Hsieh, Chih-Sheng; König, Michael; Liu, Xiaodong; Zimmermann, Christian

Working Paper

Collaboration in Bipartite Networks, with an Application to Coauthorship Networks

Tinbergen Institute Discussion Paper, No. TI 2020-056/VIII

TI 2020-056/VIII
Tinbergen Institute Discussion Paper

Collaboration in Bipartite Networks, with an Application to Coauthorship Networks

Chih-Sheng Hsieh[1]
Michael D. König[2]
Xiaodong Liu[3]
Christian Zimmermann[4]

[1] National Taiwan University
[2] VU Amsterdam, Spatial Economics and Tinbergen Institute
[3] University of Colorado Boulder
[4] Federal Reserve Bank of St. Louis

Collaboration in Bipartite Networks, with an Application to Coauthorship Networks*

Chih-Sheng Hsieh† Michael D. König‡ Xiaodong Liu§

Christian Zimmermann¶

August 17, 2020

*This paper was previously circulated under the title "Superstar economists: coauthorship networks and research output". We thank Pierre-Philippe Combes, Lorenzo Ductor, Bauke Visser, Kenan Huremovic, Marco Van Der Leij, Matt Jackson, David Miller, Joachim Voth, Francis Bloch, Pietro Biroli, Ralph Ossa, Hannes Schwandt, Seth Richards-Shubik, Fabrizio Zilibotti and various seminar participants at the following institutions: University of Zurich, Laval, Amsterdam, and Munich; Jean Monnet University in St-Étienne, Chinese University of Hong Kong, Baptist University of Hong Kong, National Taiwan University, the NSF Conference on Network Science and Economics at Washington University, the workshop on the Economics of Scientific Research at Erasmus University Rotterdam and the Econometric Society Meeting in Barcelona for their helpful comments. Moreover, we thank Adrian Etter and Marc Biedermann for excellent research assistance. Michael D. König acknowledges financial support from the Swiss National Science Foundation through research grant PZ00P1_154957 /1. The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

†Department of Economics, National Taiwan University, Taipei 10617, Taiwan. Email: csh-sieh@ntu.edu.tw.

‡Centre for Economic Policy Research (CEPR), London, United Kingdom. ETH Zurich, Swiss Economic Institute (KOF), Zurich, Switzerland. Department of Spatial Economics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: m.d.konig@vu.nl.

§Department of Economics, University of Colorado Boulder, Boulder, Colorado 80309-0256, United States. Email: xiaodong.liu@colorado.edu.

¶Department of Economic Research, Federal Reserve Bank of St. Louis, St. Louis MO 63166-0442, United States. Email: zimmermann@stlouisfed.org.

Abstract

This paper studies the impact of collaboration on research output. First, we build a micro-founded model for scientific knowledge production, where collaboration between researchers is represented by a bipartite network. The equilibrium of the game incorporates both the complementarity effect between collaborating researchers and the substitutability effect between concurrent projects of the same researcher. Next, we develop a Bayesian MCMC procedure to estimate the structural parameters, taking into account the endogenous matching of researchers and projects. Finally, we illustrate the empirical relevance of the model by analyzing the coauthorship network of economists registered in the RePEc Author Service.

Keywords: bipartite networks, coauthorship networks, research collaboration, spillovers, economics of science.

JEL: C31, C72, D85, L14

1 Introduction

Collaboration between researchers in economics has become significantly more important in recent decades. In 1996 multi-authored papers accounted for 50% of all articles published in economics. This number increased to over 75% in 2014 (Kuld and O'Hagan, 2018).[1] Through a complex network of collaborations, researchers generate spillovers not only to their coauthors but also to other researchers indirectly connected to them. The aim of this paper is to develop a structural model that helps us to understand how collaboration affects research output.

First, we build a micro-founded model for scientific knowledge production. The collaboration between researchers is characterized by a bipartite network with two types of nodes: researchers and research projects. The effort that a researcher spends in a project is represented by an edge in the bipartite network, and collaborating researchers are connected

[1] Additional evidence can be found in Ductor (2014).

through the project they work on together. We characterize the equilibrium of the game where researchers choose efforts in multiple and possibly overlapping projects to maximize utility. The equilibrium takes into account both the complementarity (or spillover) effect between collaborating researchers and the substitutability (or congestion) effect between concurrent projects of the same researcher.

Next, we propose an estimation procedure to recover the structural parameters of the model. There are three main challenges in estimating this model. First, the effort level of a researcher in the production function is unobservable. To overcome this problem, we substitute the equilibrium effort level derived from the theoretical model into the production function. The resulting equilibrium production function is highly nonlinear with a likelihood function involving high-dimensional integrals. This leads to the second challenge of the estimation, i.e., it is computationally cumbersome to apply a frequentist maximum likelihood method, even when resorting to a simulation approach. To bypass this difficulty, we adopt a Bayesian Markov Chain Monte Carlo (MCMC) approach to estimate the equilibrium production function. Lastly, the matching between researchers and projects is likely to be endogenous. Estimating the production function without taking into account this potential endogeneity may incur a selection bias. We introduce a participation function to model the endogenous selection of researchers into projects. Then we jointly estimate the production function and the participation function allowing for both researcher and project unobserved heterogeneity.[2]

Finally, we bring our model to the data by analyzing the coauthorship network of economists registered in the Research Papers in Economics (RePEc) Author Service. We find that the spillover effect and the congestion effect are statistically significant with the expected signs. The estimates are biased when the endogenous matching between researchers and projects is ignored. The direction of the bias is compatible with the intuition and consistent with the Monte Carlo simulation results. We also conduct a series of robustness checks

[2]As pointed out in Bonhomme (2020), a key feature of bipartite networks is two-sided heterogeneity.

to explore the sensitivity of our results to alternative specifications and samples.

There exists a growing literature, both empirical and theoretical, on the formation and impact of scientific collaboration networks. On the empirical side, the structural features of scientific collaboration networks have been analyzed in Newman (2001a,b,c, 2004a,b) and Goyal et al. (2006). Fafchamps et al. (2010) study predictors for the establishment of scientific collaborations. Ductor (2014), Ductor et al. (2014), and Anderson and Richards-Shubik (2019) study how collaboration affects the research output of individual authors. In this paper, we take a structural approach by introducing a micro-founded model to characterize how collaboration facilitates scientific knowledge production.

Our paper is further related to the recent theoretical contributions by Baumann (2014) and Salonen (2016), where agents choose time to invest into bilateral relationships. Our model extends the setup considered in these papers by allowing for investments into multiple projects that could involve more than two agents. Moreover, in a related paper Bimpikis et al. (2019) analyze firms competing in quantities à la Cournot across different markets with a similar linear-quadratic payoff specification and allow firms to choose endogenously the quantities sold to each market. While the products sold by competing firms to the same market are substitutes in Bimpikis et al. (2019), the efforts spent by collaborating agents in the same project are strategic complements in our model.

The rest of the paper is organized as follows. Section 2 introduces the theoretical model and characterizes the equilibrium. Section 3 presents the econometric methodology. The empirical implications of the model are discussed in Section 4, where Section 4.1 describes the data used in the empirical study, Section 4.2 gives the main estimation results, and Section 4.3 provides robustness analysis. Section 5 briefly concludes. The proofs, technical details, and additional robustness checks can be found in the online appendix.

2 Theoretical Model

2.1 Bipartite Network, Production Function, and Utility

Consider a *bipartite* network given by $\mathcal{G} = (\mathcal{N}, \mathcal{P}, \mathcal{E})$, where $\mathcal{N} = \{1, \ldots, n\}$ denotes the set of agents, $\mathcal{P} = \{1, \ldots, p\}$ denotes the set of projects, and $\mathcal{E}$ denotes the set of edges connecting agents and projects. In our model, an edge $e_{is} \in \mathcal{E}$ is the (non-negative) effort that agent i spends in project s. Let $\mathcal{N}_s$ denote the set of agents working on project s and $\mathcal{P}_i$ denote the set of projects agent i participates in. Let $|\cdot|$ denote the cardinality of a set.

The *production function* for project $s \in \mathcal{P}$ is given by

$$y_s(\mathcal{G}) = \sum_{i \in \mathcal{N}_s} \alpha_i e_{is} + \frac{\lambda}{2} \sum_{i \in \mathcal{N}_s} \sum_{j \in \mathcal{N}_s \setminus \{i\}} g_{ij} e_{is} e_{js} + \epsilon_s, \tag{1}$$

where $y_s(\mathcal{G})$ (or simply y_s) is the output of project s, α_i represents individual heterogeneity in productivity, $g_{ij} \in [0, 1]$ measures the degree of complementarity between collaborating agents, and ϵ_s is a random shock. The spillover effect is captured by the coefficient λ.

We assume that the *utility* of agent i is given by

$$U_i(\mathcal{G}) = \underbrace{\sum_{s \in \mathcal{P}_i} \delta_s y_s}_{\text{payoff}} - \underbrace{\frac{1}{2} \left(\sum_{s \in \mathcal{P}_i} e_{is}^2 + \phi \sum_{s \in \mathcal{P}_i} \sum_{t \in \mathcal{P}_i \setminus \{s\}} e_{is} e_{it} \right)}_{\text{cost}}. \tag{2}$$

The utility function has a payoff/cost structure. The payoff is the weighted total output of the projects agent i participates in, with the weights given by $\delta_s \in (0, 1]$.[3] The cost is quadratic in efforts, with the coefficient ϕ measuring the degree of substitutability of an agent's efforts in different projects.[4] This cost is convex if and only if the $|\mathcal{P}_i| \times |\mathcal{P}_i|$ matrix Φ_i, with diagonal elements equal to one and off-diagonal elements equal to ϕ, is positive

[3] For example, if $\delta_s = 1/|\mathcal{N}_s|$, then the individual payoff is discounted by the number of agents participating in project s (cf. Kandel and Lazear, 1992; Jackson and Wolinsky, 1996; Hollis, 2001).

[4] For example, Ductor (2014) finds evidence for a congestion externality proxied by the average number of coauthors' papers that has a negative effect on individual academic productivity.

definite. The quadratic cost specification includes the convex separable cost specification as a special case with $\phi = 0$. A theoretical model with a similar cost specification but allowing for only two activities is studied in Belhaj and Deroïan (2014) and an empirical analysis is provided in Liu (2014) and Cohen-Cole et al. (2018). In addition, a convex separable cost specification can be found in the model studied in Adams (2006).

2.2 Game and Equilibrium

The underlying game has two stages. In the first stage, agents decide which projects to participate in. The outcome of the first stage are characterized by indicator variables d_{is}, such that $d_{is} = 1$ if agent i participates in project s and $d_{is} = 0$ otherwise. Given the outcome of the first stage, agents simultaneously choose research efforts $e_{is} \geq 0$ to maximize utility in the second stage.

The following proposition provides an equilibrium characterization of the agents' effort portfolio $e = (e_1', \cdots, e_p')'$, with $e_s = (e_{1s}, \cdots, e_{ns})'$ for $s = 1, \cdots, p$, in the projects they participate in. Let

$$W = D(\mathrm{diag}_{s=1}^{p}\{\delta_s\} \otimes G)D, \qquad \text{and} \qquad M = D(J_p \otimes I_n)D, \tag{3}$$

where $\otimes$ denotes the Kronecker product, D is an np-dimensional diagonal matrix given by $D = \mathrm{diag}_{s=1}^{p}\{\mathrm{diag}_{i=1}^{n}\{d_{is}\}\}$, G is an $n \times n$ zero-diagonal matrix with the (i,j)th $(i \neq j)$ element being g_{ij}, and J_p is an $p \times p$ zero-diagonal matrix with off-diagonal elements equal to one. Let $\rho_{\max}(\cdot)$ denote the spectral radius of a square matrix.

Proposition 1. *Suppose the production function for each project $s \in \mathcal{P}$ is given by Equation (1) and the utility function for each agent $i \in \mathcal{N}$ is given by Equation (2). Let $L := L(\lambda, \phi) = \lambda W - \phi M$. Given the outcome of the first stage of the game, if*

$$\rho_{\max}(L) < 1, \tag{4}$$

then the equilibrium effort portfolio is given by

$$e^* = (I_{np} - L)^{-1} D(\delta \otimes \alpha), \tag{5}$$

where $\delta = (\delta_1, \cdots, \delta_p)'$ *and* $\alpha = (\alpha_1, \cdots, \alpha_n)'$.

The matrix L represents a weight matrix of the *line graph* $\mathcal{L}(\mathcal{G})$ for the bipartite network $\mathcal{G}$,[5] where each link between nodes sharing a project has weight $\lambda\delta_s g_{ij}$, and each link between nodes sharing an author has weight $-\phi$. An example can be found in Figure 1 with $g_{ij} = 1$ for all $i \neq j$ and $\delta_s = 1$ for all s. We illustrate the equilibrium characterization of Proposition 1 in the following example corresponding to the bipartite network in Figure 1.

Example 1. *Consider a bipartite network with* 3 *agents and* 2 *projects, where agents* 1 *and* 2 *are collaborating in the first project and agents* 1 *and* 3 *are collaborating in the second project. An illustration can be found in Figure 1. For expositional purposes, let* $g_{ij} = 1$ *for all* $i \neq j$ *and* $\delta_s = 1$ *for all* s. *Following Equation (3),*

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad \textit{and} \quad M = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

[5] Given a network $\mathcal{G}$, its line graph $\mathcal{L}(\mathcal{G})$ is a graph such that each node of $\mathcal{L}(\mathcal{G})$ represents an edge of $\mathcal{G}$, and two nodes of $\mathcal{L}(\mathcal{G})$ are connected if and only if their corresponding edges share a common endpoint in $\mathcal{G}$ (cf. e.g., West, 2001).

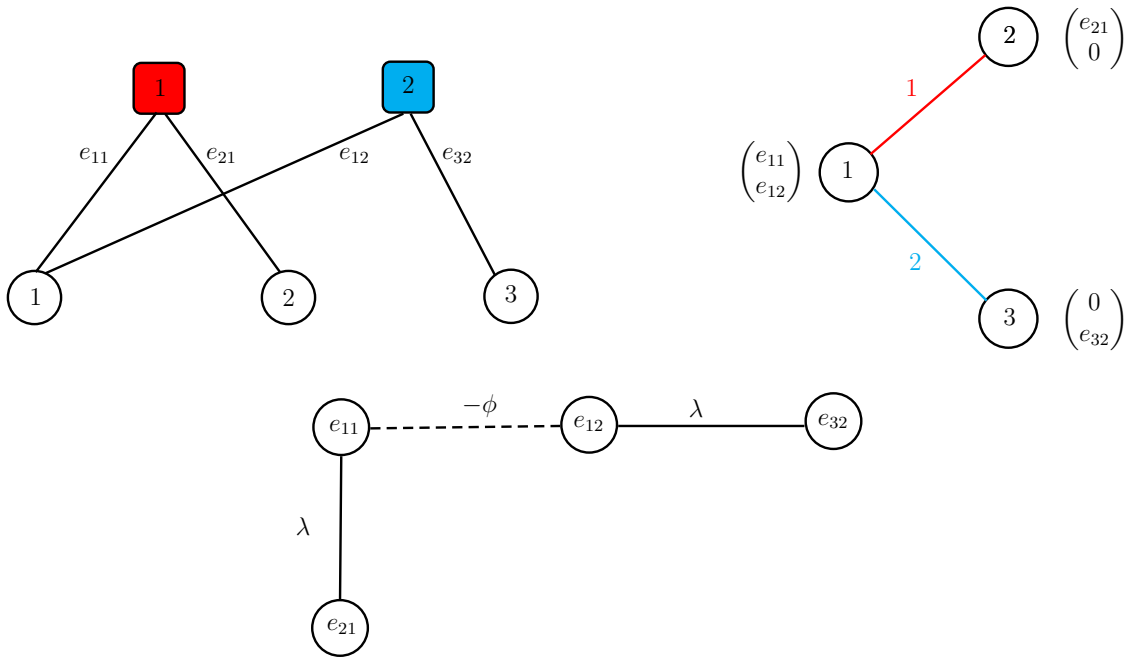

Figure 1: Top left panel: the bipartite collaboration network $\mathcal{G}$ of authors and projects analyzed in Example 1, where circles represent authors and squares represent projects. Top right panel: the projection of the bipartite network $\mathcal{G}$ on the set of coauthors. The effort levels of the individual agents for each project they are involved in are indicated next to the nodes. Bottom panel: the line graph $\mathcal{L}(\mathcal{G})$ associated with the collaboration network $\mathcal{G}$, in which each node represents the effort an author invests into different projects. Solid lines indicate nodes sharing a project while dashed lines indicate nodes with the same author.

and hence

$$L = \lambda W - \phi M = \begin{bmatrix} 0 & \lambda & 0 & -\phi & 0 & 0 \\ \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -\phi & 0 & 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 \end{bmatrix}.$$

The nonzero entries of the matrices W and M correspond to, respectively, the solid lines and the dashed lines in the line graph depicted in the bottom panel of Figure 1. Thus, the $(1,2)$th and $(2,1)$th elements of the matrix L represent the link between e_{11} and e_{21} with weight λ in the line graph, the $(4,6)$th and $(6,4)$th elements represent the link between e_{12} and e_{32} with weight λ, and the $(1,4)$th and $(4,1)$th elements represent the link between e_{11} and e_{12} with weight $-\phi$.

In this example, the sufficient condition (4) for the existence of a unique equilibrium holds if $|\phi| < 1 - \lambda^2$. From Equation (5) the equilibrium effort portfolio is

$$e^* = \begin{bmatrix} e^*_{11} \\ e^*_{21} \\ e^*_{31} \\ e^*_{12} \\ e^*_{22} \\ e^*_{32} \end{bmatrix} = \frac{1}{(1-\lambda^2)^2 - \phi^2} \begin{bmatrix} (1-\lambda^2-\phi)\alpha_1 + \lambda(1-\lambda^2)\alpha_2 - \lambda\phi\alpha_3 \\ \lambda(1-\lambda^2-\phi)\alpha_1 + (1-\lambda^2-\phi^2)\alpha_2 - \lambda^2\phi\alpha_3 \\ 0 \\ (1-\lambda^2-\phi)\alpha_1 - \lambda\phi\alpha_2 + \lambda(1-\lambda^2)\alpha_3 \\ 0 \\ \lambda(1-\lambda^2-\phi)\alpha_1 - \lambda^2\phi\alpha_2 + (1-\lambda^2-\phi^2)\alpha_3 \end{bmatrix}.$$

Observe that

$$\begin{aligned}
\frac{\partial e_{11}^*}{\partial \alpha_1} &= \frac{\partial e_{12}^*}{\partial \alpha_1} = \frac{1}{1-\lambda^2+\phi} > 0 \\
\frac{\partial e_{21}^*}{\partial \alpha_1} &= \frac{\partial e_{32}^*}{\partial \alpha_1} = \frac{\lambda}{1-\lambda^2+\phi} > 0 \\
\frac{\partial e_{21}^*}{\partial \alpha_2} &= \frac{\partial e_{32}^*}{\partial \alpha_3} = \frac{1-\lambda^2-\phi^2}{(1-\lambda^2)^2-\phi^2} > 0 \\
\frac{\partial e_{11}^*}{\partial \alpha_2} &= \frac{\partial e_{12}^*}{\partial \alpha_3} = \frac{\lambda(1-\lambda^2)}{(1-\lambda^2)^2-\phi^2} > 0
\end{aligned}$$

which suggest that more-productive agents raise not only their own effort levels but also the effort levels of their collaborators. On the other hand,

$$\begin{aligned}
\frac{\partial e_{11}^*}{\partial \alpha_3} &= \frac{\partial e_{12}^*}{\partial \alpha_2} = -\frac{\lambda\phi}{(1-\lambda^2)^2-\phi^2} < 0 \\
\frac{\partial e_{21}^*}{\partial \alpha_3} &= \frac{\partial e_{32}^*}{\partial \alpha_2} = -\frac{\lambda^2\phi}{(1-\lambda^2)^2-\phi^2} < 0
\end{aligned}$$

which suggest that more-productive agents induce lower effort levels spent by agents on other projects. An illustration can be seen in the top panels of Figure 2.

The marginal change of the equilibrium effort e_{11}^* *of agent* 1 *in project* 1 *with respect to the spillover coefficient* λ *is given by*

$$\frac{\partial e_{11}^*}{\partial \lambda} = \frac{2\lambda(1-\lambda^2-\phi)^2\alpha_1 + [(1-\lambda^4-\phi^2)(1-\lambda^2)+2\lambda^2\phi^2]\alpha_2 - \phi[(1+3\lambda^2)(1-\lambda^2)-\phi^2]\alpha_3}{[(1-\lambda^2)^2-\phi^2]^2}.$$

Observe that the coefficient of α_3 *is negative. Thus, when* α_3 *is large enough,* $\partial e_{11}^*/\partial\lambda$ *could be negative. The reason is that, with increasing* λ*, the complementarity effects between collaborating agents become stronger, and this effect is more pronounced for the collaboration of agent* 1 *with the more-productive agent* 3*, than with the less-productive agent* 2*. Moreover, when the substitution effect parameter* ϕ *is large, agent* 1 *may spend even less effort in the project with agent* 2*, indicating congestion effects across projects. An illustration can be seen in the bottom panels of Figure 2.*

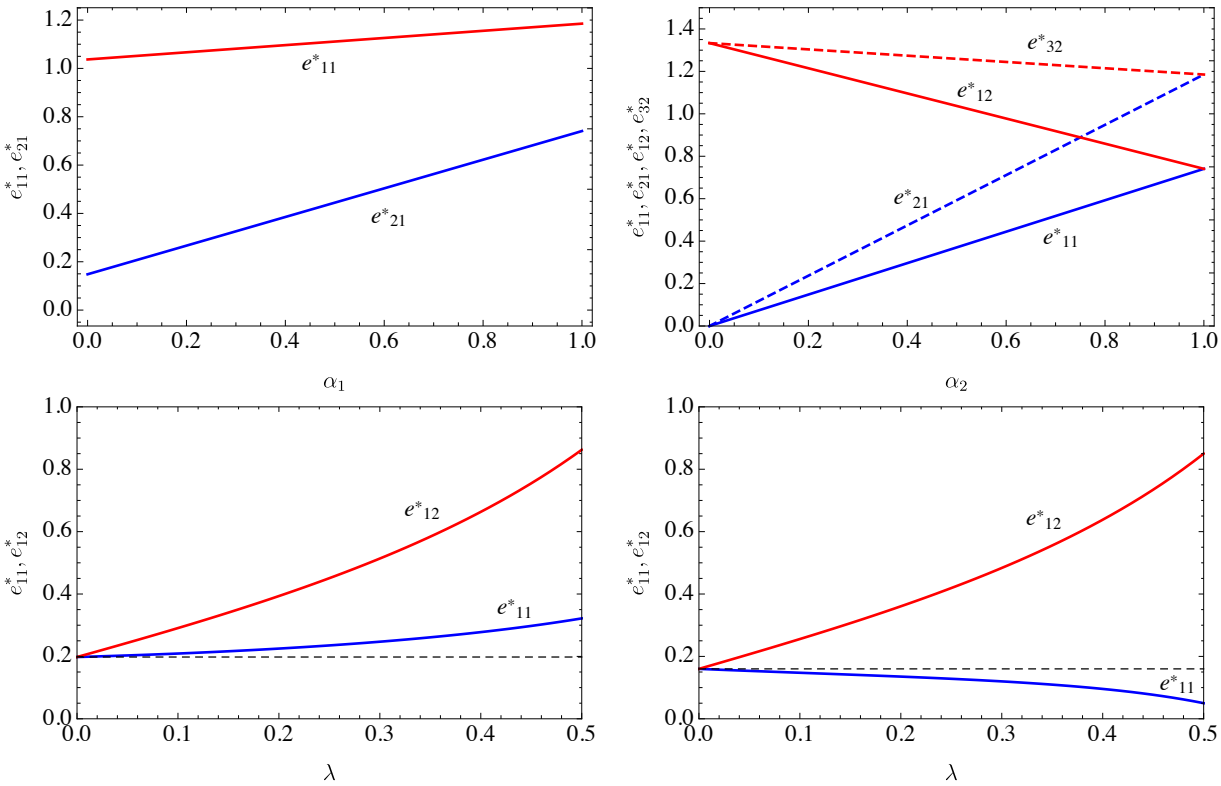

Figure 2: Top left panel: equilibrium effort levels for agents 1 and 2 in project 1 for $\phi = 0.75$, $\lambda = 0.25$, $\alpha_2 = \alpha_3 = 1$ (where $e^*_{11} = e^*_{12}$ and $e^*_{21} = e^*_{32}$) and varying values of α_1. Top right panel: equilibrium effort levels for agents 1, 2 and 3 in projects 1 and 2 for $\alpha_1 = \alpha_3 = 1$, $\phi = 0.75$, $\lambda = 0.25$ and varying values of α_2. Bottom panels: equilibrium effort levels for agent 1 with $\alpha_1 = 0.2$, $\alpha_2 = 0.1$, $\alpha_3 = 0.9$, $\phi = 0.05$ (bottom left panel) and $\phi = 0.25$ (bottom right panel) for varying values of λ. The dashed lines in the bottom panels indicate the effort level for $\lambda = 0$.

3 Estimation

Let $d_{is} = \mathbb{1}(i \in \mathcal{N}_s)$, where $\mathbb{1}(\cdot)$ denotes an indicator function. Equation (1) can be rewritten as

$$y_s = \sum_{i \in \mathcal{N}} \alpha_i d_{is} e_{is} + \frac{\lambda}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} g_{ij} d_{is} d_{js} e_{is} e_{js} + \epsilon_s, \tag{6}$$

where ϵ_s is i.i.d.$(0, \sigma_\epsilon^2)$. In the empirical model, we assume agent i's productivity is given by

$$\alpha_i = \exp(x_i' \beta + \zeta \mu_i) \tag{7}$$

where x_i is a vector of observable individual attributes and μ_i is an i.i.d.$(0, 1)$ random component capturing unobservable individual heterogeneity.

There are three main challenges in estimating this model. First, the effort level e_{is} is usually unobservable to the econometrician. To overcome this problem, we replace e_{is} in Equation (6) with the equilibrium effort level e_{is}^* given by Equation (5) and estimate the parameter vector $\vartheta_y = (\lambda, \phi, \beta', \zeta, \sigma_\epsilon^2)'$ in the equilibrium production function

$$y_s = \sum_{i \in \mathcal{N}} \alpha_i d_{is} e_{is}^* + \frac{\lambda}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} g_{ij} d_{is} d_{js} e_{is}^* e_{js}^* + \epsilon_s. \tag{8}$$

Second, the likelihood function of Equation (8) involves high-dimensional integrals and is computationally cumbersome to evaluate. As the equilibrium effort level e_{is}^* depends on individual productivities $\alpha_1, \cdots, \alpha_n$, which in turn depend on individual random components $\mu_1, \cdots, \mu_n$, according to Equations (5) and (7), the equilibrium production function (8) is nonlinear in $\mu = (\mu_1, \cdots, \mu_n)'$. The joint density of $y = (y_1, \cdots, y_p)'$ is given by

$$f(y|\vartheta_y) = \int f(y|\mu, \vartheta_y) f(\mu) d\mu$$

where $f(y|\mu, \vartheta_y)$ is the conditional density of y given μ and $f(\mu)$ is the joint density of μ. As $f(y|\vartheta_y)$ involves n-dimensional integrals, it is computational intensive to estimate ϑ_y by the

frequentist maximum likelihood method. To bypass this difficulty, we follow the approach of Zeger and Karim (1991) to sample μ together with ϑ_y from their joint posterior density $p(\mu, \vartheta_y|y) \propto f(y|\mu, \vartheta_y)\pi(\mu)\pi(\vartheta_y)$ with the priors $\pi(\mu)$ and $\pi(\vartheta_y)$.

Third, d_{is} is likely to be endogenous. For example, in a coauthorship network, high-ability researchers tend to work on many projects at the same time, and high-potential projects are usually more demanding for researchers. Estimating Equation (8) without taking into account potential endogeneity of d_{is} may incur a selection bias. To control for the endogenous selection, we assume that, in the first stage of the game, agent i decides whether to participate in project s according to

$$d_{is} = \mathbb{1}(z_{is}'\gamma + \xi\mu_i + \psi\eta_s + v_{is} > 0), \tag{9}$$

where z_{is} is a vector of observables measuring compatibility between agent i and project s,[6] μ_i is an i.i.d.$(0, 1)$ agent-specific random component, η_s is an i.i.d.$(0, 1)$ project-specific random component, and v_{is} is an i.i.d.$(0, 1)$ error term independent of μ_i and η_s. Conditional on observables, d_{is} and y_s are correlated due to the agent-specific random component μ_i that appears in both Equations (7) and (9). Furthermore, to model the correlation between d_{is} and ϵ_s in Equation (8), we rewrite ϵ_s as

$$\epsilon_s = \varsigma\eta_s + u_s,$$

where u_s is an i.i.d.$(0, \sigma_u^2)$ error term independent of η_s. In this specification, if $\zeta > 0$ and $\xi > 0$, then a researcher with higher ability (given by a higher μ_i) tends to participate in more projects; and if $\varsigma > 0$ and $\psi < 0$, then a project with higher potential (given by a higher η_s) has a higher threshold for researchers to participate in.

Let $\theta_d = (\gamma', \xi, \psi)'$ and $\theta_y = (\lambda, \phi, \beta', \zeta, \varsigma, \sigma_u^2)'$. Let $f(d|\mu, \eta, \theta_d)$ denote the joint probabil-

[6] Equation (9) can be considered as a reduced form of a game, where an agent's participation decision may depend on the attributes of other agents. In the empirical illustration, z_{is} includes terms capturing the similarity between agent i and other agents collaborating in project s in terms of affiliation, alma mater, etc.

ity of $d = [d_{is}]$ given $\mu = (\mu_1, \cdots, \mu_n)'$ and $\eta = (\eta_1, \cdots, \eta_p)'$, and $f(y|d, \mu, \eta, \theta_y)$ denote the conditional density of y given d, μ, and η. Then, μ, η, and $\theta = (\theta_y', \theta_d')'$ can be sampled from the joint posterior density $p(\mu, \eta, \theta|y, d) \propto f(y|d, \mu, \eta, \theta_y) f(d|\mu, \eta, \theta_d)\pi(\mu)\pi(\eta)\pi(\theta_y)\pi(\theta_d)$ with the priors $\pi(\mu)$, $\pi(\eta)$, $\pi(\theta_y)$ and $\pi(\theta_d)$. The details of Bayesian estimation can be found in Appendix B.

4 Empirical Study: Coauthorship Networks

4.1 Data

The data used for this study make extensive use of the metadata assembled by the RePEc initiative and its various projects. RePEc assembles information about publications relevant to economics from over 2,000 publishers, including all major commercial publishers and university presses, policy institutions, and pre-prints (working papers) from academic institutions.[7]

In addition, we make use of the data made available by various projects that build on these RePEc data and enhance it in various ways. First, we take the publication profiles of economists registered with the RePEc Author Service, which include what they have published and where they are affiliated.[8] Second, we extract information about their advisors, students, and alma mater, as recorded in the RePEc Genealogy project.[9] This academic genealogy data has been complemented with some of the data used in Colussi (2017).[10] Third, we use the New Economics Papers (NEP) project to identify the field-specific mailing lists through which the papers have been disseminated.[11] NEP has human editors who determine the field in which new working papers belong. We obtain 99 distinct NEP fields. Fourth, we

[7] See `http://repec.org/` for a general description of RePEc.
[8] RePEc Author Service: `https://authors.repec.org/`
[9] RePEc Genealogy project: `https://genealogy.repec.org/`
[10] We would like to thank Tommaso Colussi for sharing the data with us.
[11] NEP project: `https://nep.repec.org/`

use citations to the papers and articles as extracted by the CitEc project.[12] Finally, we use journal impact factors, as well as author and institution rankings from IDEAS.[13]

Compared with other data sources, RePEc has the advantage of linking these various datasets in a seamless way that is verified by the respective authors. Author identification is superior to any other dataset as homonyms are disambiguated by the authors themselves as they register and maintain their accounts. While not every author is registered, most are. Indeed, 90% of the top 1000 economists as measured by their publication records for the 1990-2000 period are registered.[14] We believe that the proportion is higher for the younger generation that is more familiar with social networks and online tools and thus more likely to register with online services.

In terms of publications, RePEc covers all important outlets and over 3,000 journals are listed, most of them with extensive coverage. References are extracted for about 30% of their articles (in addition to working papers) to compute citation counts and impact factors. The missing references principally come from publishers refusing to release them for reasons related to copyright protection. While the resulting gap is unfortunate, it is unlikely to result in a bias against particular authors, fields, or journals. The exception may be authors who are significantly cited in outlets outside of economics that may or may not be indexed in RePEc (note that several top management, statistics, and political science journals are also indexed).

To obtain a sample from RePEc that is appropriate for our analysis, we apply a series of filters as follows. First, we select papers that had a first pre-print version in 2010-2012. We choose 2010-2012 because it is old enough to give all authors a chance to have added the papers to their profiles and for the papers to have been eventually published in journals; but not too old for a good data coverage, as the coverage of RePEc becomes slimmer with older vintages. Furthermore, we require all authors of the papers to be registered with RePEc and

[12] CitEc project: `http://citec.repec.org/`

[13] IDEAS: `https://ideas.repec.org/top/`. For a detailed description of the factors and rankings, see Zimmermann (2013).

[14] `https://ideas.repec.org/coupe.html`

all authors to have the RePEc Genealogy information on where they studied. We drop all duplicate or older versions of each paper from our sample. This gives us a sample of 6,673 papers written by 3,700 distinct authors for which we have complete data.

Next, as we use citations to measure research output, we drop 2,463 papers that do not have any citations up to November 2018 when the data is extracted from the RePEc database, as well as 658 authors who only work on these dropped papers without any citations. This reduces to the sample size to 4,210 papers and 3,042 authors.[15]

Finally, as we are interested in collaborations between researchers, we drop 621 authors who wrote only single-authored papers in the sample period. This results in a final sample of 3,589 papers written by 2,421 distinct authors.[16]

In the empirical study, research output is measured by the number of citations of the paper weighted by recursive discounted impact factors of the citing outlet.[17] To capture an author's productivity, we use an author's log lifetime citations (at the point of sample collection), decades after receiving their Ph.D., dummy variables for being a male, having an NBER affiliation, graduating from the Ivy League, and being a journal editor. Descriptive statistics of the variables of interest can be found in Appendix C.

[15]In Appendix F, we conduct a robustness check by estimating the empirical model with a sample including the 2,463 papers without any citations. The main result is qualitatively unchanged.

[16]In Appendix F, we conduct a robustness check by estimating the empirical model with a sample including the 621 authors who wrote only single-authored papers in the sample period. The main result is qualitatively unchanged.

[17]The recursive impact factor R_i of journal i is computed as the fixed point of the following system of equations

$$R_i = \frac{\sum_{j\in\mathcal{J}} R_j C_{ij}}{P_i} \frac{\sum_{j\in\mathcal{J}} P_j}{\sum_{j\in\mathcal{J}} R_j P_j}, \ \forall i \in \mathcal{J}, \tag{10}$$

where $\mathcal{J}$ denotes the set of journals, C_{ij} counts the number of citations in journal j to journal i, P_i is the number of all papers/articles in journal i. It is an impact factor where every citation has the weight of the recursive impact factor of the citing journal. All R_i are normalized such that the average paper has an R_i of one. For the recursive discounted impact factor, each citation is further weighted by $1/T$, where T is the age of the citation in years.

4.2 Main Results

In the benchmark empirical model, we assume that the complementarity between researchers is homogeneous, i.e., $g_{ij} = 1$ for $i \neq j$ in Equation (1), and the payoff from a coauthored paper is not discounted, i.e., $\delta_s = 1$ in Equation (2). Table 1 collects the estimation results of Equations (8) and (9), where column (A) reports the estimates of the production function ignoring endogenous project participation, column (B) reports the joint estimates of the production and participation functions with an author-specific random component, and column (C) reports the joint estimates of the production and participation functions with both author- and project-specific random components.

Across all columns, we find that the estimated spillover effect (λ) is significant and positive. When endogenous project participation is ignored, the estimated congestion effect (ϕ) reported in column (A) is statistically insignificant. When endogenous project participation is controlled for, the estimated congestion effect (ϕ) becomes significant and positive. More specifically, in column (B), we incorporate an author-specific random component to control for endogenous project participation. The estimated coefficients (ζ and ξ) of the author-specific random component suggest that a researcher with high ability is more likely to participate in a project. Comparing the estimates of ϕ between columns (A) and (B) indicates ignoring the participation equation with an author-specific random component tends to *underestimate* the congestion effect because it fails to take into account that the researchers simultaneously working on multiple projects are more likely to be high-ability ones. In column (C), we also include a project-specific random component to further control for endogenous project participation. The estimated coefficients (ς and ψ) of the project-specific random component suggest that a high-potential project holds a higher threshold for a researcher to participate in. Compared with column (C), column (B) slightly *overestimates* the congestion effect because it does not account for unobserved heterogeneity in project potential or quality. In appendix D, we conduct some Monte Carlo simulation experiments and observe the same pattern of bias as reported here.

Table 1: Main Results

		(A) Exogenous Participation	(B) Endogenous Participation w/ Author RE	(C) Endogenous Participation w/ Author & Project RE
Production				
Spillover	(λ)	0.0923***	0.0979***	0.0958***
		(0.0165)	(0.0182)	(0.0185)
Congestion	(ϕ)	0.0248	0.2238***	0.1447***
		(0.0174)	(0.0360)	(0.0236)
Constant	(β_0)	-2.2987***	-3.0576***	-3.2404***
		(0.1642)	(0.1873)	(0.1728)
Log life-time citat.	(β_1)	0.4411***	0.5507***	0.5706***
		(0.0254)	(0.0256)	(0.0242)
Decades after grad.	(β_2)	-0.3838***	-0.4256***	-0.4742***
		(0.0310)	(0.0303)	(0.0289)
Male	(β_3)	-0.1626***	-0.0137	-0.0178
		(0.0548)	(0.0506)	(0.0497)
NBER connection	(β_4)	0.2238***	0.3535***	0.4088***
		(0.0397)	(0.0400)	(0.0364)
Ivy League connect.	(β_5)	0.3870***	0.2385***	0.2688***
		(0.0369)	(0.0416)	(0.0385)
Editor	(β_6)	0.1277**	0.0250	-0.0337
		(0.0546)	(0.0590)	(0.0577)
Author effect	(ζ)	2.2966***	2.6478***	2.8023***
		(0.0765)	(0.0931)	(0.0996)
Project effect	(ς)	–	–	1.2752**
		–	–	(0.5074)
Error term variance	(σ^2)	89.8957***	98.2534***	97.8927***
		(2.2057)	(2.3896)	(2.4867)
Participation				
Constant	(γ_0)	–	-9.9857***	-10.6201***
			(0.1009)	(0.1207)
Same NEP	(γ_1)	–	1.3360***	1.5678***
			(0.1020)	(0.1033)
Affiliation	(γ_2)	–	6.8666***	6.8409***
			(0.2955)	(0.2883)
Gender	(γ_3)	–	1.5567***	1.8669***
			(0.0940)	(0.1033)
Past coauthors	(γ_4)	–	6.3132***	6.7095***
			(0.0970)	(0.1073)
Common co-authors	(γ_5)	–	6.9941***	7.8773***
			(0.0602)	(0.0897)
Author effect	(ξ)	–	1.0490***	1.0002***
			(0.0936)	(0.0983)
Project effect	(ψ)	–	–	-2.9287**
			–	(0.1120)
Sample size		3,589 papers and 2,421 authors		

Notes: Column (A) estimates the production function ignoring endogenous project participation. Column (B) jointly estimates the production and participation functions with author random effects. Column (C) jointly estimates the production and participation functions with both author and project random effects. We implement MCMC sampling for 30,000 iterations and leave the first 1000 draws for burn-in and use the rest of draws for computing the posterior mean (as the point estimate) and the posterior standard deviation (in the parenthesis). The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density range does not cover zero.

Regarding the effect of author characteristics on project output, we find that the number of lifetime citations is a positive and significant predictor of research output (cf. e.g., Ductor, 2014), while experience (measured by decades after receiving Ph.D.) is significantly negative.[18] This finding mirrors Ductor (2014), who shows that career time has a negative impact on productivity and it is consistent with the academics' life-cycle effects documented in Levin and Stephan (1991). Being affiliated with the NBER positively and significantly impacts research output. Similarly, having attended an Ivy League university also positively affects output.

From the estimation of project participation, we find that similarities in the research (NEP) fields positively and significantly affect the matching between authors and projects (Ductor, 2014). In terms of assortative matching between coauthors, belonging to the same affiliation, having the same gender, being coauthors in the past, and sharing common coauthors all make matching more likely (cf. Freeman and Huang, 2015).[19]

Finally, we check the condition given by Equation (4) holds for the estimated λ and ϕ. In Appendix E, Figure E.1 plots the empirical distribution of equilibrium efforts given by Equation (5) in Proposition 1. We find that the predicted equilibrium efforts are all positive, alleviating the concern of corner solutions.

4.3 Robustness Analysis

We also consider two alternative specifications of the empirical model. First, we allow complementarity between researchers to be heterogeneous. Researchers differ in their knowledge bases and these differences can affect their complementarity when collaborating on a joint project. In order to capture these heterogeneous complementarities, we define g_{ij} in Equation (1) based on the Jaffe proximity measure of research fields (NEP) between each pair

[18] Following Rauber and Ursprung (2008) we have also estimated a polynomial of order five in decades after Ph.D. graduation. The result shows that the coefficient of the first order is significantly negative, while the remaining higher orders are insignificant.

[19] In Appendix F, we also experiment with alternative specifications of the participation equation. The main result is qualitatively unchanged.

of authors.[20,21] The estimation results with heterogeneous complementarity are reported in Table 2. We find the results are comparable with those reported in Table 1. In particular, the spillover effect is positive and significant, and the bias of the congestion effect follows the same pattern as the homogeneous complementarity case. It is worth pointing out that the estimates of λ are a little larger than those reported in Table 1. This is because g_{ij} based on the Jaffe proximity measure is smaller than one and thus a larger spillover coefficient is obtained in compensation.

In the second specification, we assume that the payoff is discounted by the number of coauthors in a project, i.e., $\delta_s = 1/|\mathcal{N}_s|$ in Equation (2).[22] The estimation results are reported in Table 3. Although the estimated spillover effects are larger than those reported in Table 1 due to the smaller value of δ_s, the main results are qualitatively unchanged.

In Appendix F, we perform additional robustness checks to gauge the sensitivity of the estimation results. In Table F.1, we experiment with an alternative specification of the participation equation. In Tables F.2 and F.3, we estimate the benchmark empirical model with samples which also include authors who wrote only single-authored papers in the sample period and papers without any citations. We find that the estimates are similar to those reported in Table 1, indicating the robustness of our findings.

[20] Jaffe (1986) introduces this measure for the analysis of technological proximity between patents. More recently, Bloom et al. (2013) illustrates how "Jaffe similarity" affects firms' profits with different patent portfolios.

[21] From the authors' NEP fields, we computed their research field proximity following Jaffe (1986) as

$$g_{ij} = \frac{F_i^\top F_j}{\sqrt{F_i^\top F_i}\sqrt{F_j^\top F_j}},$$

where F_i represents the NEP fields of author i and is a vector whose kth component P_{ik} counts the number of papers author i has in NEP field k divided by the total number of papers of that author with an attributed field.

[22] However, Kuld and O'Hagan (2018) argue that the available empirical evidence suggests that the number of co-authors causes very limited discounting of a published article.

Table 2: Robustness Check: Heterogeneous Complementarity

		(A) Exogenous Participation	(B) Endogenous Participation w/ Author RE	(C) Endogenous Participation w/ Author & Project RE
Production				
Spillover	(λ)	0.1679***	0.1888***	0.1840***
		(0.0264)	(0.0248)	(0.0275)
Congestion	(ϕ)	0.0417**	0.2869***	0.1958***
		(0.0157)	(0.0405)	(0.0248)
Constant	(β_0)	-2.2801***	-3.5719***	-3.5168***
		(0.1268)	(0.2108)	(0.1988)
Log life-time citat.	(β_1)	0.4271***	0.6231***	0.5937***
		(0.0186)	(0.0282)	(0.0241)
Decades after grad.	(β_2)	-0.4015***	-0.5162***	-0.4798***
		(0.0212)	(0.0288)	(0.0274)
Male	(β_3)	0.0353	0.0781	0.0742
		(0.0490)	(0.0507)	(0.0559)
NBER connection	(β_4)	0.2890***	0.4584***	0.5000***
		(0.0341)	(0.0342)	(0.0359)
Ivy League connect.	(β_5)	0.3076***	0.2444***	0.2506***
		(0.0412)	(0.0318)	(0.0278)
Editor	(β_6)	-0.0557	0.0627	0.0144
		(0.0453)	(0.0465)	(0.0546)
Author effect	(ζ)	2.1473***	2.6694***	2.9050***
		(0.0709)	(0.0985)	(0.1197)
Project effect	(ς)	–	–	1.5678***
				(0.5625)
Error term variance	(σ^2)	89.4745***	96.3344***	96.9670***
		(2.1668)	(2.3263)	(2.3472)
Participation				
Constant	(γ_0)	–	-10.0378***	-10.8167***
			(0.1028)	(0.1148)
Same NEP	(γ_1)	–	1.3428***	1.6481***
			(0.1083)	(0.1040)
Affiliation	(γ_2)	–	6.9382***	6.8972***
			(0.2997)	(0.3159)
Gender	(γ_3)	–	1.5850***	1.9603***
			(0.0964)	(0.1028)
Past coauthors	(γ_4)	–	6.3553****	6.8359***
			(0.0972)	(0.1111)
Common co-authors	(γ_5)	–	7.0509***	8.1579***
			(0.0628)	(0.0812)
Author effect	(ξ)	–	1.2847***	1.1889***
			(0.0895)	(0.0958)
Project effect	(ψ)	–	–	-3.1438***
			–	(0.1127)
Sample size		3,589 papers and 2,421 authors		

Notes: Column (A) estimates the production function ignoring endogenous project participation. Column (B) jointly estimates the production and participation functions with author random effects. Column (C) jointly estimates the production and participation functions with both author and project random effects. We implement MCMC sampling for 30,000 iterations and leave the first 1000 draws for burn-in and use the rest of draws for computing the posterior mean (as the point estimate) and the posterior standard deviation (in the parenthesis). The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density range does not cover zero.

Table 3: Robustness Check: Discounted Payoffs

		(A) Exogenous Participation	(B) Endogenous Participation w/ Author RE	(C) Endogenous Participation w/ Author & Project RE
Production				
Spillover	(λ)	0.2999***	0.3515***	0.3354***
		(0.0371)	(0.0544)	(0.0563)
Congestion	(ϕ)	0.0225	0.2144***	0.1512***
		(0.0116)	(0.0391)	(0.0246)
Constant	(β_0)	-2.5092***	-3.9907***	-3.1635***
		(0.1627)	(0.2174)	(0.2102)
Log life-time citat.	(β_1)	0.4433***	0.6511***	0.5567***
		(0.0198)	(0.0270)	(0.0283)
Decades after grad.	(β_2)	-0.3900***	-0.5265***	-0.4026***
		(0.0268)	(0.0278)	(0.0332)
Male	(β_3)	0.0318	0.1424***	-0.0788
		(0.0470)	(0.0519)	(0.0450)
NBER connection	(β_4)	0.2650***	0.5040***	0.4829***
		(0.0365)	(0.0387)	(0.0356)
Ivy League connect.	(β_5)	0.2687***	0.2104***	0.1939***
		(0.0349)	(0.0422)	(0.0360)
Editor	(β_6)	-0.0542	0.0194	-0.0605
		(0.0572)	(0.0512)	(0.0487)
Author effect	(ζ)	2.2626***	2.8629***	2.8346***
		(0.0741)	(0.1197)	(0.1049)
Project effect	(ς)	–	–	1.0654**
				(0.5662)
Error term variance	(σ^2)	89.4745***	95.2468***	102.5049***
		(2.1668)	(2.3049)	(2.5033)
Participation				
Constant	(γ_0)	–	-10.0555***	-10.4989***
			(0.1024)	(0.1141)
Same NEP	(γ_1)	–	1.3444***	1.5049***
			(0.1027)	(0.1069)
Affiliation	(γ_2)	–	6.9548***	6.7633***
			(0.3133)	(0.2807)
Gender	(γ_3)	–	1.5938***	1.7993***
			(0.0960)	(0.1016)
Past coauthors	(γ_4)	–	6.3785***	6.6654***
			(0.0994)	(0.1045)
Common co-authors	(γ_5)	–	7.0664***	7.8171***
			(0.0628)	(0.0809)
Author effect	(ξ)	–	1.3102***	1.0334***
			(0.0887)	(0.1031)
Project effect	(ψ)	–	–	-2.8904***
			–	(0.1107)
Sample size		3,589 papers and 2,421 authors		

Notes: Column (A) estimates the production function ignoring endogenous project participation. Column (B) jointly estimates the production and participation functions with author random effects. Column (C) jointly estimates the production and participation functions with both author and project random effects. We implement MCMC sampling for 30,000 iterations and leave the first 1000 draws for burn-in and use the rest of draws for computing the posterior mean (as the point estimate) and the posterior standard deviation (in the parenthesis). The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density range does not cover zero.

5 Conclusion

In this paper, we analyze the equilibrium efforts of researchers who seek to maximize their utility when involved in multiple, possibly overlapping projects in a bipartite network. We show that both the spillover effect between collaborating researchers and the congestion effect between concurrent projects play an important role in determining the equilibrium effort level. To estimate the structural parameters of the model, we develop a Bayesian MCMC procedure that accounts for endogenous selection of researchers into research projects. We then bring our model to the data by analyzing the coauthorship network of economists registered in the RePEc Author Service and find empirical evidence for both spillover and congestion effects.

As our model has an explicit micro-foundation, it provides a formal framework for counterfactual analysis. One could evaluate the importance of a researcher by hypothetically removing him/her from the coauthorship network to see the resulting loss in aggregate research output as in Ballester et al. (2006), or design an optimal funding scheme to maximize aggregate research output as in König et al. (2019). We leave these counterfactual exercises to future research.

References

Adams, C. P. (2006). Optimal team incentives with CES production. *Economics Letters*, 92(1):143-148.

Anderson, K. A. and Richards-Shubik, S. (2019). Collaborative production in science: An empirical analysis of coauthorships in economics. *Working Paper, Lehigh University.*

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. *Econometrica*, 74(5):1403-1417.

Baumann, L. (2014). Time allocation in friendship networks. *Available at SSRN 2533533.*

Belhaj, M. and Deroïan, F. (2014). Competing activities in social networks. *The BE Journal of Economic Analysis & Policy*, 14(4):1431-1466.

Bimpikis, K., Ehsani, S., and Ilkilic, R. (2019). Cournot competition in networked markets. *Management Science*, 65(6):2467-2481.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347-1393.

Bonhomme, S. (2020). Econometric analysis of bipartite networks. To appear in B. Graham and A. de Paula (ed.) *The Econometric Analysis of Network Data.*

Cohen-Cole, E., Liu, X., and Zenou, Y. (2018). Multivariate choices and identification of social interactions. *Journal of Applied Econometrics* , 33(2):165-178.

Colussi, T. (2017). Social ties in academia: A friend is a treasure. *Review of Economics and Statistics*, 100(1):45-50.

Ductor, L. (2014). Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics*, 77(3):385-407.

Ductor, L., Fafchamps, M., Goyal, S., and Van der Leij, M. J. (2014). Social networks and research output. *Review of Economics and Statistics*, 96(5):936-948.

Fafchamps, M., Van der Leij, M. J., and Goyal, S. (2010). Matching and network effects. *Journal of the European Economic Association*, 8(1):203-231.

Freeman, R. B. and Huang, W. (2015). Collaborating with people like me: Ethnic coauthor-

ship within the united states. *Journal of Labor Economics*, 33(S1):289-318.

Goyal, S., Van der Leij, M. J., and Moraga-Gonzalez, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, 114(2):403-412.

Hess, A. M. and Rothaermel, F. T. (2011). When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry. *Strategic Management Journal*, 32(8):895-909.

Hollis, A. (2001). Co-authorship and the output of academic economists. *Labour Economics*, 8(4):503-530.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44-74.

Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value. *The American Economic Review*, 76(5):pp. 984-1001.

Kandel, E. and Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of political Economy*, 100(4):801-817.

König, M. D., Liu, X., and Zenou, Y. (2019). R&D networks: Theory, empirics and policy implications. *Review of Economics and Statistics*, 101(3):476-491.

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian Econometric Methods*. Cambridge University Press.

Kuld, L. and O'Hagan, J. (2018). Rise of multi-authored papers in economics: Demise of the "lone star" and why? *Scientometrics*, 114(3):1207-1225.

Levin, S. G. and Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, 81(1):114-132.

Liu, X. (2014). Identification and efficient estimation of simultaneous equations network models. *Journal of Business & Economic Statistics*, 32(4):516-536.

Newman, M. E. J. (2001a). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404-409.

Newman, M. E. J. (2001b). Scientific collaboration networks i. Network construction and fundamental results. *Physical Review E*, 64(1):016131.

Newman, M. E. J. (2001c). Scientific collaboration networks. ii. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):016132.

Newman, M. E. J. (2004a). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(90001):5200-5205.

Newman, M. E. J. (2004b). Who is the best connected scientist? A study of scientific coauthorship networks. *Complex networks*, Springer, Berlin, Heidelberg, 337-370.

Rauber, M. and Ursprung, H. W. (2008). Life cycle and cohort productivity in economic research: The case of Germany. *German Economic Review*, 9(4):431-456.

Salonen, H. (2016). Equilibria and centrality in link formation games. *International Journal of Game Theory*, 45(4):1133-1151.

West, D. B. (2001). *Introduction to Graph Theory*. Prentice-Hall.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American statistical association*, 86(413):79-86.

Zimmermann, C. (2013). Academic rankings with RePEc. *Econometrics*, 1(3):249-280.

Online Appendix for "Collaboration in Bipartite Networks, with an Application to Coauthorship Networks"

by Chih-Sheng Hsieh, Michael D. König, Xiaodong Liu, and Christian Zimmermann

A Proof of Proposition 1

Proof of Proposition 1. Let $e_{is} = d_{is}\varepsilon_{is}$, where ε_{is} is the latent effort level. Substitution of Equation (1) into Equation (2) gives

$$\begin{aligned} U_i(\mathcal{G}) &= \sum_{s\in\mathcal{P}} d_{is}\delta_s \left(\sum_{j\in\mathcal{N}} \alpha_j d_{js}\varepsilon_{js} + \frac{\lambda}{2} \sum_{j\in\mathcal{N}} \sum_{k\in\mathcal{N}\setminus\{j\}} g_{jk} d_{js} d_{ks} \varepsilon_{js}\varepsilon_{ks} + \epsilon_s \right) \\ &\quad -\frac{1}{2} \left(\sum_{s\in\mathcal{P}} d_{is}\varepsilon_{is}^2 + \phi \sum_{s\in\mathcal{P}} \sum_{t\in\mathcal{P}\setminus\{s\}} d_{is} d_{it} \varepsilon_{is}\varepsilon_{it} \right). \end{aligned} \tag{11}$$

The first-order condition of maximizing utility in Equation (11) with respect to ε_{is} gives

$$d_{is} \left(\delta_s \alpha_i + \lambda \delta_s \sum_{j\in\mathcal{N}\setminus\{i\}} g_{ij} d_{js} \varepsilon_{js} - \varepsilon_{is} - \phi \sum_{t\in\mathcal{P}\setminus\{s\}} d_{it}\varepsilon_{it} \right) = 0.$$

In matrix form, the first-order condition can be written as

$$D(\delta \otimes \alpha) - (I_{np} - L)e = 0.$$

If $\rho_{\max}(L) < 1$, the matrix $I_{np} - L$ is positive definite. It follows by Lemmas 2 and 3 in Bimpikis et al. (2019) that the unique equilibrium is given by the solution to the linear complementarity problem and the inactive links ($d_{is} = 0$) are strategically redundant and play no role in determining the equilibrium. Hence, it follows by a similar arguments as in the proof of Theorem 1 in Bimpikis et al. (2019) that the game has a unique equilibrium with the equilibrium effort levels are given by Equation (5). ∎

B Bayesian Estimation

Since the likelihood function based on Equations (8) and (9) involves high-dimensional integrals, it is computationally cumbersome to apply a frequentist maximum likelihood method even when resorting to a simulation approach. As an alternative estimation method, the Bayesian Markov Chain Monte Carlo (MCMC) approach can be more efficient for estimating latent variable models (cf. Zeger and Karim, 1991). We divide the parameter vector θ and other unknown latent variables into blocks and assign the prior distributions as follows:

$$\begin{aligned}
\lambda &\sim \mathcal{N}(0, \sigma_\lambda^2), \\
\phi &\sim \mathcal{N}(0, \sigma_\phi^2), \\
\beta &\sim \mathcal{N}(0, \Sigma_\beta), \\
\zeta &\sim \mathcal{N}(0, \sigma_\zeta^2), \\
\varsigma &\sim \mathcal{N}(0, \sigma_\varsigma^2), \\
\gamma &\sim \mathcal{N}(0, \Sigma_\gamma), \\
\xi &\sim \mathcal{N}(0, \sigma_\xi^2), \\
\psi &\sim \mathcal{N}(0, \sigma_\psi^2), \\
\sigma_u^2 &\sim \mathcal{IG}\left(\frac{\tau_0}{2}, \frac{\nu_0}{2}\right),
\end{aligned}$$

and $\mu_i \sim \mathcal{N}(0, 1)$ for $i \in \mathcal{N}$. We consider the normal and inverse gamma ($\mathcal{IG}$) conjugate priors, which are widely used in the Bayesian literature (Koop et al., 2007). The hyper parameters are chosen to make the prior distribution relatively flat and cover a wide range of the parameter space, i.e., we set $\sigma_\lambda^2 = \sigma_\phi^2 = 10$, $\Sigma_\beta = 10I$, $\sigma_\zeta^2 = \sigma_\varsigma^2 = 10$, $\Sigma_\gamma = 1000I$, $\sigma_\xi^2 = \sigma_\psi^2 = 1000$, $\tau_0 = 2.2$, and $\nu_0 = 0.1$.

The MCMC sampling procedure combines the Gibbs sampling and the Metropolis-Hastings (M-H) algorithm. It consists of the following steps:

1. Draw the latent variable μ_i using the M-H algorithm based on $f(\mu_i|y, d, \theta, \mu_{-i}, \eta)$, for

$i = 1, \ldots, n$.

2. Draw the latent variable η_s using the M-H algorithm based on $f(\eta_s|y, d, \theta, \mu, \eta_{-s})$, for $s = 1, \ldots, p$.

3. Draw γ using the M-H algorithm based on $f(\gamma|y, d, \theta\backslash\{\gamma\}, \mu, \eta)$.

4. Draw ξ using the M-H algorithm based on $f(\xi|y, d, \theta\backslash\{\xi\}, \mu, \eta)$.

5. Draw ψ using the M-H algorithm based on $f(\psi|y, d, \theta\backslash\{\psi\}, \mu, \eta)$.

6. Draw λ using the M-H algorithm based on $f(\lambda|y, d, \theta\backslash\{\lambda\}, \mu, \eta)$.

7. Draw ϕ using the M-H algorithm based on $f(\phi|y, d, \theta\backslash\{\phi\}, \mu, \eta)$.

8. Draw β using the M-H algorithm based on $f(\beta|y, d, \theta\backslash\{\beta\}, \mu, \eta)$.

9. Draw ζ using the M-H algorithm based on $f(\zeta|y, d, \theta\backslash\{\zeta\}, \mu, \eta)$.

10. Draw ς using the M-H algorithm based on $f(\varsigma|y, d, \theta\backslash\{\varsigma\}, \mu, \eta)$.

11. Draw σ_u^2 using the conjugate inverse gamma conditional posterior distribution.

We collect the draws from iterating the above steps and compute the posterior mean and the posterior standard deviation as our estimation results.

C Data Description

To obtain a sample from RePEc that is appropriate for our analysis, we apply a series of filters as follows.

First, we select papers that had a first pre-print version in 2010-2012. Furthermore, we require all authors of the papers to be registered with RePEc and all authors to have the RePEc Genealogy information on where they studied. We drop all duplicate or older versions of each paper from our sample. This gives us a sample of 6,673 papers written by

Table C.1: Summary Statistics of Sample (I)

	Min	Max	Mean	S.D.	Sample size
Papers					
Weighted citation	0.0000	317.9515	3.7587	12.3858	6673
Number of authors (in each paper)	1	5	1.4160	0.6421	6673
Authors					
Log lifetime citations	0	10.7634	5.3176	1.8428	3700
Decades after Ph.D. graduation	-0.7	6.2000	1.0642	1.0676	3700
Male	0	1	0.8154	0.3880	3700
NBER connection	0	1	0.0889	0.2847	3700
Ivy League connection	0	1	0.1268	0.3327	3700
Editor	0	1	0.0476	0.2129	3700
Number of papers (for each author)	1	63	2.5538	2.7762	3700

Notes: This sample is constructed based on works that were released as working papers in 2010-2012. We drop papers in which not all of their authors were registered with RePEc. We also drop authors who do not have the RePEc Genealogy information on where they studied.

Table C.2: Summary Statistics of Sample (II)

	Min	Max	Mean	S.D.	Sample size
Papers					
Weighted citation	0.0000	317.9515	5.9577	15.1682	4210
Number of authors (in each paper)	1	5	1.5124	0.6820	4210
Authors					
Log lifetime citations	0	10.7634	5.5445	1.7358	3042
Decades after Ph.D. graduation	-0.7	6.2000	1.0701	1.0447	3042
Male	0	1	0.8222	0.3824	3042
NBER connection	0	1	0.1019	0.3026	3042
Ivy League connection	0	1	0.1341	0.3408	3042
Editor	0	1	0.0516	0.2213	3042
Number of papers (for each author)	1	19	2.0930	1.7079	3042

Notes: This sample is constructed based on works that were released as working papers in 2010-2012. We drop papers in which not all of their authors were registered with RePEc. We also drop authors who do not have the RePEc Genealogy information on where they studied. In this sample, we further drop papers which do not have any citations up to November 2018.

Table C.3: Summary Statistics of Sample (III)

	Min	Max	Mean	S.D.	Sample size
Papers					
Weighted citation	1e-04	317.9515	6.4578	16.0868	3589
Number of authors (in each paper)	1	5	1.6010	0.7017	3589
Authors					
Log lifetime citations	0	10.7634	5.7441	1.6782	2421
Decades after Ph.D. graduation	-0.7	6.2000	1.1056	1.0372	2421
Male	0	1	0.8228	0.3819	2421
NBER connection	0	1	0.1136	0.3174	2421
Ivy League connection	0	1	0.1450	0.3522	2421
Editor	0	1	0.0566	0.2311	2421
Number of papers (for each author)	1	19	2.3734	1.8112	2421

Notes: This sample is constructed based on workss that were released as working papers in 2010-2012. We drop papers in which not all of their authors were registered with RePEc. We also drop authors who do not have the RePEc Genealogy information on where they studied. In this sample, we further drop papers which do not have any citations up to November 2018 and the authors who only wrote a single-authored paper in the sampling period.

3,700 distinct authors for which we have complete data. We call this sample: Sample (I). This is the sample we used to obtain the estimates reported in Table F.2 in Appendix F. Descriptive statistics of the variables of interest in Sample (I) are reported in Table C.1.

Next, we drop 2,463 papers that do not have any citations up to July 2018 when the data is extracted from the RePEc database, as well as 658 authors who only work on these dropped papers without any citations. This reduces to the sample size to 4,210 papers and 3,042 authors. We call this sample: Sample (II). This is the sample we used to obtain the estimates reported in Table F.3 in Appendix F. Descriptive statistics of the variables of interest in Sample (II) are reported in Table C.2.

Finally, we drop 621 authors who only wrote a single-authored paper in the sample period. This results in a sample of 3,589 papers written by 2,421 distinct authors. We call this sample: Sample (III). This is the sample we used to obtain the main results reported in Section 4.2. Descriptive statistics of the variables of interest in Sample (III) are reported in Table C.3.

D Monte Carlo Simulation

To show that the proposed Bayesian MCMC estimation approach in Appendix B can effectively recover the true parameters in Equations (8) and (9), we conduct a Monte Carlo simulation with 100 repetitions. In each repetition, we generate an artificial bipartite collaboration network of 200 authors ($n = 200$) and 400 projects ($p = 400$). The data generating process (DGP) runs as follows: we first simulate dyadic binary exogenous variables $z_{is} \in \{0, 1\}$ randomly with the probability $P(z_{is} = 1) = 0.64$; individual exogenous variable x_i from normal distribution $N(0, 4)$; and both author and project latent variables μ_i and η_s from $N(0, 1)$. Then, we generate the artificial collaboration network and project output based on the participation function of Equation (9) and the production function of Equation (8) with DGP parameters listed in Table D.1. In these artificial collaboration networks, each author on average participates in 3.195 projects and the standard deviation equals 2.604. Each project on average has 1.598 authors and the standard deviation is 1.621. The average of artificial project outputs is 6.238 and the standard deviation is 26.502.

Following the empirical study in Section 4, we also explore three different model specifications to examine the misspecification biases on the parameter estimates. The simulation results are summarized in Table D.1. In Column (A) of Table D.1, we intentionally ignore endogenous project participation by estimating the production function of Equation (8) alone. The simulation result shows that, while the estimated spillover effect (λ) is essentially unbiased, the estimated congestion effect (ϕ) is significantly downward biased by more than 100%, which is consistent with the our empirical finding. Moreover, omitting the project latent variable η_s leads to a huge upward bias on the estimated error variance σ^2. In Column (B), we take into account endogenous project participation by estimating Equations (8) and (9) jointly, but still ignoring project latent variable η_s. The simulation result shows that the estimation bias on congestion effect (ϕ) turns to positive from Column (A) to Column (B), which is also in line with the observation in our empirical study. Finally, in Column (C) we estimate the true DGP model and the simulation result confirms that the employed

Table D.1: Simulation results.

		(A) Exogenous Participation		(B) Endogenous Participation w/ Author RE		(C) Endogenous Participation w/ Author & Project RE	
	DGP	Est.	S.D.	Est.	S.D.	Est.	S.D.
Production							
λ	0.1000	0.0973	0.0098	0.0977	0.0087	0.0996	0.0011
ϕ	0.1000	-0.0161	0.0442	0.1550	0.0823	0.1031	0.0139
β_0	-0.5000	-1.1855	0.2889	-0.9935	0.2693	-0.4909	0.0464
β_1	0.5000	0.6728	0.0876	0.6524	0.0860	0.4979	0.0114
ζ	0.5000	0.9596	0.3140	1.1193	0.3139	0.5019	0.0196
ς	4.0000					4.0300	0.0655
σ^2	0.5000	15.1201	1.1774	15.3512	1.3009	0.3316	0.0431
Participation							
γ_0	-5.5000			-5.3512	0.0876	-5.5317	0.0655
γ_1	1.0000			0.9988	0.0794	1.0090	0.0746
ξ	0.5000			0.6910	0.1808	0.5266	0.0437
ψ	-0.5000					-0.4960	0.0372

Bayesian MCMC approach can effectively recover all true model parameters.

E Estimated Equilibrium Efforts

Figure E.1 plots the empirical distribution of equilibrium efforts given by Equation (5) in Proposition 1.

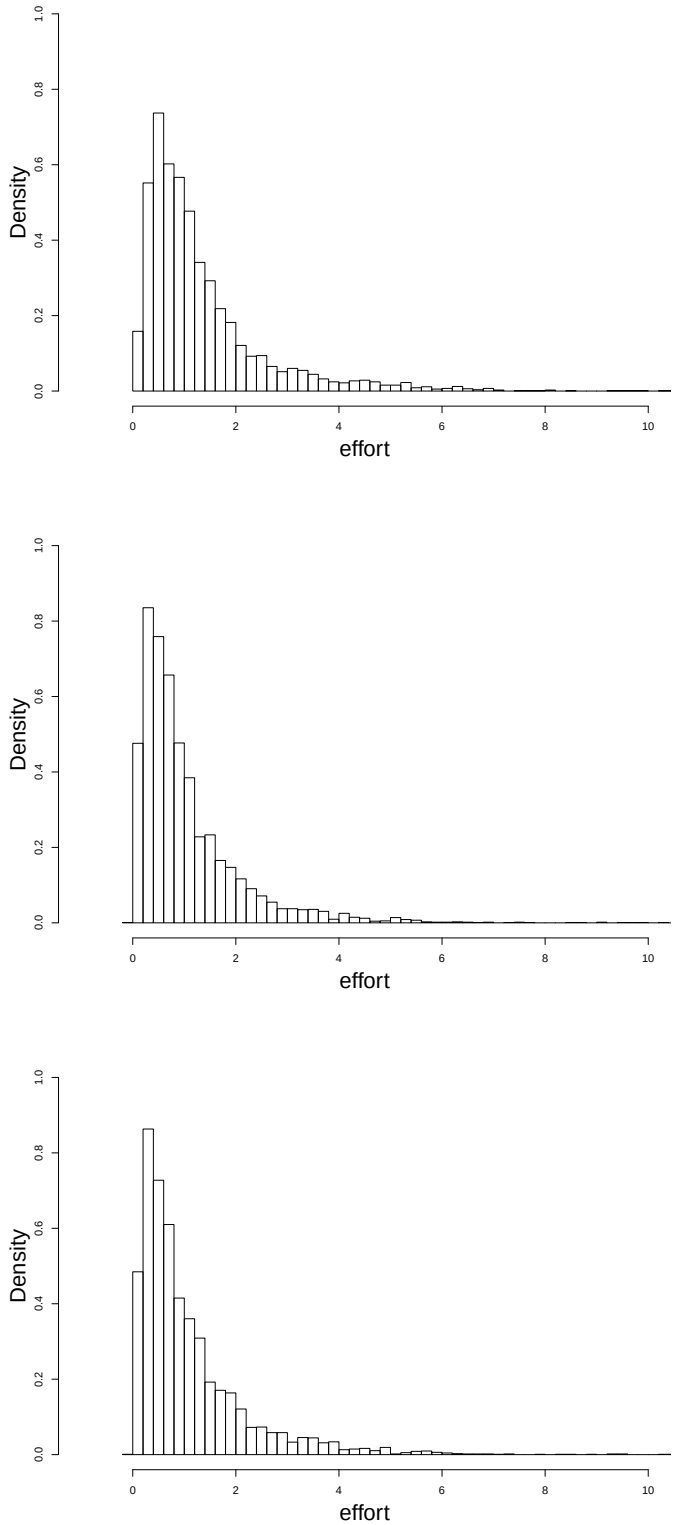

Figure E.1: Distributions of equilibrium efforts. The top graph is based on Column A of Table 1; the middle graph is based on Column B of Table 1; and the bottom graph is based on Column C of Table 1.

F Additional Robustness Checks

In this section, we perform additional robustness checks to gauge the sensitivity of the estimation results. In Table F.1, we experiment with an alternative specification of the participation equation. In Tables F.2 and F.3, we estimate the benchmark empirical model with Sample (I) and Sample (II) respectively (see Appendix C). We find that the estimates are similar to those reported in Table 1, indicating the robustness of our findings.

Table F.1: Robustness Check: Alternative Participation Equations

		(A) Homogeneous Complementarity	(B) Heterogeneous Complementarity	(C) Discounted Payoffs
Production				
Spillover	(λ)	0.0944***	0.2020***	0.3616***
		(0.0189)	(0.0216)	(0.0625)
Congestion	(ϕ)	0.1550***	0.2289***	0.1911***
		(0.0273)	(0.0153)	(0.0389)
Constant	(β_0)	-2.9330***	-3.4404***	-3.2743***
		(0.2062)	(0.1849)	(0.2009)
Log life-time citat.	(β_1)	0.5257***	0.5928***	0.5746***
		(0.0284)	(0.0264)	(0.0242)
Decades after grad.	(β_2)	-0.4370***	-0.5163***	-0.4020***
		(0.0298)	(0.0286)	(0.0284)
Male	(β_3)	-0.0320	0.0015	-0.1216**
		(0.0541)	(0.0485)	(0.0521)
NBER connection	(β_4)	0.4165***	0.5636***	0.4804***
		(0.0327)	(0.0386)	(0.0381)
Ivy League connect.	(β_5)	0.2692***	0.2619***	0.1751***
		(0.0385)	(0.0315)	(0.0404)
Editor	(β_6)	0.0023	0.0495	-0.0351
		(0.0539)	(0.0455)	(0.0610)
Author effect	(ζ)	2.6626***	2.7133***	2.7678***
		(0.1177)	(0.1023)	(0.1066)
Project effect	(ς)	1.4802***	1.9140***	1.1847**
		(0.5150)	(0.4689)	(0.5435)
Error term variance	(σ^2)	97.1196***	91.0082***	98.0885***
		(2.4054)	(2.2493)	(2.4668)
Participation				
Constant	(γ_0)	-7.8888***	-7.9011***	-7.8910***
		(0.0452)	(0.0440)	(0.0783)
Same NEP	(γ_1)	1.5496***	1.5685***	1.5562***
		(0.0976)	(0.0942)	(0.0966)
Author effect	(ξ)	0.2921***	0.4066***	0.3438***
		(0.0804)	(0.0720)	(0.0805)
Project effect	(ψ)	-0.4620***	-0.5256***	-0.5138***
		(0.0814)	(0.0733)	(0.0791)
Sample size		3,589 papers and 2,421 authors		

Notes: Column (A) assumes homogeneous complementarity. Column (B) allows for heterogeneous complementarity using Jaffe's similarity measure for the research fields of collaborating authors. Column (C) considers the case where the payoff is discounted by the number of coauthors in a project. We implement MCMC sampling for 30,000 iterations and leave the first 1000 draws for burn-in and use the rest of draws for computing the posterior mean (as the point estimate) and the posterior standard deviation (in the parenthesis). The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density range does not cover zero.

Table F.2: Robustness Check: Sample (I)

		(A) Exogenous Participation	(B) Endogenous Participation w/ Author RE	(C) Endogenous Participation w/ Author & Project RE
Production				
Spillover	(λ)	0.1743***	0.1601***	0.1612***
		(0.0082)	(0.0085)	(0.0097)
Congestion	(ϕ)	0.2274***	0.4274***	0.4145***
		(0.0370)	(0.0345)	(0.0316)
Constant	(β_0)	-3.0893***	-4.1731***	-4.2633***
		(0.1595)	(0.1835)	(0.1972)
Log life-time citat.	(β_1)	0.5289***	0.6856***	0.7142***
		(0.0219)	(0.0242)	(0.0258)
Decades after grad.	(β_2)	-0.4367***	-0.5526***	-0.5989***
		(0.0287)	(0.0273)	(0.0252)
Male	(β_3)	-0.1894***	0.1363**	0.0619*
		(0.0403)	(0.0508)	(0.0355)
NBER connection	(β_4)	0.2352***	0.4422***	0.3981***
		(0.0343)	(0.0265)	(0.0282)
Ivy League connect.	(β_5)	0.4867***	0.2906***	0.3393***
		(0.0372)	(0.0278)	(0.0293)
Editor	(β_6)	-0.0080	0.1374***	0.2583***
		(0.0513)	(0.0412)	(0.0354)
Author effect	(ζ)	2.4832***	2.0141***	2.0216***
		(0.0757)	(0.0555)	(0.0629)
Project effect	(ς)	–	–	-0.1661
		–	–	(0.7486)
Error term variance	(σ^2)	60.3747***	65.7751***	65.9307***
		(1.0923)	(1.1697)	(1.1781)
Participation				
Constant	(γ_0)	–	-10.6645***	-10.6869***
			(0.0837)	(0.1570)
Same NEP	(γ_1)	–	1.4354***	1.4589***
			(0.0780)	(0.0769)
Affiliation	(γ_2)	–	6.6128***	6.6178***
			(0.2292)	(0.2441)
Gender	(γ_3)	–	1.7284***	1.7320***
			(0.0794)	(0.0793)
Past coauthors	(γ_4)	–	6.3181***	6.3392***
			(0.0748)	(0.1122)
Common co-authors	(γ_5)	–	7.1381***	7.1922***
			(0.0436)	(0.1003)
Author effect	(ξ)	–	0.7854***	0.7606***
			(0.0458)	(0.0478)
Project effect	(ψ)	–	–	-1.8862***
			–	(0.2191)
Sample size		6,673 papers and 3,700 authors		

Notes: Column (A) estimates the production function ignoring endogenous project participation. Column (B) jointly estimates the production and participation functions with author random effects. Column (C) jointly estimates the production and participation functions with both author and project random effects. We implement MCMC sampling for 30,000 iterations and leave the first 1000 draws for burn-in and use the rest of draws for computing the posterior mean (as the point estimate) and the posterior standard deviation (in the parenthesis). The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density range does not cover zero.

Table F.3: Robustness Check: Sample (II)

		(A) Exogenous Participation	(B) Endogenous Participation w/ Author RE	(C) Endogenous Participation w/ Author & Project RE
Production				
Spillover	(λ)	0.0890***	0.0976***	0.0949***
		(0.0165)	(0.0135)	(0.0179)
Congestion	(ϕ)	0.0450***	0.2917***	0.1211***
		(0.0147)	(0.0332)	(0.0194)
Constant	(β_0)	-1.9979***	-2.6058***	-2.6224***
		(0.1255)	(0.1358)	(0.1454)
Log life-time citat.	(β_1)	0.4035***	0.4765***	0.4860***
		(0.0174)	(0.0173)	(0.0201)
Decades after grad.	(β_2)	-0.3851***	-0.3510***	-0.3684***
		(0.0213)	(0.0225)	(0.0253)
Male	(β_3)	-0.0620	0.1258***	-0.0426
		(0.0452)	(0.0410)	(0.0373)
NBER connection	(β_4)	0.2215***	0.3113***	0.1912***
		(0.0266)	(0.0344)	(0.0333)
Ivy League connect.	(β_5)	0.3619***	0.2691***	0.2299***
		(0.0318)	(0.0329)	(0.0342)
Editor	(β_6)	-0.0656	-0.0022	0.0399
		(0.0510)	(0.0475)	(0.0476)
Author effect	(ζ)	1.8939***	1.5781***	2.0467***
		(0.0535)	(0.0518)	(0.0850)
Project effect	(ς)	–	–	1.3828***
		–	–	(0.4739)
Error term variance	(σ^2)	76.1799***	86.3026***	89.0732***
		(1.6805)	(1.9230)	(3.0510)
Participation				
Constant	(γ_0)	–	-10.3408**	-10.7291***
			(0.1013)	(0.1371)
Same NEP	(γ_1)	–	1.3330***	1.5519***
			(0.0958)	(0.1042)
Affiliation	(γ_2)	–	7.0405***	6.8927***
			(0.2998)	(0.2984)
Gender	(γ_3)	–	1.6649***	1.8465***
			(0.0959)	(0.1101)
Past coauthors	(γ_4)	–	6.3929***	6.5954***
			(0.0915)	(0.1073)
Common co-authors	(γ_5)	–	7.1009***	7.6119***
			(0.0551)	(0.1187)
Author effect	(ξ)	–	0.8591***	0.5138***
			(0.0595)	(0.0749)
Project effect	(ψ)	–	–	-2.1268***
			–	(0.1067)
Sample size		4,210 papers and 3,042 authors		

Notes: Column (A) estimates the production function ignoring endogenous project participation. Column (B) jointly estimates the production and participation functions with author random effects. Column (C) jointly estimates the production and participation functions with both author and project random effects. We implement MCMC sampling for 30,000 iterations and leave the first 1000 draws for burn-in and use the rest of draws for computing the posterior mean (as the point estimate) and the posterior standard deviation (in the parenthesis). The asterisks ***(**,*) indicate that the 99% (95%, 90%) highest posterior density range does not cover zero.