Team Networks with Partially Observed Links *

Yang Xu †

Department of Economics and Nuffield College, University of Oxford

February 27, 2025

Abstract

This article studies a linear production model in team networks with missing links. In the model, heterogeneous workers produce jointly and repeatedly within various teams. However, a subset of links is truncated and hence unobserved. To account for the fact that the network is only partially observed, we provide a Generalized Method of Moments estimator under normal errors. Furthermore, we propose a test for link truncation, which is free from distribution assumption on the error term. Applying the truncation-robust estimator to academic publication data reveals and corrects a substantial downward bias in the scaling factor that aggregates individual fixed effects into team-specific fixed effects, suggesting an underestimated collaboration premium.

Keywords: Team Production, Endogenous Network, Link Truncation

1 Introduction

In economic networks, nodes or agents generally have individual-specific heterogeneity that can be modeled through a fixed-effect approach. At the same time, links or edges also exhibit unobserved heterogeneity that may relate to the node-level heterogeneity. In settings such as team production, networks provide a structured approach to studying complex relational dependencies between heterogeneous workers and heterogeneous teams (Bonhomme (2021)). However, as is the case with panel models, the standard random sampling assumption sometimes fails to hold in network models (Chandrasekhar and Lewis (2011)). Notably, in a partially observed network, only a subset of links may be selectively sampled and observed, while the rest remain unobserved. Therefore, we may need new econometric tools to correct for the sampling and truncation bias that are present in this endogenous network. This article proposes a truncation-robust estimator in a linear production model under normal errors, highlighting the importance of distinguishing the observed network from the latent network.

A stylized example of a network before and after link truncation is presented in Figure 1. We focus on truncation at the link level. Multiple links are allowed between pairs of vertices to allow for repeated teamwork within teams, resulting in a multigraph representation. In an ideal world, one could observe the latent network $\mathcal{G}^*$. However, in reality, one does not observe the dashed links that are missing from the truncated graph $\mathcal{G}$. This is different from censoring: if an observation Y_ℓ^* falls under (or above) a censoring threshold. a censored value (for example, zero

*We are extremely grateful for the generous guidance of Martin Weidner. We also thank Isaiah Andrews, Helmut Farbmacher, Raffaella Giacomini, Peter Hull, Martin Mugnier, Taisuke Otsu, Xiaoxia Shi, and Frank Windmeijer for helpful discussion and comments. We are responsible for all errors.

†yang.xu@economics.ox.ac.uk

dollar in the case of censored wage) is entered to replace its latent value Y_ℓ^*. In our setting, the truncated links are omitted completely and the number of the missing links is also unknown.

(a) Latent Graph $\mathcal{G}^*$ (b) Observed / Truncated Graph $\mathcal{G}$

Notes: In this stylized example, the network has three nodes - A, B, and C. Solid lines represent links with non-negative outcome. Dashed lines represent links that have negative outcomes and are hence unobserved. The left figure plots a latent network with five links, whereas the right figure plots a partially observed network consisted of three links only, a strict sub-graph of the latent graph.

Figure 1: Stylized Example

Teams arise naturally in economic production, in which heterogeneous workers collaborate within different teams (Jones (2021)). Examples include patent production (Jaravel, Petkova, and Bell (2018); S. P. Kerr and W. R. Kerr (2018)), software development (El-Komboz, Fackler, and Goldbeck (2024)), and professional sports (Devereux (2018)), among others. In academia, collaboration has increasingly become an integral part of scientific research in many fields spanning from physical sciences to social sciences (Wuchty, Jones, and Uzzi (2007)). Researchers work both individually and collaboratively with fellow academics in teams of various sizes. As an innovative production process, research production is almost invariably subject to uncertainty as researchers ask and study novel questions. This means that while some research projects result in publication, some projects do not bear the fruits. As a result, many unpublished projects are systematically omitted from the publication data. Among these unpublished papers, neither the edge weight (paper quality) nor the edge itself (team composition) is observed.[1]. Figure 2 shows the histogram of the journal impact factors for published economic papers. The histogram shows clear left truncation at zero, with a substantial probability mass accumulated in the left tail, raising the question of whether truncation is the underlying mechanism. In Section 4, we apply the J-test from Section 2.5 to the publication data, and provide further evidence for missing links. This presents an econometric challenge for studying the *collaboration* network, as only its sub-graph—the *publication* network—is observed.

Estimation and inference on partially observed networks lead to misspecification error and bias if the conventional estimators rely on fully observing all links. There are three main econometric challenges. First, the truncation mechanism introduces selection bias to the data, creating an *endogenous* network $\mathcal{G}$. This means that we can no longer use linear regression to estimate the fixed effects and the scaling factor, as in Abowd, Kramarz, and Margolis (1999) and Bonhomme (2021), the latter implicitly assuming fully observed team networks. Second, workers and teams typically have substantial heterogeneity (Bonhomme (2021)) that is typically unobserved. Third, since many real-life networks are sparse (Graham (2017)), the dimension of individual fixed effects grows at the same rate as the number of observations (links), a challenge further exacerbated by the link truncation. Therefore, the main contribution of this article is to provide a Generalized Method of Moment (GMM) estimator and a J-test for partially observed team networks. We also contribute to the empirical literature on teamwork productivity by revealing and correcting the plausibly underestimated premiums in academic collaboration network. Measuring the premium is important, because forming teams of the optimal size could enhance the team productivity, ceteris paribus.

[1] Omission of either the edge or edge weight constitutes a partially observed network.

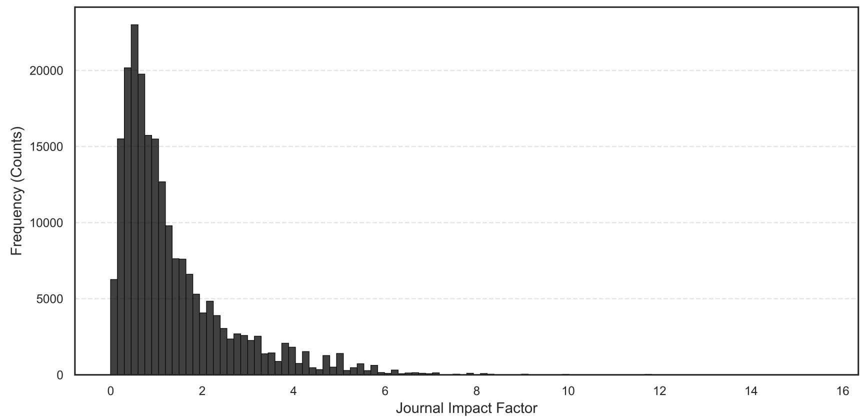

Figure 2: Empirical Distribution of Paper Quality

Notes: This figure presents the histogram of the paper quality of articles published in economic journals. Paper quality is matched with journal-year impact factor through Clarivate Web of Science Journal Citation Reports (1997 to 2020). The vertical axis measures the article count whereas the horizontal axis measures the impact factor. See Section 4 for details on the data and the sample selection.

1.1 Literature Review

Our article relates to the discussion on network sampling in the rapidly growing literature on networks. It is often challenging to accurately collect data on social and economic networks. On the other hand, various data constraints on networks can introduce econometric complexities. When networks are constructed by a partial sample of nodes, Chandrasekhar and Lewis (2011) recommend analytical correction or graphical reconstruction to alleviate estimation biases. Furthermore, Chandrasekhar et al. (2024) show that the estimation of diffusion effect may not be robust even with a vanishing share of mis-measured local links. Lewbel, Qu, and Tang (2024) show that the two-stage estimation and inference of the linear social effects model (Manski (1993)) remain valid if the measurement errors in the adjacency matrix grow sufficiently slow compared to the sample size. Boucher and Houndetoungan (2023) probabilistic reconstruct partially observed networks to correct the downward bias in the estimated peer effects. For network survey data capping the maximum number of reported links, Griffith (2022) proposes both correction and bounding methods.

This article also relates to the literature on limited dependent variable. Following the pioneering work of Tobin (1958) on censored regression (also referred to as Tobit model), Amemiya (1973) establishes the consistency of Tobin's maximum likelihood estimator and provides a consistent initial estimator. Generalizing the regression bias under the limited dependent variable as a special omitted variable bias, Heckman (1979) constructs a consistent two-stage estimator. Moving from standard linear regression to quantile regression, the censored least absolute deviation (LAD) estimator exploits symmetry to build orthogonality conditions that are also free of parametric assumptions on the i.i.d. errors (see Powell (1986a) and Powell (1986b)). Honoré (1992) incorporates fixed effects into the censored and truncated panel models, using sample trimming to restore symmetry and build orthogonality conditions. Hu (2002) focuses on a dynamic censored model in which a lagged dependent variable is introduced to the censored panel model.

The remainder of this article is organized as follows. Section 2 explains the econometric model, introducing a naive estimator for fully observed networks, a GMM estimator for partially observed networks, and a test for missing links. Section 3 shows the Monte Carlo results. Section 4 applies the model to academic publication data. Section 5 concludes.

2 The Model

2.1 Data Structure

We use the superscript star * to indicate latent variables. The latent outcome of project ℓ is $Y_\ell^* \in \mathbb{R}$. The econometrician observes Y_ℓ^* only if $Y_\ell^* \geq 0$, and we denote the observed outcome as Y_ℓ. The choice of the cut-off point at zero is without loss of generality.[2] Let $\alpha_i \in \mathbb{R}$ denote the unobserved individual-specific fixed effect of worker i, capturing i's unobserved type. Let α represent the vector of all individual-specific fixed effects. We highlight the distinction between *teams*, which are consisted of individual workers, and *projects*, which are produced by individual teams. Importantly, one team may produce multiple projects, and one worker may be a member of multiple teams.

Given any team network, the mapping $i(\ell)$ pins down the worker fixed effect $\alpha_{i(\ell)}$ associated with project ℓ. The key structural parameter $\lambda \in (0, \infty)$, referred to as the scaling factor in Bonhomme (2021), aggregates the individual fixed effects $\alpha_{i(\ell)}$ at the team level into a team-specific fixed effect a_ℓ. Finally, the error term U_ℓ is scaled by an unknown standard deviation $\sigma \in (0, \infty)$. The data-generating process (DGP) is

$$
\begin{aligned}
a_\ell &= \begin{cases} \alpha_{i(\ell)}, & \text{if } \ell \text{ has one worker } i, \\ \lambda(\alpha_{i(\ell)} + \alpha_{j(\ell)}), & \text{if } \ell \text{ has two workers } i \text{ and } j, \end{cases} \\
Y_\ell^* &= a_\ell + \sigma U_\ell, \\
Y_\ell &= \begin{cases} Y_\ell^*, & \text{if } Y_\ell^* \geq 0, \\ \text{Unobserved}, & \text{if } Y_\ell^* < 0. \end{cases}
\end{aligned} \tag{1}
$$

Without loss of generality, we focus on teams of at most two workers. When the team has one contributor, its fixed effect is simply $\alpha_{i(\ell)}$ in which we implicitly normalize its scaling factor to unity. When the team has two contributors, its fixed effect becomes $\lambda(\alpha_{i(\ell)} + \alpha_{j(\ell)})$. Our specification implies that team production is symmetric between contributors - the order of workers does not matter.

Our innovation in the linear model, compared with Bonhomme (2021), is the truncation mechanism. As in the model of generalized Hölder means in Ahmadpoor and Jones (2019), the use of λ here is analogous to their use of scalar β_n, which represents the impact benefits associated with teamwork with team-size of n. It governs the sign and magnitude of teamwork premium because incorporating λ flexibly enables collaboration to have productivity gain (or loss) relative to solo production. In contrast, arithmetic means inversely scale the sum of worker fixed effects by the integer-valued team size. For example, the weights are fixed at $\frac{1}{2}$ for teams of two people, and $\frac{1}{3}$ for teams of three people, both of which imply zero teamwork premium. Our formulation also keeps the dimension of overall fixed effects tractable by decomposing team-specific fixed effects into subsets of worker-specific fixed effects.

Assumption 1 (Conditionally IID Normal Shocks)**.** *The project-specific shock U_ℓ is independently and identically drawn across ℓ from a standard normal distribution conditional on α, the vector of individual-specific fixed effects, and the latent graph $\mathcal{G}^*$:*

$$U_\ell \overset{i.i.d.}{\sim} N(0,1) \,|\, \alpha, \mathcal{G}^*.$$

This baseline model is simple, but remains flexible. Section 2.4 shows an extension to larger teams and an extension that allows heterogeneous σ_ℓ for teams of various sizes. We also discuss the incorporation of time-varying covariates $X_{i(\ell),t(\ell)}$. The model does not incorporate team formation or no social effects between workers.

[2] This article only considers one-sided truncation from the left. The extension to right truncation is similar.

Furthermore, Assumption 1 rules out serial dependence between repeated teamwork. This assumption is strong but plausible in the academic setting in which researchers only work on novel ideas that have not been studied before, ensuring that project-level shocks remain uncorrelated.

Network-specific notations Denote $\mathcal{V}$ (or $\mathcal{V}^*$) and $\mathcal{E}$ (or $\mathcal{E}^*$) as the set of (latent) nodes and that of (latent) links. Because this article focuses on partially observed links, we impose fully observability of nodes, that is, $\mathcal{V} = \mathcal{V}^*$. The latent graph is $\mathcal{G}^* = (\mathcal{V}^*, \mathcal{E}^*)$ whereas the truncated graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Here, $\mathcal{E} \subseteq \mathcal{E}^*$. Single-worker teams are represented as self-loops. To distinguish multiple projects of the same team, we use $Y_{ij(\ell)}$ to explicitly state the project index ℓ as well as the team composition (i, j). The observed data $\{Y_\ell\}_{\ell=1}^{|\mathcal{E}|}$ is a random sample drawn from $\{Y_\ell^*\}_{\ell=1}^{|\mathcal{E}^*|}$ conditioned on $Y_\ell^* \geq 0$, i.e., non-negative outcomes.

2.2 Fully Observed Network

In an ideal world, the latent network $\mathcal{G}^*$ is fully observed, so $\mathcal{G} = \mathcal{G}^*$ holds. Then, one can relax the parametric and IID assumptions in Assumption 1, and instead impose a weaker assumption on the error term.

Assumption 2 (Mean-zero shocks). *The shock U_ℓ is mean-zero, conditional on α and $\mathcal{G}^*$. That is, $\mathbb{E}[U_\ell \mid \alpha, \mathcal{G}^*] = 0$.*

To derive a consistent estimator of the scaling factor, we subtract the unconditional expectation of the sum of $Y^*_{i(\ell)}$ and $Y^*_{j(\ell)}$, weighted by λ, from the expectation of $Y^*_{ij(\ell)}$, which leads to an important moment equality:

$$\mathbb{E}[Y^*_{ij(\ell)} - \lambda(Y^*_{i(\ell)} + Y^*_{j(\ell)})] = 0 \tag{2}$$

which holds for any triplet $(Y^*_{i(\ell)}, Y^*_{j(\ell)}, Y^*_{ij(\ell)})$. Because λ enters linearly into the moment restriction, λ has a closed-form estimator as a ratio in Lemma 1. To find it, we need to aggregate equation (2) across all projects ℓ in G^*.

Lemma 1 (Naive Estimator). *If the latent network $\mathcal{G}^*$ is fully observed, and Assumption 2 holds, then*

$$\hat{\lambda}_{naive}(\mathcal{G}^*) = \frac{\sum_{i,j:\{i,j\} \in \mathcal{E}^*} Y^*_{ij(\ell)}}{\sum_{i,j:\{i,j\} \in \mathcal{E}^*} (Y^*_{i(\ell)} + Y^*_{j(\ell)})} \tag{3}$$

is a consistent estimator of λ.

This estimator compares team output relative to solo output, implicitly controlling for unobserved worker types. This idea is also used in Anderson and Richards-Shubik (2022), where researchers' individual average productivity is subtracted from the team outcome in order to isolate the impact of various team-specific factors—such as team size, information spillover, and division of time—in a regression tree model. Here, we focus exclusively on team size effect, assuming the absence of network spillovers. One of the key advantages of our naive estimator is that it does not impose parametric assumption on the error term, except that its conditional mean has to be zero.

Without link truncation, that is, $\mathcal{G} = \mathcal{G}^*$, this naive estimator is correctly specified. However, all the outcome variables in (2) are latent, meaning that one must observe projects with negative outcomes. With link truncation, however, $\mathbb{E}[U_\ell | \alpha, \mathcal{G}] \neq 0$. As a result, $\mathbb{E}[Y^*_{ij(\ell)} - \lambda(Y^*_{i(\ell)} + Y^*_{j(\ell)}) \,|\, Y^*_{i(\ell)} \geq 0,\, Y^*_{j(\ell)} \geq 0,\, Y^*_{ij(\ell)} \geq 0] \neq 0$. Intuitively, the truncation mechanism tends to select projects with larger U_ℓ; for example, conditional on the same author types, "luckier" projects are more likely to reach the publication stage. An immediate implication is the inconsistency of $\hat{\lambda}_{\text{naive}}(\mathcal{G})$, which also biases the estimation of σ, and if of additional interests, the individual fixed effects α_i.

Therefore, we need an alternative strategy. But it does not mean that this naive estimator should be discarded. We will leverage it to build a test for missing links in Section 2.5.

Remark 1. *From this section onward, all stochastic specifications will be conditioned on the latent graph $\mathcal{G}^*$ and the individual fixed effects α.*

2.3 Partially Observed Network

Given a partially observed network, our goal is to find a conditional moment function $m(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})$ such that

$$\mathbb{E}\left[m(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}) \,\middle|\, Y^*_{i(\ell)} \geq 0,\, Y^*_{j(\ell)} \geq 0,\, Y^*_{ij(\ell)} \geq 0\right] = 0$$

for all possible values of α_i and α_j. Here, the conditions $Y^*_{i(\ell)} \geq 0$, $Y^*_{j(\ell)} \geq 0$, $Y^*_{ij(\ell)} \geq 0$ guarantees that the equality holds even on partially observed networks.

Motivated by the use of normal errors for limited dependent variable models, as introduced by Tobin (1958), we use normality to structure the missing links, which we exploit to derive moment restrictions. The key insight is to leverage the normality assumption to express the higher-order conditional moments of Y_ℓ as a linear combination of the lower-order moments and the model parameters, removing the dependency of our moments on the missing data.

The moments of the truncated normal have been used in the cross-sectional and panel literature but have not been systematically applied. Amemiya (1973) uses the first four moments to propose a consistent initial estimator for Tobit models. Building on this insight, Honore (1998) uses the relation between the first two moments of truncated normal to difference out individual fixed effects, building an IV estimator for censored panel data model. Outside of the econometrics literature, Horrace (2015) derives various properties of the first four moments of the truncated normal distribution, and Orjebin (2014) derives a recursive formula for its moments. We do not take credit for the lemma presented below.

Lemma 2 (Truncated normal). *Suppose $\widetilde{Y} \sim N(\tilde{\alpha}, \tilde{\sigma}^2)$ where $\tilde{\alpha} \in \mathbb{R}$ and $\tilde{\sigma} \geq 0$. Then, for $k \in \mathbb{N}^+$,*

$$\mathbb{E}[\widetilde{Y}^{k+1} - \tilde{\alpha}\widetilde{Y}^k - k\tilde{\sigma}^2\, \widetilde{Y}^{k-1} | \widetilde{Y} \geq 0] = 0.$$

Under normal errors, we can now easily construct an arbitrary number of additional moments by raising k to higher powers. There is one more hurdle left. Because the dimension of individual fixed effects α grows at the same rate as the number of observations, we need to take care of the incidental parameter problem (discussed in the seminal paper by Neyman and Scott (1948)) in order to establish consistency of $(\lambda,\, \sigma)$. The linearity of the conditional moments is thus appealing, because it enables us to difference out both the node-level and the link-level fixed effects, allowing us to build conditional moments free of fixed effects.

Proposition 1 (System of moment restrictions). *Define the k^{th} moment condition m_k as*

$$m_k(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}) := Y^k_{i(\ell)} Y^k_{j(\ell)} Y^k_{ij(\ell)} (Y_{ij(\ell)} - \lambda(Y_{i(\ell)} + Y_{j(\ell)})) + k\sigma^2 Y^{k-1}_{i(\ell)} Y^{k-1}_{j(\ell)} Y^{k-1}_{ij(\ell)} (\lambda(Y_{i(\ell)} + Y_{j(\ell)}) Y_{ij(\ell)} - Y_{i(\ell)} Y_{j(\ell)}).$$

Then, for $k \in \mathbb{N}^+$, the following equality holds:

$$\mathbb{E}\left[m_k(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}) \,\middle|\, Y^*_{i(\ell)} \geq 0,\, Y^*_{j(\ell)} \geq 0,\, Y^*_{ij(\ell)} \geq 0\right] = 0.$$

There is, however, an alternative way to generate more moments without relying on higher powers of the outcome variable, if one observes the latent graph $\mathcal{G}^*$. Because of the network exogeneity assumption, we could interact the first moment $m_1(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})$ with arbitrary choices of network statistics, such as node degrees, to generate new moment restrictions. However, we highlight that this lemma breaks down for the partially observed graph $\mathcal{G}$.

We are now able to consistently estimate (λ, σ) using the generalized method of moments. In principle, we now have an infinite number of moment restrictions at our disposal, which would accommodate extensions of the baseline model with any finite number of parameters. However, the higher moments might be less robust than the lower ones as they are more dependent on the parametric assumption, which does not necessarily hold in real-world DGPs. Because the baseline model only has two unknown parameters, let us construct the system of conditional moments using $k = 1, 2$ as follows:

$$m(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})' := \Big(m_1(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}) \quad m_2(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})\Big)\,. \tag{4}$$

Before we dive into the asymptotic properties of the GMM estimator, let us first discuss how we should prepare the data and construct the triplets. One salient challenge is the heterogeneous collaboration pattern of the workers. For example, in academic research, some researchers may have many solo papers but very few co-authored papers, whereas others may have few solo papers but lots of co-authored papers. Therefore, if we simply enumerate all triplet combinations, it is generally not possible to ensure independence across triplets.

The correlation between triplets arises when two triplets share a common project. Consider the following scenario: worker A has two solo projects, whereas worker B has one. Additionally, they have a team project together. Combinatorially, one could construct two triplets by reusing the team project and B's solo project. Nonetheless, the two triplets are also correlated.

Therefore, the GMM estimator provided below in Theorem 1 uses *uncorrelated* triplets—triplets that are mutually independent of each other—using the algorithm outlined in Appendix A. It is possible to add correlated triplets. However, one must adjust the corresponding variance-covariance matrix, and impose regularity conditions on the graph structure $\mathcal{G}$ to rule out some extreme cases to ensure convergence.

Theorem 1 (Truncation-robust estimator)**.** *Let C be the number of uncorrelated triplets. Define the sample moments*

$$\hat{g}(\lambda, \sigma)' = \frac{1}{C} \sum_{\{i,j\} \in \mathcal{E}} \Big(m_1(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}) \quad m_2(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})\Big)\,.$$

Suppose

(i) Assumption 1 holds.

(ii) The limiting gradient and variance matrices exist. Specifically,

$$G = \operatorname*{plim}_{C \to \infty} \frac{1}{C} \sum_{\{i,j\} \in \mathcal{E}} \mathbb{E}\left[\tfrac{\partial}{\partial \lambda} m(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}), \quad \tfrac{\partial}{\partial \sigma} m(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})\right],$$

$$V = \operatorname*{plim}_{C \to \infty} \frac{1}{C} \sum_{\{i,j\} \in \mathcal{E}} Var(m(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})).$$

(iii) Global identification of (λ, σ).

Then, the baseline model (1) has a consistent GMM estimator, that is,

$$\begin{pmatrix}\hat{\lambda}\\ \hat{\sigma}\end{pmatrix} = \operatorname*{arg\,min}_{\lambda\in\mathbb{R},\sigma\in\mathbb{R}^+} \hat{g}(\lambda,\sigma)'\hat{g}(\lambda,\sigma).$$

Furthermore, as $C \to \infty$, the asymptotic distribution is

$$\sqrt{C}\left[\begin{pmatrix}\hat{\lambda}\\ \hat{\sigma}\end{pmatrix} - \begin{pmatrix}\lambda\\ \sigma\end{pmatrix}\right] \Longrightarrow N\left(0,(G'V^{-1}G)^{-1}\right).$$

This theorem establishes the consistency of $(\hat{\lambda},\, \hat{\sigma})$. One can, if additionally interested in the individual fixed effect α_i, simply plug these estimates into Lemma 2. Note that using the biased naive estimator necessarily distorts the estimation of fixed effects. This gives $|\mathcal{V}|$ moment conditions, $\mathbb{E}[Y_{i(\ell)}^2 - \alpha_{i(\ell)}Y_{i(\ell)} - \hat{\sigma}^2|Y_{i(\ell)}^* \geq 0] = 0$, one for every worker who produces alone. Furthermore, for each unique team composition (i,j), $\mathbb{E}[Y_{ij(\ell)}^2 - \hat{\lambda}(\alpha_{i(\ell)}+\alpha_{j(\ell)})Y_{ij(\ell)} - \hat{\sigma}^2|Y_{ij(\ell)}^* \geq 0] = 0$. Because we have more moments than the number of individual fixed effects $|\mathcal{V}|$, the GMM can be applied to consistently estimate $\{\alpha_i\}_{i=1}^{|\mathcal{V}|}$.

Remark 2 (Noise in the estimated fixed effects)**.** *In order to make meaningful interpretation of the fixed-effect estimates, which might be estimated with significant noises, it is necessary to verify the network is sufficiently dense and connected. A dense team network means that every worker has a relatively large number of observations, on average. In addition, greater network connectivity enhances estimation precision, as formalized by Jochmans and Weidner (2019) on the impact of connectivity on the precision of individual fixed-effect estimates in a two-way regression model.*

2.4 Extensions

2.4.1 Heteroscedasticity

We can also allow for heterogeneous variances. The baseline setup in (1) assumes that every project ℓ faces conditionally i.i.d. shocks. However, we may also wish to allow $\sigma_{s(\ell)} \in \{\sigma_1, \sigma_2\}$ to vary according to the size of the team $s(\ell) \in \{1,2\}$ where $s(\cdot)$ is a function that takes an index ℓ and returns the size of the team. For example, it could be the case that team production is more volatile than solo production (or the other way around). Let (σ_1, σ_2) denote the variances of the one-worker and two-worker projects, respectively. The extended model becomes

$$\begin{aligned}
a_\ell &= \begin{cases}\alpha_{i(\ell)}, & \text{if } \ell \text{ has one worker } i,\\ \lambda(\alpha_{i(\ell)}+\alpha_{j(\ell)}), & \text{if } \ell \text{ has two workers } i \text{ and } j,\end{cases}\\
Y_\ell^* &= a_\ell + \sigma_{s(\ell)}U_\ell,\\
Y_\ell &= \begin{cases}Y_\ell^*, & \text{if } Y_\ell^* \geq 0,\\ \text{Unobserved}, & \text{if } Y_\ell^* < 0.\end{cases}
\end{aligned} \tag{5}$$

Like the baseline model, we maintain the normalization of $\lambda_1 = 1$. To estimate $(\lambda, \sigma_1, \sigma_2)$, we can simply apply Proposition 1 with $k = 1,2,3$. Now, with three moment restrictions, the unknown parameters are just-identified.

2.4.2 Covariates

In some cases, we may want to incorporate covariates into team production. Although time-invariant variables will be absorbed into individual fixed effects α_i and hence will not be identified, time-varying covariates $X_{i(\ell),t(\ell)}$ will not be absorbed. For example, one might be interested in estimating the marginal effect $\beta \in \mathbb{R}$ of accumulating an additional year of work experience on the quality of a researcher's output. Specifically, we could define $X_{i(\ell),t(\ell)}$ as the "academic age" of researcher i at the time of writing paper ℓ—that is, the number of years that have passed since their graduation. The model becomes:

$$
\begin{aligned}
a_\ell &= \begin{cases} \alpha_{i(\ell)}, & \text{if } \ell \text{ has one worker } i, \\ \lambda(\alpha_{i(\ell)} + \alpha_{j(\ell)}), & \text{if } \ell \text{ has two workers } i \text{ and } j, \end{cases} \\
Y^*_{\ell,t} &= \begin{cases} a_\ell + \beta X_{i(\ell),t(\ell)}, & \text{if } \ell \text{ has one worker } i, \\ a_\ell + \beta\lambda\left(X_{i(\ell),t(\ell)} + X_{j(\ell),t(\ell)}\right), & \text{if } \ell \text{ has two workers } i \text{ and } j, \end{cases} \\
Y_{\ell,t} &= \begin{cases} Y^*_{\ell,t}, & \text{if } Y^*_{\ell,t} \geq 0, \\ \text{Unobserved}, & \text{if } Y^*_{\ell,t} < 0. \end{cases}
\end{aligned} \tag{6}
$$

Now, the vector of parameters becomes (λ, σ, β). And the project index ℓ is mapped to worker index i as well as the time index t. Furthermore, even though covariates enter the production function additively, they are scaled by the product of β and λ. Proposition 1 still applies, because the non-stochastic covariates simply shift the mean. However, in order to identify β, the extended model additionally requires some temporal variation such that the change in i's characteristics at the time of producing i's solo project and the team project must *not* be equal to minus one times the change in j's characteristics at the time of producing j's solo project and the team project. Otherwise, all the time-varying components cancel out, along with β. Finally, we apply Proposition 1 with k up to three to derive a similar GMM estimator.

2.4.3 Three or More Workers

Although the baseline DGP considers at most two workers per team, our model extends naturally to teams of arbitrary size $S \in \mathbb{N}^+$. Take $S = 3$, for example, the team-specific fixed effect a_ℓ becomes a weighted average of types of workers i, j and k. Specifically, $a_\ell = \lambda_3(\alpha_{i(\ell)} + \alpha_{j(\ell)} + \alpha_{k(\ell)})$ where $\lambda_3 \in \mathbb{R}$ is the scaling factor associated with three-person teams. Building from (1) , we can rewrite the model as:

$$
\begin{aligned}
a_\ell &= \begin{cases} \alpha_{i(\ell)}, & \text{if } \ell \text{ has one worker } i, \\ \lambda_3(\alpha_{i(\ell)} + \alpha_{j(\ell)} + \alpha_{k(\ell)}), & \text{if } \ell \text{ has three workers } i, j \text{ and } k, \end{cases} \\
Y^*_\ell &= a_\ell + \sigma U_\ell, \\
Y_\ell &= \begin{cases} Y^*_\ell, & \text{if } Y^*_\ell \geq 0, \\ \text{Unobserved}, & \text{if } Y^*_\ell < 0. \end{cases}
\end{aligned} \tag{7}
$$

The GMM estimator now takes $(Y_{i(\ell)}, Y_{j(\ell)}, Y_{k(\ell)}, Y_{ijk(\ell)})$, which is a 4-dimensional tuple. We maintain the normalization of $\lambda_1 = 1$. To estimate (λ_3, σ), we apply Proposition 1 to ($Y_{i(\ell)}$, $Y_{j(\ell)}$, $Y_{k(\ell)}$, $Y_{ijk(\ell)}$) individually and use a linear combination of them to find two moment restrictions that are free of the fixed effects.

2.5 Test for Missing Links

Before selecting an estimator, one may first wish to test whether the network is fully observed. However, due to the nature of omitted data, directly testing for missing links is generally infeasible, as omitted observations are unobserved by definition. Even if the empirical distribution appears to be truncated (for example, Figure 2), the observed accumulation of probability mass could arguably result from an idiosyncratic DGP with bounded support. This section outlines a test for link truncation. Let us consider the null hypothesis $\mathcal{H}_0$ that there is no link truncation and $\mathcal{G} = \mathcal{G}^*$. The alternative hypothesis $\mathcal{H}_1$ is that $\mathcal{G}$ is only partially observed.

In the GMM framework, the J-test provided in Hansen (1982) provides a simple way to test the validity of moment functions. Ideally, we want to avoid relying on parametric assumption, because error misspecification could produce false rejection of the null. Our goal is to construct a set of over-identifying moments, in addition to equation (2) as implied by the production function, that hold under $\mathcal{H}_0$ but become invalid under $\mathcal{H}_1$. We exploit the transition from the exogenous network $\mathcal{G}^*$, which holds under $\mathcal{H}_0$, to the endogenous network $\mathcal{G}$, which holds under $\mathcal{H}_1$. To see this, recall that Assumption 1 states that the latent network $\mathcal{G}^*$ is exogenous to shocks U conditional on fixed effects α. However, omitting a link ℓ for which $Y^*_\ell < 0$ induces a positive correlation between U_ℓ and the indicator for observing any link, namely $\mathbf{1}\{Y^*_\ell \geq 0\}$. In other words, $U \not\perp\!\!\!\perp \mathcal{G} \,|\, \alpha$ despite $U \perp\!\!\!\perp \mathcal{G}^* \,|\, \alpha$.

Now, we generate additional moments simply by interacting (2) with network statistics computed on $\mathcal{G}$. Note that the network statistics should be computed at the team-level to align with the triplets' granularity. Rejecting these moments in the J-test outlined below thus implies rejection of the null. Hence, we are able to test for truncation without observing the omitted links.

Lemma 3 (J-test for missing links)**.** *Let the null $\mathcal{H}_0$ be that links are fully observed. For node pair $\{i,j\} \in \mathcal{E}$, compute network statistics $f^k_{ij}(\mathcal{G})$, $k = 1, \ldots, K$. Consider $m_0(\lambda, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}) = Y_{ij(\ell)} - \lambda(Y_{i(\ell)} + Y_{j(\ell)})$, define $g_0(\lambda) = \frac{1}{C} \sum_{\{i,j\}\in\mathcal{E}} m_0(\lambda, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})$. Let the number of uncorrelated triplets be C. Define the moments as*

$$g(\lambda)' = \begin{pmatrix} g_0(\lambda) & g_1(\lambda) & \cdots & g_K(\lambda) \end{pmatrix},$$

where

$$g_k(\lambda) = \frac{1}{C} \sum_{\{i,j\}\in\mathcal{E}} f^k_{ij}(\mathcal{G}) \cdot m_0(\lambda, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}), \quad k = 1, \ldots, K.$$

Suppose $\hat{S}_{\mathcal{H}_0}$ and $\hat{\lambda}$ are consistent estimates of the asymptotic variance of the moments and of λ under the null. As $C \to \infty$,

$$T_{\mathcal{H}_0} = C\, g(\hat{\lambda})'\, \hat{S}^{-1}_{\mathcal{H}_0}\, g(\hat{\lambda}) \sim \chi^2_K.$$

To see why $f^k_{ij}(\mathcal{G}) \cdot m_0(\lambda, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})$ is a valid moment, we note that $U \perp\!\!\!\perp \mathcal{G} \,|\, \alpha$ under the null, so $f^k_{ij}(U) \perp\!\!\!\perp \mathcal{G} \,|\, \alpha$. Hence, $\mathbb{E}[f^k_{ij}(\mathcal{G}) \cdot m_0(\lambda, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})] = 0$. However, one should still carefully explore and select appropriate network statistics, as some statistics are more informative than others and can thus enhance the statistical power of the J-test. Finally, there is one caveat to the interpretation of this test. Because the over-identifying restrictions critically depend on the production function, rejection of the null could imply missing links and / or a rejection of the linear production function.

3 Simulation

To numerically test the performances of the GMM estimator, we conduct Monte Carlo simulations to test its performance under small sample, sparsity and misspecified error distributions. The focus of the simulation exercises is on the scaling factor λ. We outline the simulation algorithm below, setting the true parameters $\lambda = 0.7$ and $\sigma = 2$ and repeating the following steps 1,000 times.

Algorithm 1 Simulation

1. Fixed Effects Simulation: draw N individual-specific fixed effects $\{\alpha_i, i = 1, \ldots, N\}$ from Pareto $(0, 10)$.
2. Random Team Formation: teams of two workers are exogenously and sparsely matched. The network sparsity is governed by the average number of links per node.
3. Team Production: within each team, worker(s) produce a project of quality Y_ℓ^* independently or jointly. Since its shock is drawn i.i.d. from $N(0, \sigma^2)$, we have constructed a list of uncorrelated triplets.
4. Link Truncation: drop triplets containing any project with negative outcome ($Y_\ell^* < 0$).
5. Estimation: estimate λ using the naive estimator and the truncation-robust GMM estimator.

3.1 Sparsity

Many real-life networks are sparse. A network is sparse when the fraction of the realized links out of all potential linkages decreases to zero as the number of nodes increases. It is crucial to verify that our GMM estimator performs well under sparsity. In principle, because our estimator only uses triplets $(Y_{ij(\ell)}, Y_{i(\ell)}, Y_{j(\ell)})$, rather than the entire graph $\mathcal{G}$, it should remain robust for sparse networks. In the simulation, we manipulate the network sparsity by controlling the average number of links per dyad. The leftmost panel of Table 1 examines a highly sparse network, in which there is, on average, one-tenth of link per node. The middle panel analyzes a relatively sparse network, in which the sparsity increases to one link per node on average. Furthermore, the rightmost panel presents a less sparse case, with an average of ten links per node. Not surprisingly, the variances of both the naive and the GMM estimators decrease as the network becomes more dense, simply because of more observations. One advantage of the naive estimator is its high efficiency and unbiasedness in fully observer networks. However, when there are missing links, the significant downward bias of the naive estimator persists even as the network gets denser. In contrasts, the GMM estimator substantially reduces the median bias and the median absolute error (MAE).

		Highly Sparse (# Links = 0.1 # Nodes)				Sparse (# Links = # Nodes)				Less Sparse (# Links = 10 # Nodes)			
		True Value	Bias	MAE	SE	True Value	Bias	MAE	SE	True Value	Bias	MAE	SE
Fully Observed Network	Naive Estimator	70.00%	-0.10%	1.20%	1.78%	70.00%	0.01%	0.39%	0.58%	70.00%	0.00%	0.10%	0.16%
	GMM Estimator	70.00%	-0.25%	2.80%	4.13%	70.00%	-0.25%	1.15%	1.65%	70.00%	-0.07%	0.38%	0.54%
Partially Observed Network	Naive Estimator	70.00%	-6.96%	6.96%	1.37%	70.00%	-6.94%	6.94%	0.44%	70.00%	-6.94%	6.94%	0.15%
	GMM Estimator	70.00%	-0.16%	4.16%	6.15%	70.00%	0.43%	3.50%	6.22%	70.00%	0.08%	1.79%	2.72%

Table 1: Simulation: Sparsity

Notes: This table reports the simulation results of the naive estimator and the GMM estimator when the graph is highly sparse, sparse and less sparse. We fix the number of nodes at $10,000$ *to mimic the size of the economics publication network. At the same time, we set the total number of links to* $1,000$, $10,000$ *and* $100,000$, *to simulate highly sparse, sparse, and less sparse settings, respectively. Bias refers to the median bias, and MAE refers to the median absolute error. The standard error (SE) is computed as the interquartile range of the simulated estimates divided by* 1.35. *We simulate* $1,000$ *times.*

3.2 Size

Social and economic networks typically have various sizes, namely, the name of the nodes in the network. In this section, we study three size settings: small, medium and large. Throughout this exercise, we maintain a sparse structure in which the number of nodes equals the number of links. As the network becomes smaller, both estimators become invariably more sensitive to sampling uncertainty in terms of larger standard errors. For fully observed networks, the naive estimator has superior performances in terms of bias and efficiency. However, for partially observed networks, the GMM estimator outperforms the naive estimator by correcting the downward bias on the scaling factor. As the network size increases, the GMM estimator reduces bias by an order of magnitude when compared to the naive estimator. On the other hand, the naive estimator does not benefit at all from the increased network size.

		Small-sized (100 nodes)				Medium-sized (1,000 nodes)				Large-sized (10,000 nodes)			
		True Value	Bias	MAE	SE	True Value	Bias	MAE	SE	True Value	Bias	MAE	SE
Fully Observed Network	Naive Estimator	70.00%	-0.08%	3.68%	5.44%	70.00%	-0.03%	1.16%	1.73%	70.00%	0.01%	0.37%	0.56%
	GMM Estimator	70.00%	-1.66%	6.47%	9.15%	70.00%	-0.53%	2.81%	4.06%	70.00%	-0.20%	1.09%	1.60%
Partially Observed Network	Naive Estimator	70.00%	-7.11%	7.11%	4.75%	70.00%	-6.96%	6.96%	1.43%	70.00%	-6.94%	6.94%	0.43%
	GMM Estimator	70.00%	-4.09%	7.26%	9.12%	70.00%	0.26%	4.10%	6.29%	70.00%	0.11%	3.41%	5.75%

Table 2: Simulation: Network Size

Notes: This table reports the simulation results of the naive estimator and the GMM estimator when the graph size is small (100 *nodes), medium (*$1,000$ *nodes) and large (*$10,000$ *nodes). We set the ratio of nodes to links at* $1:1$ *to maintain sparsity. Bias refers to the median bias, and MAE refers to median absolute error. The standard error (SE) is computed as the interquartile range of the simulated estimates divided by* 1.35*. We simulate* $1,000$ *times.*

3.3 Non-Gaussian Error

		T Distribution (Degrees of Freedom = 10)				Extreme Value Distribution (Shape parameter = 1/2)			
		True Value	Bias	MAE	SE	True Value	Bias	MAE	SE
Fully Observed Network	Naive Estimator	70.00%	0.01%	0.44%	0.64%	70.00%	-3.02%	3.02%	0.42%
	GMM Estimator	70.00%	0.88%	1.52%	2.16%	70.00%	3.36%	3.43%	1.98%
Partially Observed Network	Naive Estimator	70.00%	-7.46%	7.46%	0.47%	70.00%	-7.29%	7.29%	0.38%
	GMM Estimator	70.00%	-1.30%	3.59%	5.22%	70.00%	0.40%	2.30%	3.41%

Table 3: Simulation: Non-Gaussian Error

Notes: This table reports the simulation results of the naive estimator and the GMM estimator when the error term in the DGP is misspecified as normal. We fix the number of nodes at $10,000$*. We also set the ratio of nodes to links at* $1:1$ *to maintain sparsity. In the left panel, the standardized t-distribution with* 10 *degrees of freedom is used. In the right panel, the generalized extreme value distribution with shape parameter equal to* $\frac{1}{2}$ *is used (here, we follow the parametrization used by SciPy). Bias refers to the median bias, and MAE refers to median absolute error. The standard error (SE) is computed as the interquartile range of the simulated estimates divided by* 1.35*. We simulate* $1,000$ *times.*

The simulation exercise in Table 3 is designed to examine the sensitivity of our estimator when the latent error distribution is not Gaussian. In the left panel, we set U_ℓ to follow a standardized Student's t-distribution with 10 degrees of freedom. The t-distribution provides a modest deviation from the normal distribution by allowing the error term to have heavier tails. When the network is fully observed and the latent error follows a t-distribution, the naive estimator has smaller bias and MAE than the GMM estimator.

So far, both the normal and the t-distribution are symmetric with mean zero. It is important to test other non-standard distributions with nonzero mean, skewness, and kurtosis. Therefore, in the right panel, we experiment with a generalized extreme-value distribution. We choose an extreme value distribution with a shape parameter of $\frac{1}{2}$ because its first four moments—mean (0.23), variance (0.86), skewness (-0.63), and kurtosis (0.25)—are nonzero.

When the error follows an extreme value distribution, both estimators exhibit similar biases if the network is fully observed. However, when some links are missing, the GMM estimator demonstrates far superior performance over the naive estimator. Its bias and MAE are much closer to zero than those of the naive estimator, even in the case of t-distribution. Even though the mean-zero t-distribution aligns with the naive estimator's assumption, this finding is not surprising given that the left-truncated t-distribution has a strictly positive mean.

To summarize, when the network is partially observed, the GMM estimator effectively corrects the substantial downward biases suffered by the naive estimator, albeit with some efficiency lost.

4 Application

In this section, we revisit the classical question of estimating the productivity premium (or loss) in academic collaboration. Although collaboration increases productivity, it also incurs coordination costs such as communication difficulty, shirking, free-riding and clouded credit assignment (Becker and Murphy (1992); Jones (2021)). There are many different approaches to measure productivity and to study the relationship between co-authorship and productivity. The literature reports various estimates, some of which are of opposite signs. Both Hollis (2001) and Ductor (2015)[3] construct their own measures of productivity based on factors such as page length and paper quality. The former employs an instrumental variable approach to estimate the productivity premium at approximately 70%. whereas the latter estimates a productivity loss ranging from -7% to -20%. Applying a model of generalized means, Ahmadpoor and Jones (2019) estimate the distribution of λ across fields and find that co-authorship increases paper impact in most fields, with a median premium of 105%. Using the sample from Ductor et al. (2014), Bonhomme (2021) estimates the team premium at around 34%. More recently, Anderson and Richards-Shubik (2022) fit a regression tree on journal impact score, and also find that larger teams are more productive on average.

Because the measurement of the paper quality varies substantially in the aforementioned literature, making the numeric comparison of estimates challenging. What sets our approach apart from the previous literature is our focus on how much of a *difference* it makes when missing links (unpublished research projects) are accounted for. We apply our GMM estimator to academic collaboration data in which the quality and quantity of unpublished projects are unobserved. Through our baseline model, we examine the premium associated with two-author production relative to single-author production. We find that neglecting the missing links results in a substantial downward bias, necessitating a correction.

4.1 Data and Sample Selection

Our data source is Microsoft Academic Graph (MAG), a project run by Microsoft Research that uses machine readers to crawl and collect publication records from the Internet. MAG data are exhaustive in that they include articles from (almost) all scientific fields, spanning hundreds of years of publication. To measure the quality of project ℓ, we use the impact factor published by Web of Science (WoS). The impact factor is the ratio between the total number of citations received by a given journal in the current year and the total number of published articles in the two previous years. Since the raw citation count is non-negative, the impact factor is also non-negative by construction. Another popular measure of paper quality is paper-specific citation count. However, since the MAG initiative was discontinued in 2021, articles published around 2020 may not have sufficient time to accumulate citations, relative to older articles. We therefore use the journal-level impact factor, which has also been adopted

[3] Their outcome variable is based on Ductor et al. (2014) which is primarily taken from the quality index computed by Kodrzycki and Yu (2006). For the journals included in the EconLit database but missing from Kodrzycki and Yu (2006), Ductor et al. (2014) build a predicted index for them.

by previous studies (e.g., Ductor et al. (2014) and Anderson and Richards-Shubik (2022)).

We focus on articles published in "economics journals" that are also exclusively written by "economists". First, we use Web of Science journal classification to identify economics journals. Second, since our GMM estimator relies on the triplet $(Y_{ij(\ell)}, Y_{i(\ell)}, Y_{j(\ell)})$, we focus on economists who have published at least one single-authored and one co-authored article in economics journals. While it would be interesting to study team production in other fields, our parsimonious production function may be less applicable to them. In fields such as the physical and medical sciences, the order of the authors' names matters since the first authors are usually the ones who contribute more. In economics, because most journals use alphabetical ordering[4], the name order does not provide additional information on individual inputs. This is consistent with our linear production model (1) that imposes symmetry between individual fixed effects. For the same reason, we do not include interdisciplinary collaboration between economists and non-economists.

Our sample runs from 1997 to 2020, as the WoS Journal Citation Report only commenced in 1997 and MAG project was shut down in 2021. There are 359 economic journals whose impact factors are available during at least one year between 1997 and 2020. Following the algorithm in Appendix A, our sample contains 15,875 economists, 25,047 co-authored papers and 50,094 single-authored papers. The average paper per author is 3.16, which suggests the sparse network structure of the collaboration graph.

There are two potential obstacles to identification and estimation. First, the quality of individual journal may change across time. To address this concern, we match detailed *journal-year* impact factor index from WoS to individual articles based on their publication year. The second challenge pertains to using fixed effects to model individual types that may evolve over time. Specifically, worker types might grow as workers accumulate more work experience. To mitigate this concern, within each triplet, for researchers with multiple solo publications, we pair the co-authored article with single-authored articles published around the same time where possible.

4.2 Collaboration Network with Missing Links

The GMM estimator is a consistent estimator for partially observed networks $\mathcal{G}$ as well as for fully observed networks $\mathcal{G}^*$. However, it remains important to test for missing links as the GMM estimator is less efficient and relies on parametric assumption on the error term. We consider the null hypothesis that the academic collaboration network is fully observed.

To perform the J-test set out in Lemma 3, we generate additional valid moments by interacting (2) with arbitrary network statistics that are functions of the observed graph $\mathcal{G}$. There are numerous choices of candidates for network statistics. In our test, we choose node degree and closeness centrality[5]. Because these statistics are computed at the node level, we take a simple sum at the dyadic level (i, j) before interacting them with the linear moment (2). We then perform a two-step GMM estimation, obtaining a J-test statistic of 6.56. In particular, the asymptotic distribution of the test statistics is chi-squared with two degrees of freedom, since the number of parameters is one (i.e. λ) and the number of moments is three. This corresponds to a p-value of 0.0376. At the significance level of 0.05, we reject the null hypothesis that the collaboration network is fully observed.

[4]There are few exceptions. For instance, American Economic Association introduced a randomization tool for authors opting to randomize the name order of their coauthored papers. Nevertheless, the randomized name order remains uninformative on the individual inputs.

[5]We recommend testing different network statistics, since some statistics generate little variation at the node or dyadic level, and could lead to false negative results when used as over-identifying moments.

4.3 Correcting Underestimated Collaboration Premium

The J-test implemented in the previous section suggests the need to account for missing links. Assuming full network observability, the naive estimator is misspecified and gives $\hat{\lambda}_{\text{naive}} = 0.584$ with 90% bootstrap CI of $[0.580, 0.590]$. To address plausible truncation bias, we employ the GMM estimator from Theorem 1. We find that $\hat{\lambda}$ increases by more than 10 percent to 0.651, with a standard error of 0.04 and 90% CI of $[0.584, 0.718]$. Although the difference between the estimated λ may appear small in absolute terms, it is substantial given that the scaling factor is typically bounded between 0 and 1.

To better interpret the two estimates, we compute the average productivity gain implied by the team-size scaling factor. Suppose that two researchers of identical type work together, the average productivity gain (computed as $\hat{\lambda} \cdot 2 - 1$) is 30.2%. This implies an almost 30 percent increase in the paper quality (Figure 3), doubling the estimated premium from the naive estimator.

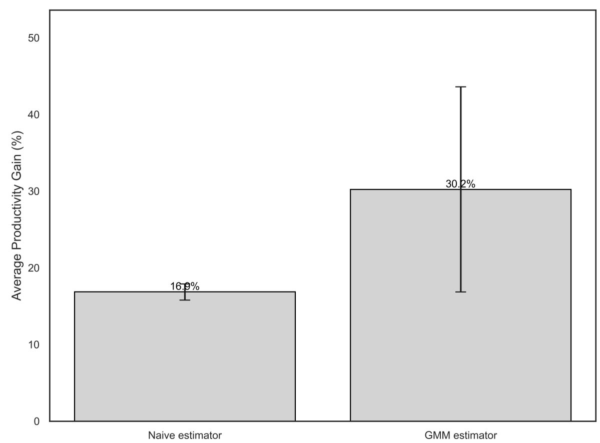

Figure 3: Average Productivity Gain

Notes: This bar plot shows the average team premium implied by the naive estimate and the GMM estimates, with 90% confidence intervals. The outcome variable is measured by the journal impact factor.

The downward truncation bias suffered by $\hat{\lambda}_{\text{naive}}$ is not surprising. The left-sided truncation at zero disproportionately favors “lucky” projects that receive large and positive shocks. This artificially inflates the workers’ input and hence deflates the estimated scaling factor.

As of now, we have used the journal impact factor index to assess paper quality. However, alternative measures of paper quality also exist. For robustness checks, we re-estimate our model using the Clarivate’s Eigenfactor Metrics (West (2017)). This eigenfactor-based index measures paper quality by assessing the extent to which a paper is cited by other important or high-quality articles, using the structure of the directed citation network to measure the journal influence (and hence article quality of those published in that journal). The naive and the GMM estimates are respectively 0.586 with 90% CI of $[0.576, 0.597]$, and 0.670, with 90% CI of $[0.569, 0.770]$. The results from two different metrics are consistent (see Appendix B for the associated bar plot).

4.4 Evidence for Time-varying Productivity Gain

As our sample spans more than two decades, we also ask if the team premium changes over time. The motivation for this question is that IT technology may enhance worker productivity (see Dulebohn and Hoch (2017) and Karl, Peluchette, and Aghakhani (2022)). Prior to the introduction of accessible video conference tools, notably Skype in 2003, email was the main communication tool for researchers. We hence partition the sample into two periods -

1997 to 2003 and 2004 to 2020. However, we emphasize that we do not make causal interpretation of the impact of IT technology on research productivity, as there could be many omitted variables, and we also do not observe researchers' individual adoption time of video conferencing tools.

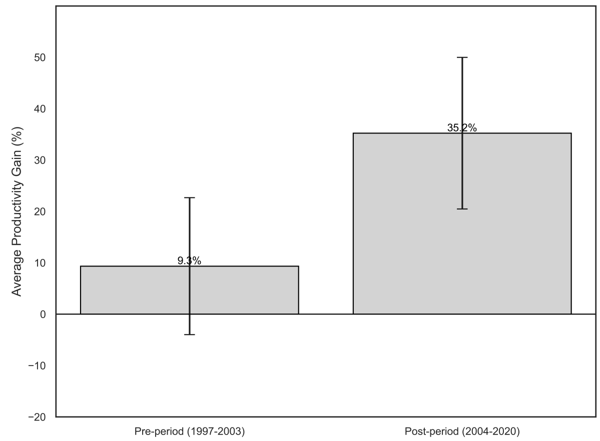

Figure 4: Time-varying Productivity Gain

Notes: This bar plot shows the average team premium estimated by the GMM estimator in the pre-2003 and post-2003 periods, with 90% confidence intervals. The outcome variable is measured by the journal impact factor.

Figure 4 reveals a stark difference between the two estimated premiums: the average productivity gain of 35.2% after 2003 appears significantly higher than the 9.3% gain for the 1997–2003 sample. It also exceeds the 30.2% gain from the pooled 1997–2020 sample. Furthermore, the confidence interval of the pre-2003 estimate includes zero whereas the post-2003 estimate does not, hinting an increase in teamwork premium in the second period.

5 Conclusion

Network sampling is not always random. Using partially observed networks could lead to a significant bias that requires correction. This article studies a linear team production network with partially observed links. Part of the estimation challenge arises from the endogenous network: partial sampling causes correlation between the observed graph $\mathcal{G}$ and the unobserved error. Therefore, standard approaches that assume a fully observed or exogenous network may not work. However, unlike the panel literature, there is currently limited discussion of partially observed network data in the network literature. Our main econometric contributions are a truncation-robust GMM estimator and a test for link truncation for team networks. Our normality assumption is restrictive, and it would be important to relax the parametric assumption on the shock term in future work. On the empirical side, we also contribute to the literature that the reported teamwork premium may be severely biased if missing links are not accounted for.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). "High Wage Workers and High Wage Firms". In: *Econometrica* 67.2, pp. 251–333.

Ahmadpoor, M. and B. F. Jones (2019). "Decoding team and individual impact in science and invention". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.28, pp. 13885–13890.

Amemiya, T. (1973). "Regression Analysis when the Dependent Variable Is Truncated Normal". In: *Econometrica* 41.6, pp. 997–1016.

Anderson, K. A. and S. Richards-Shubik (2022). "Collaborative Production in Science: An Empirical Analysis of Coauthorships in Economics". In: *The Review of Economics and Statistics* 104.6, pp. 1241–1255.

Becker, G. S. and K. M. Murphy (1992). "The Division of Labor, Coordination Costs, and Knowledge". In: *The Quarterly Journal of Economics* 107.4, pp. 1137–1160.

Bonhomme, S. (2021). "Teams: Heterogeneity, Sorting, and Complementarity". In: *Proceedings of the 15th World Congress of the Econometric Society*.

Boucher, V. and E. A. Houndetoungan (2023). "Estimating Peer Effects Using Partial Network Data".

Chandrasekhar, A. G., P. Goldsmith-Pinkham, T. H. McCormick, S. Thau, and J. Wei (2024). *Non-robustness of diffusion estimates on networks with measurement error*.

Chandrasekhar, A. G. and R. Lewis (2011). "Econometrics of Sampled Networks".

Devereux, K. (2018). "Identifying the value of teamwork: Application to professional tennis". In.

Ductor, L. (2015). "Does Co-authorship Lead to Higher Academic Productivity?" In: *Oxford Bulletin of Economics and Statistics* 77.3, pp. 385–407.

Ductor, L., M. Fafchamps, S. Goyal, and M. J. van der Leij (2014). "Social Networks and Research Output". In: *The Review of Economics and Statistics* 96.5, pp. 936–948.

Dulebohn, J. H. and J. E. Hoch (2017). "Virtual teams in organizations". In: *Human Resource Management Review*. Virtual Teams in Organizations 27.4, pp. 569–574.

Graham, B. S. (2017). "An Econometric Model of Network Formation With Degree Heterogeneity". In: *Econometrica* 85.4, pp. 1033–1063.

Griffith, A. (2022). "Name Your Friends, but Only Five? The Importance of Censoring in Peer Effects Estimates Using Social Network Data". In: *Journal of Labor Economics* 40.4, pp. 779–805.

Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators". In: *Econometrica* 50.4, pp. 1029–1054.

Heckman, J. J. (1979). "Sample Selection Bias as a Specification Error". In: *Econometrica* 47.1, pp. 153–161.

Hollis, A. (2001). "Co-authorship and the output of academic economists". In: *Labour Economics* 8.4, pp. 503–530.

Honore, B. E. (1998). "IV Estimation of Panel Data Tobit Models with Normal Errors." In.

Honoré, B. E. (1992). "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects". In: *Econometrica* 60.3, pp. 533–565.

Horrace, W. C. (2015). "Moments of the truncated normal distribution". In: *Journal of Productivity Analysis* 43.2, pp. 133–138.

Hu, L. (2002). "Estimation of a Censored Dynamic Panel Data Model". In: *Econometrica* 70.6, pp. 2499–2517.

Jaravel, X., N. Petkova, and A. Bell (2018). "Team-Specific Capital and Innovation". In: *American Economic Review* 108.4-5, pp. 1034–1073.

Jochmans, K. and M. Weidner (2019). "Fixed-Effect Regressions on Network Data". In: *Econometrica* 87.5, pp. 1543–1560.

Jones, B. F. (2021). "The Rise of Research Teams: Benefits and Costs in Economics". In: *Journal of Economic Perspectives* 35.2, pp. 191–216.

Karl, K. A., J. V. Peluchette, and N. Aghakhani (2022). "Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly". In: *Small Group Research* 53.3, pp. 343–365.

Kerr, S. P. and W. R. Kerr (2018). "Global Collaborative Patents". In: *The Economic Journal* 128.612, F235–F272.

Kodrzycki, Y. K. and P. D. Yu (2006). *New Approaches to Ranking Economics Journals*. SSRN Scholarly Paper. Rochester, NY.

El-Komboz, L. A., T. Fackler, and M. Goldbeck (2024). *Productivity Spillovers Among Knowledge Workers in Agglomerations: Evidence from GitHub*. SSRN Scholarly Paper. Rochester, NY.

Lewbel, A., X. Qu, and X. Tang (2024). "Ignoring measurement errors in social networks". In: *The Econometrics Journal* 27.2, pp. 171–187.

Manski, C. F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem". In: *The Review of Economic Studies* 60.3, pp. 531–542.

Neyman, J. and E. L. Scott (1948). "Consistent Estimates Based on Partially Consistent Observations". In: *Econometrica* 16.1, pp. 1–32.

Orjebin, E. (2014). "A Recursive Formula for the Moments of a Truncated Univariate Normal Distribution". Dissertation. The University of Queensland.

Powell, J. L. (1986a). "Censored regression quantiles". In: *Journal of Econometrics* 32.1, pp. 143–155.

— (1986b). "Symmetrically Trimmed Least Squares Estimation for Tobit Models". In: *Econometrica* 54.6, pp. 1435–1460.

Tobin, J. (1958). "Estimation of Relationships for Limited Dependent Variables". In: *Econometrica* 26.1, pp. 24–36.

West, J. (2017). *A Closer Look at the Eigenfactor™ Metrics — Clarivate*.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). "The Increasing Dominance of Teams in Production of Knowledge". In: *Science* 316.5827, pp. 1036–1039.

Appendix A Triplet Construction

A triplet is a three-dimension tuple $(Y_{ij(\ell)}, Y_{i(\ell)}, Y_{j(\ell)})$ linking team project $Y_{ij(\ell)}$ between workers i and j, i's solo project $Y_{i(\ell)}$, and j's solo project $Y_{j(\ell)}$. Note the absence of asterisk $(*)$, as the truncated links are unobserved.

Since workers' collaboration pattern may be highly heterogeneous, a systematic approach is needed to construct triplets from a team network. For example, some workers may have few team projects but have many solo projects, or vice versa. This opens up to (many) different ways of building the sample of triplets. Our proposed algorithm below has two main advantages. First, it ensures independence across triplets. Second, if the time information of each project ℓ is available, which is the case in publication data, we can optimize the matching between solo and team projects. This helps mitigate potential concerns about time-varying types in a sample over long periods.

For example, suppose that A and B have a co-authored paper published in 2010. Additionally, A has two single-authored papers published in 2006 and 2011; B has one single-authored papers in 2015. Our algorithm would match A's 2011 project and B's 2015 project to their joint 2010 project.

Algorithm 2 Iterate until all team projects are either matched or dropped.

1. For each team project ℓ with outcome $Y_{ij(\ell)}$, identify its two workers i and j.
2. For i and j, list all of their unmatched solo projects.
3. Drop team project ℓ if either i or j has no *unmatched* solo project. This is to prevent reuse of the same project and to ensure independence across triplets.
4. If team project ℓ is not dropped, for i (same for j),
   - If there is only one $Y_{i(\ell)}$, simply match it to $Y_{ij(\ell)}$.
   - If there are multiple $\{Y_{i(\ell)}\}$, compute the absolute time differences between each of them with $Y_{ij(\ell)}$. Find the one with the smallest difference and match it to $Y_{ij(\ell)}$.
5. The triplet $(Y_{ij(\ell)}, Y_{i(\ell)}, Y_{j(\ell)})$ is now completed. Remove previously matched projects.

Appendix B Robustness check: Eigenfactor Index

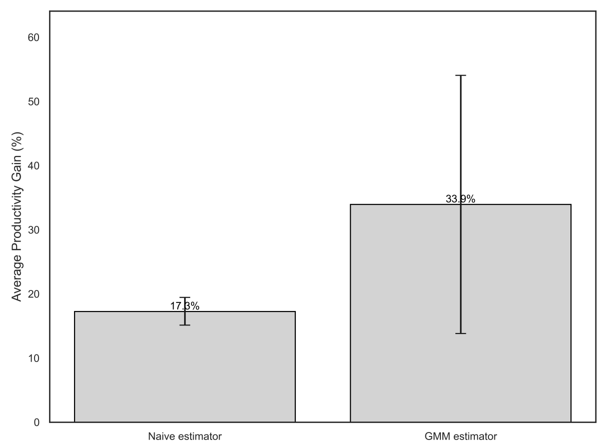

Figure 5: Average Productivity Gain (Eigenfactor Index)

Notes: This bar plot shows the average team premium that are implied by the naive estimate and the GMM estimates, with 90% confidence intervals. The outcome variable is measured by the eigenfactor index.

Appendix C Proof of Lemma 2

Lemma 2 (Truncated normal)**.** *Suppose $\widetilde{Y} \sim N(\tilde{\alpha}, \tilde{\sigma}^2)$ where $\tilde{\alpha} \in \mathbb{R}$ and $\tilde{\sigma} \geq 0$. Then, for $k \in \mathbb{N}^+$,*

$$\mathbb{E}[\widetilde{Y}^{k+1} - \tilde{\alpha}\widetilde{Y}^k - k\tilde{\sigma}^2\, \widetilde{Y}^{k-1} | \widetilde{Y} \geq 0] = 0.$$

Proof. Denote the k^{th} moment of the truncated normal Y as

$$\mu_k := \mathbb{E}[\widetilde{Y}^k | \widetilde{Y} \geq 0],$$

where $\mu_0 = 1$. Let ϕ and Φ denote the PDF and CDF of the standard normal distribution.

$$\begin{aligned}
\mu_{k+1} &= \int_0^\infty y^{k+1} \cdot \left(\frac{1}{\tilde{\sigma}} \frac{\phi(\frac{y-\tilde{\alpha}}{\tilde{\sigma}})}{1-\Phi(\frac{-\tilde{\alpha}}{\tilde{\sigma}})} \right) dy \\
&= \frac{1}{1-\Phi(\frac{-\tilde{\alpha}}{\tilde{\sigma}})} \int_0^\infty y^k \left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}^2} + \frac{\tilde{\alpha}}{\tilde{\sigma}^2} \right) \cdot \tilde{\sigma}^2 \left(\frac{1}{\tilde{\sigma}} \phi\left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}} \right) \right) dy \\
&= \tilde{\sigma} \frac{1}{1-\Phi(\frac{-\tilde{\alpha}}{\tilde{\sigma}})} \int_0^\infty y^k \left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}^2} \cdot \phi\left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}} \right) \right) dy + \tilde{\alpha} \left[\frac{1}{1-\Phi(\frac{-\tilde{\alpha}}{\tilde{\sigma}})} \int_0^\infty y^k \cdot \left(\frac{1}{\tilde{\sigma}} \phi\left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}} \right) \right) dy \right]
\end{aligned}$$

Note the second term is simply μ_k. Recall that $\phi(z)' = (-z)\phi(z)$. Applying integration by parts to the first term,

$$\begin{aligned}
\mu_{k+1} &= \tilde{\sigma} \frac{-1}{1-\Phi(\frac{-\tilde{\alpha}}{\tilde{\sigma}})} \int_0^\infty y^k \left(-\frac{y-\tilde{\alpha}}{\tilde{\sigma}} \cdot \frac{1}{\tilde{\sigma}} \phi\left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}} \right) \right) dy + \tilde{\alpha}\mu_k \\
&= -\tilde{\sigma} \left[\frac{1}{1-\Phi(\frac{-\tilde{\alpha}}{\tilde{\sigma}})} \int_0^\infty y^k \left(\frac{d}{dy} \phi\left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}} \right) \right) dy \right] + \tilde{\alpha}\mu_k \\
&= \tilde{\alpha}\mu_k - \tilde{\sigma} \left[y^k \phi\left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}} \right) \Big|_0^\infty - \int_0^\infty (ky^{k-1}) \cdot \phi\left(\frac{y-\tilde{\alpha}}{\tilde{\sigma}} \right) dy \right] \\
&= \tilde{\alpha}\mu_k + k\tilde{\sigma}^2 \mu_{k-1}.
\end{aligned}$$

Therefore, $\mu_{k+1} = \widetilde{\alpha}\mu_k + k\widetilde{\sigma}^2\mu_{k-1}$. After rearranging, we have shown that $\mathbb{E}[\widetilde{Y}^{k+1} - \widetilde{\alpha}\widetilde{Y}^k - k\widetilde{\sigma}^2\widetilde{Y}^{k-1} | \widetilde{Y}^* \geq 0] = 0$. □

Appendix D Proof of Proposition 1

Proposition 1 (System of moment restrictions)**.** *Define the k^{th} moment condition m_k as*

$$m_k(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}) := Y_{i(\ell)}^k Y_{j(\ell)}^k Y_{ij(\ell)}^k (Y_{ij(\ell)} - \lambda(Y_{i(\ell)} + Y_{j(\ell)})) + k\sigma^2 Y_{i(\ell)}^{k-1} Y_{j(\ell)}^{k-1} Y_{ij(\ell)}^{k-1} (\lambda(Y_{i(\ell)} + Y_{j(\ell)}) Y_{ij(\ell)} - Y_{i(\ell)} Y_{j(\ell)}).$$

Then, for $k \in \mathbb{N}^+$, the following equality holds:

$$\mathbb{E}\left[m_k(\lambda, \sigma, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}) \,\middle|\, Y_{i(\ell)}^* \geq 0,\, Y_{j(\ell)}^* \geq 0,\, Y_{ij(\ell)}^* \geq 0 \right] = 0.$$

Proof. First, we apply Lemma 2 three times to $Y_{i(\ell)} \sim N(\alpha_i, \sigma^2)$, $Y_{j(\ell)} \sim N(\alpha_j, \sigma^2)$, and $Y_{ij(\ell)} \sim N(\lambda(\alpha_i + \alpha_j), \sigma^2)$, respectively. This gives the following system of equalities:

$$\begin{aligned}
\mathbb{E}[Y_{i(\ell)}^{k+1} - \alpha_i Y_{i(\ell)}^k - k\sigma^2 Y_{i(\ell)}^{k-1} | Y_{i(\ell)}^* \geq 0] = 0, \\
\mathbb{E}[Y_{j(\ell)}^{k+1} - \alpha_j Y_{j(\ell)}^k - k\sigma^2 Y_{j(\ell)}^{k-1} | Y_{i(\ell)}^* \geq 0] = 0, \\
\mathbb{E}[Y_{ij(\ell)}^{k+1} - \lambda(\alpha_i + \alpha_j) Y_{ij(\ell)}^k - k\sigma^2 Y_{ij(\ell)}^{k-1} | Y_{ij(\ell)}^* \geq 0] = 0.
\end{aligned}$$

By Assumption 1, the moments above hold when we further condition on the fixed effects.

$$\begin{aligned}
\mathbb{E}[Y_{i(\ell)}^{k+1} - \alpha_i Y_{i(\ell)}^k - k\sigma^2 Y_{i(\ell)}^{k-1} | Y_{i(\ell)}^* \geq 0, Y_{j(\ell)}^* \geq 0, Y_{ij(\ell)}^* \geq 0, \alpha_i, \alpha_j] = 0, \\
\mathbb{E}[Y_{j(\ell)}^{k+1} - \alpha_j Y_{j(\ell)}^k - k\sigma^2 Y_{j(\ell)}^{k-1} | Y_{i(\ell)}^* \geq 0, Y_{j(\ell)}^* \geq 0, Y_{ij(\ell)}^* \geq 0, \alpha_i, \alpha_j] = 0, \\
\mathbb{E}[Y_{ij(\ell)}^{k+1} - \lambda(\alpha_i + \alpha_j) Y_{ij(\ell)}^k - k\sigma^2 Y_{ij(\ell)}^{k-1} | Y_{i(\ell)}^* \geq 0, Y_{j(\ell)}^* \geq 0, Y_{ij(\ell)}^* \geq 0, \alpha_i, \alpha_j] = 0.
\end{aligned}$$

First, scale both sides of each equation by λ. Second, multiply $Y^k_{j(\ell)}Y^k_{ij(\ell)}$ to the first equation, multiply $Y^k_{i(\ell)}Y^k_{ij(\ell)}$ to the second equation, and multiply $Y^k_{i(\ell)}Y^k_{j(\ell)}$ to the third equation. Equalities are preserved because shocks are distributed i.i.d. due to Assumption 1.

$$\begin{aligned}
\mathbb{E}[\lambda Y^{k+1}_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)} - \lambda\alpha_i Y^k_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)} - k\lambda\sigma^2 Y^{k-1}_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)}|Y^*_{i(\ell)} \geq 0, Y^*_{j(\ell)} \geq 0, Y^*_{ij(\ell)} \geq 0, \alpha_i, \alpha_j] = 0,\\
\mathbb{E}[\lambda Y^k_{i(\ell)}Y^{k+1}_{j(\ell)}Y^k_{ij(\ell)} - \lambda\alpha_j Y^k_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)} - k\lambda\sigma^2 Y^k_{i(\ell)}Y^{k-1}_{j(\ell)}Y^k_{ij(\ell)}|Y^*_{i(\ell)} \geq 0, Y^*_{j(\ell)} \geq 0, Y^*_{ij(\ell)} \geq 0, \alpha_i, \alpha_j] = 0,\\
\mathbb{E}[Y^k_{i(\ell)}Y^k_{j(\ell)}Y^{k+1}_{ij(\ell)} - \lambda(\alpha_i+\alpha_j) Y^k_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)} - k\sigma^2 Y^k_{i(\ell)}Y^k_{j(\ell)}Y^{k-1}_{ij(\ell)}|Y^*_{i(\ell)} \geq 0, Y^*_{j(\ell)} \geq 0, Y^*_{ij(\ell)} \geq 0, \alpha_i, \alpha_j] = 0.
\end{aligned}$$

Next, difference out (α_i, α_j) by subtracting the sum of the first two equations from the third. This gives

$$\begin{aligned}
\mathbb{E}[(Y^k_{i(\ell)}Y^k_{j(\ell)}Y^{k+1}_{ij(\ell)} - \lambda Y^{k+1}_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)} - \lambda Y^k_{i(\ell)}Y^{k+1}_{j(\ell)}Y^k_{ij(\ell)}) + k\sigma^2(\lambda Y^{k-1}_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)} + \lambda Y^k_{i(\ell)}Y^{k-1}_{j(\ell)}Y^k_{ij(\ell)} - Y^k_{i(\ell)}Y^k_{j(\ell)}Y^{k-1}_{ij(\ell)})\\
|Y^*_{i(\ell)} \geq 0, Y^*_{j(\ell)} \geq 0, Y^*_{ij(\ell)} \geq 0, \alpha_i, \alpha_j] = 0.
\end{aligned}$$

By the law of iterated expectations,

$$\begin{aligned}
\mathbb{E}[(Y^k_{i(\ell)}Y^k_{j(\ell)}Y^{k+1}_{ij(\ell)} - \lambda Y^{k+1}_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)} - \lambda Y^k_{i(\ell)}Y^{k+1}_{j(\ell)}Y^k_{ij(\ell)}) + k\sigma^2(\lambda Y^{k-1}_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)} + \lambda Y^k_{i(\ell)}Y^{k-1}_{j(\ell)}Y^k_{ij(\ell)} - Y^k_{i(\ell)}Y^k_{j(\ell)}Y^{k-1}_{ij(\ell)})\\
|Y^*_{i(\ell)} \geq 0, Y^*_{j(\ell)} \geq 0, Y^*_{ij(\ell)} \geq 0] = 0.
\end{aligned}$$

After rearranging,

$$\mathbb{E}[Y^k_{i(\ell)}Y^k_{j(\ell)}Y^k_{ij(\ell)}(Y_{ij(\ell)} - \lambda(Y_{i(\ell)}+Y_{j(\ell)})) + k\sigma^2 Y^{k-1}_{i(\ell)}Y^{k-1}_{j(\ell)}Y^{k-1}_{ij(\ell)}(\lambda(Y_{i(\ell)}+Y_{j(\ell)})Y_{ij(\ell)} - Y_{i(\ell)}Y_{j(\ell)})|Y^*_{i(\ell)} \geq 0, Y^*_{j(\ell)} \geq 0, Y^*_{ij(\ell)} \geq 0] = 0.$$

□

Appendix E Proof of Theorem 1

Theorem 1 (Truncation-robust estimator)**.** *Let C be the number of uncorrelated triplets. Define the sample moments*

$$\hat{g}(\lambda,\sigma)' = \frac{1}{C}\sum_{\{i,j\}\in\mathcal{E}}\Big(m_1(\lambda,\sigma,Y_{i(\ell)},Y_{j(\ell)},Y_{ij(\ell)}) \quad m_2(\lambda,\sigma,Y_{i(\ell)},Y_{j(\ell)},Y_{ij(\ell)})\Big).$$

Suppose

(i) Assumption 1 holds.

(ii) The limiting gradient and variance matrices exist. Specifically,

$$\begin{aligned}
G &= \operatorname*{plim}_{C\to\infty}\frac{1}{C}\sum_{\{i,j\}\in\mathcal{E}}\mathbb{E}\left[\tfrac{\partial}{\partial\lambda}m(\lambda,\sigma,Y_{i(\ell)},Y_{j(\ell)},Y_{ij(\ell)}), \quad \tfrac{\partial}{\partial\sigma}m(\lambda,\sigma,Y_{i(\ell)},Y_{j(\ell)},Y_{ij(\ell)})\right],\\
V &= \operatorname*{plim}_{C\to\infty}\frac{1}{C}\sum_{\{i,j\}\in\mathcal{E}}Var(m(\lambda,\sigma,Y_{i(\ell)},Y_{j(\ell)},Y_{ij(\ell)})).
\end{aligned}$$

(iii) Global identification of (λ, σ).

Then, the baseline model (1) has a consistent GMM estimator, that is,

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\sigma} \end{pmatrix} = \underset{\lambda \in \mathbb{R}, \sigma \in \mathbb{R}^{+}}{\arg\min}\ \hat{g}(\lambda, \sigma)' \hat{g}(\lambda, \sigma).$$

Furthermore, as $C \to \infty$, the asymptotic distribution is

$$\sqrt{C} \left[\begin{pmatrix} \hat{\lambda} \\ \hat{\sigma} \end{pmatrix} - \begin{pmatrix} \lambda \\ \sigma \end{pmatrix} \right] \Longrightarrow N\left(0, (G'V^{-1}G)^{-1}\right).$$

Proof. We verify the conditions discussed in Hansen (1982). First, we check that the limiting gradient vector $G = \mathbb{E}\left[\frac{\partial m(\lambda, \sigma^2, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})}{\partial \lambda} \frac{\partial m(\lambda, \sigma^2, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)})}{\partial \sigma^2}\right]$ exists. Second, we check that the sample variance of the moments converges to a well-defined $V = \mathrm{Var}(m(\lambda, \sigma^2, Y_{i(\ell)}, Y_{j(\ell)}, Y_{ij(\ell)}))$. The final rank condition implied by the assumption guarantees the local identification of the parameters. Therefore, the asymptotic distribution of the sample GMM estimator is

$$\sqrt{C} \left[\begin{pmatrix} \hat{\lambda} \\ \hat{\sigma}^2 \end{pmatrix} - \begin{pmatrix} \lambda \\ \sigma^2 \end{pmatrix} \right] \Longrightarrow N\left(0, (G'V^{-1}G)^{-1}\right).$$

□