x) = a'Qa$ for an $m \times m$ symmetric matrix Q , for m the dimension of A . For this purpose we rely on [Proposition 2](#). We start by noting that (7) is equivalent to

$$\int \psi_\theta(y, x) \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left(-\frac{1}{2\sigma^2} (y - x_1a - x_2\beta)'(y - x_1a - x_2\beta) \right) dy = a'Qa.$$

Using the above reparameterization in terms of (v_1, v_2) , we equivalently have

$$\begin{aligned} & \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \iint \psi_\theta(x_2\beta + u_1v_1 + u_2v_2, x) \\ & \times \exp \left(-\frac{1}{2\sigma^2} (v_1 - u_1'x_1a)'(v_1 - u_1'x_1a) \right) \exp \left(-\frac{1}{2\sigma^2} v_2'v_2 \right) dv_1 dv_2 = a'Qa. \end{aligned} \quad (11)$$

We then have the following result, shown in [Appendix A](#).

Proposition 4 *Consider model (8), and suppose that X_1 has full column rank almost surely. Then the following two statements are equivalent:*

(i) $\int \psi_\theta(y, x) f_\theta(y \mid x, a) dy = a'Qa.$

(ii) $\frac{1}{(2\pi\sigma^2)^{\frac{n_2}{2}}} \int \psi_\theta(x_2\beta + u_1v_1 + u_2v_2, x) \exp \left(-\frac{1}{2\sigma^2} v_2'v_2 \right) dv_2 = v'(x_1^\dagger)'Qx_1^\dagger v - \sigma^2 \text{Trace}((x_1^\dagger)'Qx_1^\dagger).$

²The assumption that the elements of ε be mutually independent may be empirically restrictive. In applications of AKM, it is common to only rely on between-job-spell variation in log wages in estimation, in order not to restrict the within-job-spell correlation in ε ([Kline, Saggio, and Sølvssten, 2020](#), [Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler, 2023](#)).

By Proposition 2, Proposition 4 characterizes all the available moment restrictions on $\mu = \mathbb{E}[A'QA]$.³ The proof relies on Fourier transforms. A special case of Proposition 4 is obtained when $\psi_\theta(y, x)$ is a function of $u'_1(y - x_2\beta)$ and x only, which implies

$$\psi_\theta(y, x) = (y - x_2\beta)'(x_1^\dagger)'Qx_1^\dagger(y - x_2\beta) - \sigma^2\text{Trace}((x_1^\dagger)'Qx_1^\dagger). \quad (12)$$

The trace correction in (12) is a well-known formula to obtain unbiased estimators of quadratic forms. Andrews, Gill, Schank, and Upward (2008), and Kline, Saggio, and Solvsten (2020) in a heteroskedastic context, apply such corrections to estimate variance components in log wage variance decompositions.

Moreover, given the particular solution (12), any other solution is of the form

$$\tilde{\psi}_\theta(y, x) = \psi_\theta(y, x) + \varphi_\theta(u'_1(y - x_2\beta), u'_2(y - x_2\beta), x),$$

where, as in Proposition 3, the function φ_θ satisfies

$$\int \varphi_\theta(v_1, v_2, x) \exp\left(-\frac{1}{2\sigma^2}v_2'v_2\right) dv_2 = 0.$$

Remark 1 Proposition 2 can be applied to find estimators of other average effects beyond quadratic forms. Consider the quantity $\mu = \mathbb{E}[m_\theta(A, X)]$, for some function $m_\theta(\cdot)$. Examples are higher-order moments of A , its distribution function, or some nonlinear moments of A . Using similar arguments to the ones leading to Proposition 4, one can derive a formula for all moment restrictions on μ ,

$$\mathbb{E}[\psi_\theta(Y, X)] = \mu. \quad (13)$$

However, the expression of ψ_θ , which we derive in Appendix B, depends on the inverse of the Fourier transform operator, and it does not generally admit a closed-form expression when $m_\theta(a, x)$ is not polynomial in a . In addition, estimating μ based on (13) generally requires regularizing ψ_θ (as shown in a panel data setting in the working paper version of Bonhomme, 2012).

³When X_1 is a network matrix, ensuring the assumption that X_1 has full column rank often requires to restrict the sample to a connected subnetwork. See Abowd, Kramarz, and Margolis (1999) and Abowd, Creecy, and Kramarz (2002) for methods to compute connected subnetworks in settings with workers and firms.

5 Logit network models

In this section and the next two we study logit models for network data. Letting $X = (X_1, X_2)$, we assume

$$Y = \mathbf{1} \{X_1 A + X_2 \theta + \varepsilon > 0\}, \quad (\varepsilon | X, A) \sim iidLogistic. \quad (14)$$

Model (14) contains several popular binary choice models as special cases. A first example is a static panel data logit model, which obtains when the columns of X_1 are individual indicators. Another example is the logistic network formation model of [Graham \(2017\)](#), see (5), which obtains when Y are link outcomes and the elements of $X_1 A$ are the individual sums $A_i + A_j$, for $i \neq j$.

Equation (14) also covers models of binary outcomes on a network. To illustrate, we will consider the following binary choice counterpart to the AKM model of [Abowd, Kramarz, and Margolis \(1999\)](#):

$$Y_{it} = \mathbf{1} \{X'_{it} \theta + \alpha_i + \psi_{j(i,t)} + \varepsilon_{it} > 0\}, \quad (\varepsilon_{it} | X_{11}, \dots, X_{NT}) \sim iidLogistic, \quad (15)$$

where, as in the linear case, i are workers, t are time periods, and $j(i, t)$ denotes the firm where i is employed at t . It is easy to see that (15) can be written in the form (14) for a suitable definition of X_1 . In this setting, the network is a bipartite multigraph: there may be multiple edges pointing from a worker i to a firm j indicating that worker i was in an employment relationship with firm j over multiple periods.

While, in many applications of AKM, outcomes Y_{it} are log earnings or log wages, it may be of interest to account for worker and firm heterogeneity when studying other labor market outcomes. For example, [Lachowska, Mas, Saggio, and Woodbury \(2023\)](#) apply AKM to the analysis of log hours worked. In this setting, the AKM logit specification (15) could be employed to analyze the determinants of part-time and full-time work using a binary measure of working time. In other applications, one may be interested in applying AKM to study determinants of worker promotions (e.g., [Benson, Li, and Shue, 2019](#)) or of the type of labor contract such as fixed-term or permanent contract (e.g., [Güell and Petrongolo, 2007](#)). The logit specification (15) can also be useful in applications of AKM to other fields (including education, innovation, urban economics, trade, and empirical

finance) where binary outcomes are common.

In the next section we will first focus on the covariates' coefficients θ in model (14). For example, Margolis (1996) studies the earnings returns to seniority in France while accounting for worker and firm heterogeneity. A binary specification such as (15) allows one to document the effects of seniority on binary labor market outcomes, such as working part-time or full-time, being awarded a promotion, or working under a permanent or temporary contract. In Section 7 we will study average effects, which are functions of worker and firm heterogeneity. We will see that deriving non-trivial moment restrictions on average effects seems more challenging than obtaining informative restrictions about the θ parameter.

6 Model parameters in logit network models

In this section we first provide a characterization of all moment restrictions available on θ in model (14), and then discuss several examples.

6.1 Characterization

We start by noting that, in model (14), (6) can be equivalently written as

$$\sum_{y \in \{0,1\}^n} \phi_\theta(y, x) \prod_{i=1}^n \left(\frac{\exp(x'_{i1}a + x'_{i2}\theta)}{\exp(x'_{i1}a + x'_{i2}\theta) + 1} \right)^{y_i} \left(\frac{1}{\exp(x'_{i1}a + x'_{i2}\theta) + 1} \right)^{1-y_i} = 0,$$

where we have denoted as y_i the i th element of y , x'_{i1} the i th row of x_1 , and x'_{i2} the i th row of x_2 , for $i \in \{1, \dots, n\}$.

Equivalently, we have

$$\sum_{y \in \{0,1\}^n} \phi_\theta(y, x) \prod_{i=1}^n \exp(y_i x'_{i1}a + y_i x'_{i2}\theta) = 0,$$

that is,

$$\sum_{y \in \{0,1\}^n} \phi_\theta(y, x) \exp(y'x_2\theta) \exp(y'x_1a) = 0.$$

Letting, for all x_1 , $\mathcal{S}_{x_1} = \{x'_1 y : y \in \{0,1\}^n\}$ denote the set of possible values of $x'_1 y$,

this implies that (6) can be equivalently written as

$$\sum_{s \in \mathcal{S}_{x_1}} \sum_{y \in \{0,1\}^n} \mathbf{1}\{x_1' y = s\} \phi_\theta(y, x) \exp(y' x_2 \theta) \exp(s' a) = 0.$$

Finally, since $\exp(s'a)$, for $s \in \mathcal{S}_{x_1}$, are linearly independent functions of a , we obtain the following characterization.

Proposition 5 *In model (14), the following two statements are equivalent:*

- (i) $\int \phi_\theta(y, x) f_\theta(y | x, a) dy = 0.$
- (ii) $\sum_{y \in \{0,1\}^n} \mathbf{1}\{x_1' y = s\} \phi_\theta(y, x) \exp(y' x_2 \theta) = 0,$ for all $s \in \mathcal{S}_{x_1}.$

Proposition 5 provides an exhaustive characterization of available moment restrictions in the logit model (14). For a non-zero ϕ to exist, it is necessary that, for some $s \in \mathcal{S}_{x_1}$, $y' x_2$ varies given that $x_1' y = s$. In this model, $S = X_1' Y$ is sufficient for A . Below we will illustrate Proposition 5 using several examples: the static panel data logit model, the logistic network formation model, and the AKM logit model.

6.2 Panel data: conditional logit

Consider first the panel data model

$$Y_{it} = \mathbf{1}\{A_i + X_{it}' \theta + \varepsilon_{it} > 0\}, \quad (\varepsilon_{it} | A, X) \sim iid Logistic, \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (16)$$

Let $i \in \{1, \dots, N\}$, $y_i = (y_{i1}, \dots, y_{iT})'$ and $x_i = (x_{i1}', \dots, x_{iT}')'$. For simplicity we search for functions $\phi_\theta(y_i, x_i)$ that only depend on (y, x) through (y_i, x_i) . We thus look for $\phi_\theta(y_i, x_i)$ such that

$$\sum_{y_i \in \{0,1\}^T} \mathbf{1}\left\{\sum_{t=1}^T y_{it} = s\right\} \phi_\theta(y_i, x_i) \exp\left(\sum_{t=1}^T y_{it} x_{it}' \theta\right) = 0, \quad s = 0, 1, \dots, T. \quad (17)$$

Consider the $T = 2$ case, and take $s = 1$. We obtain

$$\phi_\theta(1, 0, x_i) \exp(x_{i1}' \theta) + \phi_\theta(0, 1, x_i) \exp(x_{i2}' \theta) = 0,$$

which coincides with the moment function of conditional logit (Rasch, 1960, Andersen, 1970). When $T > 2$ we recover additional moment restrictions, as in Davezies,

D'Haultfoeuille, and Mugnier (2020).

6.3 Network formation: tetrad logit

Consider next the logistic network formation model (5) introduced in Graham (2017). Links are undirected,⁴ and there are $n = N(N - 1)/2$ observations (one for each dyad), where N is the number of agents. In this model, the sufficient statistic $s = x'_1 y$ in Proposition 5 is the vector of degrees, i.e., the degree sequence of the network.

For simplicity we focus on functions of tetrads formed by four agents (i, j, k, ℓ) . We thus look for $\phi_\theta(y_{ij}, y_{ik}, y_{il}, y_{jk}, y_{jl}, y_{kl}, x)$ that satisfy

$$\sum_{\substack{y_{ij}, y_{ik}, y_{il} \\ y_{jk}, y_{jl}, y_{kl}}} \mathbf{1}\{y_{ij} + y_{ik} + y_{il} = s_1, y_{ij} + y_{jk} + y_{jl} = s_2, y_{jk} + y_{ik} + y_{kl} = s_3, y_{il} + y_{jl} + y_{kl} = s_4\} \\ \times \phi_\theta(y_{ij}, y_{ik}, y_{il}, y_{jk}, y_{jl}, y_{kl}, x) \exp\left(\left(y_{ij}x_{ij} + y_{ik}x_{ik} + y_{il}x_{il} + y_{jk}x_{jk} + y_{jl}x_{jl} + y_{kl}x_{kl}\right)' \theta\right) = 0.$$

Taking first $s_1 = s_2 = s_3 = s_4 = 1$, we obtain

$$\begin{aligned} & \phi_\theta(1, 0, 0, 0, 0, 1, x) \exp\left(\left(x_{ij} + x_{kl}\right)' \theta\right) + \phi_\theta(0, 1, 0, 0, 1, 0, x) \exp\left(\left(x_{ik} + x_{jl}\right)' \theta\right) \\ & + \phi_\theta(0, 0, 1, 1, 0, 0, x) \exp\left(\left(x_{il} + x_{jk}\right)' \theta\right) = 0. \end{aligned} \quad (18)$$

Considering next $s_1 = s_2 = s_3 = s_4 = 2$, we obtain

$$\begin{aligned} & \phi_\theta(1, 1, 0, 0, 1, 1, x) \exp\left(\left(x_{ij} + x_{ik} + x_{jl} + x_{kl}\right)' \theta\right) \\ & + \phi_\theta(1, 0, 1, 1, 0, 1, x) \exp\left(\left(x_{ij} + x_{il} + x_{jk} + x_{kl}\right)' \theta\right) \\ & + \phi_\theta(0, 1, 1, 1, 1, 0, x) \exp\left(\left(x_{ik} + x_{il} + x_{jk} + x_{jl}\right)' \theta\right) = 0. \end{aligned} \quad (19)$$

Finally, for $s_1 = s_2 = 2, s_3 = s_4 = 1$, we obtain

$$\begin{aligned} & \phi_\theta(1, 1, 0, 0, 1, 0, x) \exp\left(\left(x_{ij} + x_{ik} + x_{jl}\right)' \theta\right) \\ & + \phi_\theta(1, 0, 1, 1, 0, 0, x) \exp\left(\left(x_{ij} + x_{il} + x_{jk}\right)' \theta\right) = 0, \end{aligned} \quad (20)$$

and there will be analogous restrictions associated with permutations of the degree sequence (2,2,1,1). All other possible degree sequences have no identifying content for θ .

⁴Logit models with directed links (Charbonneau, 2017) have a similar structure.

Together, taking ϕ as in (18), (19), and (20) (alongside its permutations), implies the moment restrictions underpinning the “tetrad logit” estimator in Graham (2017). However, Proposition 5 clarifies that these restrictions may not be unique, and it provides all available moment restrictions in the logistic network formation model.

6.4 Binary choice on a network: AKM logit

In this subsection we derive moment restrictions on θ in model (15). We focus on the case where X_{it} does not vary within job spells. For example, when controlling for the worker’s age, job seniority only varies between spells.⁵ We focus the analysis on the $T = 2$ case, and we consider several subnetwork configurations of the data displayed in Figure 2. Given a subnetwork configuration, we verify if moment conditions on θ exist, and what form they take.

Configuration A: one worker in the same firm. Suppose worker i stays in the same firm j in both periods. We look for $\phi_\theta(y_{it}, y_{i,t+1}, x)$ such that, for $s \in \{0, 1, 2\}$,

$$\sum_{y_{it}, y_{i,t+1}} \mathbf{1}\{y_{it} + y_{i,t+1} = s\} \phi_\theta(y_{it}, y_{i,t+1}, x) \exp((y_{it}x_{2it} + y_{i,t+1}x_{2i,t+1})'\theta) = 0.$$

Given that $x_{2it} = x_{2i,t+1}$ (since the covariate does not vary within spell), this implies

$$\left(\sum_{y_{it}, y_{i,t+1}} \mathbf{1}\{y_{it} + y_{i,t+1} = s\} \phi_\theta(y_{it}, y_{i,t+1}, x) \right) \exp(sx'_{2it}\theta) = 0.$$

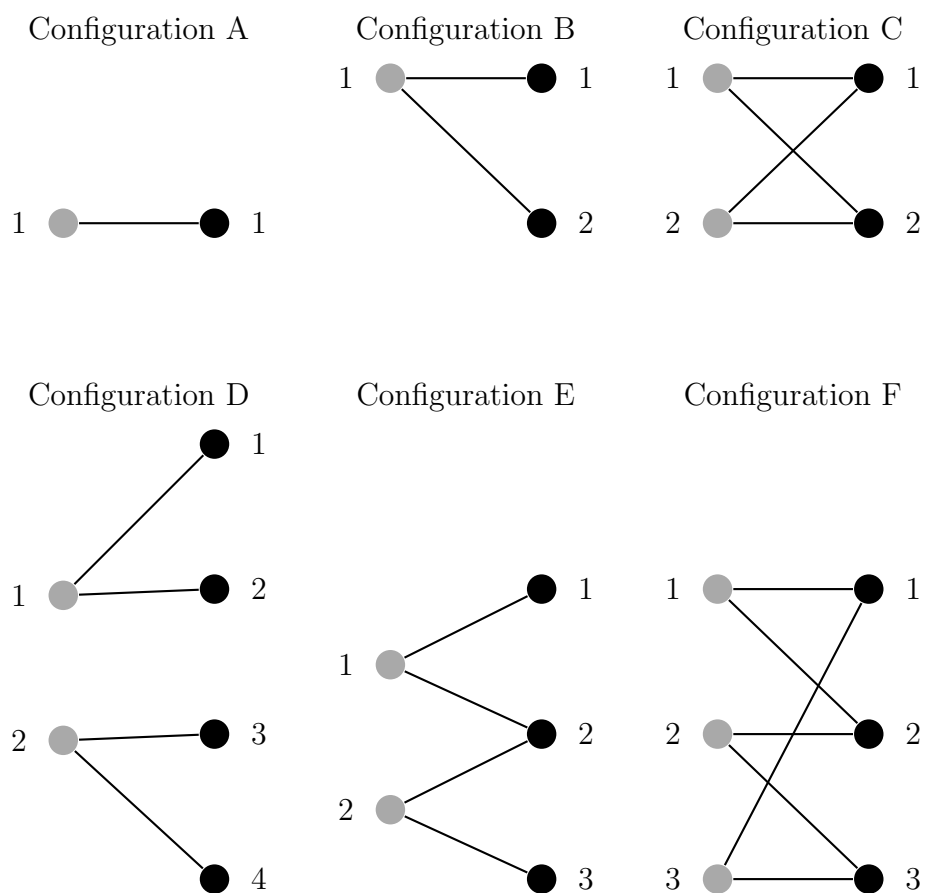
Hence θ drops out from the equation, and there is no information to estimate θ in this configuration.

Configuration B: one worker moving between two firms. Suppose worker i moves between firms j and j' . We look for $\phi_\theta(y_{it}, y_{i,t+1}, x)$ such that

$$\sum_{y_{it}, y_{i,t+1}} \mathbf{1}\{y_{it} + y_{i,t+1} = s_1, y_{it} = s_2\} \phi_\theta(y_{it}, y_{i,t+1}, x) \exp((y_{it}x_{2it} + y_{i,t+1}x_{2i,t+1})'\theta) = 0.$$

⁵If X_{it} does vary within spells, then the conditional logit estimator can be used for consistent estimation of θ .

Figure 2: Configurations of worker-firm subnetworks in the logit model with worker and firm heterogeneity



Notes: Worker nodes are in the left columns, indicated in grey. Firm nodes are in the right columns, indicated in black. In panels C and F we derive non-trivial moment restrictions on the parameter θ in model (15), while no such restrictions exist in panels A,B,D,E.

However, in this case, (s_1, s_2) fully determines $(y_{it}, y_{i,t+1})$. Hence, for each (s_1, s_2) we obtain $\phi_\theta(y_{it}, y_{i,t+1}, x) = 0$, which shows there is no information about θ in this configuration.

Configuration C: two workers moving between the same two firms. Suppose workers i and i' both move between the same firms j and j' . We look for a function $\phi_\theta(y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, x)$ such that

$$\sum_{y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}} \mathbf{1} \{y_{it} + y_{i,t+1} = s_1, y_{i't} + y_{i',t+1} = s_2, y_{it} + y_{i't} = s_3, y_{i,t+1} + y_{i',t+1} = s_4\} \\ \times \phi_\theta(y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, x) \exp((y_{it}x_{2it} + y_{i,t+1}x_{2i,t+1} + y_{i't}x_{2i't} + y_{i',t+1}x_{2i',t+1})'\theta) = 0.$$

It turns out that, in this subnetwork configuration, there exist non-trivial moment restrictions on θ . To see this, take $s_1 = s_2 = s_3 = s_4 = 1$. We obtain

$$\phi_\theta(1, 0, 0, 1, x) \exp((x_{2it} + x_{2i',t+1})'\theta) + \phi_\theta(0, 1, 1, 0, x) \exp((x_{2i,t+1} + x_{2i't})'\theta) = 0.$$

This implies the conditional moment restriction

$$\mathbb{E} \left[Y_{it}(1 - Y_{i,t+1})(1 - Y_{i't})Y_{i',t+1} \exp((X_{2i,t+1} + X_{2i't})'\theta) \right. \\ \left. - (1 - Y_{it})Y_{i,t+1}Y_{i't}(1 - Y_{i',t+1}) \exp((X_{2it} + X_{2i',t+1})'\theta) \mid X \right] = 0.$$

Configuration D: two workers moving between different firms. Suppose worker i moves between j and j' , and worker i' moves between different firms j'' and j''' . We look for $\phi_\theta(y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, x)$ such that

$$\sum_{y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}} \mathbf{1} \{y_{it} + y_{i,t+1} = s_1, y_{i't} + y_{i',t+1} = s_2, y_{it} = s_3, y_{i,t+1} = s_4, y_{i't} = s_5, y_{i',t+1} = s_6\} \\ \times \phi_\theta(y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, x) \exp((y_{it}x_{2it} + y_{i,t+1}x_{2i,t+1} + y_{i't}x_{2i't} + y_{i',t+1}x_{2i',t+1})'\theta) = 0.$$

It is easy to see there is no non-trivial ϕ function in this case. Intuitively, since workers never share a firm, it is not possible to “difference out” the firm component of heterogeneity.

Configuration E: two workers moving to different firms from the same firm.

Suppose worker i moves between j and j' , and worker i' moves from the same firm j to a different firm j'' . We look for $\phi_\theta(y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, x)$ such that

$$\sum_{y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}} \mathbf{1}\{y_{it} + y_{i,t+1} = s_1, y_{i't} + y_{i',t+1} = s_2, y_{it} + y_{i't} = s_3, y_{i,t+1} = s_4, y_{i',t+1} = s_5\} \\ \times \phi_\theta(y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, x) \exp((y_{it}x_{2it} + y_{i,t+1}x_{2i,t+1} + y_{i't}x_{2i't} + y_{i',t+1}x_{2i',t+1})'\theta) = 0.$$

It is easy to see there is no information about θ in this configuration.

Configuration F: three workers in a loop. There are many other subnetwork configurations providing information beyond configuration C. Indeed, consider three workers who move as follows: i moves between firms j and j' , i' moves between j' and j'' , and i'' moves between j'' and j . We look for $\phi_\theta(y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, y_{i''t}, y_{i'',t+1}, x)$ such that

$$\sum_{y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, y_{i''t}, y_{i'',t+1}} \mathbf{1}\left\{y_{it} + y_{i,t+1} = s_1, \right. \\ \left. y_{i't} + y_{i',t+1} = s_2, y_{i''t} + y_{i'',t+1} = s_3, y_{it} + y_{i'',t+1} = s_4, y_{i't} + y_{i,t+1} = s_5, y_{i''t} + y_{i',t+1} = s_6\right\} \\ \times \phi_\theta(y_{it}, y_{i,t+1}, y_{i't}, y_{i',t+1}, y_{i''t}, y_{i'',t+1}, x) \\ \times \exp((y_{it}x_{2it} + y_{i,t+1}x_{2i,t+1} + y_{i't}x_{2i't} + y_{i',t+1}x_{2i',t+1} + y_{i''t}x_{2i''t} + y_{i'',t+1}x_{2i'',t+1})'\theta) = 0.$$

Taking $s_1 = s_2 = s_3 = s_4 = s_5 = s_6 = 1$, one obtains

$$\phi_\theta(1, 0, 1, 0, 1, 0, x) \exp((x_{2it} + x_{2i't} + x_{2i''t})'\theta) \\ + \phi_\theta(0, 1, 0, 1, 0, 1, x) \exp((x_{2i,t+1} + x_{2i',t+1} + x_{2i'',t+1})'\theta) = 0.$$

This implies the conditional moment restriction

$$\mathbb{E}\left[Y_{it}(1 - Y_{i,t+1})Y_{i't}(1 - Y_{i',t+1})Y_{i''t}(1 - Y_{i'',t+1}) \exp((X_{2i,t+1} + X_{2i',t+1} + X_{2i'',t+1})'\theta) \right. \\ \left. - (1 - Y_{it})Y_{i,t+1}(1 - Y_{i't})Y_{i',t+1}(1 - Y_{i''t})Y_{i'',t+1} \exp((X_{2it} + X_{2i't} + X_{2i''t})'\theta) \mid X\right] = 0.$$

7 Average effects in logit network models

In this section we again consider model (14), and we study average effects of the form

$$\mu = \mathbb{E}[m_\theta(A, X)],$$

for some known function m_θ . In this case, (7) can be equivalently written as

$$\frac{\sum_{y \in \{0,1\}^n} \psi_\theta(y, x) \prod_{i=1}^n \exp(y_i x'_{i1} a + y_i x'_{i2} \theta)}{\prod_{i=1}^n (\exp(x'_{i1} a + x'_{i2} \theta) + 1)} = m_\theta(a, x_1, x_2). \quad (21)$$

This equation characterizes the set of available moment restrictions on μ , i.e., the set of ψ functions such that (13) holds.

As a simple example, consider the case where $T = 2$ in the static panel logit model (16), with a binary covariate X_{it} , and consider

$$\begin{aligned} m_\theta(a, x) &= \Pr(Y_{i1} = 1 \mid X_{i1} = 1, A = a, \theta) - \Pr(Y_{i1} = 1 \mid X_{i1} = 0, A = a, \theta) \\ &= \frac{\exp(\theta + a)}{1 + \exp(\theta + a)} - \frac{\exp(a)}{1 + \exp(a)}, \end{aligned}$$

so that μ is an average partial effect. We show in Appendix C that no function ψ satisfies (21). Intuitively, this comes from the fact that the distribution of A given $X_{i1} = X_{i2}$ (i.e., for “stayers”) is unidentified.

In contrast, as we also show in Appendix C, the average partial effect of “movers”, corresponding to

$$m_\theta(a, x) = \mathbf{1}\{x_1 \neq x_2\} \left[\frac{\exp(\theta + a)}{1 + \exp(\theta + a)} - \frac{\exp(a)}{1 + \exp(a)} \right],$$

admits a characterization as in (21), whenever ψ satisfies

$$\psi_\theta(y_1, y_2, x) = 0, \text{ for all } (y_1, y_2), \text{ if } x_1 = x_2, \quad (22)$$

$$\psi_\theta(1, 0, x) \exp(\theta x_1) + \psi_\theta(0, 1, x) \exp(\theta x_2) = \exp(\theta) - 1, \text{ if } x_1 \neq x_2. \quad (23)$$

A simple example satisfying those conditions is

$$\psi_\theta(y_1, y_2, x) = (x_2 - x_1)(y_2 - y_1),$$

as pointed out (in a more general nonparametric model) by [Chernozhukov, Fernández-Val, Hahn, and Newey \(2013\)](#). However, (22) and (23) imply additional moment restrictions. For example, one can take

$$\psi_{\theta}(y_1, y_2, x) = \mathbf{1}\{x_1 \neq x_2\} [y_1(1 - y_2) \exp(\theta(1 - x_1)) - (1 - y_1)y_2 \exp(-\theta x_2)],$$

which provides an additional moment restriction on μ under the logit model's assumptions.

It appears difficult to obtain moment equality restrictions on average partial effects in logit models on networks outside of the panel data case. As an example, consider the subnetwork configuration C in [Figure 2](#). In this case we have seen in the previous section how to obtain moment restrictions on θ . However, we show in [Appendix C](#) that no function ψ satisfies (21) for the average partial effect corresponding to

$$m_{\theta}(a, x) = \frac{\exp(\theta + a_1 + a_3)}{1 + \exp(\theta + a_1 + a_3)} - \frac{\exp(a_1 + a_3)}{1 + \exp(a_1 + a_3)},$$

where, in this model, a_1 is worker i 's fixed effect and a_3 is firm j 's fixed effect.

In models where no functional differencing restrictions are available, one may still be able to construct bounds on the average effect of interest. In panel data settings, this strategy was pursued by [Chernozhukov, Fernández-Val, Hahn, and Newey \(2013\)](#), [Davezies, D'Haultfoeuille, and Laage \(2021\)](#), and [Dobronyi, Gu, and Kim \(2021\)](#), among others. However, implementing bounds approaches often requires estimating conditional moments given X . When X represents a network matrix, conditional moment estimation may be especially challenging. In a panel data setting, [Pakel and Weidner \(2021\)](#) propose a bounding strategy that avoids the curse of dimensionality associated with conditioning covariates. Extending their approach to network settings is an interesting question for future work.

Lastly, in this section we have focused on binary choice models. The situation may be more favorable, in the sense of there existing informative functions ϕ and ψ , in models with continuous outcomes such as the CES specification (4).

8 Remarks on estimation

To close our discussion, we briefly outline some possibilities for estimation of parameters and average effects, without providing details.

Given a moment function ϕ as in Proposition 1, and a realization (y, x) from the joint distribution of (Y, X) , one can estimate θ based on

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \|\phi_{\theta}(y, x)\|,$$

for some norm $\|\cdot\|$. In some models, this approach will deliver familiar estimators. For example, in the linear model (8), an estimator of β based on (9) is the “quasi-differencing” estimator

$$\hat{\beta} = \left[x_2'(I_n - x_1 x_1^{\dagger}) x_2 \right]^{-1} x_2'(I_n - x_1 x_1^{\dagger}) y, \quad (24)$$

and an estimator of σ^2 based on (10) is the “degree-of-freedom-corrected” estimator

$$\hat{\sigma}^2 = \frac{(y - x_2 \hat{\beta})' [I_n - x_1 x_1^{\dagger}] (y - x_2 \hat{\beta})}{\operatorname{Trace}(I_n - x_1 x_1^{\dagger})}. \quad (25)$$

When constructing a function ϕ using the entire data is impractical, one can construct a set of functions $\phi_{\theta}^{(k)}(y, x)$ that depend on y and x only through a subset of the data. An estimator of θ is then

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left\| \sum_{k=1}^K \phi_{\theta}^{(k)}(y, x) \right\|.$$

In the logit network formation model (5), taking ϕ as in (18), (19), and (20) (alongside its permutations), leads to estimators in the spirit of the tetrad logit estimator of [Graham \(2017\)](#).

When focusing on average effects, a possible estimation approach based on Proposition 2 consists in setting

$$\hat{\mu} = \psi_{\hat{\theta}}(y, x),$$

for some estimator $\hat{\theta}$. For example, in the linear model (8), an estimator of the quadratic form $\mu = \mathbb{E}[A'QA]$ based on (12) is

$$\hat{\mu} = (y - x_2 \hat{\beta})' (x_1^{\dagger})' Q x_1^{\dagger} (y - x_2 \hat{\beta}) - \hat{\sigma}^2 \operatorname{Trace}((x_1^{\dagger})' Q x_1^{\dagger}),$$

where $\hat{\beta}$ and $\hat{\sigma}^2$ are given by (24) and (25), respectively. This corresponds to the bias-corrected estimator of [Andrews, Gill, Schank, and Upward \(2008\)](#). For other average effects, regularization is typically needed for reliable estimation.

For all these estimators, there are important questions that remain to be addressed. What are their asymptotic properties (under suitable assumptions on how the network grows with the sample size)? How to conduct feasible inference on the population parameters? And, out of the available functions ϕ and ψ , how to choose a small subset of those (for tractability) without sacrificing too much precision (for efficiency)? Answering these questions will be an important task for future work.

References

- ABOWD, J. M., R. H. CREECY, AND F. KRAMARZ (2002): “Computing person and firm effects using linked longitudinal employer-employee data,” Discussion paper, Center for Economic Studies, US Census Bureau.
- ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): “High wage workers and high wage firms,” *Econometrica*, 67(2), 251–333.
- AGUIRREGABIRIA, V., AND J. M. CARRO (2021): “Identification of average marginal effects in fixed effects dynamic discrete choice models,” *arXiv preprint arXiv:2107.06141*.
- AHMADPOOR, M., AND B. F. JONES (2019): “Decoding team and individual impact in science and invention,” *Proceedings of the National Academy of Sciences*, 116(28), 13885–13890.
- ANDERSEN, E. B. (1970): “Asymptotic properties of conditional maximum-likelihood estimators,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(2), 283–301.
- ANDREWS, M. J., L. GILL, T. SCHANK, AND R. UPWARD (2008): “High wage workers and low wage firms: negative assortative matching or limited mobility bias?,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3), 673–697.
- ARELLANO, M., AND S. BOND (1991): “Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations,” *The review of economic studies*, 58(2), 277–297.
- ARELLANO, M., AND S. BONHOMME (2012): “Identifying distributional characteristics in random coefficients panel data models,” *The Review of Economic Studies*, 79(3), 987–1020.
- BECKER, G. S. (1973): “A theory of marriage: Part I,” *Journal of Political economy*, 81(4), 813–846.
- BENSON, A., D. LI, AND K. SHUE (2019): “Promotions and the peter principle,” *The Quarterly Journal of Economics*, 134(4), 2085–2134.

- BICKEL, P. J., AND A. CHEN (2009): “A nonparametric view of network models and Newman–Girvan and other modularities,” *Proceedings of the National Academy of Sciences*, 106(50), 21068–21073.
- BONHOMME, S. (2012): “Functional differencing,” *Econometrica*, 80(4), 1337–1385.
- (2020): “Econometric analysis of bipartite networks,” in *The Econometric Analysis of Network Data*, pp. 83–121. Elsevier.
- (2021): “Teams: Heterogeneity, sorting, and complementarity,” *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2021-15).
- BONHOMME, S., K. DANO, AND B. S. GRAHAM (2023): “Identification in a Binary Choice Panel Data Model with a Predetermined Covariate,” Discussion paper, National Bureau of Economic Research.
- BONHOMME, S., K. HOLZHEU, T. LAMADON, E. MANRESA, M. MOGSTAD, AND B. SETZLER (2023): “How much should we trust estimates of firm effects and worker sorting?,” *Journal of Labor Economics*, 41(2), 291–322.
- BONHOMME, S., T. LAMADON, AND E. MANRESA (2019): “A distributional framework for matched employer employee data,” *Econometrica*, 87(3), 699–739.
- CARD, D., J. HEINING, AND P. KLINE (2013): “Workplace heterogeneity and the rise of West German wage inequality,” *The Quarterly journal of economics*, 128(3), 967–1015.
- CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2007): “Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization,” *Handbook of econometrics*, 6, 5633–5751.
- CHAMBERLAIN, G. (1992): “Efficiency bounds for semiparametric regression,” *Econometrica: Journal of the Econometric Society*, pp. 567–596.
- CHARBONNEAU, K. B. (2017): “Multiple fixed effects in binary response panel data models,” *The Econometrics Journal*, 20(3), S1–S13.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, J. HAHN, AND W. NEWEY (2013): “Average and quantile effects in nonseparable panel models,” *Econometrica*, 81(2), 535–580.

- DANO, K. (2023): “Transition Probabilities and Identifying Moments in Dynamic Fixed Effects Logit Models,” *arXiv preprint arXiv:2303.00083*.
- DAVEZIES, L., X. D’HAULTFOEUILLE, AND L. LAAGE (2021): “Identification and estimation of average marginal effects in fixed effects logit models,” *arXiv preprint arXiv:2105.00879*.
- DAVEZIES, L., X. D’HAULTFOEUILLE, AND M. MUGNIER (2020): “Fixed Effects Binary Choice Models with Three or More Periods,” *arXiv preprint arXiv:2009.08108*.
- DE PAULA, Á., S. RICHARDS-SHUBIK, AND E. TAMER (2018): “Identifying preferences in networks with bounded degree,” *Econometrica*, 86(1), 263–288.
- DHAENE, G., AND K. JOCHMANS (2015): “Split-panel jackknife estimation of fixed-effect models,” *The Review of Economic Studies*, 82(3), 991–1030.
- DHAENE, G., AND M. WEIDNER (2023): “Approximate Functional Differencing,” *arXiv preprint arXiv:2301.13736*.
- DOBRONYI, C., J. GU, AND K. I. KIM (2021): “Identification of dynamic panel logit models with fixed effects,” *arXiv preprint arXiv:2104.04590*.
- ENGL, H. W., M. HANKE, AND A. NEUBAUER (1996): *Regularization of inverse problems*, vol. 375. Springer Science & Business Media.
- FERNÁNDEZ-VAL, I., AND M. WEIDNER (2016): “Individual and time effects in nonlinear panel models with large N, T,” *Journal of Econometrics*, 192(1), 291–312.
- GRAHAM, B. S. (2017): “An econometric model of network formation with degree heterogeneity,” *Econometrica*, 85(4), 1033–1063.
- (2020): “Sparse network asymptotics for logistic regression,” Discussion paper, National Bureau of Economic Research.
- GUALDANI, C. (2021): “An econometric model of network formation with an application to board interlocks between firms,” *Journal of Econometrics*, 224(2), 345–370.
- GÜELL, M., AND B. PETRONGOLO (2007): “How binding are legal limits? Transitions from temporary to permanent work in Spain,” *Labour economics*, 14(2), 153–183.

- HAHN, J., AND W. NEWEY (2004): “Jackknife and analytical bias reduction for nonlinear panel models,” *Econometrica*, 72(4), 1295–1319.
- HONORÉ, B. E., C. MURIS, AND M. WEIDNER (2021): “Dynamic ordered panel logit models,” *arXiv preprint arXiv:2107.03253*.
- HONORÉ, B. E., AND M. WEIDNER (2020): “Moment conditions for dynamic panel logit models with fixed effects,” *arXiv preprint arXiv:2005.05942*.
- HUGHES, D. W. (2022): “Estimating Nonlinear Network Data Models with Fixed Effects,” *arXiv preprint arXiv:2203.15603*.
- KLINE, P., R. SAGGIO, AND M. SØLVSTEN (2020): “Leave-out estimation of variance components,” *Econometrica*, 88(5), 1859–1898.
- KUERSTEINER, G. M., AND I. R. PRUCHA (2020): “Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity,” *Econometrica*, 88(5), 2109–2146.
- LACHOWSKA, M., A. MAS, R. SAGGIO, AND S. A. WOODBURY (2023): “Work hours mismatch,” Discussion paper, National Bureau of Economic Research.
- LENTZ, R., S. PIYAPROMDEE, AND J.-M. ROBIN (2022): “The Anatomy of Sorting-Evidence from Danish Data,” Discussion paper, HAL.
- MARGOLIS, D. N. (1996): “Cohort effects and returns to seniority in France,” *Annales d’Economie et de Statistique*, pp. 443–464.
- PAKEL, C., AND M. WEIDNER (2021): “Bounds on average effects in discrete choice panel data models,” Discussion paper, Technical report, Working paper.
- POSTEL-VINAY, F., AND J.-M. ROBIN (2002): “Equilibrium wage dispersion with worker and employer heterogeneity,” *Econometrica*, 70(6), 2295–2350.
- RASCH, G. (1960): *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.
- SHENG, S. (2020): “A structural econometric analysis of network formation games through subnetworks,” *Econometrica*, 88(5), 1829–1858.

SHIMER, R., AND L. SMITH (2000): “Assortative matching and search,” *Econometrica*, 68(2), 343–369.

SONG, J., D. J. PRICE, F. GUVENEN, N. BLOOM, AND T. VON WACHTER (2019): “Firming up inequality,” *The Quarterly journal of economics*, 134(1), 1–50.

WOODCOCK, S. D. (2008): “Wage differentials in the presence of unobserved worker, firm, and match heterogeneity,” *Labour Economics*, 15(4), 771–793.

APPENDIX

A Proofs

A.1 Proofs of Propositions 1 and 2

Propositions 1 and 2 follow directly from the following elementary lemma whose proof we include for completeness.

Lemma 1 *Let $g : \mathcal{Z} \rightarrow \mathbb{R}^p$ be a function such that $\sup_{z \in \mathcal{Z}} \|g(z)\|_\infty < \infty$, where $\|\cdot\|_\infty$ denotes the sup norm on \mathbb{R}^p . Suppose that, for all non-negative functions $h : \mathcal{Z} \rightarrow \mathbb{R}^+$ such that $\int_{\mathcal{Z}} h(z) dz = 1$, we have $\int_{\mathcal{Z}} g(z)h(z) dz = 0$. Then $g(z) = 0$ almost everywhere on \mathcal{Z} .*

Proof. Let $k \in \{1, \dots, p\}$, and, for all $z \in \mathcal{Z}$, let $g_k(z)$ denote the k -th element of $g(z)$. For all $\ell \in \mathcal{L}^1(\mathcal{Z})$, let $\|\ell\|_1 = \int_{\mathcal{Z}} |\ell(z)| dz$, $\ell^+(z) = \max(\ell(z), 0)$, and $\ell^-(z) = -\min(\ell(z), 0)$. For all $\ell \in \mathcal{L}^1(\mathcal{Z})$ we have

$$\int_{\mathcal{Z}} g_k(z) \ell(z) dz = \|\ell^+\|_1 \int_{\mathcal{Z}} g_k(z) \frac{\ell^+(z)}{\|\ell^+\|_1} dz - \|\ell^-\|_1 \int_{\mathcal{Z}} g_k(z) \frac{\ell^-(z)}{\|\ell^-\|_1} dz = 0,$$

where we have used that $\frac{\ell^+(z)}{\|\ell^+\|_1}$ and $\frac{\ell^-(z)}{\|\ell^-\|_1}$ are non-negative and integrate to one (with the convention $0/0 = 0$ whenever $\|\ell^+\|_1 = 0$ or $\|\ell^-\|_1 = 0$). Since g_k is bounded, it follows that

$$\sup_{z \in \mathcal{Z}} |g_k(z)| = \sup_{\ell \in \mathcal{L}^1(\mathcal{Z}) : \|\ell\|_1 = 1} \left| \int_{\mathcal{Z}} g_k(z) \ell(z) dz \right| = 0,$$

so $g_k = 0$ almost everywhere on \mathcal{Z} . Lastly, since this holds for all $k \in \{1, \dots, p\}$, it follows that $g = 0$ almost everywhere on \mathcal{Z} . ■

Proof of Proposition 1. It is sufficient to show that (i) implies (ii). Suppose that (i) holds. Let $Z = (A, X)$, and $g(Z) = \mathbb{E}[\phi_\theta(Y, X) | A, X]$. It follows from (i) and Lemma 1 that $g = 0$ almost everywhere on \mathcal{Z} . This shows (ii) and completes the proof.

Proof of Proposition 2. Let $Z = (A, X)$, and $g(Z) = \mathbb{E}[\psi_\theta(Y, X) | A, X] - m_\theta(A, X)$. It follows from (i) and Lemma 1 that $g = 0$ almost everywhere on \mathcal{Z} . This shows (ii) and completes the proof.

A.2 Proof of Proposition 4

Let $n_1 = n - n_2$. Then u_1 is an $n \times n_1$ matrix, and v_1 is an $n_1 \times 1$ vector.

Let

$$\varphi_\theta(v_1, x) = \frac{1}{(2\pi\sigma^2)^{\frac{n_2}{2}}} \int \psi_\theta(x_2\beta + u_1v_1 + u_2v_2, x) \exp\left(-\frac{1}{2\sigma^2}v_2'v_2\right) dv_2.$$

It follows from (11) that (7) is equivalent to

$$\frac{1}{(2\pi\sigma^2)^{\frac{n_1}{2}}} \int \varphi_\theta(v_1, x) \exp\left(-\frac{1}{2\sigma^2}(v_1 - u_1'x_1a)'(v_1 - u_1'x_1a)\right) dv_1 = a'Qa.$$

As a result, (7) is equivalent to

$$\frac{1}{(2\pi\sigma^2)^{\frac{n_1}{2}}} \int \varphi_\theta(v_1, x) \exp\left(-\frac{1}{2\sigma^2}(v_1 - b)'(v_1 - b)\right) dv_1 = b'((u_1'x_1)^\dagger)'Q(u_1'x_1)^\dagger b,$$

where $b = u_1'x_1a$, and we have used that x_1 has full column rank.

Let \mathcal{F} denote the Fourier transform operator. For any integrable function $f : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$ we have, for all $s \in \mathbb{R}^{n_1}$,

$$\mathcal{F}[f](s) = \int f(x) \exp(\mathbf{i}s'x) dx,$$

where \mathbf{i} is a complex root of -1 , and the integral is over \mathbb{R}^{n_1} .

We have

$$\mathcal{F}[\varphi_\theta(v_1, x)](s) \exp\left(-\frac{\sigma^2}{2}s's\right) = \mathcal{F}[b'((u_1'x_1)^\dagger)'Q(u_1'x_1)^\dagger b](s).$$

Let $C = ((u_1'x_1)^\dagger)'Q(u_1'x_1)^\dagger$. We have, for $\delta(\cdot)$ the Dirac delta function,

$$\begin{aligned} \mathcal{F}[b'Cb](s) &= \sum_{i,j} c_{ij} \mathcal{F}[b_i b_j](s) \\ &= \sum_i c_{ii} \mathcal{F}[b_i^2](s) + \sum_{i \neq j} c_{ij} \mathcal{F}[b_i b_j](s) \\ &= \sum_i c_{ii} \mathcal{F}[b_i^2](s) + \sum_{i \neq j} c_{ij} \mathcal{F}[b_i](s) \mathcal{F}[b_j](s) \\ &= \sum_i c_{ii} [-(2\pi)\delta''(s_i)] (2\pi)^{n_1-1} \prod_{j \neq i} \delta(s_j) \\ &\quad + \sum_{i \neq j} c_{ij} [-i(2\pi)\delta'(s_i)] [-i(2\pi)\delta'(s_j)] (2\pi)^{n_1-2} \prod_{k \neq (i,j)} \delta(s_k) \\ &= -(2\pi)^{n_1} \left(\sum_i c_{ii} \delta''(s_i) \prod_{j \neq i} \delta(s_j) + \sum_{i \neq j} c_{ij} \delta'(s_i) \delta'(s_j) \prod_{k \neq (i,j)} \delta(s_k) \right). \end{aligned}$$

Since this holds for all s , the Fourier inversion theorem gives

$$\begin{aligned}\varphi_\theta(v_1, x) &= \frac{1}{(2\pi)^{n_1}} \int \mathcal{F}[b'Cb](s) \exp\left(\frac{\sigma^2}{2}s's\right) \exp(-iv_1's) ds \\ &= -\sum_i c_{ii} \int \delta''(s_i) \exp\left(\frac{\sigma^2}{2}s_i^2\right) \exp(-iv_{1i}s_i) ds_i \\ &\quad - \sum_{i \neq j} c_{ij} \iint \delta'(s_i)\delta'(s_j) \exp\left(\frac{\sigma^2}{2}(s_i^2 + s_j^2)\right) \exp(-i(v_{1i}s_i + v_{1j}s_j)) ds_i ds_j.\end{aligned}$$

Now, by integration by parts, we have

$$\begin{aligned}\int \delta'(s_i) \exp\left(\frac{\sigma^2}{2}s_i^2\right) \exp(-iv_{1i}s_i) ds_i &= iv_{1i}, \\ \int \delta''(s_i) \exp\left(\frac{\sigma^2}{2}s_i^2\right) \exp(-iv_{1i}s_i) ds_i &= \sigma^2 - v_{1i}^2.\end{aligned}$$

Hence,

$$\begin{aligned}\varphi_\theta(v_1, x) &= -\sum_i c_{ii}(\sigma^2 - v_{1i}^2) + \sum_{i \neq j} c_{ij}v_{1i}v_{1j} \\ &= v_1' C v_1 - \sigma^2 \text{Trace}(C),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{(2\pi\sigma^2)^{\frac{n_2}{2}}} \int \psi_\theta(x_2\beta + u_1v_1 + u_2v_2, x) \exp\left(-\frac{1}{2\sigma^2}v_2'v_2\right) dv_2 \\ &= v'u_1((u_1'x_1)^\dagger)'Q(u_1'x_1)^\dagger u_1'v - \sigma^2 \text{Trace}((x_1^\dagger)'Qx_1^\dagger) \\ &= v'(x_1^\dagger)'Qx_1^\dagger v - \sigma^2 \text{Trace}((x_1^\dagger)'Qx_1^\dagger).\end{aligned}$$

This concludes the proof of Proposition 4.

B Other average effects in the linear model

Following the arguments in the proof of Proposition 4, we have

$$\mathcal{F}[\varphi_\theta(v_1, x)](s) \exp\left(-\frac{\sigma^2}{2}s's\right) = \mathcal{F}[m_\theta((u_1'x_1)^\dagger b, x)](s).$$

Hence, provided the Fourier inversion theorem can be applied, we have

$$\varphi_\theta(v_1, x) = \mathcal{F}^{-1} \left[\mathcal{F} [m_\theta((u'_1 x_1)^\dagger b, x)](s) \exp\left(\frac{\sigma^2}{2} s' s\right) \right] (v_1),$$

and thus

$$\begin{aligned} & \frac{1}{(2\pi\sigma^2)^{\frac{n_2}{2}}} \int \psi_\theta(x_2\beta + u_1 v_1 + u_2 v_2, x) \exp\left(-\frac{1}{2\sigma^2} v_2' v_2\right) dv_2 \\ &= \mathcal{F}^{-1} \left[\mathcal{F} [m_\theta((u'_1 x_1)^\dagger b, x)](s) \exp\left(\frac{\sigma^2}{2} s' s\right) \right] (v_1). \end{aligned}$$

As a special case, suppose $\psi_\theta(y, x)$ is a function of $u'_1(y - x_2\beta)$ and x only. Then

$$\psi_\theta(y, x) = \mathcal{F}^{-1} \left[\mathcal{F} [m_\theta((u'_1 x_1)^\dagger b, x)](s) \exp\left(\frac{\sigma^2}{2} s' s\right) \right] (u'_1(y - x_2\beta)).$$

C Average effects in logit models

Let

$$m_\theta(a, x) = \frac{\exp(\theta + a)}{1 + \exp(\theta + a)} - \frac{\exp(a)}{1 + \exp(a)}.$$

In the panel data case with $T = 2$, (21) can equivalently be written as

$$\begin{aligned} & \frac{\exp(\theta + a)}{1 + \exp(\theta + a)} - \frac{\exp(a)}{1 + \exp(a)} \\ &= \frac{\sum_{(y_{i1}, y_{i2}) \in \{0,1\}^2} \psi_\theta(y_{i1}, y_{i2}, x_i) \exp((y_{i1} + y_{i2})a + (y_{i1}x_{i1} + y_{i2}x_{i2})\theta)}{(\exp(a + x_{i1}\theta) + 1)(\exp(a + x_{i2}\theta) + 1)}. \end{aligned}$$

That is,

$$\begin{aligned} & (\exp(\theta) - 1) \exp(a) (\exp(a + x_{i1}\theta) + 1) (\exp(a + x_{i2}\theta) + 1) \\ &= (1 + \exp(\theta + a))(1 + \exp(a)) \\ & \quad \times \sum_{(y_{i1}, y_{i2}) \in \{0,1\}^2} \psi_\theta(y_{i1}, y_{i2}, x_i) \exp((y_{i1} + y_{i2})a + (y_{i1}x_{i1} + y_{i2}x_{i2})\theta). \end{aligned} \quad (26)$$

Let $Z = \exp(a)$. The coefficient of Z^0 on the left-hand side of (26) is equal to zero, and the coefficient on the right-hand side is $\psi_\theta(0, 0, x_i)$. It thus follows that $\psi_\theta(0, 0, x_i) = 0$.

The coefficient of Z^4 on the left-hand side of (26) is equal to zero, and the coefficient on

the right-hand side is $\exp((x_{i1} + x_{i2} + 1)\theta)\psi_\theta(1, 1, x_i)$. It thus follows that $\psi_\theta(1, 1, x_i) = 0$.

Hence,

$$\begin{aligned} & (\exp(\theta) - 1) (\exp(a + x_{i1}\theta) + 1) (\exp(a + x_{i2}\theta) + 1) \\ &= (1 + \exp(\theta + a))(1 + \exp(a)) [\psi_\theta(1, 0, x_i) \exp(x_{i1}\theta) + \psi_\theta(0, 1, x_i) \exp(x_{i2}\theta)]. \end{aligned} \quad (27)$$

So the existence of valid moment functions requires that the ratio

$$\frac{(\exp(\theta) - 1) (\exp(a + x_{i1}\theta) + 1) (\exp(a + x_{i2}\theta) + 1)}{(1 + \exp(\theta + a))(1 + \exp(a))}$$

does not depend on a . This is not possible if $x_{i1} = x_{i2}$. However, if $x_{i1} \neq x_{i2}$ then (27) simplifies to

$$\exp(\theta) - 1 = \psi_\theta(1, 0, x_i) \exp(x_{i1}\theta) + \psi_\theta(0, 1, x_i) \exp(x_{i2}\theta),$$

which is indeed identical to (23).

Next, consider Configuration C in Figure 2, and let

$$m_\theta(a, x) = \frac{\exp(\theta + a_1 + a_3)}{1 + \exp(\theta + a_1 + a_3)} - \frac{\exp(a_1 + a_3)}{1 + \exp(a_1 + a_3)}.$$

By (21), we look for ψ such that

$$\frac{\exp(\theta + a_1 + a_3)}{1 + \exp(\theta + a_1 + a_3)} - \frac{\exp(a_1 + a_3)}{1 + \exp(a_1 + a_3)} = \frac{\sum_{y \in \{0,1\}^4} \psi_\theta(y, x) \prod_{i=1}^4 \exp(y_i x'_{i1} a + y_i x'_{i2} \theta)}{\prod_{i=1}^4 (\exp(x'_{i1} a + x'_{i2} \theta) + 1)}.$$

That is,

$$\begin{aligned} & (\exp(\theta) - 1) \exp(a_1 + a_3) (\exp(a_1 + a_3 + x_{12}\theta) + 1) (\exp(a_1 + a_4 + x_{22}\theta) + 1) \\ & \quad \times (\exp(a_2 + a_3 + x_{32}\theta) + 1) (\exp(a_2 + a_4 + x_{42}\theta) + 1) \\ &= \sum_{y \in \{0,1\}^4} \psi_\theta(y, x) \exp \left(y_1(a_1 + a_3 + x_{12}\theta) + y_2(a_1 + a_4 + x_{22}\theta) + y_3(a_2 + a_3 + x_{32}\theta) \right. \\ & \quad \left. + y_4(a_2 + a_4 + x_{42}\theta) \right) (1 + \exp(\theta + a_1 + a_3))(1 + \exp(a_1 + a_3)). \end{aligned} \quad (28)$$

Let $Z_k = \exp(a_k)$ for $k \in \{1, \dots, 4\}$. The left-hand side and right-hand side in (28) are polynomials in (Z_1, \dots, Z_4) . Since they are equal to each other for all a_1, \dots, a_4 in \mathbb{R} , hence for all Z_1, \dots, Z_4 in $\mathbb{R}_{>0}$, they are equal to each other for all Z_1, \dots, Z_4 in \mathbb{R} as well.

Suppose next that $\theta \neq 0$. The right-hand side in (28) is a multiple of $(1 + Z_1 Z_3)$ and $(1 + \exp(\theta) Z_1 Z_3)$. However, the left-hand side in (28) is either a multiple of $(1 + Z_1 Z_3)$ or a multiple of $(1 + \exp(\theta) Z_1 Z_3)$ (depending on the value of x_{12}) but it is not a multiple of both terms. It follows that, when $\theta \neq 0$, there is no ψ satisfying (28). A direct comparison of the monomial terms of the polynomials on both sides of equation (28) confirms this argument.