Identification, Estimation, and Inference in Two-Sided Interaction Models*

Federico Crippa
Department of Economics
Northwestern University
federicocrippa2025@u.northwestern.edu

This version: October 29, 2025.
Newest version here.

Abstract

This paper studies a class of models for two-sided interactions, where outcomes depend on latent characteristics of two distinct agent types. Models in this class have two core elements: the matching network, which records which agent pairs interact, and the interaction function, which maps latent characteristics of these agents to outcomes and determines the role of complementarities. I introduce the Tukey model, which captures complementarities with a single interaction parameter, along with two extensions that allow richer complementarity patterns. First, I establish an identification trade-off between the flexibility of the interaction function and the density of the matching network: the Tukey model is identified under mild conditions, whereas the more flexible extensions require dense networks that are rarely observed in applications. Second, I propose a cycle-based estimator for the Tukey interaction parameter and show that it is consistent and asymptotically normal even when the network is sparse. Third, I use its asymptotic distribution to construct a formal test of no complementarities. Finally, an empirical illustration shows that the Tukey model recovers economically meaningful complementarities.

*I am deeply grateful to Federico Bugni, Ivan Canay, and Chuck Manski for their guidance throughout this project. I thank Devis Decet, Anastasiia Evdokimova, and Amilcar Velez for many helpful discussions that shaped the ideas in this paper. I am further thankful to Eric Auerbach, Bruno Fava, Danil Fedchenko, Joel Horowitz, Giacomo Marcolin, Zhen Xie, and the participants of the Econometrics Seminar, the HELP Workshop and the Econometrics Reading Group at Northwestern University for their valuable comments and suggestions. All errors are my own.

1 Introduction

In many economic settings, outcomes result from the interaction of two distinct types of agents and depend on the latent characteristics each side brings to the match. For example, wages reflect both worker skills and firm attributes (Abowd et al., 1999); corporate performance depends on managerial ability together with company-specific features (Bertrand and Schoar, 2003); and the productivity of public offices is shaped by local conditions and the capacity of the bureaucrats in charge (Fenizia, 2022).

Analyzing such settings requires using observable outcomes to (i) disentangle the contributions of the agents involved in the interaction, quantifying their latent characteristics, and (ii) understand how these characteristics interact to generate outcomes. Doing so provides the foundation for addressing a wide range of questions. In the labor market, for instance: are high-productivity workers more likely to match with high-productivity firms? How much of wage variation is due to worker heterogeneity versus firm heterogeneity? Which observable characteristics correlate with latent productivities? What is the role of complementarities in this setting? If complementarities matter and wages proxy for output, can counterfactual reallocation of matches raise aggregate productivity?

To study these questions, this paper introduces the Bipartite Interaction (BI) framework, which flexibly models two-sided interactions. Any model in this framework has two components: (i) a matching network, where nodes represent agents and edges capture which pairs are observed, and (ii) an interaction function, which maps the latent characteristics of matched agents into observed outcomes. The matching network summarizes the pattern of available data, while restrictions on the interaction function reflect assumptions about complementarities and generate distinct models within the framework. A well-known example is the two-way fixed effects (TWFE) model, which assumes an additively separable interaction function and therefore rules out complementarities. This model has become widely used by applied economists, particularly following its influential application by Abowd et al. (1999).

Building on this framework, I introduce the Tukey model, named after John Tukey, who proposed an analogous functional form in the context of nonlinear ANOVA (Tukey, 1949). The Tukey model enriches the TWFE specification by allowing for complementarities in a simple and interpretable way: it relies on the assumption that the cross-partial derivative of the interaction function is determined by a single scalar parameter. This interaction parameter entirely captures the presence, strength, and direction of complementarities; when it equals zero, complementarities vanish, and the model coincides with TWFE.

Next, I study two extensions of the Tukey model that accommodate richer complementarity structures by relaxing the restriction of a constant cross-partial derivative. The first allows a

firm-specific interaction parameter, permitting heterogeneity in complementarity across firms. This specification uses the functional form in Bonhomme et al. (2019) but drops their grouping assumption, treating each agent individually without clustering. The second extension imposes no parametric restriction on the cross-partial derivative. It serves as a fully flexible nonparametric benchmark, requiring only that the interaction function be monotone in both arguments.

Taken together, the BI framework, the Tukey model, and these extensions lead to three main contributions. First, I show that identification in the BI framework reveals a trade-off between the flexibility of the interaction function and the structure of the matching network: weaker assumptions on the interaction function require richer structure in the matching network. For example, in the TWFE model, point identification requires only that the matching network be connected, meaning any pair of nodes can be joined by a finite path of edges. This condition is no longer sufficient in general. In the nonparametric specification, for instance, point identification requires that any two nodes of the same type be linked by a path of length at most two. Importantly, the trade-off is asymmetric: while the interaction function is unobserved and must be restricted by assumption, the matching network is directly observed, and restrictions on it are straightforward to check.

In empirical applications, researchers typically begin by choosing the restrictions to impose on the interaction function. The results in this paper can then be used to verify whether the observed matching network satisfies the corresponding conditions for parameter identification. For the Tukey model, point identification requires the same connectedness condition as the TWFE model, plus one additional requirement: the presence of at least one informative cycle of length four in the matching network (a closed path involving four distinct edges and nodes). This condition is often satisfied in practice, making the Tukey model a flexible and empirically applicable alternative to the TWFE specification. By contrast, the requirements for point identification in the two extensions are much more stringent and rarely met in empirical settings where two-sided interactions are typically studied.

Second, I propose a new estimator in the Tukey model for the interaction parameter corresponding to the cross-partial derivative of the interaction function, and study its properties in a large-graph asymptotic analysis that does not require any agent to be observed many times. As identification relies on the presence of a cycle in the matching network, the estimator uses multiple such cycles to average out noise and consistently estimate the parameter. Unlike alternative methods in similar settings, the estimator does not require estimating latent characteristics. It is consistent under the mild requirement that the number of cycles scales with graph size; in employer–employee matched data, I show that such cycles are typically abundant. The estimator is asymptotically normal, with its variance shaped by three elements: (i) the variance of the outcome error terms, (ii) heterogeneity in latent characteristics within cycles, and (iii) the ordering of

agents in each cycle. Since the ordering depends on labels chosen by the researcher, I propose an instrument-based procedure that assigns them using observable characteristics correlated with latent productivities, ensuring consistency and asymptotic normality.

Third, I develop a formal test for the absence of complementarities in the BI framework. The test exploits the nesting of the TWFE model within the Tukey model: since the TWFE specification corresponds to the Tukey model with the interaction parameter equal to zero, testing this null provides a direct test for the absence of complementarities. The assumption of no complementarities implied by the TWFE model is often discussed in empirical work, and several informal diagnostics are in use, but a formal test has not, to my knowledge, been studied. The cycle-based estimator for the interaction parameter does not require estimating any individual latent characteristics, which allows construction of a test whose asymptotic properties can be studied by an asymptotic analysis aligned with common data patterns, including cases where nodes in the matching network have no more than two links.

To illustrate how the Tukey model can provide richer insights into two-sided interactions, I revisit the application in Limodio (2021) on the interaction between public managers and tasks, focusing on the implementation of World Bank projects. The success of a project is modeled as the outcome of the interaction between the ability of the manager in charge and the characteristics of the country where they operate. Original estimates using the TWFE model found negative sorting, with high-performing managers more likely to be matched with low-performing countries. Estimates from the Tukey model add further insight: the interaction parameter is negative and statistically different from zero, indicating negative complementarities, where high-performing managers have larger value added when matched with low-performing countries. Given this interaction function, negative sorting is optimal for maximizing average project success, which may explain the allocation pattern documented in Limodio (2021) as a rational response to the structure of complementarities.

1.1 Related Literature

The BI framework proposed in this paper is related to a wide range of earlier approaches to two-sided interactions, many of which are reviewed in Bonhomme (2020). The benchmark in this class is the TWFE, which assumes a modular (additively separable) interaction function. While most empirical applications focus on worker-firm interactions (Abowd et al., 1999; Card et al., 2013; Kline, 2024), TWFE-style methods have also been applied to other two-sided settings, including managers and firms (Bertrand and Schoar, 2003), teachers and students (Jackson et al., 2014; Chetty et al., 2014a,b), patients and healthcare providers (Finkelstein et al., 2016), and bureaucrats and geographic postings (Fenizia, 2022; Limodio, 2021). Despite its versatility, the model relies

on the strong modularity assumption, which rules out complementarities: the marginal productivity of each agent is assumed constant and independent of the other side of the match. In many contexts, this contrasts with the emphasis placed on matching, where significant resources are devoted to finding "the right match".

To relax modularity, Bonhomme et al. (2019) propose a model in which the interaction function allows for complementarities across the latent characteristics of agents. Their approach assumes that agents can be partitioned into a finite number of groups, with all members of a group sharing the same latent characteristics. In practice, this requires researchers to assign agents to groups, typically via a clustering procedure, and then estimate group-specific productivities and intergroup complementarities. A similar approach is adopted in Lei and Ross (2024). These models introduce greater flexibility in the interaction function relative to TWFE and have delivered valuable empirical insights (Weigel et al., 2024; Mourot, 2025), but rely on the grouped heterogeneity assumption. By contrast, the approach developed in this paper offers a feasible way to introduce complementarities into two-sided interactions without relying on grouping.

Concerns about complementarities have also surfaced in many TWFE applications. A common strategy to justify their exclusion has been to estimate a saturated version of the model, compare R^2 values, and conclude, based on the typically small changes observed, that complementarities are not quantitatively important (Card et al., 2013; Song et al., 2019; Fenizia, 2022; Adhvaryu et al., 2024). However, Kline (2024) shows that these R^2 comparisons can be misleading and argues that complementarity patterns are better detected by focusing on cycles in the matching network. Building on this insight, this paper develops a procedure to formally test the absence of complementarities, providing a formal way to assess the assumption at the core of the TWFE model.

Worker-firm interactions are the most common application for this TWFE model (Abowd et al., 1999; Card et al., 2013; Kline, 2024), yet modeling labor markets in this "reduced form" way has faced criticism (Eeckhout and Kircher, 2011; Hagedorn et al., 2017; Lopes de Melo, 2018; Eeckhout, 2018). These concerns extend to the BI framework as well. Still, the framework remains valuable: both in other domains where its structure may be more credible (as in the empirical illustration) and as a foundation for re-examining labor markets under alternative, potentially more realistic, restrictions than those implied by the TWFE model.

Related developments in balanced unit-time panel settings show the value of modeling interactions beyond modularity. The model in Tukey (1949), for example, already incorporated complementarities more richly, and many recent contributions extend this idea using factor models or nonparametric interaction structures (Bai, 2009; Freyberger, 2018; Freeman and Weidner, 2023; Sbai Sassi, 2024; Armstrong et al., 2025). Two differences distinguish the BI framework from classical panels. First, panel models typically focus on estimating coefficients on observed regressors,

whereas the BI framework aims to recover the interaction function and the fixed effects themselves. Second, panel analysis usually relies on observing all unit-time combinations, while the BI framework is defined over incomplete matching networks in which only a subset of possible matches is observed. As a result, the structure of the matching network is central to both identification and inference in the BI framework, in contrast to panel settings where it is fixed by design and excluded from the modeling analysis.

Because of this crucial role of the matching network, this paper also contributes to the literature on network econometrics. For identification, as in Bramoullé et al. (2009), Graham (2017), and De Paula et al. (2018), I impose assumptions on the observed graph to ensure that parameters can be identified, using a strategy that eliminates nuisance parameters to identify the interaction coefficient, paralleling the approach of Jochmans (2017) for multiplicative models. For inference, I consider an asymptotic regime in which the graph grows without requiring the number of realized links to increase proportionally with the number of potential links, and allowing the number of edges per node to remain bounded. Most existing inference results for sparse networks assume degrees (numbers of edges per node) that may diverge with network size (Jochmans and Weidner, 2019; Cai, 2022). Exceptions that accommodate bounded-degree graphs, such as Verdier (2020) and Auerbach et al. (2025), do not focus on parameters governing interactions. By explicitly allowing for bounded degree, the present analysis better reflects the structure of many real-world datasets on two-sided interactions and clarifies how the matching network affects the precision of BI estimators.

1.2 Paper Structure and Notation

The remainder of the paper is organized as follows. Section 2 introduces the BI framework, the Tukey model, and its extensions studied in this paper. Section 3 derives necessary and sufficient conditions for parameter identification in the Tukey model and its extensions. Section 4 turns to estimation in the Tukey model, mostly focusing on the estimator for the interaction parameter and its asymptotic properties. Section 5 presents an empirical illustration, showing how the Tukey model can be used in practice and how it provides additional insights into the interaction function. Section 6 reports Monte Carlo simulations assessing the finite-sample performance of the estimator. Section 7 concludes.

Throughout the paper, lowercase letters denote scalar parameters or quantities specific to an individual agent, typically scalars but possibly vectors when productivities are multidimensional (e.g., α_i represents the latent productivity of worker i). Uppercase letters denote sets (e.g., $\alpha_i \in A$, where A is typically a compact metric space). Bold symbols denote tuples: for example, $\boldsymbol{\alpha}$ denotes the collection $(\alpha_1, \ldots, \alpha_I)$.

2 Models

I present the Bipartite Interaction (BI) framework in the context of the labor market, where interactions involve workers and firms. This setting serves as a concrete example for exposition, but the framework itself is general and applies to any two-sided environment.

2.1 Bipartite Interaction Framework

Let $I \in \mathbb{N}$ and $J \in \mathbb{N}$ denote the number of workers and firms, respectively. Each worker i has a latent deterministic productivity $\alpha_i \in A$, and each firm j has a latent deterministic productivity $\psi_j \in \Psi$, where A and Ψ are compact subsets of $\mathbb{R}^d$. Let $\boldsymbol{\alpha} \coloneqq (\alpha_1, \ldots, \alpha_I)$ and $\boldsymbol{\psi} \coloneqq (\psi_1, \ldots, \psi_J)$ denote the collections of these productivities.

The potential outcome of the interaction between worker i and firm j, for example the wage worker i would receive if employed by firm j, is

$$y_{ij} = \underbrace{f(\alpha_i, \psi_j)}_{\theta_{ij}} + \eta_{ij},$$

where $f \colon A \times \Psi \to \mathbb{R}$ is the interaction function and η_{ij} is a mean-zero random term. Define $\theta_{ij} \coloneqq f(\alpha_i, \psi_j)$ as the deterministic component of the outcome, equal to $\mathbb{E}[y_{ij}]$, determined by the interaction function and the productivities of the matched agents. To focus on f, I omit covariates. Incorporating covariates in the BI framework is a valuable extension not considered in this paper.

To study the properties of f, it is useful to consider the role of its cross-partial derivative, which leads to the notions of modularity and complementarities. I formally define modularity in Appendix A.1. Intuitively, when α_i and ψ_j are scalars and f is differentiable, the interaction function is *modular* when the cross-partial derivative of f is zero, *supermodular* when it is nonnegative, and *submodular* when it is nonpositive. This is closely related to complementarities. I say that an interaction exhibits *complementarities* when the cross-partial derivative is nonzero, which I call *positive* or *negative* when this derivative is positive or negative, respectively.

The potential outcome y_{ij} is defined for all worker-firm pairs, representing the outcome that would be realized if match (i, j) occurred, whether or not it is observed in the data. In practice, only a subset of matches is realized, and y_{ij} is observed only for some pairs. Define

$$D_{ij} = \begin{cases} 1, & \text{if } y_{ij} \text{ is observed,} \\ 0, & \text{otherwise,} \end{cases}$$

as the indicator of observed matches, treated as fixed in the analysis, and collect them in $O_{IJ} = \{(i, j) : D_{ij} = 1\}$.

Let $G_{IJ} = ([I], [J], O_{IJ})$ be the bipartite network linking worker i to firm j whenever $D_{ij} = 1$. G_{IJ} is bipartite because its nodes are partitioned into two disjoint sets, workers and firms, with no edges within the same set.

Definition 1. (Matching Network) *The matching network G_{IJ} is the bipartite graph with node sets $[I]$ and $[J]$ and edge set $O_{IJ} = \{(i, j) : D_{ij} = 1\}$. An edge (i, j) indicates that y_{ij} is observed.*

Figure 1 shows an example with 5 workers (purple), 3 firms (green), and 6 edges, each representing an observed outcome.

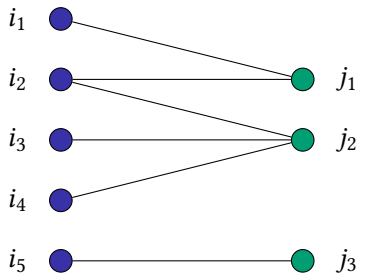


Figure 1: Matching network with nodes colored by type: workers in purple and firms in green. In this example, $I = 5$, $J = 3$, and the set of observed matches is $O_{IJ} = \{(i_1, j_1), (i_2, j_1), (i_2, j_2), (i_3, j_2), (i_4, j_2), (i_5, j_3)\}$.

The fact that worker i_2, for example, is linked to two firms (j_1 and j_2) does not imply that the matches occur simultaneously. The matching network is constructed over a chosen time window, which may span several years, and each edge can correspond to a different period: i_2 may be employed by j_1 in one year and by j_2 in another. The BI framework is static, abstracting from the timing of moves, so it does not matter whether the match with j_1 occurs before or after the one with j_2.

The matching network is deterministic: the BI framework does not model its formation, and the structure of G_{IJ} is taken as given. The only source of randomness in the model are the terms $\{\eta_{ij}\}_{(i,j)\in O_{IJ}}$, one for each observed outcome. These terms are mutually independent, mean-zero, and may have pair-specific distributions, allowing for heteroskedasticity. Since G_{IJ} is non-random, whether the match (i, j) is observed does not depend on η_{ij}, making the random term exogenous.

For any realized match ($D_{ij} = 1$), the model considers only a single outcome y_{ij}. If repeated observations of the same worker-firm pair are available, they can be averaged to yield a single y_{ij}, leaving the analysis unchanged.

Parameters of interest. The primitive parameters of the model are the interaction function f and the productivity vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$. Recovering f allows researchers to characterize how worker and firm productivities map into the outcome. A central question is whether an agent's marginal productivity depends on the characteristics of the other agent in the match, and, when this occurs, to describe the nature and pattern of such complementarities. The productivity vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$, once linked to additional information on workers and firms, shed light on the determinants of productivity differences. For example, regressing $\boldsymbol{\alpha}$ on worker demographics reveals which observable traits drive variation in worker productivity.

Together, $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ permit the study of assortative matching, the tendency of workers and firms with similar relative productivities to pair together. A common measure of sorting is the correlation between α_i and ψ_j across observed matches: one constructs a vector containing α_i for each match (i, j), a second vector with the corresponding ψ_j, and then computes their correlation. A positive value indicates positive sorting, whereas a value near zero or negative suggests weak or reverse sorting.

Joint knowledge of f, $\boldsymbol{\alpha}$, and $\boldsymbol{\psi}$ also enables decomposing outcome variance, quantifying the shares attributable to worker heterogeneity, firm heterogeneity, and their interaction. These quantities further allow measuring factor misallocation by comparing observed matches to the efficient allocation, and support counterfactual analyses in which workers are reassigned across firms.

These examples illustrate the wide range of parameters of interest that can be derived from f, $\boldsymbol{\alpha}$, and $\boldsymbol{\psi}$ as primitive inputs. A comprehensive treatment of all such parameters is beyond the scope of this paper. The focus here will be on primitive f, $\boldsymbol{\alpha}$, and $\boldsymbol{\psi}$ under different assumptions on f and G_{IJ}, occasionally using selected derived quantities to highlight key features of the results.

Functional specifications. Imposing restrictions on the interaction function within the BI framework yields distinct models. I focus on three specifications: the Tukey model, the BLM model, and the seriation model. The next sections describe each of these models and discuss their economic interpretation. Appendix A formalizes their connection to shape restrictions on f, showing how their functional forms emerge from broader assumptions.

2.2 Tukey Model

The cross-partial derivative of f captures the complementarity structure of the interaction function in the BI framework. A natural way to introduce flexibility is to assume that this cross-partial

derivative is constant, summarized by a single parameter. This yields the specification

$$\theta_{ij} = \alpha_i + \psi_j + \beta_0 \alpha_i \psi_j, \qquad \text{(Tukey model)}$$

with $\beta_0 \in B \subset \mathbb{R}$ compact. I refer to this as the Tukey model, after the statistician John Tukey, who studied an analogous functional form in two-way ANOVA to test whether two categorical factors affect the response additively (Tukey, 1949; Ward and Dick, 1952; Šimeček and Šimečková, 2013). In that setting, the focus is on testing the null hypothesis $\beta_0 = 0$ under restrictive assumptions (homoskedastic, normally distributed errors η_{ij} and complete matching network G_{IJ}), and no attention is given to the estimation of the parameter itself.

Conversely, in the BI framework, the interaction parameter β_0 in the Tukey model acquires a direct economic meaning. It is the constant cross-partial $\partial^2 f / \partial\alpha\partial\psi$, capturing all departures from modularity and governing the complementarity pattern between the two productivities.

Despite its parsimony, where the entire complementarity structure is summarized by a single parameter, the Tukey model is flexible enough to encompass supermodular ($\beta_0 \geq 0$), submodular ($\beta_0 \leq 0$), and modular ($\beta_0 = 0$) interaction functions. The sign of β_0 determines the direction of complementarities, while its magnitude reflects the relative weight of the multiplicative component compared to the additive one: as $|\beta_0|$ grows, the role of the match itself becomes more important.

2.2.1 Tukey as Extension of TWFE Model

When $\beta_0 = 0$, the Tukey model reduces to the widely used TWFE specification, in its baseline form without covariates:

$$\theta_{ij} = \alpha_i + \psi_j. \qquad \text{(TWFE model)}$$

In the TWFE model, the cross-partial derivative $\partial^2 f / \partial\alpha\partial\psi$ is zero, so the interaction function is modular and complementarities are ruled out by assumption. The marginal contribution of a worker (firm) is independent of the firm (worker) they are matched with and remains constant across all matches. Under modularity, for a fixed set of matched agents, total output depends only on individual productivities and is invariant to the assignment of matches: any allocation is efficient.

While the assumption of no complementarities is restrictive, especially given the emphasis in economic theory on complementarities as a driver of sorting patterns such as positive assortative matching (Becker, 1973; Shimer and Smith, 2000), its appropriateness depends on the empirical setting and the research question. In practice, the TWFE model is often viewed less as a literal

description of interactions and more as a tractable approximation to richer structures (Abowd et al., 1999; Card et al., 2013). The nesting of TWFE within the Tukey model makes it possible to assess when this approximation is likely to be informative and, conversely, when ignoring complementarities may lead to misleading conclusions; see Remark 2 for details.

2.3 Beyond Tukey Model

The Tukey model introduces complementarities in a simple and interpretable way. To capture richer complementarity patterns, one can relax its restrictions and consider more flexible interaction functions: the BLM model allows the interaction parameter to be firm-specific, accommodating heterogeneity in complementarities across firms; the seriation model provides a fully nonparametric alternative, imposing only monotonicity and leaving the cross-partial derivative unrestricted.

2.3.1 BLM Model

Rather than modeling firm productivity as a scalar, suppose instead that each firm is characterized by a productivity vector $\psi_j = (b_j, a_j)$. This leads to the specification

$$\theta_{ij} = a_j + b_j \alpha_i, \qquad \text{(BLM model)}$$

where each firm is endowed with both an intercept and a slope, each varying across firms.

This interaction function is the one studied by Bonhomme et al. (2019), which motivates the label BLM model. In their approach, the functional form is embedded in a grouped setting, where all firms and workers within a group share the same latent characteristics. In the BI framework, by contrast, the BLM model refers to the case in which each group consists of a single worker or a single firm.

The BLM model nests the Tukey model as a special case: when $b_j = 1 + \beta_0 a_j$ for all j, or equivalently $\beta_0 = \frac{b_j - 1}{a_j}$, the specification reduces to the Tukey one, with $\psi_j = a_j$. The BLM model can thus be viewed as a generalization of the Tukey model, where the slope parameter capturing complementarities varies across firms. This additional flexibility allows researchers to distinguish a firm's intrinsic productivity a_j, which enters the model in levels, from its capacity to extract value from workers and exploit complementarities, governed by b_j. Unlike the Tukey model, which implicitly ties these two roles together, the BLM model permits them to differ, thereby enabling empirical investigation of whether more productive firms are also those in which workers' marginal productivity is higher.

2.3.2 Seriation Model

The BLM model imposes a specific parametric form on the interaction function. As a benchmark, it is useful to also consider a specification that removes all parametric restrictions on the cross-partial derivative and allows for fully flexible complementarities. The natural analogue comes from nonparametric regression: when the interest is in learning $\mathbb{E}[Y|X]$ nonparametrically from observations of continuous random variables Y and X, some regularity assumptions are needed, typically smoothness. Similarly, in the bipartite interaction setting, some regularity condition on f is necessary to ensure that the data are informative.

I retain scalar productivities and impose a monotonicity assumption on f, leaving the cross-partial unrestricted:

$$\theta_{ij} = f_m(\alpha_i, \psi_j), \qquad \text{(Seriation model)}$$

with $f_m : A \times \Psi \to \mathbb{R}$ increasing in both arguments. The monotonicity assumption in a bipartite network setting connects directly to seriation problems in statistics (Flammarion et al., 2019) and to matrix-completion approaches for bivariate isotonic matrices under unknown permutations (Mao et al., 2020). Outside the bipartite setting, it is also related to nonparametric latent-space models for network formation; see, for example, Gao (2020) and references therein.

Monotonicity is a common shape restriction in economics (see the survey in Chetverikov et al. (2018)). In this context, it requires that the ranking of average outcomes produced by two workers (firms), when matched with the same firm (worker), is preserved across all matches. This property holds, for example, in the TWFE model.

The seriation model imposes no parametric restriction on f_m: when it is twice differentiable, monotonicity ensures constant signs for the partial derivatives but imposes no constraint on the cross-partial derivatives that capture complementarities. As a result, the complementarity pattern is highly flexible: the same firm can exhibit supermodular or submodular behavior depending on the productivity of the worker it is matched with. This is not possible in the BLM model, where the cross-partial derivative is fixed for each firm and does not vary across workers.

2.4 Trade-off in BI Framework

The three models described above, together with the TWFE model, trace out a spectrum within the BI framework. At one end lies the fully modular TWFE model, which rules out complementarities altogether. At the other end is the seriation model, which only imposes monotonicity and allows complementarities to vary freely across matches. The Tukey and BLM models sit between these extremes: they restrict the complementarity structure parametrically, but retain

more flexibility than the TWFE model.

A second dimension of variation arises from the matching network G_{IJ}, which encodes the observed links between workers and firms. In the most informative case, G_{IJ} is complete, with $D_{ij} = 1$ for every pair. At the opposite extreme, the graph may be sparse (with far fewer links than the maximum possible) and fragmented, with each firm linked to only a handful of workers, and vice versa.

Together, the interaction function f and the network G_{IJ} determine what can be learned from the data. The function f and the dimensions I and J dictate how many parameters must be recovered, while the structure of G_{IJ} governs how much information is available. This creates a trade-off: richer flexibility in f requires stronger assumptions on G_{IJ}, while sparser networks can only be informative under more restrictive assumptions on f.

This trade-off is asymmetric. The function f is unobserved and summarizes the latent interaction between workers and firms; G_{IJ} is observed and can be inspected directly. Restrictions on G_{IJ} can therefore be verified in the data, while restrictions on f cannot. For the TWFE model, Abowd et al. (1999) and Jochmans and Weidner (2019) provide conditions on G_{IJ} for identifying and, under additional assumptions, estimating the productivity parameters. For the Tukey, BLM, and seriation models, there are no results that allow the matching network to be incomplete. In the next section, I derive the necessary and sufficient conditions that link the structure of f with the shape of G_{IJ}, extending identification results to incomplete and possibly sparse graphs. These conditions involve some requirements on the matching network G_{IJ}. For each requirement, I will also briefly assess its plausibility in the context of employer-employee data.

In applications, the researcher can proceed in two steps. First, select the restriction on f that best fits the economic environment. Then, use these results to check whether the observed G_{IJ} satisfies the corresponding conditions. When the focus is on point-identification, three outcomes are possible: (i) the parameters are not identified; (ii) they are identified but not consistently estimable; (iii) they are both identified and consistently estimable. Clarifying which case applies is essential for understanding what questions the data can credibly answer. Naturally, the richer and more connected the graph, the wider the scope of questions that can be addressed.

3 Identification

I analyze identification in the Tukey, BLM and seriation models. Section 3.1 introduces the notion of identification adopted for the BI framework. Sections 3.2, 3.3, and 3.4 present the identification results for the three models. Finally, Section 3.5 revisits the trade-off between the flexibility of the interaction function and the structure of the matching network in light of these results.

3.1 Identification in BI Framework

I adopt the classical notion of point identification from Koopmans (1949), which requires that the model parameters be uniquely recoverable from the distribution of observables. In the BI framework, the observables are the collection $\{y_{ij}\}_{(i,j)\in\mathcal{O}_{IJ}}$. Hence, for identification purposes, I can treat $\mathbb{E}[y_{ij}] = \theta_{ij}$ as known for each observed match.

Formally, I study identification through the noiseless map

$$(f, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ}) \mapsto \boldsymbol{\theta}_{\mathcal{O}} \coloneqq \{\theta_{ij}\}_{(i,j)\in\mathcal{O}_{IJ}},$$

where $(f, \boldsymbol{\alpha}, \boldsymbol{\psi})$ are unknown and the graph G_{IJ} is known. The parameter vector $(f, \boldsymbol{\alpha}, \boldsymbol{\psi})$ is *point-identified* if the map

$$(f, \boldsymbol{\alpha}, \boldsymbol{\psi}) \mapsto \boldsymbol{\theta}_{\mathcal{O}}$$

is injective: that is, whenever parameters $(f, \boldsymbol{\alpha}, \boldsymbol{\psi})$ and $(f', \boldsymbol{\alpha}', \boldsymbol{\psi}')$ generate the same $\boldsymbol{\theta}_{\mathcal{O}}$, they must coincide exactly.

In practice, a parameter is identified if and only if it can be expressed as a function of $\boldsymbol{\theta}_{\mathcal{O}}$: this is the equivalence I will use to establish identification in the proofs.

This definition deliberately abstracts from sampling noise and treats the distribution of each y_{ij} as known, even though in applications each match is typically observed only once. While such a notion does not distinguish between parameters that can or cannot be consistently estimated in the presence of error terms η_{ij}, it provides a fundamental benchmark: if a parameter is not identified in the noiseless model, then no estimator can recover it. Conversely, whenever a parameter is identified in this sense, it warrants further analysis to determine whether, and under what conditions, consistent estimation is feasible.

A location normalization is often necessary, since the absolute levels of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ are not uniquely determined by the observables. In the TWFE model, for example, adding a constant to all α_i and subtracting it from all ψ_j leaves $\boldsymbol{\theta}_{\mathcal{O}}$ unchanged; the same invariance holds in richer specifications with complementarities, where shifts in $\boldsymbol{\alpha}$ can be offset by changes in $\boldsymbol{\psi}$ and f without altering $\boldsymbol{\theta}_{\mathcal{O}}$. To fix the reference level and ensure point identification, a normalization such as $\sum_i \alpha_i = 0$ or $\alpha_1 = 0$ is imposed. Whenever such a normalization is required, I state it explicitly and adopt the form that yields the clearest formulas.

Remark 1. (Partial Identification) *This paper focuses on point identification as a starting point for studying flexible complementarity patterns in the BI framework. The same noiseless approach can, however, be extended to partial identification, where the identified set for $(f, \boldsymbol{\alpha}, \boldsymbol{\psi})$ is given by the set of counterimages of the observed $\boldsymbol{\theta}_{\mathcal{O}}$. Because matching networks are often sparse and the*

conditions for point identification difficult to satisfy, partial identification results provide a way to extract meaningful information about the underlying economic structure even when point identification fails. Extending the BI framework in this direction appears hence especially promising, and future work could build on recent advances such as Crippa and Fedchenko (2025), who study partial identification in the distinct but related pairwise interaction model.

3.2 Identification in Tukey Model

The Tukey model introduces complementarities in the interaction function through a single parameter, β_0, which represents the constant cross-partial derivative and fully characterizes f. I present the identification results in two steps: first, the identification of β_0; second, the identification of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$.

3.2.1 Identification of β_0

Identifying β_0 requires additional structure on the matching network G_{IJ}. I begin by recalling the notion of a cycle in a bipartite graph.

Definition 2. (Cycle in the Matching Network) *In the bipartite graph $G_{IJ} = ([I], [J], O_{IJ})$, a cycle of length $2K$ ($K \geq 2$) is a closed alternating sequence of workers and firms*

$$i_1, j_1, i_2, j_2, \ldots, i_K, j_K, i_1$$

such that, for each $k = 1, \ldots, K$, (i_k, j_k) and (i_{k+1}, j_k) belong to O_{IJ} (with $i_{K+1} = i_1$), and all $i_1, \ldots, i_K$ and $j_1, \ldots, j_K$ are distinct.

The graph in Figure 2a contains no cycles: no path starts and ends at the same node while traversing distinct edges. By contrast, adding an edge between i_1 and j_2, as in Figure 2b, creates the closed path i_1, j_1, i_2, j_2, i_1, which forms a cycle of length 4.

The key condition required for identification of β_0 is given in Assumption 1.

Assumption 1. (Informative cycle) *The matching network G_{IJ} contains a length-4 cycle i_1, j_1, i_2, j_2 such that $\alpha_{i_1} \neq \alpha_{i_2}$ and $\psi_{j_1} \neq \psi_{j_2}$.*

Assumption 1 requires the presence of a 4-cycle in which both workers and firms differ in productivity. Such heterogeneity is crucial: without it, the outcomes along the cycle would not vary, and the cycle would provide no information about β_0. When heterogeneity is present, however, the cycle reveals contrasts across matches that make the interaction parameter point-identified.

Theorem 1. (Identification of β_0) *Under the Tukey model, Assumption 1 is sufficient for identification of β_0. If the matching network contains no more than one cycle, it is also necessary.*

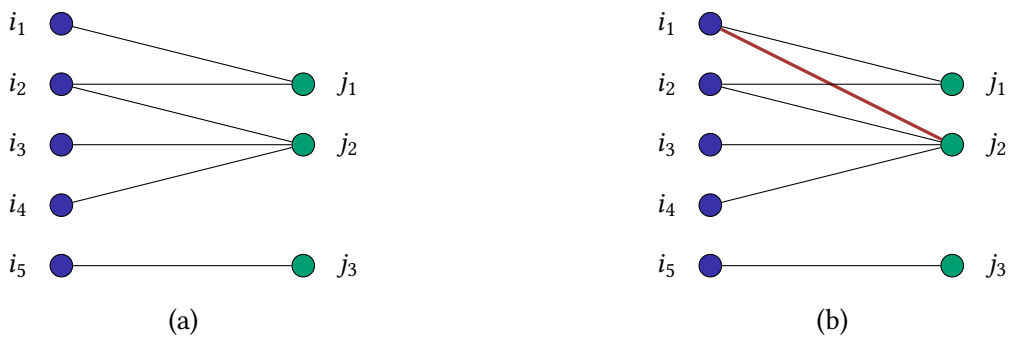


Figure 2: Two examples of matching networks: (a) a network without cycles, and (b) a network containing a cycle of length four. In panel (b), the cycle is formed by the sequence of nodes $i_1 \to j_1 \to i_2 \to j_2 \to i_1$.

Identification of β_0 hence requires the presence of a cycle in the matching network, and if that cycle is unique, it must be of length four. The proof (Appendix H.1) shows that outcomes from a cycle of length $2K$ generate a degree-$(K-1)$ polynomial in β_0, with coefficients given by known functions of $\boldsymbol{\theta}_O$. The identification set consists of the roots of this polynomial. When the network contains at most one cycle, point identification requires $K = 2$, i.e. a cycle of length 4. More generally, a cycle of length $2K$ yields an identification set with $K-1$ elements. If multiple longer cycles are present, the intersection of their identification sets can reduce to a singleton, achieving point identification.

The role of cycles in detecting departures from modularity was previously noted by Card et al. (2013) and discussed by Kline (2024), though in those cases cycles were used as a diagnostic device rather than as a source of identification for an interaction parameter.

In the labor market setting, Assumption 1 requires some degree of worker mobility across firms. Because each worker can be matched with only one firm at a time, the multiple links needed for a cycle arise only when workers change employers. The condition therefore, requires that, for some pair of firms, at least two movers exist, and that both the workers and the firms involved differ in productivity. This is not especially restrictive: labor markets are typically segmented into local or sectoral clusters, and when one worker moves between two firms, the likelihood of additional movers between the same firms is higher than under random matching. Evidence supports this: in the application of Kline (2024), for instance, about 55% of firms belong to at least one cycle.

3.2.2 Identification of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$.

Once β_0 is identified, the productivity vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ in the Tukey model can be identified under an additional condition on G_{IJ}.

Assumption 2. (Connectedness) *The matching network G_{IJ} is connected: for any two nodes (workers or firms) in G_{IJ}, there exists a path (i.e., a sequence of nodes linked by edges) joining them.*

The graph in Figure 3a violates Assumption 2, since no path exists, for example, from node i_4 to node i_5. Adding an edge between i_4 and j_3, as in Figure 3b, makes the graph connected.

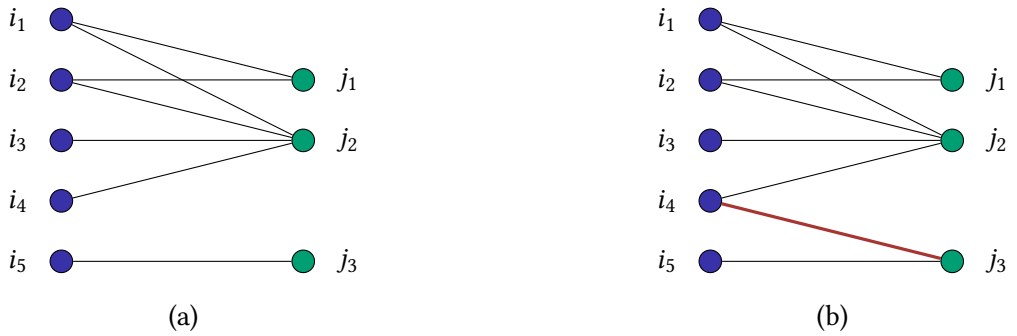


Figure 3: Two examples of matching networks: (a) a non-connected network, and (b) a connected network, which satisfies Assumption 2.

The identification result for $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ in the Tukey model is reported below.

Theorem 2. (Identification of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ in the Tukey model) *Under the Tukey model and the normalization $\alpha_{i_0} = 0$, if β_0 is identified, Assumption 2 is necessary and sufficient for identification of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$.*

Theorem 2 extends the classical identification result for the TWFE model: it requires knowledge of β_0, but allows this parameter to differ from zero. The proof proceeds by first showing that, for each observed match, one can construct a function of $\boldsymbol{\theta}_O$ and the identified β_0 that equals the product of known functions of the corresponding worker and firm productivities. Connectedness of G_{IJ} then guarantees that all worker and firm effects can be recovered up to the normalization, completing the identification argument.

In labor market applications, the matching network is rarely fully connected, especially over short time horizons. A common practice is therefore to restrict the analysis to the largest connected component. For example, in the West German labor market studied by Card et al. (2013), the largest component contains over 95% of workers and 90% of firms. Similar firm coverage is reported by Bonhomme et al. (2023) for Austria, Italy, Sweden, Norway, and the United States, although worker coverage is often much lower, in some cases falling below 50%.

Theorems 1 and 2 highlight the additional requirements introduced by allowing complementarities. Relative to the TWFE model, the Tukey model imposes only a modest cost: the same connectedness condition is needed, with the sole extra requirement that the graph contain at least one informative cycle to identify β_0. This result exploits the fact that β_0 is a global parameter: it governs the entire interaction function f and does not vary across specific workers or firms.

Hence, it can be identified using local information in G_{IJ} and then applied across all matches to recover $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$.

The Tukey model, therefore, offers a simple and flexible way to incorporate complementarities into the BI framework, while requiring only minimal conditions on the matching network for identification. By contrast, as I show in the following sections, more general specifications such as the BLM and seriation models demand substantially stronger network requirements, which are typically stringent and rarely satisfied in applications.

Remark 2. (TWFE as Approximation) *The TWFE model is often defended on the grounds that the benefits of departing from modularity are limited, making it a useful approximation even when interactions exhibit some complementarities. The nesting of TWFE within the Tukey model, combined with the focus on the noiseless case, allows this claim to be evaluated formally. In Appendix B, I study when the TWFE specification provides a good approximation to interactions governed by the Tukey model. The results show that relying on TWFE can yield misleading conclusions once complementarities play a non-negligible role. In particular, I derive the bias implied by the TWFE specification and show that, for example, in settings with a supermodular interaction function and strong positive sorting, the sorting captured by TWFE can be zero. In such cases, hence, the TWFE model would incorrectly suggest the absence of sorting.*

3.3 Identification in BLM Model

The BLM model extends the BI framework by allowing each firm to have its own interaction parameter. Bonhomme et al. (2019) study this specification under the assumption that workers and firms are partitioned into groups, with all agents in a group sharing the same productivity. Grouping substantially reduces the effective number of nodes in the matching network, since each group can be represented as a single node connected to many others. Under the assumption that the matching network is complete, Bonhomme et al. (2019) show that α_i, a_j, and b_j are point-identified. Completeness, however, is stronger than necessary. In what follows, I establish a weaker connectivity condition on G_{IJ} that still ensures identification of the productivity parameters.

To formalize this condition, I introduce the following property of the bipartite graph G_{IJ}.

Definition 3. (Seed-and-Snowballs Connectivity) *A bipartite graph G_{IJ} satisfies Seed-and-Snowballs connectivity if there exists a "seed" firm $j_0 \in J$ from which one can reach all nodes by iterating the following steps:*

1. ***Seed.*** *Set $S_0^J = \{j_0\}$.*
2. ***Snowball.*** *For $n = 0, 1, 2, \ldots$:*

(a) Add to S_n^I all workers connected to at least one firm in S_n^J:

$$S_n^I = \{i \in I : \deg_{S_n^J}(i) \geq 1\}.$$

(b) Add to S_{n+1}^J all firms connected to at least two workers in S_n^I:

$$S_{n+1}^J = S_n^J \cup \{j \in J : \deg_{S_n^I}(j) \geq 2\},$$

with $\deg_S(i)$ indicating the number of links between node i and nodes in S. If for some N the process yields $S_N^J = J$ and $S_N^I = I$, the graph satisfies the property.

Intuitively, the procedure alternates between (i) adding all workers linked to any firm already in the snowball and (ii) adding all firms connected to at least two of the workers in it. The "two-worker" condition guarantees that each newly added firm lies on a cycle, and that these cycles overlap so the snowball can propagate through the graph. In Figure 4a, the property fails because firm j_3 is not part of any cycle. Adding the edge (i_5, j_1), as in Figure 4b, creates overlapping cycles and makes the graph Seed-and-Snowballs connected.

To verify this, run the iterative procedure with j_1 as the seed, so that $S_0^J = \{j_1\}$. First, include the workers linked to the seed: $S_0^I = \{i_1, i_2, i_5\}$. Then, add firms linked to at least two of these workers: $S_1^J = \{j_1, j_2\}$. Repeating the process yields $S_1^I = \{i_1, i_2, i_5, i_3, i_4\}$ and $S_2^J = \{j_1, j_2, j_3\}$. Since the snowball eventually reaches every node, the graph satisfies Seed-and-Snowballs connectivity.

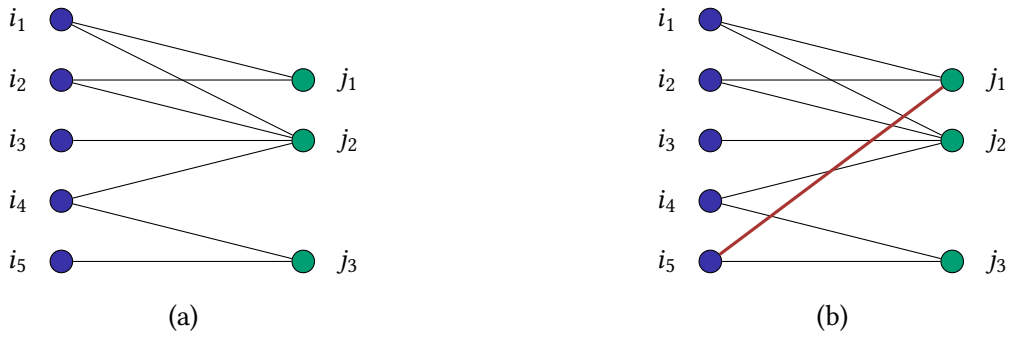


Figure 4: Two examples of matching networks: (a) a network that does not satisfy Seed-and-Snowballs connectivity, and (b) a network that does.

To the best of my knowledge, this property has not been discussed in the existing network literature. Appendix C shows that it is sufficient for the form of connectivity required by Kline et al. (2020) to ensure the validity of their variance estimator in the TWFE model, where the matching network must remain connected after removing any single worker node along with its incident edges.

As in the Tukey model, Seed-and-Snowballs connectivity must be complemented by sufficient heterogeneity in the productivities appearing in the restrictions to obtain the key condition for identification.

Assumption 3. (Informative Seed-and-Snowballs) *The matching network G_{IJ} satisfies Seed-and-Snowballs connectivity, with the additional requirements that: (i) when computing $\deg_{S_n^J}(i)$, only edges to firms with non-zero slopes are counted; and (ii) when computing $\deg_{S_n^I}(j)$, only edges to workers with distinct productivities are counted.*

Intuitively, since the BLM model generalizes the Tukey model by allowing a firm-specific interaction parameter, Assumption 3 extends Assumptions 1 and 2. It requires that each firm lie on a cycle involving heterogeneous productivities, and that these cycles overlap at least at one node. This overlap enables information to propagate through the network and ensures identification of the parameters in the BLM model, as formalized in the next theorem.

Theorem 3. (Identification in the BLM model) *Under the BLM model and the normalization $a_{j_0} = b_{j_0} = 1$, Assumption 3 is necessary and sufficient for identification of $\boldsymbol{\alpha}$, $\boldsymbol{a}$, and $\boldsymbol{b}$.*

While Assumption 3 is necessary for identification, it demands a matching network far richer than what is typically observed in applications. In the labor market context, for example, Kline (2024) finds that nearly half of the firms in their data are not part of any cycle, implying that their productivities cannot be identified. This is only a lower bound: being part of a cycle is not enough, since the cycles must also share at least one node. As a result, even under favorable conditions, the BLM model would fail to identify the productivity of a large share of firms, limiting its empirical applicability.

Theorem 3 thus highlights that the additional flexibility of firm-specific complementarities comes at a steep cost: the matching network must satisfy a stringent requirement that is rarely met in practice. Point identification in the BLM model is therefore challenging in most empirical settings with two-sided interactions, unless one imposes additional dimension-reduction restrictions such as those in Bonhomme et al. (2019).

3.4 Identification in Seriation Model

One might ask whether the restrictive graph condition found for identification with the BLM model is driven not by its richer interaction function, but by the use of multidimensional firm productivity, which introduces additional parameters.

The seriation model provides an alternative: it allows fully flexible complementarities while retaining scalar productivities. Here, the interaction function is left entirely nonparametric. As

the next proposition shows, this flexibility comes with a sharp limitation: f_m, $\boldsymbol{\alpha}$, and $\boldsymbol{\psi}$ can be recovered only up to strictly monotonic reparameterizations.

Proposition 1. (Lack of cardinal identification in the seriation model) *Model* $(f_m, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ *can be identified only up to any strictly increasing reparameterization of* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\psi}$.

Proposition 1 implies that only the ordinal information (the ranking) of worker and firm productivities can be identified. Cardinal differences are not preserved under strictly monotonic transformations, so f_m, $\boldsymbol{\alpha}$, and $\boldsymbol{\psi}$ cannot be separately identified.

While this prevents recovery of productivity levels, the ranks of $\{\alpha_i\}$ and $\{\psi_j\}$, and thus the rankings of the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$, are still identified. Rank-based methods can therefore be employed to study sorting patterns and the determinants of productivity ranks without imposing cardinal structure.

To state an identification condition for the seriation model, I first introduce the notion of within-side diameters.

Definition 4. (Within-side Diameter) *The within-side diameters of a graph* G_{IJ} *are the largest shortest-path distances between any two nodes on the same side of the bipartition: letting* $d(u, v)$ *be the number of edges in the shortest path between nodes* u *and* v,

$$\text{diam}_{\text{I}}(G_{IJ}) = \max_{i,i' \in I} d(i, i'), \quad \text{diam}_{\text{J}}(G_{IJ}) = \max_{j,j' \in J} d(j, j').$$

In the graph of Figure 4a, the within-side diameter for I is 4, as the shortest path from i_1 to i_5 spans four edges. For J, the diameter is 4, corresponding to the path between j_1 and j_3. In Figure 4b, the diameter for J falls to 2 thanks to the shorter path $j_1 \longrightarrow i_5 \longrightarrow j_3$, while the diameter for I remains 4.

The following condition for identification in the seriation model directly involves the within-side diameters.

Assumption 4. (Diameter 2) *The matching network* G_{IJ} *has within-side diameters equal to two.*

Assumption 4 requires that every pair of workers shares at least one common firm, and every pair of firms shares at least one common worker. This ensures that all nodes on the same side of the bipartition are linked through a single intermediary on the opposite side. The graph in Figure 5a fails this property, as workers i_3 and i_5 have no firm in common. Adding the edge (i_3, j_3), as in Figure 5b, creates the necessary connections and yields within-side diameters of two.

The next result shows that a within-side diameter of two is exactly the connectivity needed to recover the rankings of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$.

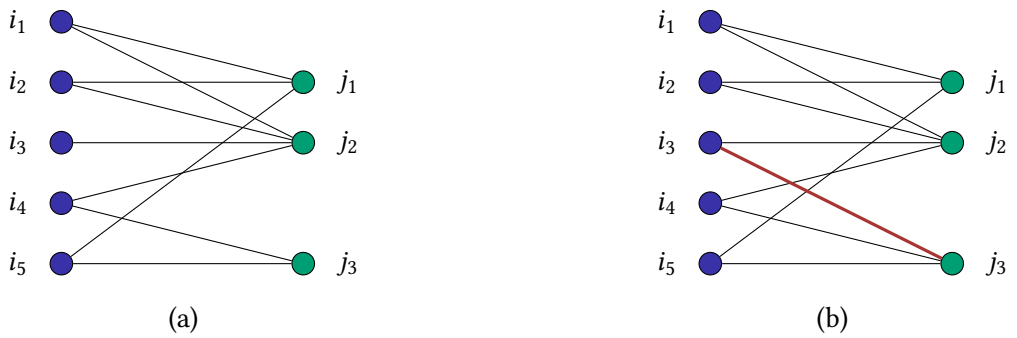


Figure 5: Two examples of matching networks: (a) a network that does not satisfy Assumption 4, and (b) a network that does, with within-side diameter equal to 2.

Theorem 4. (Identification of rankings in the seriation model) *Under the Seriation model, Assumption 4 is necessary and sufficient for identification of the rankings of* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\psi}$.

From an empirical standpoint, Assumption 4 is strong: it requires that every pair of workers have at least one firm in common and that every pair of firms share at least one worker. In labor market data, this would mean that any two workers have worked for the same firms, an unlikely occurrence outside of small or highly interconnected markets. As with the BLM model, the seriation model therefore has limited empirical applicability when the focus is on point identification.

That said, point identification is not always essential for extracting useful information from the data. Even when Assumption 4 fails, the seriation model can still deliver partial identification of the productivity rankings. Although I do not study partial identification in this paper, it remains a promising direction to explore. Following the approach by Crippa and Fedchenko (2025), for example, one could derive informative sets for the rankings of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$.

3.5 Identification Trade-off

When the identification conditions for the TWFE, Tukey, BLM, and seriation models are compared, the trade-off introduced in Section 2.4 becomes clear: weaker restrictions on the interaction function require richer structure in the matching network.

Theorems 3 and 4 show that the conditions for point identification in the BLM and seriation models are unlikely to be satisfied in most empirical settings involving two-sided interactions. By contrast, Theorems 1 and 2 demonstrate that the interaction parameter β_0, together with $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$, can be identified under assumptions that are plausibly met in practice.

The Tukey model thus emerges as a more flexible yet tractable alternative to the TWFE model. It is flexible because it accommodates complementarities, through the parameter β_0. It is tractable because it imposes only mild requirements on the network, unlike its extensions and similarly to the TWFE. In the next section, I show that the additional interaction parameter can be consistently

estimated under assumptions that are commonly satisfied by matching networks observed in applications.

4 Estimation and Inference in Tukey Model

I study estimation and inference for the parameters of the Tukey model. Section 4.1 presents the asymptotic setting used to analyze estimators in the BI framework. Section 4.2 introduces an estimator for β_0, establishes its consistency, derives its asymptotic distribution, and shows how it can be used to construct a test for the absence of complementarities. Section 4.3 then considers the estimation of productivities $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$. While I propose an estimator, I do not study its properties; instead, I show how its analysis and the limitations it faces connect to those of the existing productivity estimators in the TWFE model.

4.1 Large Sample in BI Framework

To study the large-sample properties of the estimators introduced below, I consider an asymptotic setting in which both the number of workers I and the number of firms J grow large ($I \to \infty$, $J \to \infty$). Information accumulates by adding new nodes to the bipartite graph, rather than by repeatedly sampling matches along existing edges.

The analysis is hence based on an asymptotic setting in which the network expands. The sequences of matching networks G_{IJ}, worker productivities $\boldsymbol{\alpha}_I$, and firm productivities $\boldsymbol{\psi}_J$ are taken as deterministic, and the set of observed matches O_{IJ} is allowed to vary with (I, J). Randomness arises solely from the error terms $\{\eta_{ij}\}_{(i,j)\in O_{IJ}}$, which are assumed independent across matches.

This setup contrasts with stochastic network formation models, where the graph itself is random. Instead, the analysis reflects empirical applications in which the observed labor market network is treated as fixed, and inference concerns the role of unobserved shocks given this network structure. Equivalently, the setting can be viewed as one where the network and productivities are random, but inference is conducted conditional on their realization.

The role of the matching network in the asymptotic analysis mirrors its role in identification: conditions on the sequence of graphs G_{IJ} ensure the validity of the asymptotic results. These conditions cannot be verified from a single observed graph. Instead, what matters is how the observed structure relates to the properties required of the asymptotic sequence, to assess whether the asymptotic results provide a good approximation to the finite-sample behavior.

In labor market applications, the large-I, large-J asymptotic setting is well-suited, as available datasets often include millions of workers and firms. At the same time, the data are sparse: each

worker is observed with only a handful of firms, and each firm with only a modest number of workers. For example, in the U.S. labor market over a one-year horizon, a typical worker is employed by one or two firms, while a typical firm employs only a few dozen workers. To capture this structure, the asymptotic setting allows node degrees to remain bounded as I and J grow, rather than requiring any worker to be linked with many firms or any firm with many workers.

4.2 Interaction Parameter β_0

Recall the Tukey model:

$$y_{ij} = \alpha_i + \psi_j + \beta_0 \alpha_i \psi_j + \eta_{ij},$$

with $\mathbb{E}[\eta_{ij}] = 0$, and $\{\eta_{ij}\}_{(i,j)\in O_{IJ}}$ independent.

This section introduces an estimator for β_0, and analyzes its properties in the large sample setting described above. Note that, despite Tukey (1949) and the following literature on non-additivity in ANOVA consider this same model, they do not discuss any estimator for β_0, rather focusing on directly testing the hypothesis $\beta_0 = 0$.

Theorem 1 shows that a single informative four-cycle suffices to identify β_0. In the presence of noise, however, estimation requires pooling information across many four-cycles present in G_{IJ}. It is therefore convenient to treat each four-cycle as an observational unit and impose conditions on the sequence of matching networks that guarantee the number of distinct four-cycles grows with I and J.

4.2.1 Estimator for β_0

Index the four-cycles in G_{IJ} by $\ell = 1, \ldots, L$. For clarity of exposition, I restrict attention to edge-disjoint cycles, assuming that no cycles share an edge (but they can share one or two nodes). This restriction is not essential, and information from overlapping cycles can also be aggregated, but focusing on edge-disjoint cycles keeps the notation tractable.

Each cycle ℓ consists of two distinct workers $\{i_\ell, i'_\ell\}$ and two distinct firms $\{j_\ell, j'_\ell\}$. For expositional purposes, suppose labels are ordered so that $\alpha_{i_\ell} \geq \alpha_{i'_\ell}$ and $\psi_{j_\ell} \geq \psi_{j'_\ell}$. Of course, these labels are unknown to the researcher: they only observe the pairs in the cycle, not the underlying productivities.

To work with the formulas below, the researcher must nonetheless assign distinct labels to workers and firms in each cycle. At this stage, I leave the rule for assigning labels unspecified; later, I propose a procedure that uniquely determines them.

Formally, the researcher assigns labels $(i_{\ell,\pi_\ell}, i'_{\ell,\pi_\ell})$ and $(j_{\ell,\pi_\ell}, j'_{\ell,\pi_\ell})$ to the worker and firm pairs

$\{i_\ell, i'_\ell\}$ and $\{j_\ell, j'_\ell\}$. Given this labeling, define

$$\pi_\ell \coloneqq \pi_\ell^\alpha \pi_\ell^\psi,$$

where

$$\pi_\ell^\alpha := \begin{cases} 1, & \text{if } (i_{\ell,\pi_\ell}, i'_{\ell,\pi_\ell}) = (i_\ell, i'_\ell), \\ -1, & \text{if } (i_{\ell,\pi_\ell}, i'_{\ell,\pi_\ell}) = (i'_\ell, i_\ell), \end{cases} \qquad \pi_\ell^\psi := \begin{cases} 1, & \text{if } (j_{\ell,\pi_\ell}, j'_{\ell,\pi_\ell}) = (j_\ell, j'_\ell), \\ -1, & \text{if } (j_{\ell,\pi_\ell}, j'_{\ell,\pi_\ell}) = (j'_\ell, j_\ell). \end{cases}$$

In words, π_ℓ^α (π_ℓ^ψ) equals 1 when the assigned label i_{ℓ,π_ℓ} (j_{ℓ,π_ℓ}) corresponds to the higher-productivity worker (firm), and -1 otherwise. Because productivities are unobserved, π_ℓ cannot be directly chosen, but each labeling rule uniquely determines its value. For this reason, I refer to the label assignment chosen by the researcher simply as π_ℓ.

For each cycle ℓ with labeling π_ℓ, define

$$\hat{\Delta}_{1,\ell,\pi_\ell} := y_{i_{\ell,\pi_\ell} j_{\ell,\pi_\ell}} - y_{i'_{\ell,\pi_\ell} j_{\ell,\pi_\ell}} - y_{i_{\ell,\pi_\ell} j'_{\ell,\pi_\ell}} + y_{i'_{\ell,\pi_\ell} j'_{\ell,\pi_\ell}},$$

$$\hat{\Delta}_{2,\ell,\pi_\ell} := y_{i_{\ell,\pi_\ell} j_{\ell,\pi_\ell}} y_{i'_{\ell,\pi_\ell} j'_{\ell,\pi_\ell}} - y_{i'_{\ell,\pi_\ell} j_{\ell,\pi_\ell}} y_{i_{\ell,\pi_\ell} j'_{\ell,\pi_\ell}}.$$

Intuitively, $\hat{\Delta}_{1,\ell,\pi_\ell}$ compares the difference in outcomes between workers across firms, while $\hat{\Delta}_{2,\ell,\pi_\ell}$ contrasts two cross-products, each involving all four nodes in the cycle. Both statistics depend on how workers and firms are labeled within the cycle, as shown by the following example.

Example 1. (Computing $\hat{\Delta}_{1,\ell,\pi_\ell}$ and $\hat{\Delta}_{2,\ell,\pi_\ell}$). *Cycle ℓ includes workers Alice and Bob, and firms Canon and Dell. Alice earns 120 with Canon and 100 with Dell, while Bob earns 100 with Canon and 90 with Dell.*

To compute $\hat{\Delta}_{1,\ell,\pi_\ell}$ and $\hat{\Delta}_{2,\ell,\pi_\ell}$, the researcher must first assign labels $(i_{\ell,\pi_\ell}, i'_{\ell,\pi_\ell})$ to Alice and Bob, and $(j_{\ell,\pi_\ell}, j'_{\ell,\pi_\ell})$ to Canon and Dell. Each of the four possible assignments induces a value for π_ℓ and determines the outcomes $y_{i_{\ell,\pi_\ell} j_{\ell,\pi_\ell}}$, $y_{i'_{\ell,\pi_\ell} j_{\ell,\pi_\ell}}$, $y_{i_{\ell,\pi_\ell} j'_{\ell,\pi_\ell}}$, and $y_{i'_{\ell,\pi_\ell} j'_{\ell,\pi_\ell}}$, and thus the corresponding values of $\hat{\Delta}_{1,\ell,\pi_\ell}$ and $\hat{\Delta}_{2,\ell,\pi_\ell}$. The rows in the following table summarize each assignment:

i_{ℓ,π_ℓ}	i'_{ℓ,π_ℓ}	j_{ℓ,π_ℓ}	j'_{ℓ,π_ℓ}	$y_{i_{\ell,\pi_\ell} j_{\ell,\pi_\ell}}$	$y_{i'_{\ell,\pi_\ell} j_{\ell,\pi_\ell}}$	$y_{i_{\ell,\pi_\ell} j'_{\ell,\pi_\ell}}$	$y_{i'_{\ell,\pi_\ell} j'_{\ell,\pi_\ell}}$	$\hat{\Delta}_{1,\ell,\pi_\ell}$	$\hat{\Delta}_{2,\ell,\pi_\ell}$	π_ℓ
Alice	Bob	Canon	Dell	120	100	100	90	10	800	1
Alice	Bob	Dell	Canon	100	90	120	100	-10	-800	-1
Bob	Alice	Canon	Dell	100	120	90	100	-10	-800	-1
Bob	Alice	Dell	Canon	90	100	100	120	10	800	1

Although there are four possible labelings, they generate only two distinct pairs $(\hat{\Delta}_{1,\ell,\pi_\ell}, \hat{\Delta}_{2,\ell,\pi_\ell})$, corresponding to the two possible values of π_ℓ (which remains unknown to the researcher).

With these statistics in hand, define the estimator

$$\hat{\beta}_{L,\pi} \coloneqq -\frac{\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{1,\ell,\pi_\ell}}{\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi_\ell}}.$$

The estimator $\hat{\beta}_{L,\pi}$ does not require estimating $\boldsymbol{\alpha}$ or $\boldsymbol{\psi}$. As a result, β_0 can be estimated consistently even when individual productivities cannot be, for example when each worker is observed only a few times; see Remark 3 for details.

The estimator depends on the labelings $(\pi_1, \ldots, \pi_L)$. In a network with L cycles, different labeling combinations can generate up to 2^L distinct estimators of β_0. Such dependence on arbitrary labeling is undesirable, since two researchers analyzing the same data could obtain different estimates solely because they chose different labelings. To remove this ambiguity, I later introduce a procedure that selects a particular combination of labelings.

To see why $\hat{\beta}_{L,\pi}$ is an estimator for β_0, decompose $\hat{\Delta}_{1,\ell,\pi_\ell}$ and $\hat{\Delta}_{2,\ell,\pi_\ell}$ as:

$$\begin{aligned}\hat{\Delta}_{1,\ell,\pi_\ell} &= \Delta_{1,\ell,\pi_\ell} + \epsilon_{\Delta_1,\ell,\pi_\ell},\\ \hat{\Delta}_{2,\ell,\pi_\ell} &= \Delta_{2,\ell,\pi_\ell} + \epsilon_{\Delta_2,\ell,\pi_\ell},\end{aligned}$$

where

$$\begin{aligned}\Delta_{1,\ell,\pi_\ell} &= \beta_0\,(\alpha_{i_\ell} - \alpha_{i'_\ell})(\psi_{j_\ell} - \psi_{j'_\ell})\,\pi_\ell,\\ \Delta_{2,\ell,\pi_\ell} &= -(\alpha_{i_\ell} - \alpha_{i'_\ell})(\psi_{j_\ell} - \psi_{j'_\ell})\,\pi_\ell,\end{aligned}$$

and

$$\epsilon_{\Delta_1,\ell,\pi_\ell} = \big(\eta_{i_\ell j_\ell} - \eta_{i'_\ell j_\ell} - \eta_{i_\ell j'_\ell} + \eta_{i'_\ell j'_\ell}\big)\pi_\ell, \tag{4.1}$$

$$\epsilon_{\Delta_2,\ell,\pi_\ell} = \big(\theta_{i_\ell j_\ell}\eta_{i'_\ell j'_\ell} + \theta_{i'_\ell j'_\ell}\eta_{i_\ell j_\ell} - \theta_{i'_\ell j_\ell}\eta_{i_\ell j'_\ell} - \theta_{i_\ell j'_\ell}\eta_{i'_\ell j_\ell} + \eta_{i_\ell j_\ell}\eta_{i'_\ell j'_\ell} - \eta_{i'_\ell j_\ell}\eta_{i_\ell j'_\ell}\big)\pi_\ell. \tag{4.2}$$

Here, the absolute values of Δ_{1,ℓ,π_ℓ} and Δ_{2,ℓ,π_ℓ} depend only on the model parameters, while their sign is determined by the labeling π_ℓ. The random terms $\epsilon_{\Delta_1,\ell,\pi_\ell}$ and $\epsilon_{\Delta_2,\ell,\pi_\ell}$ are mean-zero combinations of the four errors η_{ij}, with their sign again determined solely by π_ℓ. This decomposition highlights why averaging across many cycles reduces sampling variability: the error terms $\{\eta_{ij}\}$ are independent and mean zero across edges. Taking the ratio of these averages then cancels the common dependence on latent productivities $(\boldsymbol{\alpha}, \boldsymbol{\psi})$ and on labelings, leaving only the parameter β_0.

The next section formalizes this intuition and establishes that, under suitable conditions on the sequence of graphs, error terms, and labelings, $\hat{\beta}_{L,\pi}$ converges almost surely to β_0.

4.2.2 Consistency

To establish the consistency of $\hat{\beta}_{L,\pi}$, I require the following conditions.

Assumption 5.

5.1 (Error Regularity) *The error terms η_{ij} are independent across all (i, j) and satisfy $\mathbb{E}[\eta_{ij}] = 0$, $\mathrm{Var}(\eta_{ij}) \geq C_\eta > 0$, and $\mathbb{E}[|\eta_{ij}|^{2+\delta}] \leq M < \infty$ for some $\delta > 0$ and uniform bounds C_η and M.*

5.2 (Cycles Growth) *The sequence of matching networks G_{IJ} is such that, as $I \longrightarrow \infty$ and $J \longrightarrow \infty$, $L \longrightarrow \infty$.*

5.3 (Cycles Heterogeneity) *The sequences of latent productivities $\boldsymbol{\alpha}_I$ and $\boldsymbol{\psi}_J$, and the sequence of matching networks G_{IJ} are such that*

$$\mu_L = \frac{1}{L}\sum_{\ell=1}^{L}(\alpha_{i_\ell} - \alpha_{i'_\ell})(\psi_{j_\ell} - \psi_{j'_\ell})$$

is bounded away from zero ($\mu_L > C_\mu > 0$).

5.4 (Labeling Regularity) *The sequence of cycle labelings $\{\pi_\ell\}_{\ell=1}^{L}$ satisfies*

$$\mu_L c_\pi - \frac{1}{L}\sum_{\ell=1}^{L}(\alpha_{i_\ell} - \alpha_{i'_\ell})(\psi_{j_\ell} - \psi_{j'_\ell})\,\pi_\ell \xrightarrow{a.s.} 0,$$

for some constant c_π with $|c_\pi| > C_\pi > 0$. Moreover, the labeling signs $\{\pi_\ell\}$ are independent of the error terms $\{\eta_{ij}\}$.

Assumption 5.1 imposes standard regularity conditions on the error terms. Importantly, it does not require the η_{ij} to be identically distributed and allows for heteroskedasticity: each error may follow its own distribution, provided it has mean zero, is non-degenerate, and admits $(2+\delta)$ moments. While the non-degeneracy and moment conditions are not strictly necessary for consistency, they are needed for deriving the asymptotic distribution of the estimator in subsequent results.

Assumption 5.2 concerns the sequence of matching networks and requires that the number of cycles grows with the overall size of the network. The condition does not restrict node degrees and is compatible with graphs G_{IJ} where degrees remain bounded: for example, no worker or firm needs to appear in more than two matches. Since $\hat{\beta}_{L,\pi}$ treats cycles as the fundamental units of observation, the assumption guarantees that the number of such units increases sufficiently to justify asymptotic analysis.

Assumption 5.3 requires that the average, across cycles, of the products of worker and firm productivity differences remains bounded away from zero. Since this condition is formulated using ordered labels, it does not depend on the specific choice of labeling. It is closely related to Assumption 1, which underlies identification, and guarantees the presence of systematic heterogeneity in productivities across cycles. Put differently, because μ_L is an average of nonnegative terms, requiring μ_L to stay strictly positive implies that a non-vanishing fraction of the summands must themselves be bounded away from zero. This excludes degenerate cases in which heterogeneity vanishes asymptotically, requiring enough variation across workers and firms to make the cycles informative.

Assumption 5.4 restricts the sequence of labelings. It rules out labeling schemes that drive the average of Δ_{2,ℓ,π_ℓ} toward zero, since in that case both the numerator and denominator of the estimator would vanish, making β_0 possibly unrecoverable. It also excludes labelings that are systematically correlated with the error terms, as such dependence would invalidate the averaging argument underpinning the law of large numbers. The assumption does not prescribe a unique labeling rule; rather, it specifies the conditions that any labeling must satisfy to be admissible, allowing for both deterministic and stochastic labeling rules. In Section 4.2.4, I present one such rule, based on observable instruments, which ensures the condition and provides a practically implementable estimator.

Strong consistency of $\hat{\beta}_{L,\pi}$ for β_0 then follows from the strong law of large numbers, as stated in the following theorem.

Theorem 5. (Strong Consistency) *Under Assumption 5, as $I \to \infty$ and $J \to \infty$, the estimator $\hat{\beta}_{L,\pi}$ is strongly consistent for β_0, i.e., $\hat{\beta}_{L,\pi} \to^{a.s.} \beta_0$.*

The assumptions on G_{IJ} required for consistency of $\hat{\beta}_L$ are relatively weak. For instance, under the bipartite Erdős–Rényi random graph model, Assumption 5.2 holds whenever the link probability p_{IJ} satisfies $\sqrt{IJ}p_{IJ} \to \infty$ (see Appendix D for the formal definition of the model and the derivation of this threshold). By contrast, ensuring that the graph is connected requires the stronger condition $\frac{\sqrt{IJ}p_{IJ}}{\log(\sqrt{IJ})} > 1$. Hence, under the Erdős–Rényi model, the condition on the matching network needed for identification of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ automatically guarantees the one required for consistency of $\hat{\beta}_{L,\pi}$.

The effective sample size for estimating β_0 is L, the number of cycles, rather than the number of observed matches. This parallels other settings in econometrics: in local linear regression, the effective sample size is proportional to the number of observations times the bandwidth, while in clustered data it is given by the number of clusters rather than the number of units.

In labor market applications, Assumption 5.2 is satisfied whenever the number of cycles grows proportionally with the number of workers and firms, even if the graph consists of many disconnected local subgraphs. For instance, in the empirical setting of Kline (2024), the data contain

roughly 750,000 workers, 70,000 firms, and 5,000 cycles: the number of cycles is large enough that asymptotic approximations are likely to provide a good guide to finite-sample behavior.

The next step is to characterize the asymptotic distribution of $\hat{\beta}_{L,\pi}$, which provides the basis for valid inference and for testing the absence of complementarities.

Remark 3. (No Need to Estimate Productivities) *A key feature of $\hat{\beta}_{L,\pi}$ is that it achieves consistency without requiring estimation of the productivity components $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$, which would demand many observations of each worker and firm. This is in contrast to more standard approaches, such as the iterative least squares procedure of Bai (2009), which estimates β_0 jointly with unit-specific effects. In that setting, the asymptotic behavior of the estimator for β_0 depends on the properties of the estimators for $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$. By avoiding their estimation altogether, and directly isolating β_0, the cycle-based estimator sidesteps this difficulty.*

4.2.3 Asymptotic Distribution

The estimator $\hat{\beta}_{L,\pi}$ is a ratio of averages, which makes its asymptotic behavior amenable to analysis via the Lyapunov Central Limit Theorem. To set up the argument, define

$$u_{\ell,\pi_\ell} \coloneqq \epsilon_{\Delta_1,\ell,\pi_\ell} + \beta_0 \epsilon_{\Delta_2,\ell,\pi_\ell},$$

with $\epsilon_{\Delta_1,\ell,\pi_\ell}$ and $\epsilon_{\Delta_2,\ell,\pi_\ell}$ defined in Equations 4.1 and 4.2. The mean-zero composite error u_{ℓ,π_ℓ} depends on β_0, the four productivity terms, the labeling, and the four error terms associated with cycle ℓ.

With this notation in place, the asymptotic distribution of $\hat{\beta}_{L,\pi}$ can be derived.

Theorem 6. (Asymptotic Normality) *Under Assumption 5, as $I \to \infty$ and $J \to \infty$,*

$$\frac{\sqrt{L}(\hat{\beta}_{L,\pi} - \beta_0)}{\frac{\sigma_{u,L}}{\mu_L c_\pi}} \xrightarrow{d} \mathcal{N}\Big(0, 1\Big),$$

where $\sigma_{u,L} \coloneqq \sqrt{\frac{1}{L}\sum_{\ell=1}^{L} \mathrm{Var}(u_{\ell,\pi_\ell})}$ is the square root of the average variance of the composite errors $\{u_{\ell,\pi_\ell}\}$ across cycles, and μ_L and c_π are as defined in Assumptions 5.3 and 5.4.

The estimator $\hat{\beta}_{L,\pi}$ converges at rate $\sqrt{L}$, where L, the number of cycles, acts as the effective sample size. In a complete bipartite graph, this corresponds to the rate $\sqrt{IJ}$, since the edges can be partitioned into $\frac{I}{2} \times \frac{J}{2}$ edge-disjoint four-cycles. An important implication is that $\hat{\beta}_{L,\pi}$ remains consistent even when one dimension, either I or J, is held fixed while the other grows.

In addition to the sample size, three scaling terms appear in the asymptotic distribution. The first, $\sigma_{u,L}$, is the square root of the average variance of the composite error terms $u_{\ell,\pi}$. Although

these errors depend on the chosen labeling, their variances do not; hence $\sigma_{u,L}$ is invariant to labeling. It reflects only the variability of the underlying noise terms η_{ij}: greater noise in y_{ij} increases $\sigma_{u,L}$ and reduces estimator precision.

The second, μ_L, summarizes the heterogeneity in worker and firm productivities across cycles. If workers and firms within cycles are too similar, μ_L is small and the estimator becomes imprecise. By contrast, stronger heterogeneity makes μ_L larger, yielding sharper estimates.

The third, c_π, isolates the effect of labeling choice. It shows explicitly how different labeling rules can affect the variance of the estimator, with some labelings increasing its precision relative to others.

Feasible Scaling Factor Estimation. To make the asymptotic normality result operational for inference, the scaling factor $\frac{\sigma_{u,L}}{\mu_L c_\pi}$ must be estimated. A consistent estimator is $\frac{\sqrt{\hat{\sigma}^2_{u,L,\pi}}}{-\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi}}$, where $-\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi}$ consistently estimates $\mu_L c_\pi$ (as established in the consistency proof), and $\hat{\sigma}^2_{u,L,\pi}$ estimates $\sigma^2_{u,L}$. A convenient choice for $\hat{\sigma}^2_{u,L,\pi}$ is the average of the squares of the estimated residuals $\hat{u}_{\ell,\pi}$:

$$\hat{\sigma}^2_{u,L,\pi} \coloneqq \frac{1}{L}\sum_{\ell=1}^{L}\hat{u}^2_{\ell,\pi} = \frac{1}{L}\sum_{\ell=1}^{L}\left(\hat{\Delta}_{1,\ell,\pi} + \hat{\beta}_{L,\pi}\hat{\Delta}_{2,\ell,\pi}\right)^2,$$

whose consistency is established in the following proposition.

Proposition 2. (Consistent Variance Estimator) *Under Assumption 5, as $I \to \infty$ and $J \to \infty$, the estimator $\hat{\sigma}^2_{u,L,\pi}$ is consistent for $\sigma^2_{u,L}$*

$$\hat{\sigma}^2_{u,L,\pi} - \sigma^2_{u,L} \xrightarrow{p} 0.$$

Theorem 6, together with these feasible estimators for the scaling terms, yields valid asymptotic confidence intervals for β_0. The next section introduces a practical procedure for selecting labelings in each cycle, ensuring Assumption 5.4 holds, uniquely defining the estimator, and guaranteeing the validity of inference.

4.2.4 Rank-Based Labeling

Assumption 5.4 restricts how labels can be assigned within each cycle but does not prescribe a specific rule. In this section, I propose a rank-based procedure for assigning labels $(i_{\ell,\pi_\ell}, i'_{\ell,\pi_\ell})$ and $(j_{\ell,\pi_\ell}, j'_{\ell,\pi_\ell})$ in each cycle, and show that it satisfies Assumption 5.4. As a result, the estimator with this rank-based labeling is consistent and asymptotically normal, and not dependent on arbitrary labeling choices.

Suppose the researcher observes some characteristics of workers and firms, denoted by $\boldsymbol{z}^{\boldsymbol{\alpha}} = (z_1^{\alpha}, \ldots, z_I^{\alpha})$ and $\boldsymbol{z}^{\boldsymbol{\psi}} = (z_1^{\psi}, \ldots, z_J^{\psi})$, and treated as nonrandom. Labels in each cycle are then assigned using these instruments according to the following rule.

Definition 5. (Rank-Based Labeling) *In each cycle ℓ, the labels $(i_{\ell,\pi_{\ell,z}}, i'_{\ell,\pi_{\ell,z}})$ and $(j_{\ell,\pi_{\ell,z}}, j'_{\ell,\pi_{\ell,z}})$ are assigned so that*

$$z^{\alpha}_{i_{\ell,\pi_{\ell,z}}} > z^{\alpha}_{i'_{\ell,\pi_{\ell,z}}} \quad \textit{and} \quad z^{\psi}_{j_{\ell,\pi_{\ell,z}}} > z^{\psi}_{j'_{\ell,\pi_{\ell,z}}}.$$

Under this rule, the worker and firm with larger instrument values are always labeled $i_{\ell,\pi_{\ell,z}}$ and $j_{\ell,\pi_{\ell,z}}$, respectively. This induces the signs

$$\pi^{\alpha}_{\ell,z} = \operatorname{sign}\left(z^{\alpha}_{i_\ell} - z^{\alpha}_{i'_\ell}\right), \quad \pi^{\psi}_{\ell,z} = \operatorname{sign}\left(z^{\psi}_{j_\ell} - z^{\psi}_{j'_\ell}\right).$$

For expositional clarity, I exclude the possibility of ties in the instruments; when ties occur, they can be resolved at random.

Let $\hat{\beta}_{L,z}$ denote the Rank-Based Labeling (RBL) estimator, defined using the rank-based labeling:

$$\hat{\beta}_{L,z} \coloneqq -\frac{\frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{1,\ell,\pi_{\ell,z}}}{\frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{2,\ell,\pi_{\ell,z}}}. \qquad \text{(RBL estimator)}$$

The properties of $\hat{\beta}_{L,z}$ depend on the choice of instruments $\boldsymbol{z}^{\alpha}$ and $\boldsymbol{z}^{\psi}$.

Consider first the oracle case in which the instruments available to the researcher coincide exactly with the latent productivities $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$. In this case, the induced labelings satisfy $\pi_{\ell,z} = 1$ for all ℓ, so that Assumption 5.4 holds with $c_\pi = 1$. Such oracle instruments are, of course, infeasible in practice, but Assumption 5.4 does not require $c_\pi = 1$, just that it remains bounded away from zero. Intuitively, this occurs whenever the instruments correctly rank the higher-productivity worker and firm more often than not.

In applications, instruments should therefore be observable characteristics plausibly associated with latent productivities. For instance, years of schooling and firm size can serve as instruments for $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$, as they are often strongly correlated with worker and firm productivity. By contrast, outcomes themselves, though mechanically related to productivities, violate the exogeneity requirement, since they also depend on the error terms $\{\eta_{ij}\}$. This means that, as shown in Appendix E, using outcomes as instruments leads to biased estimates and invalid inference.

The next example provides a practical illustration of how labels are assigned.

Example 2. (Rank-Based Labeling in Practice). *Consider two cycles. The first involves workers Alice and Bob and firms Canon and Dell; the second involves workers Elizabeth and Fred and firms*

General Motors and Honda. To assign labels in each cycle, use years of schooling as an instrument for workers, and the number of employees as an instrument for firms. The instrument values are:

Worker	z_i^α
Alice	16
Bob	14
Elizabeth	18
Fred	12

Firm	z_j^ψ
Canon	170,000
Dell	110,000
General Motors	160,000
Honda	190,000

Assign labels in each cycle according to the rank-based labeling, so that the worker with higher schooling receives label $i_{\ell,\pi_{\ell,z}}$ and the firm with more employees receives label $j_{\ell,\pi_{\ell,z}}$. The resulting labels and the corresponding cycle statistics are:

$i_{\ell,\pi_{\ell,z}}$	$i'_{\ell,\pi_{\ell,z}}$	$j_{\ell,\pi_{\ell,z}}$	$j'_{\ell,\pi_{\ell,z}}$	$z^\alpha_{i_{\ell,\pi_{\ell,z}}}$	$z^\alpha_{i'_{\ell,\pi_{\ell,z}}}$	$z^\psi_{j_{\ell,\pi_{\ell,z}}}$	$z^\psi_{j'_{\ell,\pi_{\ell,z}}}$	$\hat{\Delta}_{1,\ell,\pi_{\ell,z}}$	$\hat{\Delta}_{2,\ell,\pi_{\ell,z}}$
Alice	Bob	Canon	Dell	16	14	170k	110k	10	800
Elizabeth	Fred	Honda	General Motors	18	12	190k	160k	20	900

This table, where each row represents a distinct cycle, is the analysis dataset: the estimator $\hat{\beta}_{L,z}$ is the ratio of the averages of the last two columns.

To formalize the conditions that instruments must satisfy in addition to exogeneity, define the averages

$$\overline{\pi^\alpha} \coloneqq \frac{1}{L}\sum_{\ell=1}^{L} \pi^\alpha_{\ell,z}, \qquad \overline{\pi^\psi} \coloneqq \frac{1}{L}\sum_{\ell=1}^{L} \pi^\psi_{\ell,z}, \qquad \overline{\pi^\alpha \pi^\psi} \coloneqq \frac{1}{L}\sum_{\ell=1}^{L} \pi^\alpha_{\ell,z}\pi^\psi_{\ell,z},$$

and impose the following conditions on the sequences $\{\pi^\alpha_{\ell,z}\}$ and $\{\pi^\psi_{\ell,z}\}$.

Assumption 6.

6.1 (Relevance) *Instruments correctly rank workers and firms sufficiently often: there exist constants $c_\alpha, c_\psi > 0$ such that, for all sufficiently large L,*

$$\overline{\pi^\alpha} \geq c_\alpha, \qquad \overline{\pi^\psi} \geq c_\psi.$$

6.2 (No Negative Association) *Worker and firm instruments are not systematically opposed:*

$$\frac{1}{L}\sum_{\ell=1}^{L} (\pi^\alpha_{\ell,z} - \overline{\pi^\alpha})\,(\pi^\psi_{\ell,z} - \overline{\pi^\psi}) \quad \geq 0.$$

6.3 (No Large Gap Penalty) *Instrument ranks are not systematically misaligned with the magnitude of productivity differences:*

$$\frac{1}{L}\sum_{\ell=1}^{L}\Big((\alpha_{i_\ell}-\alpha_{i'_\ell})(\psi_{j_\ell}-\psi_{j'_\ell})-\mu_L\Big)\Big(\pi^{\alpha}_{\ell,z}\pi^{\psi}_{\ell,z}-\overline{\pi^{\alpha}\pi^{\psi}}\Big)\quad\geq 0.$$

Assumption 6.1 requires that the instruments contain useful information about the latent productivities: the rankings they induce must align with the true rankings often enough. Perfect accuracy is not necessary, and occasional misorderings are allowed, as long as the instruments select the correct ordering in a sufficiently large fraction of cycles.

Assumption 6.2 rules out a systematic negative association between worker-side and firm-side labelings. That is, it excludes the case in which the worker instrument tends to misorder exactly when the firm instrument orders correctly (or vice versa). The restriction applies only on average: isolated instances of such behavior are admissible provided they do not dominate.

Assumption 6.3 prevents a systematic association between large productivity gaps and mislabeling. Specifically, it rules out the possibility that cycles with large differences in worker-firm productivities are disproportionately associated with incorrect rankings. This condition is mild: in practice, misclassifications are more likely when productivity gaps are small, not when they are large.

The next proposition establishes that, when the instruments satisfy Assumption 6, the rank-based labeling $\pi_{\ell,z}$ satisfies the condition required by Assumption 5.4. Consequently, the RBL estimator $\hat{\beta}_{L,z}$ achieves the asymptotic properties established in Theorems 5 and 6.

Proposition 3. (Rank-Based Labeling Validity) *Under Assumptions 5.1, 5.2, 5.3, and 6, the rank-based labeling $\pi_{\ell,z}$ satisfies Assumption 5.4.*

Proposition 3 ensures that the rank-based labeling delivers a well-defined estimator $\hat{\beta}_{L,z}$ with the aforementioned asymptotic properties. This result allows $\hat{\beta}_{L,z}$ to serve not only as an estimator of the interaction parameter but also as the basis for the construction of a formal test for the absence of complementarities and hence for modularity of the interaction function in the BI framework. The next section develops this test and studies its properties.

4.2.5 Test for Absence of Complementarities

The TWFE model is nested within the Tukey model as the special case $\beta_0 = 0$. Hence, the asymptotic distribution of the RBL estimator $\hat{\beta}_{L,z}$ can be used to test the null hypothesis $H_0 : \beta_0 = 0$. Rejecting H_0 not only rejects the TWFE specification, but also rejects modularity of the interaction function f, since any modular function can be written in the additive form assumed by the

TWFE model (see Appendix A). To the best of my knowledge, no formal test of modularity has previously been available in the BI framework, even though empirical discussions often revolve around whether complementarities are present.

Testing $H_0 : \beta_0 = 0$ is the central focus of Tukey (1949), although no estimator for β_0 is proposed there. Relying on restrictive assumptions (homoskedastic normally distributed errors η_{ij} and complete matching network G_{IJ}), Tukey's procedure first estimates the additive effects $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ under the null through two-way fixed effects regression, and then test whether including the interaction term $\alpha_i \psi_j$ significantly improves model fit, for example through regression-based tests or changes in R^2.

Similar heuristic approaches have been adopted in the TWFE literature (e.g., Card et al. (2013); Fenizia (2022)), but the sparsity of the matching network G_{IJ} makes consistent estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ infeasible, so that any procedure based on regression residuals seem to be justifiable only under strong assumptions on the error terms and the matching network. By contrast, the estimator $\hat{\beta}_{L,\pi}$ remains consistent even under heteroskedasticity and in sparse matching structures.

Test Statistic. Consider the t-statistic

$$\hat{T}_{L,z} \coloneqq \frac{\sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{1,\ell,\pi_{\ell,z}}}{\frac{\hat{\sigma}_{u,L,z}}{\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi_{\ell,z}}}} = \frac{\sum_{\ell=1}^{L}\hat{\Delta}_{1,\ell,\pi_{\ell,z}}}{\sqrt{\sum_{\ell=1}^{L}\left(\hat{\Delta}_{1,\ell,\pi_{\ell,z}} - \frac{\sum_{\ell=1}^{L}\hat{\Delta}_{1,\ell,\pi_{\ell,z}}}{\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi_{\ell,z}}}\hat{\Delta}_{2,\ell,\pi_{\ell,z}}\right)^2}},$$

and define the test $\phi_{L,z}$ with size γ that rejects the null according to

$$\phi_{L,z}(\hat{T}_{L,z}, \gamma) = \mathbf{1}\left\{|\hat{T}_{L,z}| \geq c_{\gamma/2}\right\},$$

where $c_{\gamma/2}$ is the $\gamma/2$ quantile of the standard normal distribution. The asymptotic validity and consistency of this test follow directly from the asymptotic normality result in Theorem 6, as summarized in the following corollary.

Corollary 1. (Test for Modularity) *Under Assumptions 5 and 6, as $I \to \infty$ and $J \to \infty$, if the interaction function f is modular, and hence the null hypothesis $H_0 : \beta_0 = 0$ is true, the test $\phi_{L,z}(\hat{T}_{L,z}, \gamma)$ is asymptotically valid:*

$$\lim_{L\to\infty} \mathbb{E}[\phi_{L,z}(\hat{T}_{L,z}, \gamma)] = \gamma.$$

Under the Tukey model, when $\beta_0 \neq 0$ and hence the interaction function is not modular, the test is consistent:

$$\lim_{L\to\infty} \mathbb{E}[\phi_{L,z}(\hat{T}_{L,z}, \gamma)] = 1.$$

The test $\phi_{L,z}$ controls asymptotic size for any modular function, but Theorem 6 guarantees its consistency only under correct specification of the Tukey model. In particular, a non-modular function f can admit a representation with $\beta_0 = 0$, in which case the test has no power. Thus, $\phi_{L,z}$ tests a necessary but not sufficient condition for modularity. The situation is analogous to testing independence between two variables using the correlation coefficient: while a nonzero correlation implies dependence, a correlation of zero does not rule out dependence. In practical terms, rejection of the null provides strong evidence against modularity, but failure to reject should be interpreted with caution, as it does not imply that the interaction function is modular.

4.3 Productivities $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$

Theorem 2 shows that, once β_0 is known, connectedness of the matching graph guarantees identification of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$. Two cases arise. When $\beta_0 = 0$, the model reduces to the standard TWFE specification, and existing estimators can be applied. I briefly review the available results and highlight how they rely on strong conditions that are rarely satisfied in labor market data. When $\beta_0 \neq 0$, I propose a method that uses the consistent estimator $\hat{\beta}_{L,\pi}$ as input for estimating the productivity parameters. A full analysis of its properties is left for future work. As with the TWFE model, this approach ultimately requires denser graphs than are typically observed in applications, reflecting a shared limitation between the two cases.

4.3.1 TWFE model: $\beta_0 = 0$

When $\beta_0 = 0$, the Tukey model reduces to:

$$y_{ij} = \alpha_i + \psi_j + \eta_{ij},$$

and the productivity components can be estimated using the TWFE estimator:

$$(\hat{\boldsymbol{\alpha}}^{twfe}, \hat{\boldsymbol{\psi}}^{twfe}) = (C'C)^{-1}C'\boldsymbol{y}_O,$$

where the design matrix C is defined as in Definition 8.

Jochmans and Weidner (2019) study the asymptotic properties of the TWFE estimator. For inference on a single productivity value, they derive its asymptotic distribution under the condition

that the degree of the corresponding node diverges. More generally, inference on functionals of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ requires that the degrees of many nodes grow. As the authors emphasize, this setting is far from typical labor market applications, where node degrees usually remain bounded: the number of workers per firm or firms per worker does not increase proportionally with I and J.

Kline et al. (2020) also study inference for functionals of the productivities, focusing on quadratic forms such as variances and covariances. Their asymptotic setting requires the matching network G_{IJ} to grow in a uniformly connected manner, without fragmenting into weakly linked subgraphs. As their Table IV shows, however, this condition is rarely satisfied in labor market data, where matching networks often exhibit considerable fragmentation.

These results highlight that in most labor market applications, and indeed in other two-sided settings as well, the network is too sparse to justify asymptotic arguments for productivity estimation. This is not surprising: information on individual productivity depends only on the relatively few edges involving that node, unlike global parameters such as β_0, which influence all observed matches.

4.3.2 Case with complementarities: $\beta_0 \neq 0$

The estimator $\hat{\beta}_{L,\pi}$ can be used to construct least-squares estimators for $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ when $\beta_0 \neq 0$. I outline a simple procedure for doing so, which turns out to involve the same computational problem as estimating interactive fixed effects.

Starting from the Tukey model in Equation (4.2), multiply both sides by β_0 and add 1:

$$1 + \beta_0 y_{ij} = (1 + \beta_0 \alpha_i)(1 + \beta_0 \psi_j) + \beta_0 \eta_{ij}.$$

Now substitute β_0 with its consistent estimator $\hat{\beta}_{L,\pi}$ and define:

$$y'_{ij} = 1 + \hat{\beta}_{L,\pi} y_{ij}, \quad \alpha'_i = 1 + \hat{\beta}_{L,\pi} \alpha_i, \quad \psi'_j = 1 + \hat{\beta}_{L,\pi} \psi_j, \quad \eta'_{ij} = \hat{\beta}_{L,\pi} \eta_{ij}.$$

From these transformed variables, estimates for the original productivity terms can be recovered from estimates of $\boldsymbol{\alpha}'$ and $\boldsymbol{\psi}'$. Since

$$y'_{ij} = \alpha'_i \psi'_j + \eta'_{ij},$$

a natural approach to estimate $\boldsymbol{\alpha}'$ and $\boldsymbol{\psi}'$ is to solve the least-squares problem:

$$(\hat{\boldsymbol{\alpha}}', \hat{\boldsymbol{\psi}}') \in \arg\min_{\boldsymbol{\alpha}', \boldsymbol{\psi}'} \sum_{(i,j) \in O_{IJ}} \left(y'_{ij} - \alpha'_i \psi'_j\right)^2 \quad \text{s.t.} \quad \|\boldsymbol{\psi}'\| = 1.$$

This is the interactive fixed-effects problem, and can be solved via alternating least squares; see

Appendix F for implementation details.

A full analysis of the statistical properties of the productivity estimators derived from $(\hat{\boldsymbol{\alpha}}', \hat{\boldsymbol{\psi}}')$ lies beyond the scope of this paper and is left for future research. For this reason, in the next section, when I illustrate how the Tukey model can be applied in practice, the focus will be on β_0, the new parameter introduced by the Tukey specification, for which I developed an estimator with formally studied asymptotic properties.

5 Empirical Illustration

In this section, I revisit the application in Limodio (2021)[1] to illustrate how to estimate the interaction parameter in the Tukey model and how the resulting estimates can be used to draw additional insights about the two-sided interaction.

Limodio (2021) studies the interaction between managers and tasks in the public sector, focusing on the allocation of World Bank bureaucrats (hereafter, managers) to development projects in low- and middle-income countries. Each manager is responsible for designing, supervising, and overseeing project implementation. Project success is measured using ratings from the World Bank's Independent Evaluation Group, which assess the extent to which key objectives were achieved. In the original analysis, project success is modeled as a function of a manager-specific ability α_i and a country-specific characteristic ψ_j, using the TWFE model. An administrative dataset records manager-country assignments over time, along with project-level characteristics and evaluations, making it possible to construct the matching network and observe the outcome corresponding to each edge.

The combination of assignment data and standardized performance evaluations provides an ideal setting to study the structure and consequences of bureaucrat-task matching in an international organization. The main result in Limodio (2021) is the presence of negative sorting: high-performing managers are disproportionately allocated to low-performing countries. Possible explanations include the Bank's strategic objective of assigning stronger managers to weaker countries, internal career incentives and promotion dynamics, the demand for specialized skills in more difficult environments, and reallocations following adverse shocks such as natural disasters.

For this illustration, I focus on a version of the model without controls. This differs from the main specification in Limodio (2021), which includes year and sector fixed effects in the two-way fixed effects regression. The discussion here should therefore be viewed purely as an application of the methods developed in this paper, not as a critique of or challenge to their findings. Those results rely on additional assumptions that are not addressed in the following analysis.

The data contain 3,385 projects, corresponding to 1,876 distinct manager-country pairs. When

[1]The data used for this illustration exercise are publicly available on the author's website.

the same match y_{ij} is observed multiple times, I use the average outcome. The resulting matching network consists of 697 manager nodes, 127 country nodes, and 1,876 edges. Within this graph, there are 228 edge-distinct four-cycles, involving 369 managers and 114 countries. Thus, the information effectively used to estimate the Tukey interaction parameter comes from approximately half of the edges and managers, and from about 90% of the countries.

I consider the RBL estimator $\hat{\beta}_{L,z}$, which assigns labels in each cycle using auxiliary information on managers and countries. This requires instruments, observable characteristics of managers and countries satisfying Assumption 6. I use the average project size as the instrument for managers and the Public Infrastructure Management Index (PIMI) as the instrument for countries. Limodio (2021) documents that project size, measured by the average loan amount overseen, is predictive of managerial ability: more capable managers tend to supervise larger loans. Similarly, the PIMI developed by Dabla-Norris et al. (2012) is predictive of institutional productivity, with higher-performing countries scoring higher on this index. Assumptions 6.2 and 6.3 hold provided that the rankings induced by these instruments are not systematically opposed within cycles and not systematically misaligned when productivity gaps are large. Since such violations would require counterintuitive patterns, the assumptions appear plausible in practice. Using average loan size for managers and the PIMI for countries, therefore, offers a feasible way to implement Assumption 5.4.

Figure 6 reports the value of the RBL estimator, together with the corresponding confidence interval for nominal coverage of 0.9. The estimate is negative, indicating negative complementarities between managers and countries: the relative contribution of a high-ability manager, compared to a lower-ability one, is greater when working in a lower-productivity country. The magnitude of the estimate (0.196) can be interpreted as the ratio of the multiplicative to the additive component. This suggests that, while smaller in importance than additive effects, the multiplicative part of the interaction function still plays a meaningful role.

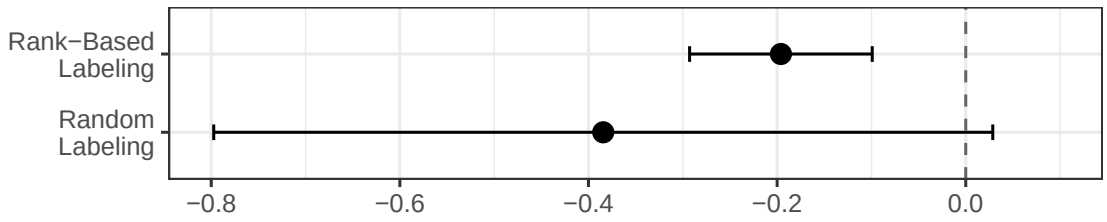


Figure 6: Estimates for the Tukey model. The confidence interval is constructed to correctly cover β_0 with probability 0.9.

The confidence interval $(-0.293, -0.099)$, centered around $\hat{\beta}_{L,z}$, does not include zero. The p-value for the null $\beta_0 = 0$ equals 0.001, providing strong evidence against the absence of complementarities in this interaction.

To underscore the importance of label choice, the figure also reports estimates obtained under random labeling. These are presented only to illustrate how reliance on uninformed labels affects the estimator and should not be interpreted as informative about the underlying interaction. As expected, random labeling inflates the variance, producing confidence intervals nearly four times wider than those based on rank-based labeling. In this case, the p-value of 0.126 would fail to reject the null of no complementarities, highlighting how proper labeling is crucial for test power, a theme further explored in the Monte Carlo simulations in the next section.

Figure 6 shows the estimate for a single random labeling, but many such estimates can be computed. Figure 7 plots the distributions of $\hat{\beta}_{L,\pi}$ and of $\hat{\mu}_\pi$ (the estimator of $\mu_L c_\pi$, the scaling factor in the asymptotic distribution) across 50,000 random labels assignments, with the values obtained under rank-based labeling highlighted in red.

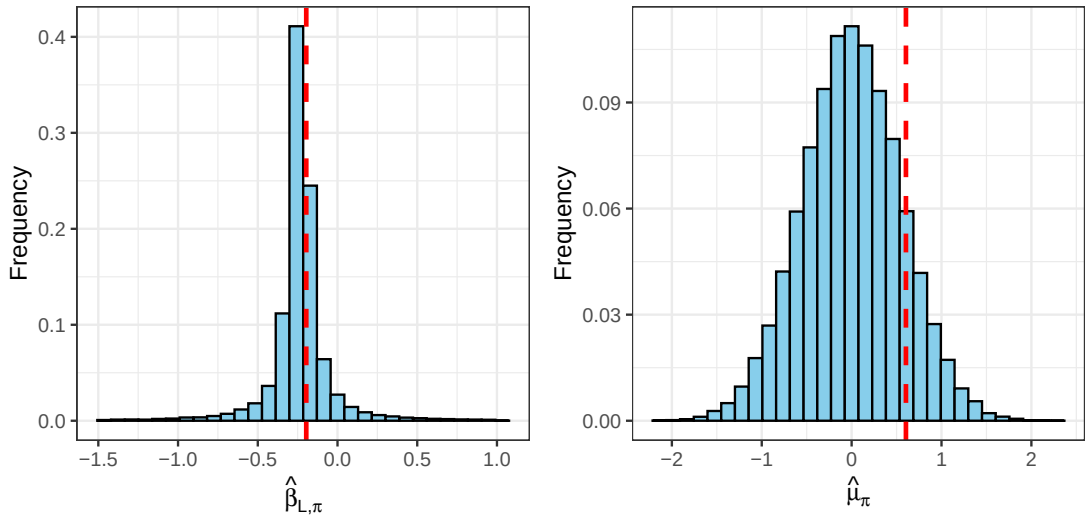


Figure 7: Empirical distribution of $\hat{\beta}_{L,\pi}$ and $\hat{\mu}_\pi$. For the distribution of $\hat{\beta}_{L,\pi}$, to make the histogram informative, I excluded 2.8% of observations out of $(-1.5, 1)$. For reference, the minimum and the maximum values I obtained are -413.589 and 1543.647.

The distribution of $\hat{\beta}_{L,\pi}$ shows that most labelings yield negative estimates, though with some variation in magnitude. The distribution of $\hat{\mu}_\pi$, symmetric around zero, indicates that the instruments are informative about the underlying characteristics: the rank-based labeling produces an estimate of $\hat{\mu}_\pi$ larger in absolute value than about 80% of random labelings. At the same time, in 20% of cases random labeling delivers even larger values, reflecting that, as expected, the instruments capture only part of the variation in latent rankings.

Figures 6 and 7 thus highlight the importance of selecting labels correctly. The analogy with instrumental variables is direct: valid instruments must be informative about the latent characteristics, not mere noise, in order to deliver consistent and precise estimates.

The negative estimate, statistically different from zero, implies that the skills of high-ability managers matter most in low-productivity countries, where managerial capacity has a greater marginal impact. This explanation was already suggested by Limodio (2021), but could not be formally assessed under the modularity restriction of the TWFE model. In this view, assigning high-performing bureaucrats to low-performing countries is efficient and raises average project success.

Overall, the empirical illustration shows that the Tukey model is straightforward to implement and provides insights into the structure of interactions that the TWFE model, by construction, rules out. The exercise also underscores the practical relevance of Assumption 5.4: only when instruments induce informative labelings do the resulting estimates convey reliable evidence about complementarities. In the next section, I turn to Monte Carlo simulations to evaluate how these insights carry over to controlled settings and to study the finite-sample behavior of the estimator.

6 Monte Carlo Simulations

In this section, I investigate the finite-sample behavior of the estimator $\hat{\beta}_{L,z}$. Because estimation of β_0 comes entirely from the cycles, I abstract from the rest of the network: I fix a number of cycles L and simulate outcomes only on those cycles. This design lets me isolate how performance depends on the effective sample size (the number of cycles), the relevance of the instruments, and the relative magnitude of productivity variation to noise.

Since in simulations the true values of $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ are known, I can examine the role of the instruments by directly controlling the labeling within each cycle. Concretely, I set the values of $\pi^{\alpha}_{\ell,z}$ and $\pi^{\psi}_{\ell,z}$, which is equivalent to choosing instruments that induce those labelings. This approach cleanly maps instrument relevance into the labelings that matter for $\hat{\beta}_{L,z}$.

The simulation design proceeds in two steps.

Step 1. Generation of productivities. For each of the L cycles, the vector of productivities of two workers and two firms is drawn from a multivariate normal distribution with mean vector μ and covariance matrix Σ:

$$\mu = (1, 3, 1, 3), \quad \Sigma = \begin{bmatrix} 1 & 0 & 0.5 & 0.5 \\ 0 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0 \\ 0.5 & 0.5 & 0 & 1 \end{bmatrix}.$$

Labels are then assigned by fixing $\pi^{\psi}_{\ell,z} = 1$ and setting $\pi^{\alpha}_{\ell,z}$ equal to 1 with probability $p \geq 0.5$

and to -1 with probability $1-p$. This corresponds to using different instruments for $\boldsymbol{\alpha}$: the larger the value of p, the stronger the instrument. Equivalently, the same exercise could be conducted using instruments for $\boldsymbol{\psi}$ or for both sides simultaneously.

Step 2. Generation of outcomes. In each Monte Carlo replication, the four error terms for every cycle are drawn independently from a normal distribution with mean zero and variance σ_{ij}^2. Combined with β_0, these errors generate the $4L$ outcomes, which form the simulated data observed in that replication.

I explore the finite-sample properties of the estimator by varying the number of cycles L, the error variance σ_{ij}, and the relevance of the instrument p:

$$L \in \{100, 500, 1000, 5000\}, \quad \sigma_{ij} \in \{0.5, 1, 2\}, \quad p \in \{1, 0.85, 0.65, 0.5\}.$$

These choices allow me to examine the role of the assumptions in Theorems 5 and 6, which require L to be large, σ_{ij} to remain bounded, and $p \neq 0.5$. The case $p = 0.5$ corresponds to non-informative labels, where $c_\pi = 2p - 1 = 0$ and Assumption 5.4 fails.

For each parameter combination, Step 1 is implemented once, while Step 2 is repeated 10,000 times. Fixing productivities, their labels, and their allocation in cycles across replications mimics the identification and inference analysis, where $\boldsymbol{\alpha}$, $\boldsymbol{\psi}$, G_{IJ}, $\boldsymbol{z}^{\boldsymbol{\alpha}}$, and $\boldsymbol{z}^{\boldsymbol{\psi}}$ are treated as deterministic.

Table 1 reports the mean squared errors of $\hat{\beta}_{L,\pi}$ as an estimator of β_0 across simulations. Table 2 presents the average widths of the corresponding 90% confidence intervals, while Table 3 displays the rejection rates of the test for absence of complementarities under the null hypothesis $H_0 : \beta_0 = 0$, for $\beta_0 \in \{0, 1\}$ and nominal significance level $\gamma = 0.1$.

		Mean squared errors			
		p			
σ_{ij}	L	1	0.85	0.65	0.5
0.5	100	0.0004	0.001	0.068	104.760
0.5	500	0.0001	0.0003	0.001	209.599
0.5	1000	0.0001	0.0001	0.001	202.407
0.5	5000	0	0	0.0001	16.723
1	100	0.002	0.470	9.173	225.307
1	500	0.0005	0.001	30.133	15.068
1	1000	0.0002	0.0005	0.005	371.678
1	5000	0	0.0001	0.0005	390.308
2	100	0.348	53.365	586.414	371.702
2	500	0.003	0.018	229.784	101.623
2	1000	0.001	0.002	5.009	165.683
2	5000	0.0002	0.0003	0.003	633.327

Table 1: Mean squared error across different values of σ_{ij}, L, and p under $\beta = 0$.

		Average confidence interval width			
		p			
σ_{ij}	L	1	0.85	0.65	0.5
0.5	100	0.070	0.110	0.890	2330.210
0.5	500	0.030	0.060	0.100	8171.620
0.5	1000	0.020	0.030	0.080	5022.450
0.5	5000	0.010	0.020	0.040	342.600
1	100	0.150	2.740	84.360	9941.750
1	500	0.070	0.110	553.890	203.280
1	1000	0.050	0.080	0.200	5879.220
1	5000	0.020	0.030	0.080	16986.770
2	100	2.980	730.570	15752.510	7345.480
2	500	0.160	0.310	2214.970	2782.270
2	1000	0.100	0.150	34.490	3163.600
2	5000	0.040	0.060	0.160	11696.520

Table 2: Average width of confidence intervals across different values of σ_{ij}, L, and p under $\beta = 0$, for a nominal coverage of 0.9.

When $p > 0.5$, and hence Assumption 5.4 holds, the mean squared error, the width of the confidence intervals, and the discrepancy between nominal and empirical size all increase with the error variance and decrease with the sample size. By contrast, the power of the test rises with sample size and falls as noise grows. These patterns are consistent with theoretical predictions.

The more interesting patterns concern the role of p, which captures the relevance of the instrument. When Assumption 5.4 is violated ($p = 0.5$), the estimator becomes inconsistent and the asymptotic distribution no longer applies. In this case, the mean squared error does not shrink with sample size, the test fails to control size under the null, and it is inconsistent against the alternative, regardless of sample size or error variance.

When the assumption holds but c_π is small (that is, when $p > 0.5$ but close to 0.5), finite-sample performance deteriorates, especially when L is small and σ_{ij} is large. The simulations confirm that a small c_π inflates the estimator's variance, producing large mean squared errors and wide confidence intervals. This effect is most pronounced when p is small relative to L. In such cases, the test for absence of complementarities continues to control size reasonably well, but its power drops sharply, as the inflated variance reduces the ability to reject the null when it is false. This mirrors the empirical illustration, where the null hypothesis $\beta_0 = 0$ was rejected under rank-based labeling but not under random labeling.

Overall, the simulations reinforce the theoretical results and highlight the central role of labeling within cycles: appropriate label selection is crucial not only for estimator precision, but also for ensuring that the test for the absence of complementarities retains meaningful power.

		Rejection rate under H_0				Rejection rate under H_1			
		p				p			
σ_{ij}	L	1	0.85	0.65	0.5	1	0.85	0.65	0.5
0.5	100	0.098	0.077	0.084	0.046	0.983	0.895	0.020	0
0.5	500	0.102	0.097	0.080	0.028	1	1	0.909	0
0.5	1000	0.102	0.098	0.076	0.040	1	1	0.995	0
0.5	5000	0.105	0.100	0.102	0.069	1	1	1	0
1	100	0.093	0.109	0.113	0.112	0.572	0.409	0.071	0
1	500	0.091	0.085	0.118	0.139	0.997	0.916	0.267	0
1	1000	0.094	0.090	0.093	0.132	1	0.997	0.650	0
1	5000	0.099	0.100	0.089	0.122	1	1	0.994	0
2	100	0.116	0.133	0.189	0.192	0.173	0.086	0.014	0.005
2	500	0.094	0.102	0.143	0.188	0.695	0.482	0.078	0.002
2	1000	0.088	0.086	0.116	0.168	0.930	0.681	0.179	0.002
2	5000	0.093	0.097	0.092	0.187	1	1	0.608	0.024

Table 3: Rejection rates across different values of σ_{ij}, L, and p for the null hypothesis $H_0 : \beta_0 = 0$, considering tests with nominal level $\gamma = 0.9$. In the right part of the table, simulations consider a true $\beta_0 = 1$.

7 Conclusion

This paper introduced the Bipartite Interaction (BI) framework for modeling two-sided interactions. Within this framework, I proposed notions of identification and large-sample, and studied three models defined by different restrictions on the interaction function, deriving conditions for their identification. The analysis highlighted a fundamental trade-off between flexibility in the interaction function and the density of the matching network: richer interaction structures require stronger graph conditions to achieve point identification.

Among the models, the Tukey specification emerged as a particularly useful case. By summarizing complementarities with a single parameter, it extends the TWFE model in a parsimonious yet interpretable way, and for identification it requires only mild additional conditions beyond those of the TWFE model. I developed a novel cycle-based estimator for its interaction parameter, which avoids estimating latent productivities, is consistent under bounded-degree graphs, and is asymptotically normal. Its asymptotic distribution also provides the basis for a formal test of the absence of complementarities. An empirical illustration showed that the Tukey model can be implemented in settings where the TWFE model is standard, delivering richer insights into the interaction process that would remain hidden under additively separable models. More broadly, the results demonstrate that complementarities matter in practice and can be studied with feasible methods.

Two directions for future research remain open. First, while the analysis here focused on

point identification, extending the BI framework to partial identification would be valuable. For the BLM and seriation models, even when point identification fails, the data may still deliver informative bounds. Characterizing identified sets and developing computational methods to recover them could lead to more credible conclusions without strong parametric assumptions. Second, while in this paper the main focus was on estimation of β_0 in the Tukey model, researchers are often interested also in estimating the productivity parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$. I outlined some estimators, but a full analysis of their properties is needed to assess when they yield reliable measures of worker and firm productivities. The BI framework, by clarifying the distinct roles of the interaction function and the matching network, provides a natural foundation for these further investigations.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Adhvaryu, A., V. Bassi, A. Nyshadham, and J. Tamayo (2024): "No line left behind: Assortative matching inside the firm," *The Review of Economics and Statistics*, 1–45.

Arkhangelsky, D. and G. Imbens (2024): "Causal models for longitudinal and panel data: A survey," *The Econometrics Journal*, 27, C1–C61.

Armstrong, T. B., M. Weidner, and A. Zeleneev (2025): "Robust Estimation and Inference in Panels with Interactive Fixed Effects," Working paper.

Auerbach, E., H. Guo, and M. Tabord-Meehan (2025): "The local approach to causal inference under network interference," Working paper.

Bai, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77, 1229–1279.

Becker, G. S. (1973): "A theory of marriage: Part I," *Journal of Political Economy*, 81, 813–846.

Bertrand, M. and A. Schoar (2003): "Managing with style: The effect of managers on firm policies," *The Quarterly Journal of Economics*, 118, 1169–1208.

Blum, A., J. Hopcroft, and R. Kannan (2020): *Foundations of Data Science*, Cambridge, UK: Cambridge University Press.

Bonhomme, S. (2020): "Econometric analysis of bipartite networks," in *The econometric analysis of network data*, Elsevier, 83–121.

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, 291–322.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A distributional framework for matched employer employee data," *Econometrica*, 87, 699–739.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55.

Cai, Y. (2022): "Linear regression with centrality measures," Working paper.

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly Journal of Economics*, 128, 967–1015.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a): "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American Economic Review*, 104, 2593–2632.

——— (2014b): "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *American Economic Review*, 104, 2633–2679.

Chetverikov, D., A. Santos, and A. M. Shaikh (2018): "The econometrics of shape restrictions," *Annual Review of Economics*, 10, 31–63.

Crippa, F. and D. Fedchenko (2025): "Partially Identified Rankings from Pairwise Interactions," Working paper.

Dabla-Norris, E., J. Brumby, A. Kyobe, Z. Mills, and C. Papageorgiou (2012): "Investing in public investment: an index of public investment efficiency," *Journal of Economic Growth*, 17, 235–266.

De Paula, Á., S. Richards-Shubik, and E. Tamer (2018): "Identifying preferences in networks with bounded degree," *Econometrica*, 86, 263–288.

Eeckhout, J. (2018): "Sorting in the labor market," *Annual Review of Economics*, 10, 1–29.

Eeckhout, J. and P. Kircher (2011): "Identifying sorting—in theory," *The Review of Economic Studies*, 78, 872–906.

Fenizia, A. (2022): "Managers and productivity in the public sector," *Econometrica*, 90, 1063–1084.

Finkelstein, A., M. Gentzkow, and H. Williams (2016): "Sources of geographic variation in health care: Evidence from patient migration," *The Quarterly Journal of Economics*, 131, 1681–1726.

Flammarion, N., C. Mao, and P. Rigollet (2019): "Optimal rates of statistical seriation," *Bernoulli*, 25, 623–653.

Freeman, H. and M. Weidner (2023): "Linear panel regressions with two-way unobserved heterogeneity," *Journal of Econometrics*, 237, 105498.

Freyberger, J. (2018): "Non-parametric panel data models with interactive fixed effects," *The Review of Economic Studies*, 85, 1824–1851.

Gao, W. Y. (2020): "Nonparametric identification in index models of link formation," *Journal of Econometrics*, 215, 399–413.

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

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying equilibrium models of labor market sorting," *Econometrica*, 85, 29–65.

Hessert, R. and S. Mallik (2021): "Moore-Penrose inverses of the signless Laplacian and edge-Laplacian of graphs," *Discrete Mathematics*, 344, 112451.

Jackson, C. K., J. E. Rockoff, and D. O. Staiger (2014): "Teacher effects and teacher-related policies," *Annual Review of Economics*, 6, 801–825.

Jochmans, K. (2017): "Two-way models for gravity," *The Review of Economics and Statistics*, 99, 478–485.

Jochmans, K. and M. Weidner (2019): "Fixed-effect regressions on network data," *Econometrica*, 87, 1543–1560.

Kline, P. (2024): "Firm wage effects," *Handbook of Labor Economics*, 5, 115–181.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88, 1859–1898.

Koopmans, T. C. (1949): "Identification problems in economic model construction," *Econometrica*, 125–144.

Lang, S. (1987): *Linear algebra*, Springer Science & Business Media.

Lei, L. and B. Ross (2024): "Estimating counterfactual matrix means with short panel data," Working paper.

Limodio, N. (2021): "Bureaucrat allocation in the public sector: Evidence from the world bank," *The Economic Journal*, 131, 3012–3040.

Lopes de Melo, R. (2018): "Firm wage differentials and labor market sorting: Reconciling theory and evidence," *Journal of Political Economy*, 126, 313–346.

Mao, C., A. Pananjady, and M. J. Wainwright (2020): "Towards optimal estimation of bivariate isotonic matrices with unknown permutations," *The Annals of Statistics*, 48, 3183–3205.

Mourot, P. (2025): "Should Top Surgeons Practice at Top Hospitals? Sorting and Complementarities in Healthcare," Working paper.

Nash Jr, J. F. (1950): "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, 36, 48–49.

Sbai Sassi, Y. (2024): "A Linear Regression Model for Non-Oriented Dyadic Data with Interactive Individual Effects," Working paper.

Shimer, R. and L. Smith (2000): "Assortative matching and search," *Econometrica*, 68, 343–369.

Šimeček, P. and M. Šimečková (2013): "Modification of Tukey's additivity test," *Journal of Statistical Planning and Inference*, 143, 197–201.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019): "Firming up inequality," *The Quarterly Journal of Economics*, 134, 1–50.

Topkis, D. M. (1998): *Supermodularity and Complementarity*, Princeton, NJ: Princeton University Press.

Tukey, J. W. (1949): "One degree of freedom for non-additivity," *Biometrics*, 5, 232–242.

Verdier, V. (2020): "Estimation and inference for linear models with two-way fixed effects and sparsely matched data," *The Review of Economics and Statistics*, 102, 1–16.

Ward, G. and I. Dick (1952): "Non-additivity in randomized block designs and balanced incomplete block designs," *New Zealand Journal of Science and Technology*, 33, 430–436.

Weigel, J. L., P. Bessone, A. Bergeron, G. Z. Tourek, and J. K. Kabeya (2024): "Supermodular bureaucrats: Evidence from randomly assigned tax collectors in the DRC," Working paper.

A Model Characterization via Shape Restrictions

This Appendix shows how the TWFE, Tukey, and BLM models introduced in Section 2 can be obtained from shape restrictions imposed directly on the interaction function f. In particular, I demonstrate that imposing certain shape restrictions leads to representations that are observationally equivalent to each of these three models: once the restrictions are in place, f can, up to a relabeling, be expressed in the form of the TWFE, Tukey, or BLM specification. For the seriation model, the connection with monotonicity of f was already made explicit in its definition, and is therefore not revisited here.

These equivalence results clarify the structural assumptions implicitly embedded in the models, which are often interpreted only as reduced-form specifications. The remainder of this Appendix is organized as follows. Section A.1 shows how modularity leads to the TWFE model; Section A.2 establishes the connection between bilinearity and the Tukey model; and Section A.3 links a diagonal bilinear structure to the BLM model.

A.1 Modularity and TWFE model

Recall the definition of modularity (see Topkis (1998) for an extensive discussion).

Definition 6. (Supermodularity, Submodularity, Modularity) *A function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is supermodular (submodular) if, for any quadruple (i, i', j, j') such that $\alpha_i > \alpha_{i'}$ and $\psi_j > \psi_{j'}$,*

$$f(\alpha_i, \psi_j) - f(\alpha_i, \psi_{j'}) \geq (\leq) f(\alpha_{i'}, \psi_j) - f(\alpha_{i'}, \psi_{j'}).$$

A function f is modular if it is both supermodular and submodular.

If f is twice continuously differentiable, supermodularity (submodularity) is equivalent to $\partial^2 f / \partial\alpha\, \partial\psi \geq 0$ (≤ 0), and modularity to $\partial^2 f / \partial\alpha\, \partial\psi = 0$ everywhere.

The next proposition formalizes the observational equivalence between a modular f and the TWFE model specification.

Proposition 4. (Modularity and TWFE model) *Any model $(f, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ with $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ modular is observationally equivalent to the TWFE model.*

Proof. For functions defined on product spaces, modularity is equivalent to additive separability. Thus, any modular f can be written as $f(\alpha_i, \psi_j) = g_\alpha(\alpha_i) + g_\psi(\psi_j)$ for some functions $g_\alpha : A \to \mathbb{R}$ and $g_\psi : \Psi \to \mathbb{R}$. Let $\alpha'_i \coloneqq g_\alpha(\alpha_i)$ and $\psi'_j \coloneqq g_\psi(\psi_j)$. Then, it holds that $f(\alpha_i, \psi_j) = g_\alpha(\alpha_i) + g_\psi(\psi_j) = \alpha'_i + \psi'_j$, and hence $(f, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ and the TWFE model are observationally equivalent. □

A.2 Bilinearity and Tukey model

Recall the definition of bilinearity (Lang, 1987).

Definition 7. (Bilinearity) *A function $f : U \times V \to \mathbb{R}$ is bilinear if, for each fixed $v \in V$, the map $u \mapsto f(u, v)$ is linear in u, and for each fixed $u \in U$, the map $v \mapsto f(u, v)$ is linear in v.*

Many panel data methods rely on bilinear interactions between latent fixed effects, whether scalar or vector. Examples include two-way and group fixed effects, difference-in-differences, synthetic control, factor models, low-rank approximations, and nuclear-norm regularization (see Arkhangelsky and Imbens (2024) for a review). Bilinear structures also arise in strategic settings: in the classic two-player, two-action simultaneous game with risk-neutral agents (Nash Jr, 1950), expected payoffs are bilinear in the players' mixed strategies.

In the Bipartite Interaction model, to allow for an intercept while maintaining one scalar productivity per agent, define the augmented vectors $\alpha_i^v \coloneqq (\alpha_i, 1)'$ and $\psi_j^v \coloneqq (\psi_j, 1)'$. Say that $f(\alpha, \psi)$ admits a *bilinear representation* if there exists a bilinear function f^v such that $f(\alpha, \psi) = f^v(\alpha^v, \psi^v)$.

With an additional assumption on the behavior of f at the origin, the following proposition shows that admitting a bilinear representation is observationally equivalent to the Tukey model.

Proposition 5. (Bilinear Representation and Tukey model) *Any model $(f, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ in which f admits a bilinear representation with $\alpha_i^v \coloneqq (\alpha_i, 1)'$ and $\psi_j^v \coloneqq (\psi_j, 1)'$ and satisfies*

$$f(\alpha, \psi) = 0 \iff \alpha = 0 \wedge \psi = 0,$$

is observationally equivalent to the Tukey model.

Proof. By Theorem 4.1 in Lang (1987), any bilinear function f_v can be written

$$f_v(\alpha^v, \psi^v) = (\alpha^v)' B \psi^v, \quad B = \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & b_{11} \end{pmatrix}.$$

Hence,

$$f(\alpha, \psi) = f_v(\alpha^v, \psi^v) = (\alpha_i, 1) B (\psi_j, 1)' = b_{00}\alpha\psi + b_{01}\alpha + b_{10}\psi + b_{11}.$$

Since $f(0, 0) = 0$, $b_{11} = 0$. Furthermore, since $f(\alpha, \psi) = 0$ only when both α and ψ are zero, $b_{01} \neq 0$ and $b_{10} \neq 0$. Set

$$\alpha' = b_{01}\alpha, \quad \psi' = b_{10}\psi, \quad \beta_0 = \frac{b_{00}}{b_{01}b_{10}},$$

and note that $f(\alpha, \psi) = b_{01}\alpha + b_{10}\psi + b_{00}\alpha\psi = \alpha' + \psi' + \beta_0\alpha'\psi'$, and hence $(f, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ and the Tukey model are observationally equivalent. □

A.3 Bilinearity and BLM model

Suppose now that each firm is characterized by the productivity vector $\psi_j^v = (b_j, a_j)$. Recall that a *diagonal bilinear form* is a bilinear form whose matrix representation is diagonal. Imposing that f admits a diagonal bilinear representation is observationally equivalent to the BLM model.

Proposition 6. (Bilinear Representation and BLM model) *Any model $(f, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ in which f admits a diagonal bilinear representation with $\alpha_i^v \coloneqq (\alpha_i, 1)'$ and $\psi_j^v = (b_j, a_j)'$ is observationally equivalent to the BLM model.*

Proof. By Theorem 4.1 in Lang (1987), any bilinear function f_v can be written

$$f_v(\alpha^v, \psi^v) = (\alpha^v)' B \psi^v, \quad B = \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & b_{11} \end{pmatrix}.$$

Since B diagonal, $b_{01} = b_{10} = 0$, and hence

$$f(\alpha, \psi) = f_v(\alpha^v, \psi^v) = (\alpha_i, 1)B(b_j, a_j)' = b_{00}\alpha b + b_{01}\alpha a + b_{10}b + b_{11}a = b_{00}\alpha b + b_{11}a.$$

Set

$$\alpha' = b_{00}\alpha, \quad a' = b_{11}a, \quad b' = b$$

and note that $f(\alpha, \psi) = b_{00}\alpha b + b_{11}a = \alpha' b' + a'$, and hence $(f, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ and the BLM model are observationally equivalent. □

B TWFE as Approximation of Tukey

The TWFE model is often interpreted as an approximation, analogous to using the best linear projection (the linear function of a variable X that minimizes mean squared error when approximating another variable Y) to summarize the relationship between random variables X and Y. In this Appendix, I investigate how such an approximation behaves when the true worker-firm interaction follows the Tukey model.

Define

$$n_i = \sum_j D_{ij}, \quad m_j = \sum_i D_{ij}, \quad n = \sum_i n_i,$$

where n is the total number of matches, n_i is the degree of worker i, and m_j is the degree of firm j. Using this notation, I introduce the two-way fixed effects (TWFE) projections.

Definition 8. (TWFE Projection) *Under Assumption 2 and the normalization $\alpha_1 = 0$, the TWFE projections $(\boldsymbol{\alpha}^{twfe}, \boldsymbol{\psi}^{twfe})$ are defined by*

$$(\boldsymbol{\alpha}^{twfe}, \boldsymbol{\psi}^{twfe}) = (C'C)^{-1}C'\boldsymbol{\theta}_O,$$

where $C \in \mathbb{R}^{n\times(I+J-1)}$ is the incidence (design) matrix linking the free worker productivities $(\alpha_2, \ldots, \alpha_I)$ and all firm productivities $(\psi_1, \ldots, \psi_J)$ to the observed matches.

Intuitively, $(\boldsymbol{\alpha}^{twfe}, \boldsymbol{\psi}^{twfe})$ are the coefficients from a TWFE fit in the noiseless model. The expression in Definition 8 is the ordinary least squares solution, with C determined entirely by the structure of the matching network G_{IJ}. Each row of C contains one nonzero entry in the worker block and one in the firm block, encoding the worker-firm pair observed in that match.

The matrix $C'C$ has the block form

$$C'C = \begin{pmatrix} \operatorname{diag}(n_2, \ldots, n_I) & D \\ D^T & \operatorname{diag}(m_1, \ldots, m_J) \end{pmatrix}$$

where D is the incidence block of the adjacency matrix of G_{IJ}. This is the signless Laplacian of G_{IJ}, the sum of its degree matrix and adjacency matrix, which captures the full connectivity structure of the graph.

Under the TWFE model, the TWFE projections $(\boldsymbol{\alpha}^{twfe}, \boldsymbol{\psi}^{twfe})$ coincide with the true productivities $(\boldsymbol{\alpha}, \boldsymbol{\psi})$. Under alternative interaction functions, however, the presence of $C'C$ in the projection formula implies that the bias depends on the entire structure of G_{IJ}. In the case of the Tukey model, this dependence admits a closed-form characterization.

Define the average partner productivities

$$\bar{\psi}_i = \frac{1}{n_i}\sum_j D_{ij}\,\psi_j, \quad \bar{\alpha}_j = \frac{1}{m_j}\sum_i D_{ij}\,\alpha_i,$$

where $\bar{\psi}_i$ is the average firm productivity for worker i and $\bar{\alpha}_j$ is the average worker productivity for firm j.

The next proposition provides the expressions for $(\boldsymbol{\alpha}^{twfe}, \boldsymbol{\psi}^{twfe})$ under the Tukey model,

making it possible to examine when the TWFE projection delivers a good approximation to the true interaction.

Proposition 7. (Misspecified TWFE) *Under the Tukey model, the normalization $\alpha_1 = 0$, and Assumption 2, the projections satisfy, for $i > 1$ and all j:*

$$\alpha_i^{twfe} = \alpha_i + \beta_0 \left(\sum_{i'=2}^{I} \Lambda_{i,i'}\, n_{i'} \alpha_{i'} \bar{\psi}_{i'} + \sum_{j'=1}^{J} \Lambda_{i,\, I-1+j'}\, m_{j'} \psi_{j'} \bar{\alpha}_{j'} \right),$$

$$\psi_j^{twfe} = \psi_j + \beta_0 \left(\sum_{i'=2}^{I} \Lambda_{I-1+j,\, i'}\, n_{i'} \alpha_{i'} \bar{\psi}_{i'} + \sum_{j'=1}^{J} \Lambda_{I-1+j,\, I-1+j'}\, m_{j'} \psi_{j'} \bar{\alpha}_{j'} \right),$$

where $\Lambda_{u,v}$ denotes the (u, v) entry of the inverse of the signless Laplacian of G_{IJ}.

Proof. To derive these expressions, rewrite the problem in matrix form and apply the ordinary least squares formula. Let $A \in \mathbb{R}^{IJ \times I}$ and $B \in \mathbb{R}^{IJ \times J}$ denote the worker and firm indicator matrices, respectively. Each entry $A_{(ij),\,i}$ and $B_{(ij),\,j}$ equals 1 if the match (i, j) involves worker i or firm j, and zero otherwise. Let $Y \in \mathbb{R}^{IJ \times 1}$ stack the θ_{ij} values, and let $S \in \mathbb{R}^{n \times IJ}$ be the selection matrix extracting observed matches, so that $\boldsymbol{\theta}_O = SY$.

To impose $\alpha_1 = 0$, drop the first column of A and define $C \coloneqq [SA_{-1} \;\; SB] \in \mathbb{R}^{n \times (I+J-1)}$, where A_{-1} is A without its first column. Let $c^{twfe} \coloneqq [\alpha_2^{twfe}, \ldots, \alpha_I^{twfe}, \psi_1^{twfe}, \ldots, \psi_J^{twfe}]'$. Then

$$c^{twfe} = (C'C)^{-1} C'SY = (C'C)^{-1} C' \boldsymbol{\theta}_O.$$

Define $a \coloneqq SA\boldsymbol{\alpha}$, $p \coloneqq SB\boldsymbol{\psi}$, and $h \coloneqq a \odot p$, where $\odot$ denotes element-wise multiplication. The systematic part of the outcome satisfies

$$\{\theta_{ij}\}_{(i,j) \in O_{IJ}} = SY = Cc + \beta_0 h,$$

and thus the projection becomes

$$c^{twfe} = c + \beta_0 (C'C)^{-1} C'h,$$

where $\beta_0 (C'C)^{-1} C'h$ is the bias term.

The matrix $C'C$ is the signless Laplacian. To compute $C'h$, observe that each row k of C' has a 1 in the two positions corresponding to the nodes in match k, and zeros elsewhere. Therefore, $C'h$ is a vector whose first $I - 1$ entries are $n_i \alpha_i \bar{\psi}_i$ for $i = 2, \ldots, I$, and whose next J entries are $m_j \psi_j \bar{\alpha}_j$ for $j = 1, \ldots, J$. Substituting into the bias formula yields the stated result. □

The matrix Λ, the inverse of the signless Laplacian, can be viewed as a "connectivity influence

matrix": the entry $\Lambda_{u,v}$ measures how a perturbation in node v's outcome propagates through the graph to node u. Larger values of $\Lambda_{u,v}$ indicate that u and v are more tightly linked, via short paths, multiple routes, or connections through high-degree nodes, so omitted interaction terms from node v spill over more strongly into node u's TWFE projection.

The bias in Proposition 7 is proportional to $|\beta_0|$, and vanishes when $\beta_0 = 0$. The expressions $n_{i'}\alpha_{i'}\bar{\psi}_{i'}$ and $m_{j'}\psi_{j'}\bar{\alpha}_{j'}$ correspond exactly to the interaction components that are omitted when TWFE is imposed. The weights $\Lambda_{u,v}$, given by the entries of the inverse signless Laplacian, determine how the structure of G_{IJ} channels these omitted terms into each node's projection. Because Λ depends on the global topology of the graph, the bias for any given worker or firm is shaped by the structure of the entire network. Closed-form expressions for Λ are available for certain graph families (Hessert and Mallik, 2021), but no general formula exists.

B.1 Bias Propagation in Sorting Measure

Crucially, the bias term at each node depend not only on β_0 and the corresponding productivity, but also on the full pattern of observed matches. Two workers i and i' with identical productivities, $\alpha_i = \alpha_{i'}$, may have different projections, α_i^{twfe} and $\alpha_{i'}^{twfe}$. In some cases, the ordering may even reverse: for example, it is possible to have $\alpha_i > \alpha_{i'}$ but $\alpha_i^{twfe} < \alpha_{i'}^{twfe}$. This is in contrast with the balanced panel case, where all worker-firm pairs are observed and projections preserve the productivity rankings.

The bias expressions show that the projections embed the sorting pattern in the matching network. Hence, relying on $\boldsymbol{\alpha}^{twfe}$ and $\boldsymbol{\psi}^{twfe}$ to measure sorting can be misleading. A standard approach is to compute the correlation between α_i^{twfe} and ψ_j^{twfe} across observed matches: one constructs two vectors, one with α_i^{twfe} and another with ψ_j^{twfe} for each observed pair (i, j), and then computes their correlation as a summary statistic of sorting. A positive correlation suggests sorting of high-productivity workers with high-productivity firms; a near-zero or negative correlation implies weak or negative sorting.

Proposition 7 can be used directly to construct cases in which the projections fail to capture the true sorting pattern. In the example below, I construct a setting with positive sorting, meaning that true productivities are positively correlated across matches. Although the true correlation is sizable (around 0.3), the correlation computed using TWFE projections is close to zero when the interaction function is supermodular with $\beta_0 = 3$. This case is not contrived: the matching network in Figure 8 is fairly typical, and simple theory predicts supermodularity being associated with positive sorting. Nonetheless, the TWFE approximation would incorrectly suggest the absence of sorting.

Example 3. (Sorting Pattern) *Consider the matching network with* $I = 3$, $J = 3$, *and* $O_{IJ} =$

$\{(i_1, j_1), (i_1, j_2), (i_1, j_3), (i_2, j_2), (i_2, j_3), (i_3, j_3)\}$, *as illustrated in Figure 8. Let the productivities be* $\alpha_1 = 4$, $\alpha_2 = 5$, *and* $\alpha_3 = 2$, *and* $\psi_1 = 10$, $\psi_2 = 8$, *and* $\psi_3 = 1$.

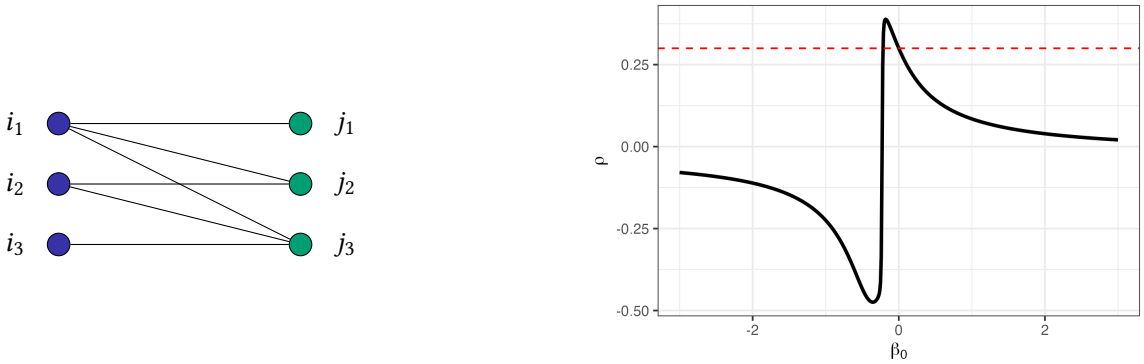


Figure 8: Left: the matching network in Example 3. Right: the correlation coefficient ρ between worker and firm TWFE projections across matches as a function of β_0.

Using Proposition 7, compute the projections $\boldsymbol{\alpha}^{twfe}$ *and* $\boldsymbol{\psi}^{twfe}$, *and let* ρ *be the correlation of productivities across matches. Figure 8 shows how* ρ *varies with* β_0*: although the true correlation is 0.3 (and is correctly captured when* $\beta_0 = 0$*), the correlation based on the projections can be close to zero or even negative.*

Consider the case $\beta_0 = 3$, *hence with a supermodular interaction function. The correlation based on the TWFE projections is only* 0.02, *with the bias nearly offsetting the underlying sorting pattern.*

Bias in the covariance also contaminates the variance decomposition, making its components difficult to interpret. In addition, because the bias can reverse the true ranking of agents, any two-stage procedure that relies on TWFE projections as outcome variables in subsequent regressions becomes unreliable.

Example 3 hence demonstrates that approximating the Tukey model with the TWFE specification, even in the noiseless case, produces parameters of limited informational value once complementarities are non-negligible. This underscores the importance of methods that allow f to feature complementarities. The Tukey model provides the minimal extension of TWFE required to capture them, offering a simple yet flexible structure for analysis.

C Seed-and-Snowballs and Leave-one-out Connectivity

In this Appendix, I examine the relationship between the Seed-and-Snowballs connectivity property introduced in Section 3.3 and the leave-one-out connectivity condition used by Kline et al. (2020) to establish the validity of their variance estimator. The latter requires that the bipartite graph remain connected after removing any single worker node together with its incident edges. The precise relationship between these two conditions is stated in the following proposition.

Proposition 8. (Seed-and-Snowballs vs. Leave-one-out Connectivity) *If G_{IJ} satisfies Seed-and-Snowballs connectivity, then it satisfies leave-one-out connectivity; the converse does not hold.*

Proof. Seed-and-Snowballs $\Rightarrow$ Leave-one-out. If $|J| = 1$ the conclusion is immediate, so assume there are at least two firms. Run the Seed–and–Snowballs algorithm using a seed j_0 and record the order in which firms enter the set S_n^J; by construction, each firm j_k with $k \geq 1$ is connected to the preceding firms through at least two distinct workers, say $i_k^{(1)}$ and $i_k^{(2)}$.

Now, remove an arbitrary worker $i^\star$ together with all incident edges. For every firm j_k there are two cases:

1. If $i^\star \notin \{i_k^{(1)}, i_k^{(2)}\}$, j_k keeps both links to the earlier firms.
2. If $i^\star$ coincides with, say, $i_k^{(1)}$, the second link $i_k^{(2)}$ remains, so j_k is still connected to the subgraph generated by $\{j_0, \ldots, j_{k-1}\}$.

Proceeding inductively from $k = 1$ to $k = |J| - 1$ shows that every firm remains connected to j_0 after $i^\star$ is deleted; hence, the whole graph stays connected. Because $i^\star$ was arbitrary, G_{IJ} is leave-one-out connected.

Leave-one-out $\not\Rightarrow$ Seed-and-Snowballs. Consider the graph in Figure 9. Deleting any single worker breaks exactly two edges, but the remaining edges still form a path that contains every node, so the graph is leave-one-out connected. However, starting for example from the seed j_0, the snowball algorithm proceeds $j_0 \to \{i_1, i_2, i_5\} \to \{j_0, j_1\} \to \{i_1, i_2, i_5, i_3\}$; remaining firms share at most one of these workers, so no further firm can be added and the procedure stops. Because the same failure occurs for any choice of the seed, Seed-and-Snowballs connectivity does not hold.

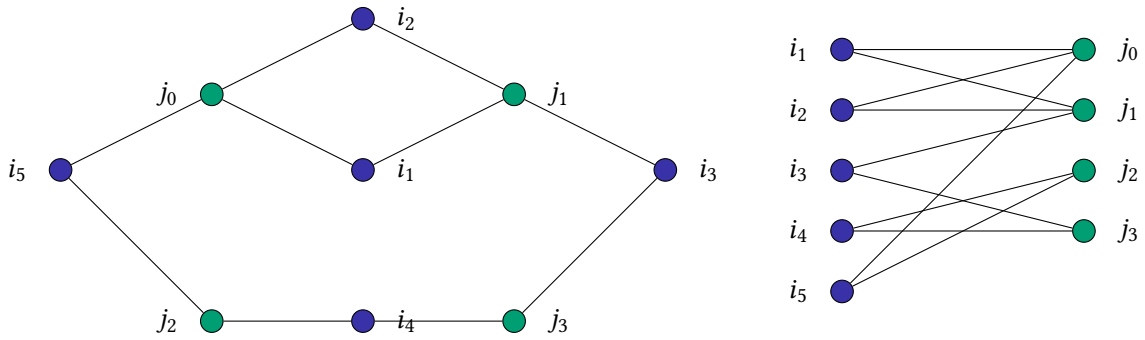


Figure 9: Two representation of the same bipartite graph satisfying leave-one-out connectivity and violating seed-and-snowball connectivity.

□

D 4-Cycles in Erdős-Rényi Model

In this Appendix, I study the conditions under which the bipartite Erdős-Rényi random graph model generates, with high probability, a diverging number of length-4 cycles, thereby satisfying Assumption 5.2. In the Bipartite Interaction model, the graph G_{IJ} is treated as fixed, so the number of cycles is deterministic. The goal here is not to provide a model of network formation, but rather to clarify the role of the "many cycles" assumption and how it relates to connectivity. While similar insights arise under other random graph models, the Erdős-Rényi model is adopted as a simple benchmark in which the entire network structure is governed by a single parameter.

In the bipartite Erdős–Rényi graph $G_{IJ}(p_{IJ})$, each pair $(i, j) \in I \times J$ is independently linked with probability $p_{IJ} \in (0, 1)$. The link probability p_{IJ} is constant across edges, but may vary with network sizes I and J. Let $C_4(I, J)$ denote the (random) number of length-4 cycles in $G_{IJ}(p_{IJ})$, and consider a setting in which the ratio $\frac{I}{J}$ remains bounded ($0 < \underline{c} < \frac{I}{J} < \overline{c} < \infty$) as $I, J \to \infty$, so that the numbers of agents on the two sides of the market grow at the same rate. The following proposition provides a sufficient condition for $G_{IJ}(p_{IJ})$ to contain a diverging number of length-4 cycles.

Proposition 9. (Cycles in Erdős-Rényi Model) *If* $Ip_{IJ} \to \infty$, *then* $C_4(I, J) \xrightarrow{p} \infty$.

Proof. Let $\mathcal{I}_\ell$ be the indicator that the quartet $\ell = (i, i', j, j')$, with $i, i' \in I$ and $j, j' \in J$, forms a cycle, and let $\mathcal{L}$ denote the set of all potential quartets. The number of length-4 cycles $C_4(I, J)$ is then given by

$$C_4(I, J) = \sum_{\ell \in \mathcal{L}} \mathcal{I}_\ell.$$

Since there are $\binom{I}{2}\binom{J}{2} = \frac{I(I-1)J(J-1)}{4}$ potential cycles, and each occurs with probability p_{IJ}^4,

$$\mathbb{E}[C_4(I, J)] = \frac{I(I-1)J(J-1)}{4} p_{IJ}^4.$$

Consider the variance of $C_4(I, J)$:

$$\mathrm{Var}(C_4(I, J)) = \mathrm{Var}\left(\sum_{\ell \in \mathcal{L}} \mathcal{I}_\ell\right) = \sum_{\ell \in \mathcal{L}} \mathrm{Var}(\mathcal{I}_\ell) + \sum_{\ell \neq \ell'} \mathrm{Cov}(\mathcal{I}_\ell, \mathcal{I}_{\ell'}),$$

and since $\mathcal{I}_\ell^2 = \mathcal{I}_\ell$,

$$\sum_\ell \mathrm{Var}(\mathcal{I}_\ell) = \sum_\ell p_{IJ}^4(1 - p_{IJ}^4) = \frac{I(I-1)J(J-1)}{4} p_{IJ}^4(1 - p_{IJ}^4),$$

which gives a term of order $O(I^4 p_{IJ}^4)$, where J is replaced by I since both grow at the same rate.

For two distinct cycles, let $k \in \{0, 1, 2\}$ be the number of shared edges. If $k = 0$, the indicators are independent, so the covariance is zero.

If $k = 1$, the two cycles share exactly one edge and involve six nodes. The number of such pairs is therefore of order $O(I^3 J^3)$. Each covariance term $\mathrm{Cov}(\mathcal{I}_\ell, \mathcal{I}_{\ell'})$ is of order $O(p_{IJ}^7)$, so the total contribution from pairs that share one edge is of order $O(I^3 J^3 p_{IJ}^7) = O(I^6 p_{IJ}^7)$.

If $k = 2$, the two cycles share exactly two edges and involve five nodes. The number of such pairs is therefore of order $O(I^5)$. Each covariance term $\mathrm{Cov}(\mathcal{I}_\ell, \mathcal{I}_{\ell'})$ is of order $O(p_{IJ}^6)$, so the total contribution from pairs that share two edges is of order $O(I^5 p_{IJ}^6)$.

Since the variance is thus

$$\mathrm{Var}(C_4(I, J)) = O(I^4 p_{IJ}^4) + O(I^6 p_{IJ}^7) + O(I^5 p_{IJ}^6),$$

I can conclude that, when $I p_{IJ} \to \infty$,

$$\frac{\mathrm{Var}(C_4(I, J))}{(\mathbb{E}[C_4(I, J)])^2} = O\left(\frac{1}{I p_{IJ}}\right) \to 0.$$

Let $\epsilon > 0$ be arbitrary. By Chebyshev's inequality,

$$\Pr\left\{\left|\frac{C_4}{\mathbb{E}[C_4]} - 1\right| \geq \epsilon\right\} = \Pr\left\{|C_4 - \mathbb{E}[C_4]| \geq \epsilon \mathbb{E}[C_4]\right\} \leq \frac{\mathrm{Var}(C_4)}{\epsilon^2 \{\mathbb{E}[C_4]\}^2} \to 0,$$

and hence $\frac{C_4}{\mathbb{E}[C_4]} \xrightarrow{p} 1$, which establishes

$$C_4(I, J) \xrightarrow{p} \infty.$$

□

The threshold for p_{IJ} required by Assumption 5.2 can be compared to the threshold needed for connectivity. For the network to be connected with high probability, it is necessary that

$$\frac{I p_{IJ}}{\log(I)} > 1$$

(see Section 8.2 of Blum et al. (2020)). Since $\log(I) \to \infty$, this condition implies $I p_{IJ} \to \infty$. In the bipartite Erdős–Rényi model, such a condition automatically guarantees that the number of 4-cycles diverges. The reverse, however, does not hold: $C_4(I, J) \to \infty$ even when $I p_{IJ}$ grows more slowly than $\log(I)$. Hence, connectivity is a strictly stronger requirement than the existence of many cycles. In particular, under the Erdős-Rényi benchmark, the graph conditions needed for

identification in the TWFE model also imply consistency of $\hat{\beta}_{L,\pi}$ for estimating β_0 in the Tukey model.

E Outcome-Based Labeling

To assign labels within each cycle, I proposed an instrument-based procedure that relies on observable characteristics informative about worker and firm productivities. A natural question is whether the outcomes themselves could be used for this purpose: since outcomes depend on productivities, one might expect that, on average, the worker (or firm) with higher productivity would also generate higher observed outcomes. While this intuition is appealing, it ignores the exogeneity requirement: the labeling choice must not depend on the errors $\{\eta_{ij}\}$, and hence the outcomes can't be used as instruments.

To see why this matters, consider the following outcome-based labeling rule, which assigns labels according to the higher average outcome within a cycle.

Definition 9. (Outcome-Based Labeling) *In each cycle ℓ, the labels $(i_{\ell,\pi_{\ell,y}}, i'_{\ell,\pi_{\ell,y}})$ and $(j_{\ell,\pi_{\ell,y}}, j'_{\ell,\pi_{\ell,y}})$ are assigned so that*

$$y_{i_{\ell,\pi_{\ell,y}} j_{\ell,\pi_{\ell,y}}} + y_{i_{\ell,\pi_{\ell,y}} j'_{\ell,\pi_{\ell,y}}} > y_{i'_{\ell,\pi_{\ell,y}} j_{\ell,\pi_{\ell,y}}} + y_{i'_{\ell,\pi_{\ell,y}} j'_{\ell,\pi_{\ell,y}}},$$
$$y_{i_{\ell,\pi_{\ell,y}} j_{\ell,\pi_{\ell,y}}} + y_{i'_{\ell,\pi_{\ell,y}} j_{\ell,\pi_{\ell,y}}} > y_{i_{\ell,\pi_{\ell,y}} j'_{\ell,\pi_{\ell,y}}} + y_{i'_{\ell,\pi_{\ell,y}} j'_{\ell,\pi_{\ell,y}}}.$$

This induces the signs

$$\pi^{\alpha}_{\ell,y} = \operatorname{sign}\left(y_{i_\ell j_\ell} + y_{i_\ell j'_\ell} - y_{i'_\ell j_\ell} - y_{i'_\ell j'_\ell}\right), \quad \pi^{\psi}_{\ell,y} = \operatorname{sign}\left(y_{i_\ell j_\ell} + y_{i'_\ell j_\ell} - y_{i_\ell j'_\ell} - y_{i'_\ell j'_\ell}\right).$$

Because these signs depend directly on the realized errors η_{ij}, the resulting estimator $\hat{\beta}_{L,y}$ is no longer consistent. To illustrate, consider the case $\beta_0 = 0$, $\alpha_{i_\ell} - \alpha_{i'_\ell} = 1$, and $\psi_{j_\ell} - \psi_{j'_\ell} = 1$, with $\eta_{i'_\ell j_\ell} = \eta_{i'_\ell j'_\ell} = 0$. In this setting, the numerator and denominator of $\hat{\beta}_{L,y}$ simplify to

$$\frac{1}{L}\sum_{\ell=1}^{L}(\eta_{i_\ell j_\ell} - \eta_{i_\ell j'_\ell})\,\pi^{\alpha}_{\ell,y}\pi^{\psi}_{\ell,y} = \frac{1}{L}\sum_{\ell=1}^{L}|\eta_{i_\ell j_\ell} - \eta_{i_\ell j'_\ell}|\,\operatorname{sign}\big(\eta_{i_\ell j_\ell} + \eta_{i_\ell j'_\ell}\big),$$
$$\frac{1}{L}\sum_{\ell=1}^{L}(2\eta_{i_\ell j_\ell} - 3\eta_{i_\ell j'_\ell} - 1)\,\pi^{\alpha}_{\ell,y}\pi^{\psi}_{\ell,y} = \frac{1}{L}\sum_{\ell=1}^{L}(2\eta_{i_\ell j_\ell} - 3\eta_{i_\ell j'_\ell} - 1)\,\operatorname{sign}\big(\eta_{i_\ell j_\ell} - \eta_{i_\ell j'_\ell}\big)\,\operatorname{sign}\big(\eta_{i_\ell j_\ell} + \eta_{i_\ell j'_\ell}\big).$$

With independent error distributions, one can construct examples where $\hat{\beta}_{L,y}$ converges al-

most surely to values different from zero, despite $\beta_0 = 0$. For instance:

$$\eta_{i_\ell j_\ell} := \begin{cases} 3, & p = \frac{1}{4}, \\ -1, & p = \frac{3}{4}, \end{cases} \qquad \eta_{i_\ell j'_\ell} := \begin{cases} 2, & p = \frac{5}{7}, \\ -5, & p = \frac{2}{7}, \end{cases} \qquad \Longrightarrow \hat{\beta}_{L,y} \to^{a.s.} \frac{5}{9},$$

while with

$$\eta_{i_\ell j_\ell} := \begin{cases} 1, & p = \frac{3}{4}, \\ -3, & p = \frac{1}{4}, \end{cases} \qquad \eta_{i_\ell j'_\ell} := \begin{cases} 6, & p = \frac{1}{4}, \\ -2, & p = \frac{3}{4}, \end{cases} \qquad \Longrightarrow \hat{\beta}_{L,y} \to^{a.s.} -\frac{3}{5}.$$

These examples show that outcome-based labeling introduces asymptotic bias into $\hat{\beta}_{L,y}$, which can be either positive or negative depending on the error distribution. The problem arises because the labeling uses information contaminated by η_{ij}, violating exogeneity. This underscores the need for external instruments to guide label assignment.

F Alternating Least Squares in the Tukey model

Consider the problem:

$$(\hat{\boldsymbol{\alpha}}', \hat{\boldsymbol{\psi}}') \in \arg\min_{\boldsymbol{\alpha}', \boldsymbol{\psi}'} \sum_{(i,j) \in O_{IJ}} \left(y'_{ij} - \alpha'_i \psi'_j\right)^2 \quad \text{s.t.} \quad \|\boldsymbol{\psi}'\| = 1. \tag{F.1}$$

Problem (F.1) is non-convex but admits a unique global optimum, up to the normalisation $\|\boldsymbol{\psi}'\| = 1$, as long as the matching network G_{IJ} is connected.[2]

In practice, solve Problem (F.1) using alternating least squares. Starting from any $\boldsymbol{\psi}^{(0)} \neq \mathbf{0}$ with $\|\boldsymbol{\psi}^{(0)}\|_2 = 1$, and iterating for $t = 0, 1, 2, \ldots$, first set

$$\alpha_i^{(t+1)} := \frac{\sum_{j: D_{ij}=1} \psi_j^{(t)} y_{ij}}{\sum_{j: D_{ij}=1} \left(\psi_j^{(t)}\right)^2}, \qquad i = 1, \ldots, I, \tag{F.2}$$

$$\psi_j^{(t+1)} := \frac{\sum_{i: D_{ij}=1} \alpha_i^{(t+1)} y_{ij}}{\sum_{i: D_{ij}=1} \left(\alpha_i^{(t+1)}\right)^2}, \qquad j = 1, \ldots, J, \tag{F.3}$$

[2] The scale can be fixed in many equivalent ways; the L^2 constraint is convenient because it preserves closed-form updates.

and then rescale

$$c \coloneqq \left\|\boldsymbol{\psi}^{(t+1)}\right\|_2, \qquad \boldsymbol{\alpha}^{(t+1)} \leftarrow c\boldsymbol{\alpha}^{(t+1)}, \qquad \boldsymbol{\psi}^{(t+1)} \leftarrow \boldsymbol{\psi}^{(t+1)}/c.$$

Equations (F.2)–(F.3) are ordinary least-squares fits with a *single* regressor: holding $\boldsymbol{\psi}$ fixed, each α_i is the slope from regressing $\{y_{ij}\}_{j:D_{ij}=1}$ on $\{\psi_j\}_{j:D_{ij}=1}$; with the updated $\boldsymbol{\alpha}$ fixed, each ψ_j is obtained symmetrically. Because (F.1) is bi-convex, the alternating least squares scheme monotonically decreases the criterion and converges to the global minimizer when G_{IJ} is connected.

G Lemmas

Lemma 7. (Regularity of Cycle-Based Error Terms) *Under Assumptions 5.1 and 5.4, the derived error terms $\epsilon_{\Delta_1,\ell,\pi_\ell}$ and $\epsilon_{\Delta_2,\ell,\pi_\ell}$ satisfy $\mathbb{E}[\epsilon_{\Delta_1,\ell,\pi_\ell}] = \mathbb{E}[\epsilon_{\Delta_2,\ell,\pi_\ell}] = 0$, $\mathrm{Var}(\epsilon_{\Delta_1,\ell,\pi_\ell}) \geq C_1 > 0$, $\mathrm{Var}(\epsilon_{\Delta_2,\ell,\pi_\ell}) \geq C_2 > 0$, and $\mathbb{E}[|\epsilon_{\Delta_1,\ell,\pi_\ell}|^{2+\delta'}] \leq M_1 < \infty$, $\mathbb{E}[|\epsilon_{\Delta_2,\ell,\pi_\ell}|^{2+\delta}] \leq M_2 < \infty$ for some $\delta > 0$ and finite constants C_1, C_2, M_1, M_2.*

Proof. Recall

$$\epsilon_{\Delta_1,\ell,\pi_\ell} = \epsilon_{\Delta_1,\ell}\pi_\ell^\alpha\pi_\ell^\psi,$$
$$\epsilon_{\Delta_2,\ell,\pi_\ell} = \epsilon_{\Delta_2,\ell}\pi_\ell^\alpha\pi_\ell^\psi,$$

with

$$\epsilon_{\Delta_1,\ell} = \eta_{i_\ell j_\ell} - \eta_{i'_\ell j_\ell} - \eta_{i_\ell j'_\ell} + \eta_{i'_\ell j'_\ell},$$
$$\epsilon_{\Delta_2,\ell} = \theta_{i_\ell j_\ell}\eta_{i'_\ell j'_\ell} + \theta_{i'_\ell j'_\ell}\eta_{i_\ell j_\ell} - \theta_{i'_\ell j_\ell}\eta_{i_\ell j'_\ell} - \theta_{i_\ell j'_\ell}\eta_{i'_\ell j_\ell} + \eta_{i_\ell j_\ell}\eta_{i'_\ell j'_\ell} - \eta_{i'_\ell j_\ell}\eta_{i_\ell j'_\ell}.$$

Since α_i, ψ_j and β_0, and hence θ_{ij}, are bounded, Assumption 5.1 and standard properties of expectation, variance, and moments guarantee $\epsilon_{\Delta_1,\ell}$ and $\epsilon_{\Delta_2,\ell}$ to have mean zero, positive variance and uniformly bounded $2+\delta$ moments.

The term $\pi_\ell^\alpha\pi_\ell^\psi$ is a random variable with support $\{-1,+1\}$, hence with uniformly bounded moments.

By Assumption 5.4, $\epsilon_{\Delta_1,\ell,\pi_\ell}$ and $\epsilon_{\Delta_2,\ell,\pi_\ell}$ are the products of two independent random variables, one with mean zero, positive variance and uniformly bounded $2+\delta$ moments and the other with uniformly bounded moments. This implies they are also mean zero, have positive variance and bounded $2+\delta$ moments. □

Lemma 8. (Properties of Composite Error u_{ℓ,π_ℓ}) *Under Assumption 5.1 and 5.4, the composite error term u_{ℓ,π_ℓ} satisfies $\mathbb{E}[u_{\ell,\pi_\ell}] = 0$ and $\mathrm{Var}(u_{\ell,\pi_\ell}) \geq C_u > 0$, for a finite constant C_u. Its moments satisfy*

$\mathbb{E}[|u_{\ell,\pi_\ell}|^{2+\delta}] \leq M_u < \infty$ *for some* $\delta > 0$ *and finite constant* M_u*. Furthermore, the variance* $\mathrm{Var}(u_{\ell,\pi_\ell})$ *does not depend on the choice of the labeling.*

Proof. Recall:

$$u_{\ell,\pi_\ell} \coloneqq \epsilon_{\Delta_1,\ell,\pi_\ell} + \beta_0 \epsilon_{\Delta_2,\ell,\pi_\ell} = (\epsilon_{\Delta_1,\ell} + \beta_0 \epsilon_{\Delta_2,\ell}) \pi_\ell^\alpha \pi_\ell^\psi = u_\ell \pi_\ell^\alpha \pi_\ell^\psi$$

with $u_\ell \coloneqq \epsilon_{\Delta_1,\ell} + \beta_0 \epsilon_{\Delta_2,\ell}$.

Since α_i, ψ_j and β_0, and hence θ_{ij}, are bounded, Assumption 5.1 and standard properties of expectation, variance, and moments guarantee $\epsilon_{\Delta_1,\ell}$ and $\epsilon_{\Delta_2,\ell}$, and hence u_ℓ, to have mean zero, positive variance and uniformly bounded $2+\delta$ moments.

The term $\pi_\ell^\alpha \pi_\ell^\psi$ is a random variable with support $\{-1,+1\}$, hence with uniformly bounded moments.

By Assumption 5.4, u_{ℓ,π_ℓ} is the product of two independent random variables, one with mean zero, positive variance, and uniformly bounded $2+\delta$ moments, and the other with uniformly bounded moments. This implies they are also mean zero, have positive variance, and bounded $2+\delta$ moments.

Finally, note that

$$\mathrm{Var}(u_{\ell,\pi_\ell}) = \mathrm{Var}(u_\ell \pi_\ell^\alpha \pi_\ell^\psi) = \mathbb{E}[(u_\ell \pi_\ell^\alpha \pi_\ell^\psi - \mathbb{E}[u_\ell \pi_\ell^\alpha \pi_\ell^\psi])^2] = \mathbb{E}[(u_\ell \pi_\ell^\alpha \pi_\ell^\psi)^2] = \mathbb{E}[u_\ell^2] = \mathrm{Var}(u_\ell),$$

and hence the variance $\mathrm{Var}(u_{\ell,\pi_\ell})$ does not depend on the labelings π_ℓ^α and π_ℓ^ψ. □

H Proofs

H.1 Proof of Theorem 1

First, I show that the presence of a cycle is necessary for identifying β_0. Suppose, by contradiction, that G_{IJ} contains no cycles. If it is connected, it is a tree. If it is disconnected, it decomposes into multiple tree components; focusing on any one component suffices, as the same argument applies to each. A tree with M nodes has $M-1$ edges, corresponding to $M-1$ observed matches. The model includes M node-specific parameters (one for each worker or firm) and the interaction parameter β_0, for a total of $M+1$ unknowns.

Consider the normalization $\alpha_{i_0} = 0$, and fix an arbitrary value of β_0. In a tree, there is always at least one edge connecting a known node to an unknown one. The equation on that edge,

$\theta_{ij} = \alpha_i + \psi_j + \beta\alpha_i\psi_j$, becomes linear in the unknown parameter. For instance, if α_i is known, then

$$\theta_{ij} - \alpha_i = \psi_j(1 + \beta\alpha_i) \quad \Longrightarrow \quad \psi_j = \frac{\theta_{ij} - \alpha_i}{1 + \beta\alpha_i}$$

which is unique provided $1 + \beta\alpha_i \neq 0$. Removing this edge shrinks the tree, and one repeats until all productivities are solved in terms of β_0. Varying β_0 yields a continuum of distinct solutions $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$, all consistent with the same observed systematic outcome. Since β_0 can be chosen freely without contradiction (apart from that one value per edge), it is not identified when G_{IJ} is a forest. Therefore, the existence of at least one cycle in G_{IJ} is necessary for the identification of β_0.

Consider such a cycle, of length $2K$ with $K \geq 2$, and relabel the corresponding outcomes $\theta_{i_1,j_1}, \theta_{i_2,j_1}, \ldots \theta_{i_K,j_K}, \theta_{i_1,j_K}$ as $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(2K)}$ in the order encountered.

Define

$$u_i \coloneqq 1 + \beta_0\, \alpha_i, \qquad v_j \coloneqq 1 + \beta_0\, \psi_j.$$

Then for each (i, j),

$$1 + \beta_0\theta_{ij} = 1 + \beta_0(\alpha_i + \psi_j + \beta_0\alpha_i\psi_j) = (1 + \beta_0\alpha_i)(1 + \beta_0\psi_j) = u_i v_j.$$

Hence, traversing the cycle gives

$$\prod_{k=1}^{K}\left(1 + \beta_0\theta^{(2k-1)}\right) = \prod_{k=1}^{K} u_{i_k} v_{j_k} = \prod_{k=1}^{K} u_{i_{k+1}} v_{j_k} = \prod_{k=1}^{K}\left(1 + \beta_0\theta^{(2k)}\right),$$

where $i_{K+1} = i_1$.

Expand the products in elementary symmetric sums:

$$\prod_{k=1}^{K}(1 + \beta_0\theta^{(2k-1)}) = \sum_{r=0}^{K} S_r^{\text{odd}}\beta_0^r,$$
$$\prod_{k=1}^{K}(1 + \beta_0\theta^{(2k)}) = \sum_{r=0}^{K} S_r^{\text{even}}\beta_0^r$$

where

$$S_r^{\text{odd}} = \sum_{1 \leq k_1 < \cdots < k_r \leq K} \theta^{(2k_1-1)} \cdots \theta^{(2k_r-1)}, \quad S_r^{\text{even}} = \sum_{1 \leq k_1 < \cdots < k_r \leq K} \theta^{(2k_1)} \cdots \theta^{(2k_r)},$$

and the notation $\sum_{1 \leq k_1 < \cdots < k_r \leq K}$ denotes the sum over all $\binom{K}{r}$ ways to choose r distinct indices

$1 \le k_1 < \cdots < k_r \le K$, each counted exactly once. Then, for each such tuple $(k_1, \ldots, k_r)$, the term $\theta^{(2k_1-1)} \ldots \theta^{(2k_r-1)}$ is the product of the corresponding r odd-indexed θ-values, and analogously for $\theta^{(2k_1)} \cdots \theta^{(2k_r)}$.

Subtracting these expansions eliminates the constant term ($r = 0$), leaving

$$\sum_{r=1}^{K} \left(S_r^{\text{odd}} - S_r^{\text{even}}\right) \beta_0^r = 0.$$

Set

$$\Delta_r = S_r^{\text{odd}} - S_r^{\text{even}}, \qquad r = 1, 2, \ldots, K.$$

to obtain the degree-K polynomial

$$\Delta_1 \beta_0 + \Delta_2 \beta_0^2 + \cdots + \Delta_K \beta_0^K = 0.$$

By construction, $\beta_0 = 0$ is always a root, but does not recover the interaction parameter. Factoring it out, one discards the trivial solution and focuses on the roots of

$$\Delta_1 + \Delta_2 \beta_0 + \Delta_3 \beta_0^2 + \cdots + \Delta_K \beta_0^{K-1} = 0.$$

If this polynomial has an additional root $\beta_0 = 0$, that is a true admissible value for the interaction parameter, differently from the purely algebraic root as above.

Uniqueness of the root is guaranteed only when $K = 2$. Any cycle with $K > 2$ does not allow for point identification of β_0, since it restricts its value to a set with $K - 1$ elements.

When $K = 2$, the derived formulas are

$$\begin{aligned} \Delta_1 &= \theta^{(1)} + \theta^{(3)} - \theta^{(2)} - \theta^{(4)} = \theta_{i_1 j_1} + \theta_{i_2 j_2} - \theta_{i_1 j_2} - \theta_{i_2 j_1}, \\ \Delta_2 &= \theta^{(1)}\theta^{(3)} - \theta^{(2)}\theta^{(4)} = \theta_{i_1 j_1}\theta_{i_2 j_2} - \theta_{i_1 j_2}\theta_{i_2 j_1}, \end{aligned}$$

and the polynomial becomes

$$(\theta_{i_1 j_1} + \theta_{i_2 j_2} - \theta_{i_1 j_2} - \theta_{i_2 j_1}) + \beta_0(\theta_{i_1 j_1}\theta_{i_2 j_2} - \theta_{i_1 j_2}\theta_{i_2 j_1}) = 0.$$

β_0 is hence identified:

$$\beta_0 = -\frac{\theta_{i_1 j_1} + \theta_{i_2 j_2} - \theta_{i_1 j_2} - \theta_{i_2 j_1}}{\theta_{i_1 j_1}\theta_{i_2 j_2} - \theta_{i_1 j_2}\theta_{i_2 j_1}}$$

where Assumption 1 guarantees the denominator being different from 0. In fact,

$$\theta_{i_1 j_1}\theta_{i_2 j_2} - \theta_{i_1 j_2}\theta_{i_2 j_1} = \alpha_{i_1}\psi_{j_1} + \alpha_{i_2}\psi_{j_2} - \alpha_{i_2}\psi_{j_1} - \alpha_{i_1}\psi_{j_2} = -(\alpha_{i_1} - \alpha_{i_2})(\psi_{j_1} - \psi_{j_2}).$$

H.2 Proof of Theorem 2

Sufficiency. Consider the model $\theta_{ij} = \alpha_i + \psi_j + \beta_0\alpha_i\psi_j$, with β_0 known. If $\beta_0 = 0$, this reduces to the usual TWFE model. If $\beta_0 \neq 0$, rewrite

$$\theta_{ij} = \alpha_i + \psi_j + \beta_0\alpha_i\psi_j \Longrightarrow 1 + \beta_0\theta_{ij} = (1 + \beta_0\alpha_i)(1 + \beta_0\psi_j)$$

and set

$$\alpha'_i = 1 + \beta_0\alpha_i, \quad \psi'_j = 1 + \beta_0\psi_j, \quad \theta'_{ij} = 1 + \beta_0\theta_{ij},$$

which give

$$\theta'_{ij} = \alpha'_i\psi'_j$$

where θ'_{ij} is identified. Clearly, $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ are identified if α'_i and ψ'_j are identified for all i and j.

Since $\alpha_{i_0} = 0$, $\alpha'_{i_0} = 1 + \beta_0\alpha_i = 1$. If workers i and i' both match with firm j, then $\frac{\theta'_{ij}}{\theta'_{i'j}} = \frac{\alpha'_i\psi'_j}{\alpha'_{i'}\psi'_j} = \frac{\alpha'_i}{\alpha'_{i'}}$, and the ratio is hence identified. Since G_{IJ} is connected, for any worker i there exists a path

$$i_0 - j_1 - i_1 - j_2 - \ldots - j_m - i.$$

Applying the ratio argument at each firm along the path:

$$\frac{\alpha'_i}{\alpha'_{i_0}} = \frac{\alpha'_i}{\alpha'_{i_{m-1}}} \cdot \frac{\alpha'_{i_{m-1}}}{\alpha'_{i_{m-2}}} \cdots \frac{\alpha'_{i_1}}{\alpha'_{i_0}} = \prod_{k=1}^{m} \frac{\theta'_{i_k j_k}}{\theta'_{i_{k-1} j_k}}$$

which pins down α_i uniquely since $\alpha'_{i_0} = 1$.

Once all α_i are known, any observed edge gives $\psi'_j = \frac{\theta'_{ij}}{\alpha'_i}$, uniquely determining each ψ_j.

Necessity. If G_{IJ} is disconnected, it splits into at least two components. On each component, one may independently rescale the local α'_i's and ψ'_j's without changing the product $\alpha'_i\psi'_j$ on that component. Thus, parameters across different components cannot be jointly identified.

H.3 Proof of Theorem 3

First, note that identification in the BLM model requires two normalizations ($a_{j_0} = b_{j_0} = 1$). To see why, consider parameter vectors $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\alpha})$ and $(\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\alpha}')$ with

$$\alpha_i' = c_1\alpha_i + c_2, \qquad b_j' = \frac{b_j}{c_1}, \qquad a_j' = a_j - \frac{b_j c_2}{c_1},$$

for $c_1, c_2 \in \mathbb{R}$ with $c_1 \neq 0$, and note that the BLM models with $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\alpha})$ and $(\boldsymbol{a}', \boldsymbol{b}', \boldsymbol{\alpha}')$ are observationally equivalent. To pin down the two degrees of freedom, two normalizations are needed.

Sufficiency. To prove sufficiency of Assumption 3, proceed along the Seed-and-Snowballs construction. Let j_0 indicate the seed with the normalization $a_{j_0} = b_{j_0} = 1$.

Assume that, at some stage $n \geq 0$, all parameters of firms in S_n^J are known. Hence, identify parameters associated with workers in S_n^I and firms in S_{n+1}^J as follows:

(i) *Firms → workers.* For each worker $i \in S_n^I$, there exists a firm $j \in S_n^J$ such that $D_{ij} = 1$ and $b_j \neq 0$, and hence

$$\alpha_i = \frac{\theta_{ij} - a_j}{b_j}.$$

(ii) *Workers → firms.* For each firm $j \in S_{n+1}^J$ there exist two workers $i, i' \in S_n^I$ such that $D_{ij} = D_{i'j} = 1$ and $\alpha_i \neq \alpha_{i'}$, and hence

$$\begin{cases} \theta_{ij} = a_j + b_j\alpha_i \\ \theta_{i'j} = a_j + b_j\alpha_{i'} \end{cases} \iff \begin{cases} b_j = \frac{\theta_{i'j} - \theta_{ij}}{\alpha_{i'} - \alpha_i} \\ a_j = \theta_{ij} - \frac{\theta_{i'j} - \theta_{ij}}{\alpha_{i'} - \alpha_i}\alpha_i. \end{cases}$$

Because the snowball reaches (I, J) after finitely many steps, every α_i, a_j, b_j is eventually determined. Together with the two normalizations, this proves sufficiency.

Necessity. If G_{IJ} is disconnected, it splits into at least two components. On each component, one may independently rescale the local α_i's, a_j's, and b_j's without changing the θ_{ij}'s on that component. Thu,s parameters across different components cannot be jointly identified.

Suppose now that G_{IJ} is connected but does not satisfy the Seed-and-Snowballs property. Fix a seed j_0 and run the algorithm until it stalls at some finite step n. Consider

$$S_n^J \subset J, \qquad S_n^I \subset I.$$

By construction, every firm in the outer set

$$J^{\text{out}} = J \setminus S_n^J$$

is linked to at most one worker already in S_n^I. If such a link exists, call its worker the bridge of the firm. Write $I^{\text{out}} = I \setminus S_n^I$ for the outer workers.

Choose an outer firm $j_\lambda \in J^{\text{out}}$ and let $i_\lambda \in S_n^I$ be its bridge (it exists because G_{IJ} is connected). Redefine

$$b_{j_\lambda}^\lambda = \lambda\, b_{j_\lambda}, \qquad a_{j_\lambda}^\lambda = \theta_{i_\lambda j_\lambda} - b_{j_\lambda}^\lambda \alpha_{i_\lambda}$$

and note that this preserves the value of $\theta_{i_\lambda j_\lambda}$ while altering $(a_{j_\lambda}, b_{j_\lambda})$.

For every worker $i \in I^{\text{out}}$ that is connected to at least one firm whose (a, b) have already been (re)defined, pick one such firm, denote it by $j(i)$, and set

$$\alpha_i^\lambda = \frac{\theta_{ij(i)} - a_{j(i)}^\lambda}{b_{j(i)}^\lambda},$$

where as long as λ is such that $b_{j(i)}^\lambda \neq 0$, α_i^λ is well-defined, and $\theta_{ij(i)}$ preserved.

Whenever a still-unprocessed firm $j \in J^{\text{out}}$ is observed with two workers whose productivities are already determined and distinct, say i and i', solve

$$\begin{cases} \theta_{ij} = a_j^\lambda + b_j^\lambda \alpha_i^\lambda \\ \theta_{i'j} = a_j^\lambda + b_j^\lambda \alpha_{i'}^\lambda \end{cases} \iff \begin{cases} b_j^\lambda = \frac{\theta_{i'j} - \theta_{ij}}{\alpha_{i'}^\lambda - \alpha_i^\lambda} \\ a_j^\lambda = \theta_{ij} - b_j^\lambda \alpha_i^\lambda \end{cases}$$

where the right-hand sides depend smoothly on λ via the α's.

Iteratively repeat these two steps until all parameters have been recovered, and in case of additional degree of freedom (because again at a certain point remain only firms with one worker whose productivity has already been determined), just set the parameters at their true value. For $\lambda = 1$ the procedure returns the original parameters. For $\lambda \neq 1$ it changes at least b_{j_λ}, so the parameter vector is different. Because the choice of λ is arbitrary, the data are compatible with infinitely many distinct $(\boldsymbol{\alpha}^\lambda, \boldsymbol{a}^\lambda, \boldsymbol{b}^\lambda)$, establishing lack of identification.

H.4 Proof of Proposition 1

Let $(f_m, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ be any representation with $\theta_{ij} = f_m(\alpha_i, \psi_j)$. Choose any strictly increasing transformations g_α and g_ψ, and define

$$\alpha_i' = g_\alpha(\alpha_i), \quad \psi_j' = g_\psi(\psi_j), \quad f_m'(\alpha, \psi) = f_m(g_\alpha^{-1}(\alpha), g_\psi^{-1}(\psi)).$$

Then,

$$f_m'(\alpha_i', \psi_j') = f_m(g_\alpha^{-1}(g_\alpha(\alpha_i)), g_\psi^{-1}(g_\psi(\psi_j))) = f_m(\alpha_i, \psi_j),$$

and hence $(f_m, \boldsymbol{\alpha}, \boldsymbol{\psi}, G_{IJ})$ and $(f_m', \boldsymbol{\alpha}', \boldsymbol{\psi}', G_{IJ})$ are observationally equivalent.

H.5 Proof of Theorem 4

I prove the result for the ranking of $\boldsymbol{\alpha}$; the argument for $\boldsymbol{\psi}$ is analogous.

Sufficiency. Fix any two workers i and i'. By Assumption 4 there exists a firm j such that $(i, j), (i', j) \in O_{IJ}$. Because f_m is strictly increasing in each argument, the sign of $\theta_{ij} - \theta_{i'j}$ reveals whether α_i exceeds $\alpha_{i'}$. Collecting these pairwise comparisons for all (i, i') constructs a complete, transitive ranking of the workers, which is therefore identified.

Necessity. Suppose, to reach a contradiction, that Assumption 4 fails. Then there exist two distinct workers i and i' such that no firm j satisfies $(i, j), (i', j) \in O_{IJ}$. Suppose $\alpha_i > \alpha_{i''}$ and $\alpha_{i'} > \alpha_{i''}$, for any $i'' \neq i$, $i'' \neq i'$. Also, suppose that for any pair of firms j and j' such that $(i, j), (i', j') \in O_{IJ}$, it holds that $\psi_j > \psi_{j'}$, and $\theta_{ij} > \theta_{i'j'}$. Now consider the bivariate isotonic matrix where workers i and i' occupy the first two rows. Swapping the first and second rows, and filling in the missing entries with appropriate values, preserves both row- and column-wise monotonicity. This yields a second isotonic matrix that is observationally equivalent to the original but reverses the relative order of i and i'. Therefore, the ranking of workers is not point-identified, implying that Assumption 4 is necessary.

H.6 Proof of Theorem 5

Rewrite the estimator as

$$\begin{aligned}\hat{\beta}_{L,\pi} &= -\frac{\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{1,\ell,\pi_\ell}}{\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi_\ell}} \\ &= -\frac{\frac{1}{L}\sum_{\ell=1}^{L}\Delta_{1,\ell,\pi_\ell}+\frac{1}{L}\sum_{\ell=1}^{L}\epsilon_{\Delta_1,\ell,\pi_\ell}}{\frac{1}{L}\sum_{\ell=1}^{L}\Delta_{2,\ell,\pi_\ell}+\frac{1}{L}\sum_{\ell=1}^{L}\epsilon_{\Delta_2,\ell,\pi_\ell}} \\ &= -\frac{\beta_0\frac{1}{L}\sum_{\ell=1}^{L}(\alpha_{i_\ell}-\alpha_{i'_\ell})(\psi_{j_\ell}-\psi_{j'_\ell})\pi_\ell^{\alpha}\pi_\ell^{\psi}+\frac{1}{L}\sum_{\ell=1}^{L}\epsilon_{\Delta_1,\ell,\pi_\ell}}{-\frac{1}{L}\sum_{\ell=1}^{L}(\alpha_{i_\ell}-\alpha_{i'_\ell})(\psi_{j_\ell}-\psi_{j'_\ell})\pi_\ell^{\alpha}\pi_\ell^{\psi}+\frac{1}{L}\sum_{\ell=1}^{L}\epsilon_{\Delta_2,\ell,\pi_\ell}}.\end{aligned}$$

Consider the two collections

$$\{\epsilon_{\Delta_1,L,\ell,\pi_\ell}:1\le\ell\le L\}\quad\text{and}\quad\{\epsilon_{\Delta_2,L,\ell,\pi_\ell}:1\le\ell\le L\}.$$

By Lemma 7, each entry in these triangular arrays has mean zero and uniformly bounded $2+\delta$ moments. Hence, by the Strong Law of Large Numbers for triangular arrays, conclude

$$\frac{1}{L}\sum_{\ell=1}^{L}\epsilon_{\Delta_1,\ell,\pi_\ell}\to^{a.s.}0\quad\text{and}\quad\frac{1}{L}\sum_{\ell=1}^{L}\epsilon_{\Delta_2,\ell,\pi_\ell}\to^{a.s.}0$$

as $L\to\infty$, a limit guaranteed by Assumption 5.2.

Applying the Continuous Mapping Theorem, and using the facts that $\mu_L\neq 0$ by Assumption 5.3 and $c_\pi\neq 0$ by Assumption 5.4, yields the desired result:

$$\hat{\beta}_L=-\frac{\beta_0(\mu_L c_\pi+o_{a.s.}(1))+o_{a.s.}(1)}{-\mu_L c_\pi+o_{a.s.}(1))+o_{a.s.}(1)}\to^{a.s.}-\frac{\beta_0\mu_L c_\pi}{-\mu_L c_\pi}=\beta_0.$$

H.7 Proof of Theorem 6

Define the functions

$$g(\Delta_1,\Delta_2,\beta):=\Delta_1+\beta\Delta_2,\qquad\text{and}\qquad\bar{g}_{L,\pi}(\beta):=\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{1,\ell,\pi_\ell}+\beta\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi_\ell}.$$

Write the estimator $\hat{\beta}_{L,\pi}$ as the minimizer of $\left[\bar{g}_{L,\pi}(\beta)\right]^2$:

$$\hat{\beta}_{L,\pi}=\arg\min_{\beta\in B}\left[\bar{g}_{L,\pi}(\beta)\right]^2=-\frac{\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{1,\ell,\pi_\ell}}{\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi_\ell}},$$

for a compact parameter space $B \subset \mathbb{R}$ containing β_0, and note that it implies $0 = \bar{g}_{L,\pi}(\hat{\beta}_{L,\pi})$.

Consider a first-order Taylor expansion of $\bar{g}_{L,\pi}(\hat{\beta}_{L,\pi})$ around the true parameter β_0:

$$0 = \bar{g}_{L,\pi}(\hat{\beta}_{L,\pi}) = \bar{g}_{L,\pi}(\beta_0) + \frac{\partial \bar{g}_{L,\pi}(\tilde{\beta})}{\partial \beta}(\hat{\beta}_{L,\pi} - \beta_0),$$

where $\tilde{\beta}$ is an intermediate value between $\hat{\beta}_L$ and β_0, and $\frac{\partial \bar{g}_{L,\pi}(\beta)}{\partial \beta} = \frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{2,\ell,\pi_\ell}$. Rearranging gives:

$$\begin{aligned}
\sqrt{L}(\hat{\beta}_{L,\pi} - \beta_0) &= \frac{\sqrt{L}\bar{g}_{L,\pi}(\beta_0)}{-\frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{2,\ell,\pi_\ell}} \\
&= \frac{\sqrt{L}(\frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{1,\ell,\pi_\ell} + \beta_0 \frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{2,\ell,\pi_\ell})}{-\frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{2,\ell,\pi_\ell}} \\
&= \frac{\sqrt{L}(\frac{1}{L}\sum_{\ell=1}^{L} \Delta_{1,\ell,\pi_\ell} + \beta_0 \frac{1}{L}\sum_{\ell=1}^{L} \Delta_{2,\ell,\pi_\ell}) + \sqrt{L}\frac{1}{L}\sum_{\ell=1}^{L}(\epsilon_{\Delta_1,\ell,\pi_\ell} + \beta_0 \epsilon_{\Delta_2,\ell,\pi_\ell})}{-\frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{2,\ell,\pi_\ell}} \\
&= \frac{\sqrt{L}(-\beta_0 \frac{1}{L}\sum_{\ell=1}^{L} \Delta_{2,\ell,\pi_\ell} + \beta_0 \frac{1}{L}\sum_{\ell=1}^{L} \Delta_{2,\ell,\pi_\ell}) + \frac{1}{\sqrt{L}}\sum_{\ell=1}^{L} u_{\ell,\pi_\ell}}{-\frac{1}{L}\sum_{\ell=1}^{L} \hat{\Delta}_{2,\ell,\pi_\ell}} \\
&= \frac{\frac{1}{\sqrt{L}}\sum_{\ell=1}^{L} u_{\ell,\pi_\ell}}{-\frac{1}{L}\sum_{\ell=1}^{L} \Delta_{2,\ell,\pi_\ell} - \frac{1}{L}\sum_{\ell=1}^{L} \epsilon_{\Delta_2,\ell}} \\
&= \frac{\frac{1}{\sqrt{L}\mu_L c_\pi}\sum_{\ell=1}^{L} u_{\ell,\pi_\ell}}{-\frac{1}{L\mu_L c_\pi}\sum_{\ell=1}^{L} \Delta_{2,\ell,\pi_\ell} - \frac{1}{L\mu_L c_\pi}\sum_{\ell=1}^{L} \epsilon_{\Delta_2,\ell}}.
\end{aligned}$$

where in the last line numerator and denominator are divided by $\mu_L c_\pi$, different from zero by Assumptions 5.3 and 5.4.

Consider the triangular array

$$\{u_{L,\ell,\pi_\ell} \coloneqq \frac{u_{\ell,\pi_\ell}}{\mu_L c_\pi} : 1 \le \ell \le L\}.$$

By Lemma 8, each term u_{L,ℓ,π_ℓ} has mean zero and uniformly bounded $2+\delta$ moments. Hence, by the Lyapunov Central Limit Theorem for triangular arrays,

$$\frac{1}{\sqrt{L}}\frac{\mu_L c_\pi}{\sigma_{u,L,\pi}}\sum_{\ell=1}^{L}\frac{u_{\ell,\pi_\ell}}{\mu_L c_\pi} \xrightarrow{d} \mathcal{N}(0,1).$$

Recall that $\sigma_{u,L,\pi} = \sqrt{\frac{1}{L}\sum_{\ell=1}^{L} \mathrm{Var}(u_{\ell,\pi_\ell})}$. Lemma 8 proves that each $\mathrm{Var}(u_{\ell,\pi_\ell})$ does not depend on the labeling of the corresponding cycle, and hence neither $\sigma_{u,L,\pi}$ does.

Hence consider:

$$\sqrt{L}\frac{\mu_L c_\pi}{\sigma_{u,L,\pi}}\left(\hat{\beta}_{L,\pi}-\beta_0\right)=\frac{\frac{1}{\sqrt{L}}\frac{\mu_L c_\pi}{\sigma_{u,L,\pi}}\sum_{\ell=1}^{L}\frac{u_\ell}{\mu_L c_\pi}}{-\frac{1}{L\mu_L c_\pi}\sum_{\ell=1}^{L}\Delta_{2,\ell,\pi_\ell}-\frac{1}{L\mu_L c_\pi}\sum_{\ell=1}^{L}\epsilon_{\Delta_2,\ell}}.$$

and note that the argument used in Theorem 5 proves that the denominator converges to 1 almost surely. Slutsky's theorem hence allows to conclude

$$\sqrt{L}\frac{\mu_L c_\pi}{\sigma_{u,L,\pi}}\left(\hat{\beta}_L-\beta_0\right)\xrightarrow{d}\mathcal{N}\Big(0,1\Big).$$

H.8 Proof of Proposition 2

Note that

$$\begin{aligned}\hat{u}_{\ell,\pi}&=\hat{\Delta}_{1,\ell,\pi}+\hat{\beta}_{L,\pi}\hat{\Delta}_{2,\ell,\pi}\\&=\Delta_{1,\ell,\pi_\ell}+\epsilon_{\Delta_1,\ell,\pi_\ell}+\hat{\beta}_{L,\pi}(\Delta_{2,\ell,\pi_\ell}+\epsilon_{\Delta_2,\ell,\pi_\ell})\\&=\underbrace{\Delta_{1,\ell,\pi_\ell}+\beta_0\Delta_{2,\ell,\pi_\ell}}_{=0}+\underbrace{\epsilon_{\Delta_1,\ell,\pi_\ell}+\beta_0\epsilon_{\Delta_2,\ell,\pi_\ell}}_{u_{\ell,\pi}}+(\hat{\beta}_{L,\pi}-\beta_0)\hat{\Delta}_{2,\ell,\pi}\end{aligned}$$

and hence

$$\hat{u}_{\ell,\pi}^2=u_{\ell,\pi}^2+2u_{\ell,\pi}(\hat{\beta}_{L,\pi}-\beta_0)\hat{\Delta}_{2,\ell,\pi}+(\hat{\beta}_{L,\pi}-\beta_0)^2\hat{\Delta}_{2,\ell,\pi}^2.$$

Since $\sigma_{u,L}^2=\frac{1}{L}\sum_{\ell=1}^{L}u_{\ell,\pi_\ell}^2$, to prove the result I need to prove that

$$\begin{aligned}&\frac{1}{L}\sum_{\ell=1}^{L}\left(2u_{\ell,\pi}(\hat{\beta}_{L,\pi}-\beta_0)\hat{\Delta}_{2,\ell,\pi}+(\hat{\beta}_{L,\pi}-\beta_0)^2\hat{\Delta}_{2,\ell,\pi}^2\right)=\\=&(\hat{\beta}_{L,\pi}-\beta_0)\frac{1}{L}\sum_{\ell=1}^{L}\left(2u_{\ell,\pi}\hat{\Delta}_{2,\ell,\pi}\right)+(\hat{\beta}_{L,\pi}-\beta_0)^2\sum_{\ell=1}^{L}\left(\hat{\Delta}_{2,\ell,\pi}^2\right)\xrightarrow{p}0.\end{aligned}$$

Assumption 5 and Theorem 5 guarantee $(\hat{\beta}_{L,\pi}-\beta_0)=o_p(1)$, $\frac{1}{L}\sum_{\ell=1}^{L}u_{\ell,\pi}\hat{\Delta}_{2,\ell,\pi}=O_p(1)$, and $\frac{1}{L}\sum_{\ell=1}^{L}\hat{\Delta}_{2,\ell,\pi}^2=O_p(1)$, and hence the result follows.

H.9 Proof of Proposition 3

Define $A_\ell \coloneqq (\alpha_{i_\ell} - \alpha_{i'_\ell})(\psi_{j_\ell} - \psi_{j'_\ell})$. It suffices to show that there exists a constant c_π with $|c_\pi| > C_\pi > 0$ such that

$$\mu_L c_\pi - \frac{1}{L}\sum_{\ell=1}^{L} A_\ell \pi^\alpha_{\ell,z}\pi^\psi_{\ell,z} \to^{a.s.} 0,$$

and that the collection $\{\pi^\alpha_{\ell,z}, \pi^\psi_{\ell,z}\}_{\ell=1}^L$ is independent of the error terms $\{\eta_{ij}\}$.

First, the instruments $\boldsymbol{z}^{\boldsymbol{\alpha}}$ and $\boldsymbol{z}^{\boldsymbol{\psi}}$ are deterministic. Hence, the induced signs $\{\pi^\alpha_{\ell,z}, \pi^\psi_{\ell,z}\}$, being deterministic functions of the instruments, are independent of the error terms $\{\eta_{ij}\}$, exogenous by Assumption 5.1.

Next, note that

$$\frac{1}{L}\sum_{\ell=1}^{L}(A_\ell - \mu_L)(\pi^\alpha_{\ell,z}\pi^\psi_{\ell,z} - \overline{\pi^\alpha\pi^\psi}) = \frac{1}{L}\sum_{\ell=1}^{L} A_\ell \pi^\alpha_{\ell,z}\pi^\psi_{\ell,z} - \mu_L\overline{\pi^\alpha\pi^\psi},$$

so that

$$\frac{1}{L}\sum_{\ell=1}^{L} A_\ell \pi^\alpha_{\ell}\pi^\psi_{\ell} = \frac{1}{L}\sum_{\ell=1}^{L}(A_\ell - \mu_L)(\pi^\alpha_{\ell,z}\pi^\psi_{\ell,z} - \overline{\pi^\alpha\pi^\psi}) + \mu_L\overline{\pi^\alpha\pi^\psi}.$$

Similarly,

$$\frac{1}{L}\sum_{\ell=1}^{L}(\pi^\alpha_{\ell,z} - \overline{\pi^\alpha})(\pi^\psi_{\ell,z} - \overline{\pi^\psi}) = \overline{\pi^\alpha\pi^\psi} - \overline{\pi^\alpha}\,\overline{\pi^\psi},$$

which implies

$$\overline{\pi^\alpha\pi^\psi}_L = \frac{1}{L}\sum_{\ell=1}^{L}(\pi^\alpha_{\ell,z} - \overline{\pi^\alpha})(\pi^\psi_{\ell,z} - \overline{\pi^\psi}) + \overline{\pi^\alpha}\,\overline{\pi^\psi}.$$

Substituting this expression yields

$$\frac{1}{L}\sum_{\ell=1}^{L} A_\ell \pi^\alpha_{\ell,z}\pi^\psi_{\ell,z} = \frac{1}{L}\sum_{\ell=1}^{L}(A_\ell - \mu_L)(\pi^\alpha_{\ell,z}\pi^\psi_{\ell,z} - \overline{\pi^\alpha\pi^\psi}) + \mu_L\frac{1}{L}\sum_{\ell=1}^{L}(\pi^\alpha_{\ell,z} - \overline{\pi^\alpha})(\pi^\psi_{\ell,z} - \overline{\pi^\psi}) + \mu_L\overline{\pi^\alpha}\,\overline{\pi^\psi}.$$

Assumption 5.2 guarantees $L \to \infty$. In the limit, the first term on the right-hand side is nonnegative by Assumption 6.3; the second is nonnegative by Assumptions 6.2 and 5.3; and the third is bounded away from zero by Assumptions 6.1 and 5.3. The entire expression is hence

bounded away from zero as well, and Assumption 5.4 holds with $c_\pi \geq c_\alpha c_\psi$.