

Teams: Heterogeneity, Sorting, and Complementarity*

Stéphane Bonhomme
University of Chicago

April 1, 2021

Abstract

How much do individuals contribute to team output? I propose an econometric framework to quantify individual contributions when only the output of their teams is observed. The identification strategy relies on following individuals who work in different teams over time. I consider two production technologies. For a production function that is additive in worker inputs, I propose a regression estimator and show how to obtain unbiased estimates of variance components that measure the contributions of heterogeneity and sorting. To estimate nonlinear models with complementarity, I propose a mixture approach under the assumption that individual types are discrete, and rely on a mean-field variational approximation for estimation. To illustrate the methods, I estimate the impact of economists on their research output, and the contributions of inventors to the quality of their patents.

JEL CODE: C13, J31.

KEYWORDS: Team production, networks, unobserved heterogeneity, sorting, complementarity.

*Presented at the World Congress of the Econometric Society, semi-plenary session “Frontiers in Modern Econometrics”, August 17, 2020. I thank an anonymous referee, Victor Chernozhukov, Bryan Graham, Chris Adams, Manuel Arellano, Eric Auerbach, Guido Imbens, Philipp Kircher, Thibaut Lamadon, Jack Light, Elena Manresa, Charles Manski, Angelo Mele, Eleonora Patacchini, Alessandra Voena, Martin Weidner, and audiences at various venues for comments. I acknowledge support from NSF grant number SES-1658920.

1 Introduction

How to infer individual contributions when only team output is observed? This question is central to a number of areas, ranging from labor economics and the economics of innovation to sports and music. My goal in this chapter is twofold: I propose an econometric framework to estimate how team performance depends on its workers, and I apply it to study the production of economic research and the impact of individual inventors on the quality of their patents.

I focus on three aspects of the relationship between workers and teams. The first one is worker *heterogeneity*, which shapes team performance. The approach I propose relies on a simple intuition: when individuals work for different teams over time, the variation in team composition will be informative about which workers contribute the most to their teams and which ones contribute the least. Hence, the network structure of the data will be crucial in order to infer worker heterogeneity.

The second aspect is the *sorting* of workers into teams. Both heterogeneity and sorting contribute to the variation in output. I show in a model where production is additive in worker inputs that the two can be separately identified. The approach I rely on generalizes the “plus-minus” metric in US team sports (see Hvattum, 2019): intuitively, when a worker i' joins a team and a worker i leaves it, the difference in output identifies the difference between the types of i' and i . More generally, I show how, in collaboration networks, worker types can be identified by using their participation in different teams, including when they work on their own.

The third aspect is the *complementarity* between workers. The presence of complementarity is important for policies that influence the reallocation of workers across teams. Complementarities are also central to theories of sorting (e.g., Becker, 1973, Garicano and Rossi-Hansberg, 2006, Eeckhout and Kircher, 2018). To identify complementarities empirically, I build and estimate nonlinear models of team production, which rely on the assumption that worker types are discrete. This framework allows me to jointly document the role of heterogeneity, sorting and complementarity.

In the analysis I will draw connections with methods for matched employer-employee data (e.g., Abowd and Kramarz, 1999). In that literature, both firm types and worker types contribute to the wage outcome, which is observed at the worker level, and workers

do not interact with one another inside the firm.¹ The worker-firm structure is thus a bipartite graph. In contrast, in the collaboration networks I study here, workers interact with one another in the team. In addition, while I assume away team effects, so the team in my framework is simply a collection of workers, the outcome is only observed at the team level. Collaboration networks follow a hypergraph structure where teams vary in size (e.g., Turnbull *et al.*, 2019, Lerner *et al.*, 2019). Despite these differences, my framework will leverage insights from the matched employer-employee data literature. As an example, when production is additive in worker inputs, I will estimate worker types using linear regression, in the spirit of Abowd *et al.* (1999).

There is a fourth important aspect in the relationship between workers and teams: the presence of *other factors*, beyond worker types, that affect output. These other factors are lumped into the error term of the model. Their presence prevents one from recovering all worker types precisely, especially in data sets where the number of productions per worker is small and collaborations are sparse. As a result, regression-based estimates of sorting and heterogeneity may be biased. To correct the bias in additive models, I implement the method of Andrews *et al.* (2008) developed for matched employer-employee data.² Bonhomme *et al.* (2020) find using samples from various countries that the bias can be substantial. Similarly, when estimating additive models in networks of collaborations, I find that bias correction is important.

A main goal of this chapter is to propose methods that account for complementarity between workers. Bonhomme *et al.* (2019, BLM) introduce a framework to allow for complementarity between workers and firms in matched employer-employee data. Here I build on this work to allow for complementarity between workers within a team. Like in BLM, I assume that heterogeneity takes a finite number of values. However, while BLM model *firm* heterogeneity as discrete and estimate firm groups using the kmeans algorithm, in my setting estimates of *worker* types are not sufficiently precise to follow a similar grouped fixed-effects approach. For this reason, I model the distribution of the discrete worker types using a random-effects approach.

The presence of unobserved worker heterogeneity in a network of teams makes estimation of nonlinear random-effects models challenging. Indeed, the likelihood function is non-

¹With some exceptions, such as the recent work by Herkenhoff *et al.* (2018).

²Kline *et al.* (2020) generalize this approach to allow for unspecified heteroskedasticity. The nature of the bias is formally studied in Jochmans and Weidner (2020).

separable, since the same worker may participate in multiple teams, and teams may contain multiple workers. To lower the computational cost, I use a mean-field variational method, which relies on a fully factored approximation. Variational estimators are widely used in networks and other complex data settings (Bishop, 2006, Blei *et al.*, 2017). In stochastic blockmodels, which are closely related to the model I study, Bickel *et al.* (2013) provide conditions for consistency. In my model I perform Monte Carlo simulations to probe the accuracy of the variational method in finite samples.³

The framework can be extended in different ways. In particular, I consider adding a model of team formation on top of the model of team production. To do so, I implement a joint random-effects approach using a stochastic blockmodel as a simple statistical model of team formation. An interesting avenue for future work will be to combine the framework with economic models of team formation. It will also be important to understand under which conditions the methods I propose are robust to the team formation model being misspecified.

I apply these methods to two empirical settings: economic research, and patents and innovation. Collaboration in scientific research is an extensively studied topic (e.g., Furman and Gaule, 2013). Aspects of the network of collaborations between economists have been studied by Goyal *et al.* (2006), Fafchamps *et al.* (2010), Hsieh *et al.* (2020), and Anderson and Richards-Shubik (2019), among others. Here I focus on the impact of individual researchers and the shape of the production function. In the literature on patents and innovation, there is increasing interest in the role of individual inventors in innovation as an engine of endogenous growth (e.g., Akcigit *et al.*, 2017, Bell *et al.*, 2019, Pearce, 2019). Relative to this literature, I propose different measures of the type of an inventor – that is, a measure of her “quality” – which I infer from the network of patent collaborations using econometric techniques.

I estimate the models using a sample of economists constructed by Ductor *et al.* (2014), and a sample of inventors and their patents constructed by Akcigit *et al.* (2016). In both samples, I find that an additive model of team production is useful to highlight the presence of heterogeneity and sorting. However, the additive model is only an imperfect tool to analyze these data. Indeed, for both economists and inventors, I find evidence of complementarity between high-type workers who collaborate in a team. These findings suggest that future research will need to account for complementarity and sorting when measuring individual

³Codes to implement the methods are available on my [webpage](#).

contributions of researchers or inventors to their teams.

Related approaches have been proposed in the literature. Relative to the analysis of peer effects and spillovers (e.g., Bramoullé *et al.*, 2009, and in particular Arcidiacono *et al.*, 2012, and Arcidiacono *et al.*, 2017), a key difference in my context is that I only observe team output. In recent work, Devereux (2018) uses fixed-effects methods to estimate the link between individual productivity and team performance in tennis games. Most closely related to this chapter, Ahmadpoor and Jones (2019) use fixed-effects methods in nonlinear models to document heterogeneity and complementarity patterns with a focus on research, innovation and patenting.⁴ While I build on this work, a specificity of my approach is that it explicitly accounts for the presence of other unobserved inputs in addition to the worker types, which motivates the use of new econometric methods. Moreover, estimates of nonlinear models with discrete types uncover novel forms of complementarity between workers. Another recent related article is Weidmann and Deming (2020), who design and analyze experiments to identify individual contributions to teamwork.

The outline of the chapter is as follows. I describe the framework in Section 2, and study additive and nonlinear models in Sections 3 and 4, respectively. I apply the methods to study economic researchers in Section 5, and inventors and their patents in Section 6. Finally, I conclude in Section 7.

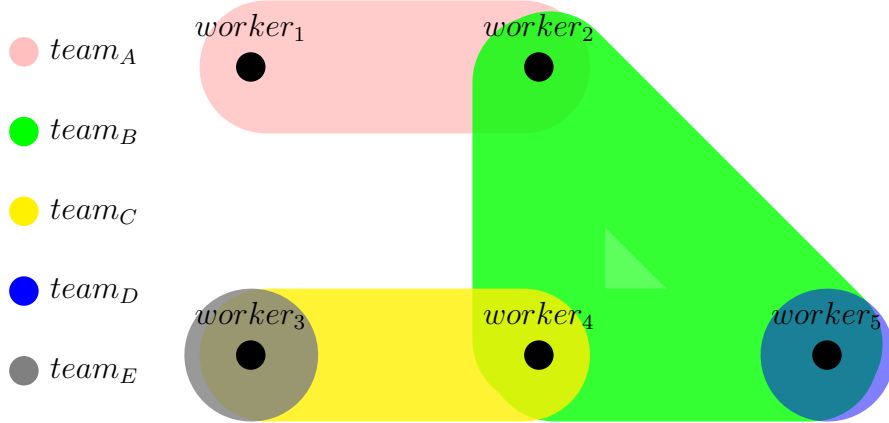
2 Framework

Consider a set of *nodes* $1, \dots, N$, which represent workers who collaborate and produce output in teams. Nodes are linked to each other by *hyperedges* that represent the collaborations between workers. In a standard graph, an edge links two nodes i and i' . In a hypergraph of collaborations, hyperedges can link a worker i to herself when i produces on her own, two workers i and i' who produce together, three workers i, i', i'' who produce together, and so on. I will refer to hyperedges as “teams”, and denote the workers in a team j of size $n_j = n$ as $(i_1(j), \dots, i_n(j))$. I denote the number of teams as J .

As an example, Figure 1 represents a hypergraph of collaborations with five teams and five workers. Workers 3 and 5 produce on their own. In addition, worker 3 produces jointly with worker 4, and workers 4 and 5 produce in a team together with worker 2. Lastly, worker 1 produces with worker 2 in a team.

⁴Related methods have been used to estimate manager fixed-effects (e.g., Bertrand and Schoar, 2003).

Figure 1: Collaboration hypergraph with five teams and five workers



Worker i contributes to the team an input quantity α_i , which is unobserved to the econometrician. I assume α_i is constant across collaborations, and I refer to α_i as the *type* of worker i . Constancy of worker types over time is a key feature of my framework. However, assuming that individual types remain constant may be restrictive.⁵ For this reason, in the empirical analysis I will use observations from at most five consecutive years. Note that, while worker types are constant, type composition varies across teams.

The output of a team j with n workers is denoted as Y_{nj} , and it is given by

$$Y_{nj} = \phi_n(\alpha_{i_1(j)}, \dots, \alpha_{i_n(j)}, \varepsilon_{nj}), \quad (1)$$

where ϕ_n is the production function of an n -worker team, and ε_{nj} represent other factors, or shocks, unobserved to the econometrician, that affect team output beyond workers' inputs α_i .^{6,7} I assume that the positions of workers in the team do not matter, so ϕ_n is symmetric with respect to its first n arguments. Hence, the framework abstracts from within-team hierarchy and organization.

I now state the first key assumption. Throughout, $\{A_k\}$ denotes the set of A_k for all k , so in particular $\{A_{nj}\}$ denotes the set of A_{nj} for all teams.

⁵To relax this assumption, one could rely on hidden Markov models, which allow types to evolve as Markov processes, or one could alternatively rely on mixed membership models (e.g., Airoldi *et al.*, 2008), which allow workers' types to depend on the team they participate in. I do not study these possibilities here.

⁶Here and in the following, $\alpha_{i_1(j)} = \sum_{i=1}^N \mathbf{1}\{i_1(j) = i\} \alpha_i$, and similarly for the other workers in team j .

⁷Since $n = n_j$ is the size of team j , an alternative notation for (1) could be

$$\tilde{Y}_j = \phi_{n_j}(\alpha_{i_1(j)}, \dots, \alpha_{i_{n_j}(j)}, \tilde{\varepsilon}_j),$$

where $\tilde{Y}_j = Y_{n_j, j}$ and $\tilde{\varepsilon}_j = \varepsilon_{n_j, j}$. I have chosen to use the subscript n in the notation given the key role of team size in the framework.

Assumption 1 (*Network exogeneity*)

$\{\varepsilon_{nj}\}$ are independent of $\{(i_1(j), \dots, i_n(j))\}$ conditional on $\{\alpha_i\}$.

Assumption 1 restricts the team formation process. It states that team formation is independent of the team-specific shocks ε_{nj} , conditional on the worker types $\alpha_1, \dots, \alpha_N$. Hence, while the probability of joining a team may depend unrestrictedly on worker types and other factors that are independent of the ε_{nj} 's, it cannot depend on the ε_{nj} 's themselves. Assumption 1 will thus fail if, before forming a team or joining one, workers have advance information about team-specific shocks. This assumption is important for the tractability of the approach. Yet it is restrictive, and in future work it will be important to relax it.⁸

I now state the second key assumption.

Assumption 2 (*Independent shocks*)

$\{\varepsilon_{nj}\}$ are mutually independent, and independent of $\{(i_1(j), \dots, i_n(j))\}$ and $\{\alpha_i\}$.

Assumption 2 rules out dependence among team-specific shocks. As an implication, shocks to a team of workers who collaborate repeatedly over time are assumed serially independent. An alternative approach, which I do not study in this chapter, would be to model the dependence, for example using a parametric process for ε_{nj} . More generally, the framework does not allow for dynamics in team production, such as state dependence effects.

Another implication of Assumption 2 is the absence of “team effects” in the model. In the framework, a team j is only a collection of workers, and there is no effect of j *per se* except through the independent shocks ε_{nj} . In particular, the model does not allow for team-specific capital (Jaravel *et al.*, 2018). In applications where firms (e.g., in R&D) or team identities (e.g., in sports) play an important role, it will be important to extend the framework to allow for team types and other team-specific factors, in addition to worker types.

Lastly, another assumption implicit in (1) is the absence of covariates, except for team size n . Covariates can be incorporated as additional inputs to production. In the empirical applications, I will account for time effects, and for age effects in robustness checks.

To specify and estimate the model, I will use two approaches. In the first one (“fixed-effects”, FE) I will treat the α_i 's as parameters and estimate them. The FE approach has

⁸A possibility to relax Assumption 1 is a joint random-effects approach, where production and team formation are estimated together (see Section 4).

the advantage of not requiring to specify a model of team formation. Moreover, when team output is additive in worker types and Assumption 1 holds, the FE approach is tractable. However, in the empirical collaboration networks that I will analyze, α_i estimates will often be very noisy. The prevalence of sample noise will require using methods for bias reduction.

In the second approach (“random-effects”, RE) I will model the joint distribution of the α_i ’s. I will use this approach to estimate nonlinear models with complementarity between worker types. The RE approach accounts naturally for sample noise. However, it requires specifying how the worker types depend on the collaborations in the network. In other words, the RE approach requires making assumptions about team formation beyond Assumption 1. I will compare several approaches based on simple statistical models.

3 Additive production

In this section and the next, I present methods for additive production and nonlinear production, in turn.

Suppose that the production function in (1) takes the following additive form

$$Y_{nj} = \lambda_n (\alpha_{i_1(j)} + \dots + \alpha_{i_n(j)}) + \varepsilon_{nj}, \quad (2)$$

where λ_n is a team-size scaling factor. I adopt the normalization $\lambda_1 = 1$. In (2), output depends on the sum of worker types, or equivalently on the mean worker type in the team (given the presence of λ_n).⁹

It is useful to write (2) in vector form as $Y_n = \lambda_n A_n \alpha + \varepsilon_n$; that is, equivalently, stacking all teams and team sizes together,

$$Y = D_\lambda A \alpha + \varepsilon, \quad (3)$$

where D_λ is a diagonal matrix, and A is a matrix of zeros and ones where a one indicates that a worker (i.e., a column) participates in a team (i.e., a row). Network exogeneity then takes the form

$$\mathbb{E}(\varepsilon_{nj} \mid A, \alpha) = \mu_n, \quad (4)$$

where μ_n is a team-size-specific constant. In the empirical analysis, I will estimate (2) under the assumption that $\mu_n = 0$. In addition, in a robustness check, I will report results from

⁹In model (2), types α_i are scalar, and their effects on output in teams of varying sizes are proportional to each other. The nonlinear model in Section 4 will allow for different patterns, possibly non-monotonic across teams of different sizes with varying worker composition.

the following alternative specification in logarithms,

$$\ln Y_{nj} = \mu_n + \alpha_{i_1(j)} + \dots + \alpha_{i_n(j)} + \varepsilon_{nj}. \quad (5)$$

3.1 Identification

To analyze identification in this setting, I treat α , λ , μ , and A as non-stochastic quantities.¹⁰ Suppose, to start with, that the λ_n 's are known, and that $\mu_n = 0$ in (4). Since model (3)-(4) is a linear regression, the identification of the worker effects $\alpha_1, \dots, \alpha_N$ is determined by the rank properties of A .

As an example, consider the matrix A corresponding to Figure 1:

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Since this matrix is non-singular, all $\alpha_1, \dots, \alpha_5$ are identified. An intuitive algorithm to show identification in this case is as follows. Let \mathcal{S} denote the set of workers i for which α_i is identified. First, focus on workers who produce on their own: since workers 3 and 5 work on their own, include them in \mathcal{S} . Then, focus on teams where the workers in \mathcal{S} collaborate with others: since workers 3 and 4 work together, include worker 4 in \mathcal{S} . Repeat this operation: since workers 2, 4 and 5 work together, include worker 2 in \mathcal{S} , and since workers 1 and 2 work together, include worker 1 in \mathcal{S} . Hence $\mathcal{S} = \{1, 2, 3, 4, 5\}$.

To generalize the approach beyond this simple example, let $B = D_\lambda A$, continuing with the case where D_λ is known. A worker type α_i , for $i \in \{1, \dots, N\}$, is identified if and only if

$$\exists v_i : B'v_i = e_i,$$

where e_i denotes the i -th canonical vector in \mathbb{R}^N . Letting $(B')^\dagger$ denote the Moore-Penrose generalized inverse of B' , and I be a conformable identity matrix, this is equivalent to

$$(I - B'(B')^\dagger)e_i = 0. \quad (6)$$

All worker types α_i for which (6) holds are identified. In general, however, not all α_i 's are identified. For example, when two workers always work together, with or without other co-workers, their respective α_i 's are not separately identified.¹¹

¹⁰Hence the conditioning on (A, α) in (4) could be removed from the notation.

¹¹However, the sum of their α_i 's is identified. Workers who always work together correspond to identical columns in B . Instead of dropping those columns, merging them would allow one to identify sums of α_i 's.

In practice, it is convenient to further restrict the sets of workers and teams such that the types of all workers in those teams are identified. To do this, I implement the following simple iterative algorithm.

Algorithm 1 *Iterate until convergence:*

1. Select all the workers such that (6) holds.
2. Select all the teams that only comprise those workers.
3. Let B^{sub} be the resulting selection of rows and columns of B . Rename B^{sub} as B .

To compute Moore-Penrose generalized inverses, I rely on sparse LU matrix decompositions. With some abuse of notation, I will still denote the resulting restricted vectors and matrices as Y , D_λ , A , and ε .¹²

Consider now the team-size parameters λ_n , under the maintained assumption that $\mu_n = 0$. Since

$$D_\lambda^{-1}Y = A\alpha + D_\lambda^{-1}\varepsilon,$$

it follows that

$$(I - AA^\dagger)D_\lambda^{-1}Y = (I - AA^\dagger)D_\lambda^{-1}\varepsilon.$$

Hence, (4) implies

$$\mathbb{E} [Z'_n(I - AA^\dagger)D_\lambda^{-1}Y] = 0, \quad \text{for all } n, \quad (7)$$

where Z_n is a $J \times 1$ vector with ones where $n_j = n$ and zeros elsewhere. System (7) is linear in the parameters λ_n^{-1} , so the conditions for identification are standard.¹³ In particular, $\lambda_1, \lambda_2, \lambda_3$ are not identified in the stylized example of Figure 1.

Lastly, when relaxing the assumption that $\mu_n = 0$ in (4), one can use a similar strategy. To see this, let μ be the vector of μ_n 's. We have

$$\mathbb{E} [Z'(I - AA^\dagger)D_\lambda^{-1}(Y - \mu)] = 0, \quad (8)$$

¹²Here my goal is to point-identify the α_i values of a set of workers, and the corresponding contributions of these workers to team output. For this reason, in the identification strategy I rely on workers who produce on their own. In other settings (e.g., team sports) this information may not be available. In such cases, it may still be possible to identify differences between α_i 's. As an example, consider a team of five players, where player i gets replaced by player i' . In this case the difference in the team's output identifies $\alpha_i - \alpha_{i'}$, though not α_i and $\alpha_{i'}$ separately. Another example is doubles tennis, as studied in Devereux (2018). In the empirical analysis, in robustness checks I will report estimates of nonlinear models that are solely based on 2-worker teams.

¹³This quasi-differencing approach is related to Holtz-Eakin *et al.* (1988) and Chamberlain (1992).

where instruments Z can be constructed as functions of collaborations in the network. As an example, the j -th row of Z may contain some functions of the team sizes of the co-workers in j , in addition to the size n_j of team j . Moreover, under assumptions on the covariance matrix of ε , one can add covariance restrictions on λ and μ to the mean restrictions in (8). I will not implement these approaches empirically in Sections 5 and 6, and proceed under the assumption that $\mu_n = 0$ for all n .

3.2 Estimation

I will focus on the following decomposition of output variance, for every team size n :¹⁴

$$\underbrace{\text{Var}_n(Y_{nj})}_{\text{Total variance}} = \underbrace{\lambda_n^2 \sum_{m=1}^n \text{Var}_n(\alpha_{im(j)})}_{\text{Heterogeneity}} + \underbrace{2\lambda_n^2 \sum_{m=1}^n \sum_{m'=m+1}^n \text{Cov}_n(\alpha_{im(j)}, \alpha_{i_{m'}(j)})}_{\text{Sorting}} + \underbrace{\text{Var}_n(\varepsilon_{nj})}_{\text{Other factors}}. \quad (9)$$

In this decomposition, the component labeled “heterogeneity” reflects the variation in worker effects on output, keeping team composition constant. Equivalently, it represents the effect of worker heterogeneity if the allocation of workers to teams were random, in which case covariances would be zero. In turn, the component labeled “sorting” reflects the variance contribution due to team composition not being random.¹⁵

To estimate the variance components in (9), I first estimate the team-size effects λ_n using an empirical counterpart to the just-identified system (7).¹⁶ Next, I estimate α using OLS as

$$\hat{\alpha} = (B' B)^{-1} B' Y,$$

for $B = D_\lambda A$. Hence, for any variance component $V_Q = \alpha' Q \alpha$, where Q is an $N \times N$ matrix, I can construct the fixed-effects (FE) estimator

$$\hat{V}_Q = \hat{\alpha}' Q \hat{\alpha}. \quad (10)$$

However, the FE estimator \hat{V}_Q is biased. To see this, note that

$$\mathbb{E}[\hat{V}_Q | A, \alpha] = V_Q + \underbrace{\mathbb{E}[\varepsilon' (B' B)^{-1} B' Q B (B' B)^{-1} \varepsilon | A, \alpha]}_{\text{Bias}}. \quad (11)$$

¹⁴Here Var_n and Cov_n denote empirical variances and covariances, respectively, for teams for size n .

¹⁵It is common in the literature on matched employer-employee data to apply a related decomposition to the variance of log-wages, as opposed to (log-) output. This raises issues for the interpretation of variance components estimates in terms of heterogeneity and sorting (Eeckhout and Kircher, 2011). In contrast, here heterogeneity and sorting can be directly interpreted as contributions to production.

¹⁶The additive specification (5) in logs can be estimated similarly to the one in levels, the only difference being a slightly different approach to estimate μ_n .

Following Andrews *et al.* (2008), I subtract an unbiased estimate of the bias to \widehat{V}_Q . To construct such an estimate, note that

$$\text{Bias} = \text{Trace}((B'B)^{-1}B'QB(B'B)^{-1}\Omega),$$

where Ω is the covariance matrix of ε . To obtain an empirical counterpart of Ω , I assume that all the ε_{nj} 's are independent (see Assumption 2), and in addition that $\text{Var}(\varepsilon_n|A, \alpha) = \sigma_n^2$ varies only with team size. I estimate σ_n^2 as

$$\widehat{\sigma}_n^2 = \frac{Y_n'(I_n - A_n A_n^\dagger)Y_n}{\text{Trace}(I_n - A_n A_n^\dagger)}. \quad (12)$$

The bias correction relies on the ε_{nj} 's being independent, and on their variance being constant within team size. While homoskedasticity could be relaxed by following Kline *et al.* (2020) and suitably restricting the sets of workers and teams, independence may be empirically strong in the applications. Under independence, Kline *et al.* (2020) derive consistency results and limiting distributions for bias-corrected variance components estimates in environments where the numbers of rows (teams) and columns (workers) in A both tend to infinity.

To implement the bias correction in large data sets, I rely on a sparse LU decomposition to compute the denominator in (12), and I approximate the trace in the bias formula using Hutchinson's approximation, as in Gaure (2014) and Kline *et al.* (2020). I use 1000 draws in the approximation.

4 Nonlinear production

I now describe a nonlinear model that allows for complementarity between worker types. To keep tractability, I model types α_i as discrete, with K points of support. I will vary K in robustness checks.

I adopt a random-effects (RE) strategy that consists in modeling the distribution of types. Specifically, I suppose that $\alpha_1, \dots, \alpha_N$ are drawn from the joint distribution $\prod_{i=1}^N \pi(\alpha_i)$, where $\pi(\alpha_i)$ are type probabilities. I will first describe an independent RE method where $\{\alpha_i\}$ are independent of collaborations $\{(i_1(j), \dots, i_n(j))\}$. I will interpret π as a prior on individual types, as in Arellano and Bonhomme (2009). While, by construction, the prior imposes that team formation is independent of types – and thus rules out sorting – the posterior estimates that I will report may still feature sorting (and empirically they will).¹⁷

¹⁷In a similar spirit, in matched employer-employee data, Woodcock (2008) points out that posterior RE

However, a natural concern is that the prior π may be empirically informative. For example, a prior that is independent of collaborations will tend to shrink sorting patterns towards zero. Unlike the FE methods of Section 3, RE methods rely implicitly or explicitly on a model of team formation, and can be sensitive to that model being misspecified. For this reason, I will also present estimates where I will model the dependence between the types α_i and the collaborations $(i_1(j), \dots, i_n(j))$.

With discrete types, a different approach would be to treat type membership indicators as discrete parameters, and use kmeans clustering methods for estimation (Bonhomme and Manresa, 2015, Bonhomme *et al.*, 2021). Bonhomme *et al.* (2019) use this grouped fixed-effects approach to incorporate discrete firm heterogeneity in a nonlinear model of wages. I will use a related approach to provide preliminary evidence of sorting and complementarity. However, accurately estimating worker types using grouped fixed-effects requires a relatively large number of observations per worker. In the data sets that I analyze in this chapter, many workers produce only a handful of times. For this reason, I instead rely on a RE approach.

4.1 Identification

Nonparametric identification of finite mixture models under independence assumptions akin to Assumption 2 has been extensively studied. When all teams consist of a single worker, the model has independent measures and finite types, and its identification has been studied in Hall and Zhou (2003), Hu (2008), and Allman *et al.* (2009), among others. When teams have multiple workers, the model has a network structure that is related to the settings studied in Allman *et al.* (2011).

Here I outline an identification argument in the case where $\{\alpha_i\}$ are independent of collaborations $\{(i_1(j), \dots, i_n(j))\}$, output is i.i.d. across teams, and team size is $n \in \{1, 2\}$. Formally, I treat worker types α_i as random, and collaborations $(i_1(j), \dots, i_n(j))$ as non-stochastic, and assume that the type distribution π is common across workers, independent of the collaborations. The goal is to identify the type distribution π , as well as the type-specific conditional distributions of output.

Focusing on workers who produce at least three times on their own gives, for $\{j_1, j_2, j_3\}$

estimates can still indicate the presence of sorting even if the prior on worker and firm types is independent of mobility patterns.

triplets such that $i_1(j_1) = i_1(j_2) = i_1(j_3)$,

$$\begin{aligned} & \Pr [Y_{1j_1} \leq y_1, Y_{1j_2} \leq y_2, Y_{1j_3} \leq y_3] \\ &= \sum_{\alpha} \pi(\alpha) \Pr [Y_{1j_1} \leq y_1 \mid \alpha_{i_1(j_1)} = \alpha] \Pr [Y_{1j_2} \leq y_2 \mid \alpha_{i_1(j_2)} = \alpha] \Pr [Y_{1j_3} \leq y_3 \mid \alpha_{i_1(j_3)} = \alpha], \end{aligned} \quad (13)$$

where I have used Assumptions 1 and 2, and that π does not depend on $\{(i_1(j), \dots, i_n(j))\}$. System (13) identifies $\pi(\alpha)$ and $\Pr [Y_{1j} \leq y \mid \alpha_{i_1(j)} = \alpha]$ up to labeling of the latent types, under suitable rank conditions (Allman *et al.*, 2009).

Next, focusing now on pairs of workers who produce once on their own and once together, we similarly have, for $\{j_1, j_2, j_3\}$ triplets such that $\{i_1(j_1), i_1(j_2)\} = \{i_1(j_3), i_2(j_3)\}$,

$$\begin{aligned} & \Pr [Y_{1j_1} \leq y_1, Y_{1j_2} \leq y_2, Y_{2j_3} \leq y_3] \\ &= \sum_{\alpha, \alpha'} \pi(\alpha)\pi(\alpha') \Pr [Y_{1j_1} \leq y_1 \mid \alpha_{i_1(j_1)} = \alpha] \Pr [Y_{1j_2} \leq y_2 \mid \alpha_{i_1(j_2)} = \alpha'] \times \\ & \quad \Pr [Y_{2j_3} \leq y_3 \mid \alpha_{i_1(j_3)} = \alpha, \alpha_{i_2(j_3)} = \alpha']. \end{aligned} \quad (14)$$

Given that $\pi(\alpha)$ and $\Pr [Y_{1j} \leq y \mid \alpha_{i_1(j)} = \alpha]$ are identified up to labeling, (14) identifies $\Pr [Y_{2j} \leq y \mid \alpha_{i_1(j)} = \alpha, \alpha_{i_2(j)} = \alpha']$ up to the same labeling, under a suitable rank condition.¹⁸ The argument requires no monotonicity assumptions about how output depends on worker types. In addition, identification can be shown using a related argument when π depends on collaborations and worker types are not independent of each other within teams, provided none of the joint probabilities $\pi(\alpha, \alpha')$, for pairs of workers who produce once on their own and once together, are positive.

While the previous argument relies on workers producing on their own, identification can also be shown in settings without 1-worker teams; see Allman *et al.* (2011) for identification results in certain random graph models.¹⁹ As robustness checks, I will report empirical results based on 2-worker teams only.

¹⁸Specifically, it suffices that there exist a set $\{y_r\}$ of values in the support of Y_{1j} such that the matrix with $((y_r, y_{r'}), (\alpha, \alpha'))$ -element $\Pr [Y_{1j} \leq y_r \mid \alpha_{i_1(j)} = \alpha] \Pr [Y_{1j} \leq y_{r'} \mid \alpha_{i_1(j)} = \alpha']$ has full column rank.

¹⁹As an example, for two workers who collaborate together at least three times, we have, for $\{j_1, j_2, j_3\}$ triplets such that $\{i_1(j_1), i_2(j_1)\} = \{i_1(j_2), i_2(j_2)\} = \{i_1(j_3), i_2(j_3)\}$,

$$\begin{aligned} \Pr [Y_{2j_1} \leq y_1, Y_{2j_2} \leq y_2, Y_{2j_3} \leq y_3] &= \sum_{\alpha, \alpha'} \pi(\alpha)\pi(\alpha') \Pr [Y_{2j_1} \leq y_1 \mid \alpha_{i_1(j_1)} = \alpha, \alpha_{i_2(j_1)} = \alpha'] \times \\ & \Pr [Y_{2j_2} \leq y_2 \mid \alpha_{i_1(j_2)} = \alpha, \alpha_{i_2(j_2)} = \alpha'] \Pr [Y_{2j_3} \leq y_3 \mid \alpha_{i_1(j_3)} = \alpha, \alpha_{i_2(j_3)} = \alpha'], \end{aligned}$$

which identifies $\pi(\alpha)$ and $\Pr [Y_{2j} \leq y \mid \alpha_{i_1(j)} = \alpha, \alpha_{i_2(j)} = \alpha']$ up to labeling of the components, under suitable rank conditions.

4.2 Estimation using a variational approach

Estimating the random-effects model of team production is challenging. To see this, consider a setup where the density $f_{\alpha_{i_1(j)}, \dots, \alpha_{i_n(j)}}^n(y)$ of Y_{nj} conditional on $\alpha_{i_1(j)}, \dots, \alpha_{i_n(j)}$ is parametric, indexed by a finite-dimensional vector θ . The RE likelihood

$$\mathcal{L}(\theta, \pi) = \sum_{\alpha_1} \dots \sum_{\alpha_N} \prod_i \pi(\alpha_i) \prod_n \prod_j f_{\alpha_{i_1(j)}, \dots, \alpha_{i_n(j)}}^n(Y_{nj}; \theta)$$

involves an intractable N -dimensional sum over all possible worker type realizations. The likelihood does not factor in simple ways, except in the special case where all teams consist of one worker.²⁰

To reduce computational complexity, I follow a mean-field variational approach that is increasingly popular in related settings in machine learning and statistics. The idea is to introduce an auxiliary distribution $\prod_{i=1}^N q_i(\alpha_i)$, and set it to be as close as possible to the posterior density of $\alpha_1, \dots, \alpha_N$. Unlike the posterior density, its variational approximation factors across i . This makes estimation feasible even in large data sets.

Formally, the variational objective function is

$$\text{ELBO}(\theta, \pi) = \max_{q_1, \dots, q_N} \ln(\mathcal{L}(\theta, \pi)) - \underbrace{\mathbb{E}_{q_1 \dots q_N} \ln \frac{\prod_i q_i(\alpha_i)}{p(\alpha_1, \dots, \alpha_N; \theta, \pi)}}_{\text{KL divergence}}, \quad (15)$$

where $p(\alpha_1, \dots, \alpha_N; \theta, \pi)$ is the (computationally intractable) posterior density of worker types given θ and π values. The second term on the right-hand side of (15) is the Kullback-Leibler (KL) divergence between the posterior density and its variational approximation. Hence, the variational objective and the RE likelihood are closer when the variational approximation is more accurate. The variational objective is a lower bound on the RE likelihood, and it is often referred to as the “evidence lower bound” (ELBO).

To see why adding the KL term can improve tractability, note that we equivalently have

$$\begin{aligned} \text{ELBO}(\theta, \pi) &= \max_{q_1, \dots, q_N} \sum_n \sum_j \mathbb{E}_{q_{i_1(j)} \dots q_{i_n(j)}} \ln f_{\alpha_{i_1(j)}, \dots, \alpha_{i_n(j)}}(Y_{nj}; \theta) + \sum_i \mathbb{E}_{q_i} \ln \pi - \sum_i \mathbb{E}_{q_i} \ln q_i \\ &= \max_{q_1, \dots, q_N} \sum_n \sum_j \sum_{\alpha_1} \dots \sum_{\alpha_n} q_{i_1(j)}(\alpha_1) \dots q_{i_n(j)}(\alpha_n) \ln f_{\alpha_1, \dots, \alpha_n}(Y_{nj}; \theta) \\ &\quad + \sum_i \sum_{\alpha} q_i(\alpha) \ln \pi(\alpha) - \sum_i \sum_{\alpha} q_i(\alpha) \ln q_i(\alpha). \end{aligned} \quad (16)$$

²⁰In 1-worker teams, the identification strategy leads to practical nonparametric estimators; see Bonhomme *et al.* (2016) for example. This could be used to nonparametrically estimate models of team production, using (14) and its extensions for $n > 2$.

Since this expression no longer involves an N -dimensional sum, it can be evaluated easily, at least when team size n is relatively low. To estimate the nonlinear models, I will maximize the evidence lower bound using the EM algorithm; see Bishop (2006) and Mariadassou *et al.* (2010). Specifically, the algorithm alternates between updates of the $q_i(\alpha)$'s given (θ, π) via Newton steps, and updates of (θ, π) given the $q_i(\alpha)$'s via weighted maximum likelihood estimation.

Computation of the evidence lower bound in (16) becomes more complex as team size increases. In the applications, I will restrict the analysis to 1-worker and 2-worker teams. It will be important to devise computational strategies to efficiently handle larger teams.²¹

Remark: statistical properties of variational estimates. An obvious issue with replacing the RE likelihood by the evidence lower bound is that this leads to optimizing a different objective function. Indeed, in the present setting as well as in most network applications, the posterior density p does not factor across workers, whereas the variational density q does. This discrepancy can make the variational estimates inconsistent even when the assumptions of the RE model hold.

The analysis in Bickel *et al.* (2013) suggests that variational estimates can be theoretically justified in network models with discrete types (see also Celisse *et al.*, 2012). Bickel *et al.* (2013) focus on a standard stochastic blockmodel (e.g., Snijders and Nowicki, 1997) with binary outcomes, where choice probabilities are functions of both agents' types. In their asymptotic environment, expected degree in the network grows at a rate faster than $\ln N$. They first show that the RE maximum likelihood estimator (MLE) is asymptotically equivalent to an infeasible MLE where the agents' types are observed.²² As a result, the RE estimator is consistent and asymptotically normal. They then show that the mean-field variational estimator (Daudin *et al.*, 2008) is asymptotically equivalent to both the infeasible MLE and the RE MLE, hence that it is consistent with the same asymptotically normal distribution.

Stochastic blockmodels with binary outcomes are closely related to the model I focus on. However, there are also important differences, since in my model workers both form teams

²¹Moreover, to estimate models with larger teams, one will need to impose additional structure on how production depends on types. A possibility, in the spirit of Ahmadpoor and Jones (2019), would be to model type-specific mean output using a CES function.

²²Related perfect classification results have been derived in panel data models under discrete heterogeneity (e.g., Hahn and Moon, 2010, Bonhomme and Manresa, 2015).

and produce, and output is continuous. In addition, the conditions in Bickel *et al.* (2013) rely on the network becoming sufficiently dense in large samples. It is unclear whether conditions akin to the ones they assume provide a good approximation to the settings I study. For this reason, in the appendix I probe the accuracy of the variational estimates using Monte Carlo simulations. In simulated samples that mimic the data on economic researchers, I find that variational estimates recover the true parameter values well when the numbers of productions and collaborations are sufficiently large. Studying the theoretical properties of mean-field variational estimators in models of team production is an important avenue for future work.

4.3 Accounting for the link between worker types and collaborations

The assumption that the distribution of worker types α_i does not depend on collaborations $\{(i_1(j), \dots, i_n(j))\}$ may be restrictive. To give intuition, consider a simple assignment model of team formation, along the lines of the roommate matching problem studied by Chiappori *et al.* (2019). In the model, workers are assigned to teams of size $n \in \{1, 2\}$ in a way that maximizes expected output. Let $\mu_1(\alpha) = \mathbb{E}(Y_{1j} | \alpha_{i_1(j)} = \alpha)$ and $\mu_2(\alpha, \alpha') = \mathbb{E}(Y_{2j} | \alpha_{i_1(j)} = \alpha, \alpha_{i_2(j)} = \alpha')$. Let $T_{1\alpha}, T_{2\alpha} \in \mathbb{N}$ be exogenous “budgets” for type- α workers, indicating the maximum numbers of teams of size 1 and 2 in which they can participate. Consider an allocation solving

$$\begin{aligned} \max \quad & \sum_{\alpha} \tau_{\alpha} \mu_1(\alpha) + \sum_{\alpha} \tau_{\alpha\alpha} \mu_2(\alpha, \alpha) + \frac{1}{2} \sum_{\alpha} \sum_{\alpha' \neq \alpha} \tau_{\alpha\alpha'} \mu_2(\alpha, \alpha') \\ \text{s.t.} \quad & \tau_{\alpha}, \tau_{\alpha\alpha'} \in \mathbb{N}, \tau_{\alpha\alpha'} = \tau_{\alpha'\alpha}, \tau_{\alpha} \leq T_{1\alpha}, 2\tau_{\alpha\alpha} + \sum_{\alpha' \neq \alpha} \tau_{\alpha\alpha'} \leq T_{2\alpha}, \end{aligned} \quad (17)$$

where τ_{α} denotes the number of teams with one worker of type α , and $\tau_{\alpha\alpha'}$ denotes the number of teams with one worker of type α and another one of type α' .

The optimization in (17) is related to the planner’s objective in the Becker (1973) two-sided marriage model.²³ As in the Becker model, in (17) the assignment of individuals to teams will be driven by complementarity patterns in the expected payoff function $\mu_2(\alpha, \alpha')$.

²³See also Carlier and Ekeland (2010) for an extension of the two-sided marriage model to a team setting. In the Becker model of marriage, the optimal allocation can be represented as a stable equilibrium in an economy with transfers. In one-sided problems, existence of a decentralized solution is not guaranteed in general (e.g., Talman and Yang, 2011). Chiappori *et al.* (2019) show existence when types are discrete and budget sizes are even. Unlike in their setup, here there are two budgets, for 1-worker and 2-worker teams, respectively. Determining optimal allocations of workers across teams of different sizes would require measuring the costs of teamwork – e.g., effort costs – which I do not consider here.

Intuitively, the higher the degree of complementarity between workers, the higher the assortativeness of the optimal allocation. I will illustrate the link between complementarity and assortativeness by computing optimal allocations of economists and inventors in Sections 5 and 6. In particular, this simple team assignment model implies that types will generally *not* be uniformly distributed across teams.

To relax the independence assumption between types and collaborations, I will rely on two approaches. In the first approach, I will maintain prior independence between the α_i 's, however I will let $\pi_i(\alpha_i)$ depend on worker i 's characteristics. This correlated random-effects approach (Chamberlain, 1984) is widely used in traditional panel data applications. However, in a network of teams, it is not *a priori* obvious which characteristics α_i should depend on. I will experiment with features of the degree distributions in the collaboration hypergraph. In addition, note that this approach imposes that worker types be conditionally independent of one another given collaborations, which is unlikely when the assignment of workers to teams is given by a team formation model such as (17).

In the second approach, I will jointly model output production and team formation. This adds another estimation step, since one needs to estimate the team formation process as well. Identification can be analyzed using related techniques,²⁴ and the mean-field approximation can be implemented similarly. A specific feature of the joint model of team formation and production is that worker types influence both the quality and the quantity of output. For example, in the case of researchers in economics, the type of a worker will influence both how many articles she publishes (i.e., quantity) and in which journals they appear (i.e., quality). To implement this approach, I will specify a simple stochastic blockmodel of team formation. Though of great interest, using an economic team formation model instead would raise several challenges, and I do not consider this possibility in this chapter.

5 Academic production in economics

In this section and the next I present two applications. I start by applying the method to study economists and their academic production.

Table 1: Descriptive statistics on economic researchers

	(a) Sample				(b) Subsample		
	All	$n = 1$	$n = 2$	$n = 3$	All	$n = 1$	$n = 2$
# Authors	6509	6194	4937	1643	4825	4724	3148
# Articles	41150	31085	8987	1078	31782	26558	5224
Mean output	5.55	4.44	7.12	7.56	5.37	4.55	7.45
Std output	11.18	9.80	12.74	12.95	10.85	9.89	12.75
p10% output	0.52	0.52	0.52	0.54	0.52	0.52	0.52
p50% output	0.76	0.66	1.81	2.14	0.67	0.66	2.06
p90% output	14.88	12.09	19.65	19.72	14.86	12.42	20.68
p95% output	24.85	22.73	35.13	37.06	24.85	23.02	35.13
p99% output	58.90	57.66	62.29	64.92	58.58	56.67	63.65
p10% #articles	5	3	2	1	5	4	2
p50% #articles	8	6	5	2	8	6	4
p90% #articles	18	12	11	7	16	13	10
p95% #articles	23	16	14	8	21	17	13
p99% #articles	39	31	20	13	35	35	18

Notes: Data from Ductor *et al.* (2014), 1995-1999. Output is a measure of journal quality, net of multiplicative year effects (reference year is 1999). Percentiles of the distribution of number of articles per author are indicated in the bottom five rows. In panel (a) the sample is restricted to at least 5 publications per author during the period. In panel (b) the subsample is restricted to collaborations between at most two authors, where all authors have at least 5 publications during 1995-1999. I use the subsample in panel (b) to estimate nonlinear models.

5.1 Data and sample selection

I use data from Ductor *et al.* (2014). These data, initially constructed by Goyal *et al.* (2006), are drawn from the EconLit database, a bibliography of journals compiled by the Journal of Economic Literature. I restrict the sample to articles published between 1995 and 1999, with at most three co-authors.

I only include authors who produced at least five articles during the period. While a natural selection when aiming to infer individual contributions, restricting the number of publications per author in this way is restrictive. Indeed, starting from 62615 authors and 89834 articles in the original 1995-1999 data, the selection restricts the sample to 6509 authors and 41150 articles. To assess how representative the results on this sample are, I will run checks using larger samples.

As measure of academic output, I follow Ductor *et al.* (2014) and use the journal quality

²⁴In this case there are two kinds of dependent variables: the output variables and the team membership indicators. See Bonhomme (2017) for an example.

variable from Kodrzycki and Yu (2006), as extended by Ductor and co-authors. From the output measure I net out multiplicative year fixed-effects, using 1999 as reference year. I show descriptive statistics in panel (a) of Table 1 for the main sample. Research output, as measured by the journal quality variable, is right-skewed. The number of collaborations per author is also skewed, with 50% of authors producing at most 8 articles, and 1% producing at least 39. The subsample in panel (b), which I will use to estimate nonlinear models of 1-author and 2-author teams, shows similar patterns.

5.2 Additive production

I start by estimating the additive model of academic production (2). I first restrict the set of authors and articles in order to ensure identification. This gives 6479 authors and 41049 articles. Hence, relative to the sample in panel (a) of Table 1, only 30 authors and 101 articles drop out.

Using this sample, I first estimate the team-size effects λ_n . I find $\lambda_2 = 0.67$ and $\lambda_3 = 0.48$. This suggests that, keeping author type(s) constant, collaborations between two co-authors increase output by $2 \times 0.67 - 1 = 34\%$, and collaborations between three co-authors increase output by $3 \times 0.48 - 1 = 44\%$. Next, I estimate the variance components in the decomposition (9). I report uncorrected estimates, which are obtained using fixed-effects, and bias-corrected estimates, which I obtain using the method described in Section 3.

Table 2: Additive production, economic researchers

	$n = 1$	$n = 2$	$n = 3$
Total variance	95.97	163.21	161.38
Heterogeneity	32.53	31.71	22.58
Heterogeneity (uncorrected)	42.80	53.26	42.70
Sorting	-	28.41	32.73
Sorting (uncorrected)	-	20.67	23.34
Other factors	63.86	82.82	86.81
Other factors (uncorrected)	53.17	92.26	99.41
Team scale λ_n	1.00	0.67	0.48

Notes: Estimates of variance components in equation (9), for different team sizes n . Descriptive statistics on the sample are given in panel (a) of Table 1.

In Table 2 I show the results of the variance decomposition (9), for the three team sizes: $n = 1, 2, 3$. The results suggest that author heterogeneity matters substantially, while playing

a smaller role in larger teams. Indeed, the variance share explained by author heterogeneity is 33% for sole-authored articles, 19% in articles with two co-authors, and 14% in articles with three co-authors. Sorting also contributes a substantial share: 17% in articles with two co-authors, and 20% in articles with three co-authors. Hence, according to this analysis, both heterogeneity and sorting contribute to the variation in academic output.

Another finding in Table 2 is the magnitude of the component due to other factors, σ_n^2 . This component explains 67% of the output variance in articles with one author, 57% in articles with two co-authors, and 62% in articles with three co-authors. The large residual variance contributes to the fixed-effects estimates of variance components being substantially biased. For example, the fixed-effects estimate of the variance contribution of heterogeneity is biased upward by 33% with one author, 69% with two co-authors, and 91% with three co-authors. This suggests that, as has been documented in matched employer-employee settings (e.g., Bonhomme *et al.*, 2020), bias correction is needed in order to obtain reliable estimates of variance components in collaboration networks.

Robustness analysis. The adequacy of the additive model (2) depends on how one measures the output variable. To explore the sensitivity to this choice, I use two alternative measures of output. First, I estimate the model in logs, accounting for additive team-size effects as in (5). The results are shown in Appendix Table C2, panel (a). Second, I use year-specific ranks of journal quality as dependent variables, instead of journal quality itself. The results are shown in Appendix Table C2, panel (b). While it is reassuring that the variance shares in the two specifications are broadly comparable to the baseline, this exercise does not fully address the issue of the sensitivity to the choice of output measure. The nonlinear model that I estimate in the next section will be less sensitive to this issue.²⁵

The estimation sample represents a relatively small share of authors and articles during the period. To probe the robustness of the findings to sample definition, I consider two less restrictive rules for inclusion in the sample, requiring every author to produce at least one or two articles, respectively, as opposed to at least five in the baseline sample. The results are shown in Appendix Table C2, panels (c) and (d). When requiring at least two articles per author, the sample contains 24555 authors and 70550 articles, and 23498 authors and 69331 articles in the sample where all author types are identified. Panel (c) of the table

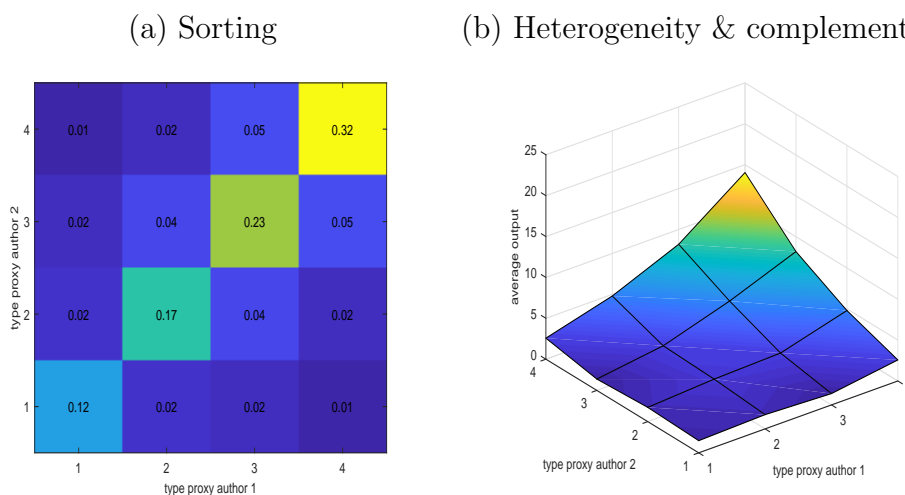
²⁵In this application, it would be interesting to link the journal quality variable to economic payoffs and career outcomes.

shows that bias-corrected variance shares are broadly similar to the baseline, and that biases are larger, so bias correction matters more. When removing all restrictions and including all authors with at least one article, the sample where all author types are identified contains 44852 authors and 79659 articles. Panel (d) of Appendix Table C2 shows larger differences compared to the baseline subsample, and a substantially higher amount of bias.

5.3 Nonlinear model

Before presenting the results of the nonlinear model, I provide some preliminary evidence of complementarity, as well as sorting and heterogeneity, in the data. To do so, I select authors who produce at least five articles *on their own*. This gives a set of 3263 authors. Then, for every author I compute the average output (i.e., average journal quality) among their sole-authored articles, and divide this measure into four quartiles. I call the resulting quartile the “type proxy” of the author. The reason for focusing on individuals with at least five sole-authored articles is to reduce noise in the proxy. At the same time, this does not fully remove the noise and requires a selected sample.

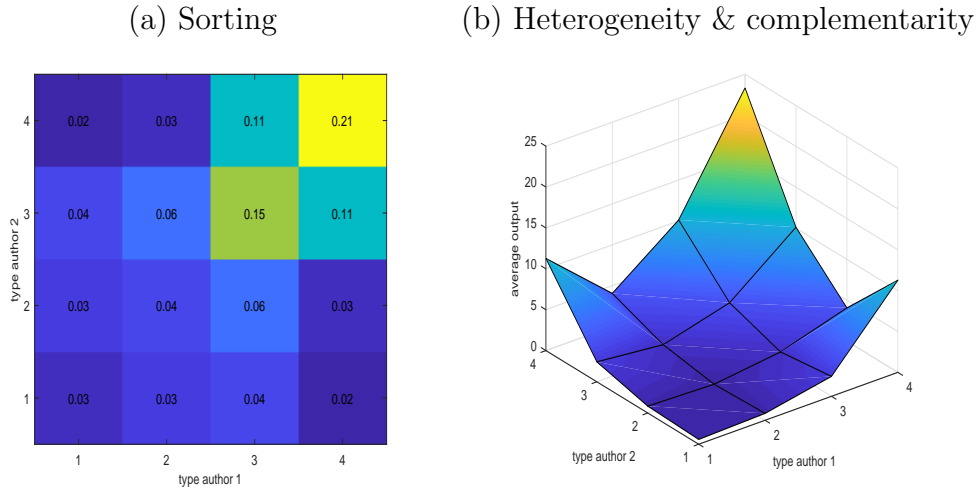
Figure 2: Preliminary evidence based on type proxies, economic researchers



Notes: Subsample of authors who produce at least 5 articles on their own. I compute type proxies as quartiles of average sole-authored output. In panel (a) I show the proportions of type proxies for authors producing together in a 2-author team. In panel (b) I show average output (i.e., average journal quality) for different combinations of the type proxies.

Given this proxy of type, I focus on 2-author teams and report two sets of results, which I interpret as reflecting sorting and complementarity, respectively. In panel (a) of Figure 2,

Figure 3: Nonlinear model estimates, economic researchers ($K = 4$)



Notes: Random-effects estimates of a finite mixture model with $K = 4$ types. In panel (a) I show the proportions of types for authors producing together in a 2-author team. In panel (b) I show average output (i.e., average journal quality) for different combinations of the types. Descriptive statistics on the sample are given in panel (b) of Table 1.

I show the proportions of every type proxy pairs in the sample. The percentages suggest that higher-type authors, who produce higher-quality articles on their own, tend to work together, implying the presence of sorting between authors of similar productivity. In panel (b) of Figure 2, I compute the average output for every type proxy pairs. Authors who produce higher-quality articles on their own tend to produce higher-quality output as well when they work in teams, which is consistent with the presence of author heterogeneity. In addition, joint output is highest when authors with the highest type proxies (i.e., in the top quartile of sole-authored production) work together, and the figure is suggestive of the presence of complementarity. Since the additive model (2) rules those out, this motivates estimating a nonlinear model of academic production.

Estimates of the nonlinear model, baseline specification. I now report estimates of a finite mixture model with K types, varying K between 2 and 6. In this analysis I only consider 1-author and 2-author productions; see panel (b) of Table 1 for summary statistics on the sample. I model the output distribution as a log-normal, with mean and variance that depend on the types of authors in the team: for 1-author teams mean and variance are functions of the type, and for 2-author teams mean and variance are symmetric functions

of the two types. The log-normal specification is restrictive, and in future work it will be important to check how robust the results are to removing this functional form assumption. I use mean-field variational EM for estimation.²⁶

For conciseness, I only comment in detail on the results for $K = 4$. In this case, the type proportions are 15%, 20%, 36%, and 29%. The means of the log-normal output distribution are, in 1-author teams:

$$\begin{pmatrix} 0.59 \\ 0.80 \\ 2.56 \\ 14.94 \end{pmatrix},$$

and in 2-author teams:

$$\begin{pmatrix} 0.59 & 0.84 & 2.44 & 11.27 \\ 0.84 & 0.59 & 1.61 & 4.07 \\ 2.44 & 1.61 & 3.81 & 10.12 \\ 11.27 & 4.07 & 10.12 & 23.34 \end{pmatrix}.$$

This suggests that authors have heterogeneous productivity levels, and these differences affect both 1-author and 2-author productions.

To visualize the implications of the estimated model for heterogeneity, sorting and complementarity, in Figure 3 I report similar quantities as in Figure 2, except that I now use the types estimated under the model instead of the type proxies. Panel (a) of Figure 3 shows that high-type authors have a stronger propensity to work together, although this evidence of sorting is less pronounced than in Figure 2.

Panel (b) of Figure 3 suggests the presence of complementarity, since the return of two high types working together is 1.5 standard deviations above the return of two low or middle types working together. This suggests gains from complementarity in economic research, which are reflected in stronger sorting at the top. In addition, the results show that sole-authored productivity and productivity in teams are not one-to-one: while the lowest type produces slightly lower quality output than the second-lower type on her own, the lowest type benefits substantially more from working with others, in particular with a high-type co-author.

In addition to documenting the patterns of heterogeneity, sorting and complementarity, the nonlinear model can be used to refine the variance decomposition in (9). Indeed, the

²⁶I declare convergence when the increment in evidence lower bound is less than 10^{-3} . To alleviate local maxima issues, I started the variational EM algorithm from different sets of parameters. The values that I obtained upon reaching the tolerance threshold differed somewhat, however they had similar implications for heterogeneity, sorting and complementarity.

Table 3: Nonlinear production, economic researchers

	$K = 2$		$K = 4$		$K = 5$		$K = 6$	
	$n = 1$	$n = 2$	$n = 1$	$n = 2$	$n = 1$	$n = 2$	$n = 1$	$n = 2$
Total variance	97.81	162.64	97.81	162.64	97.81	162.64	97.81	162.64
Heterogeneity	2.18	13.84	35.71	46.85	42.77	51.80	35.52	49.86
Sorting	-	1.65	-	16.87	-	19.66	-	17.09
Nonlinearities	-	1.36	-	3.92	-	2.80	-	3.48
Other factors	95.63	145.79	62.11	90.00	55.05	88.39	62.29	92.22

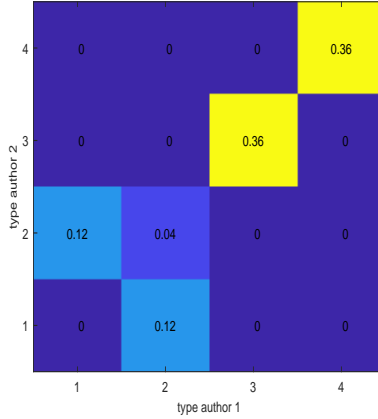
Notes: Estimates of variance components for different team sizes n in the nonlinear model with K types. Descriptive statistics on the sample are given in panel (b) of Table 1.

nonlinear model accounts for interaction effects between co-authors, over and beyond the additive effects of author types. This gives a fourth variance component, which I will refer to as “nonlinearities”, to report in the decomposition, in addition to the contributions of heterogeneity, sorting, and other factors. This component can be computed as the difference between the variance of the mean of Y_{nj} given the worker types – which depends on both additive and interactive terms in general – and the variance of its best linear approximation.

I report the results of the variance decomposition in Table 3, for $K = 2, 4, 5, 6$. While $K = 2$ seems to allow for too little heterogeneity, the results for $K = 4, 5, 6$ are broadly similar to one another. They show that author heterogeneity explains approximately 30% of output variance, and sorting explains between 10% and 12%. Despite different type modeling and form of the production function, these variance shares are not too different from those based on the additive model (see Table 2). In addition, nonlinearities explain between 1.5% and 2.5% of the variance. While this share may seem small, the presence of complementarity at the top of the productivity distribution, as shown by panel (b) of Figure 3, is likely to be important to study policies that re-allocate researchers across teams.

To assess how complementarities might affect worker allocations, I compute an allocation that maximizes total output by solving (17). I compute expected payoffs and “budgets” based on the estimates of the nonlinear model with $K = 4$. In this case, when budget sizes are even, an optimal integer-valued allocation can be computed by solving a linear program that does not impose the integral constraints. The resulting allocation in Figure 4 shows sorting at the top of the type distribution, but not at the bottom.

Figure 4: Optimal allocation, economic researchers ($K = 4$)

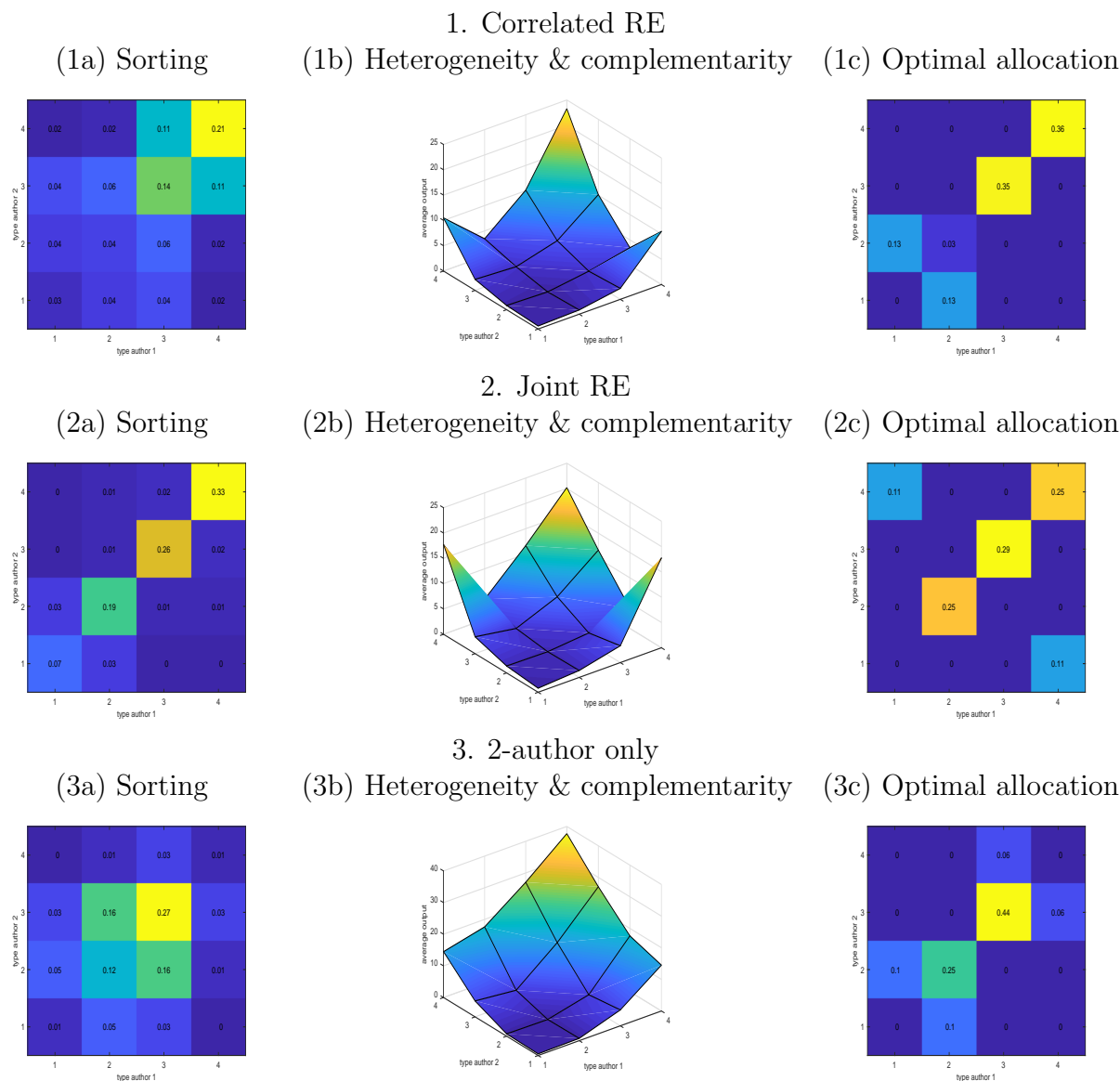


Notes: Proportions of types for authors producing together in a 2-author team, in the allocation that solves (17). Descriptive statistics on the sample are given in panel (b) of Table 1.

Other specifications. I now present the results of three exercises based on other specifications of the nonlinear model. In the baseline specification, the distribution of author types does not depend on collaborations. To assess the impact of this assumption, I estimate two alternative specifications. In the first specification, I model the type distribution as a function of author characteristics, using a multinomial logit specification. As characteristics, I use the numbers of teams in which author i participates, separately for 1-author and 2-author teams, as well as the indicators that these numbers are zero. In panels (1a) and (1b) of Figure 5, I show type-pair proportions and mean output by pairs of types, and in panel (1c) I show the optimal allocation obtained by solving (17). In Appendix Table C3 I show the variance decomposition estimates. The results are similar to the ones from the baseline specification.

In the second specification, I augment the team production model with a model of team formation, and I jointly estimate the parameters using a variational EM algorithm. I specify team formation using a stochastic blockmodel for hypergraphs, where 1-author and 2-author collaborations follow independent Poisson distributions, with a parameter that is type-specific in the first case and pair-of-types-specific in the second case. In this model, as in most models of team formation, the conditional distribution of $\{\alpha_i\}$ given $\{(i_1(j), \dots, i_n(j))\}$ does not factor across i . In addition, in such a model, author types do not only influence heterogeneity in research quality, but also in quantity. The results show some differences

Figure 5: Nonlinear model estimates and optimal allocation, economic researchers, other specifications ($K = 4$)



Notes: Random-effects estimates of a finite mixture model with $K = 4$ types. In panels (a) I show the proportions of types for authors producing together in a 2-author team. In panels (b) I show average output (i.e., average journal quality) for different combinations of the types. In panels (c) I show the proportions of types for authors producing together in a 2-author team, in the allocation that solves (17). “Correlated RE” comes from a model where the types follow a multinomial logit distribution given the numbers of 1-author and 2-author collaborations at the extensive and intensive margins. “Joint RE” comes from a model where collaborations follow independent type-specific Poisson distributions. “2-author only” only uses information from 2-author teams. Descriptive statistics on the sample are given in panel (b) of Table 1.

with the baseline. In particular, panel (2a) of Figure 5 shows stronger evidence of sorting. In panel (2b), I show mean output in 2-author teams, by type pairs. The estimates, and the implied optimal allocation in panel (2c), are again consistent with complementarity between high-type authors.²⁷ In Appendix Table C3 I report variance decomposition estimates. In 2-author teams, heterogeneity accounts for a lower share of variance compared to the baseline (22% versus 29%), sorting accounts for a larger share (19% versus 10%), and the variance share accounted for by nonlinearities is smaller than in the baseline.

In a last exercise, I report estimates based on 2-author teams only. Hence, sole-authored productions are discarded. Panel (3a) of Figure 5 shows a higher degree of sorting in the middle of the type distribution, rather than at the top. Panel (3b) again shows evidence of complementarity, although the patterns differ somewhat from the baseline, as shown by the implied optimal allocation in panel (3c). Lastly, the variance decomposition results in Appendix Table C3 show that, in this case, heterogeneity accounts for 32% of variance, and sorting accounts for 5% (as opposed to 29% and 10% in the baseline).

6 Inventors and innovation

In this section I apply the method to study patents and inventors.

6.1 Data

I use data from Akcigit *et al.* (2016). Their main source is the disambiguated inventor data of Li *et al.* (2014), which identifies unique inventors in the USPTO data. I restrict the analysis to US patents, which were granted between 1995 and 1999. Throughout, I focus on patents in the technology class “Computers and Communications”.²⁸ For the output measure, I follow Akcigit *et al.* (2016) and use Hall *et al.*’s (2001) measure of patent quality, which is a truncation-adjusted measure of forward citations of the patent. From this measure I net out multiplicative year fixed-effects, using 1999 as the reference year. In this illustration, I do not use information about the firms where inventors work, although it would be interesting to incorporate this information in future research.

²⁷However, note that the high degree of sorting prevents one from assessing with confidence the gains from collaboration between highest and lowest types (i.e., between type-1 and type-4 authors): while the point-estimates in panel (2b) are large, panel (2a) shows that such collaborations are virtually non-existent. This is likely to affect the optimal allocation numbers shown in panel (2c) of Figure 5.

²⁸Among all inventors who patent in this class, 80% patent only in this class during the period.

Table 4: Descriptive statistics on patents and inventors

	(a) Sample				(b) Subsample		
	All	$n = 1$	$n = 2$	$n \geq 3$	All	$n = 1$	$n = 2$
# Inventors	5896	4328	4267	3814	1871	1719	1011
# Patents	30068	17261	7740	5067	12793	10520	2273
Mean output	33.96	31.36	33.03	37.06	29.67	29.26	30.69
Std output	40.34	38.98	38.97	42.40	35.41	35.55	35.09
p10% output	4.43	4.23	4.43	5.46	3.96	3.96	4.23
p50% output	21.72	19.02	21.13	24.23	19.02	18.46	19.34
p90% output	76.08	69.74	73.97	81.76	65.52	65.47	65.94
p95% output	105.67	100.05	103.56	111.13	92.99	92.25	95.10
p99% output	192.83	189.46	196.47	192.83	175.42	175.41	175.41
p10% #patents	5	2	2	3	5	4	2
p50% #patents	9	5	5	6	8	7	6
p90% #patents	25	13	11	16	21	16	15
p95% #patents	32	19	16	22	29	26	17
p99% #patents	68	37	26	42	62	44	27

Notes: Sample from Akcigit et al. (2016), where the patent is from the US and granted between 1995 and 1999, and belongs to the class “Computers and Communications”. Output is the truncation-adjusted measure of forward citations of Hall et al. (2001), net of multiplicative year effects (reference year is 1999). Percentiles of the distribution of number of patents per inventor are indicated in the bottom five rows. In panel (a) the sample is restricted to at least 5 patents per inventor during the period. In panel (b) the subsample is restricted to collaborations between at most two inventors, where all inventors have at least 5 patents during 1995-1999. I use the subsample in panel (b) to estimate nonlinear models.

In panel (a) of Table 4, I provide summary statistics about the sample, where I restrict inventors to participate in at least five patents during the period. In this setting also, output and the number of collaborations are right-skewed. The original sample of US patents in the class “Computers and Communications” that were granted between 1995 and 1999 contains 65848 inventors and 62927 patents. Hence, imposing that all inventors be on at least 5 patents restricts the sample size substantially. In the sample, team size exhibits a range of variation but teams tend to be small: 57% of teams are 1-inventor teams, 26% of teams have 2 members, and less than 1% of teams have more than 6 members. In panel (b) of Table 4, I show summary statistics for the subsample of 1-inventor and 2-inventor teams that I will use to estimate the nonlinear models.

Table 5: Additive production, patents and inventors

	$n = 1$	$n = 2$	$n = 3$
Total variance	1519.3	1450.0	1388.5
Heterogeneity	578.2	467.6	261.3
Heterogeneity (uncorrected)	782.1	710.7	559.3
Sorting	-	90.2	169.6
Sorting (uncorrected)	-	-25.9	-12.0
Other factors	941.1	733.1	712.8
Other factors (uncorrected)	737.2	727.3	792.0
Team scale λ_n	1.00	0.54	0.39

Notes: Estimates of variance components in equation (9), for different team sizes n . Descriptive statistics on the sample are given in panel (a) of Table 4.

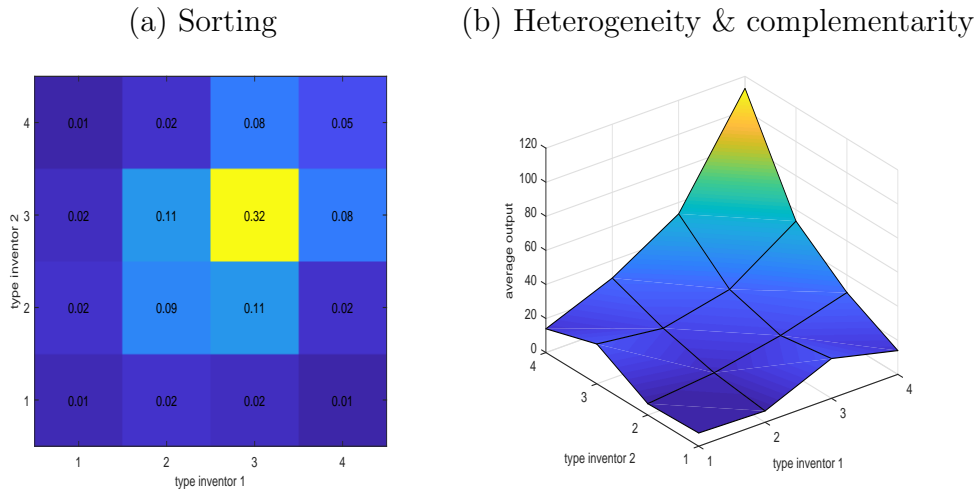
6.2 Additive model

I first estimate the additive model (2). Restricting the set of inventors and patents to ensure identification gives 5547 inventors and 29101 patents. I estimate the model by allowing for four team sizes n in λ_n , where $n = 4$ corresponds to all teams with at least four inventors. I find $\lambda_2 = 0.54$, $\lambda_3 = 0.39$, and $\lambda_4 = 0.29$. Hence, keeping inventor type(s) constant, patents with two inventors are cited 8% more than patents with one inventor, 3-inventor patents are cited 17% more, and patents with at least 4 inventors are cited 16% more.

In Table 5, I show the estimates of the variance components in (9), for $n = 1, 2, 3$. The variance share explained by inventor heterogeneity is 38% in patents with one inventor, 32% in 2-inventor patents, and 19% in 3-inventor patents. Sorting contributes 6% of variance in 2-inventor patents, and 12% in 3-inventor patents. Here also, the other factors ε_{nj} account for the main share of the output variance. Overall, the variance decomposition estimates in the patent sample are thus not very different from those in the sample of economists, with a somewhat larger contribution of worker heterogeneity and a smaller contribution of sorting.

In Appendix Table C4, I augment the sample and require every inventor to produce at least one or two patents, as opposed to at least five in the baseline sample. In addition, as in all the other robustness checks in this application, using a Poisson regression I net out from the output the effects of year and “inventor age”, as measured by the difference between the year of observation and the first year where the inventor produced a patent. Compared to the baseline estimates, the results in Appendix Table C4 show that, in these larger samples, inventor fixed-effects are estimated with more noise, and uncorrected variance components

Figure 6: Nonlinear model estimates, patents and inventors ($K = 4$)



Notes: Random-effects estimates of a finite mixture model with $K = 4$ types. In panel (a) I show the proportions of types for inventors producing together in a 2-inventor team. In panel (b) I show average output (i.e., average truncation-adjusted forward citations) for different combinations of the types. Descriptive statistics on the sample are given in panel (b) of Table 4.

are very large in magnitude. The bias-corrected estimates indicate a larger role of inventor heterogeneity than in the baseline, and a negative sorting contribution, although, given the amount of noise, these findings should be interpreted with caution.

6.3 Nonlinear model

To specify the nonlinear model,²⁹ I model the output distribution as a negative binomial with mean and variance that depend on the types of inventors in the team. This parametric form allows for a convenient treatment of the zeros in the dependent variable. I now comment in detail on the results for $K = 4$. In this case, the type proportions are 6%, 25%, 52%, and 17%. The means of the negative binomial output distribution are, in 1-inventor teams:

$$\begin{pmatrix} 9.80 \\ 12.31 \\ 27.03 \\ 65.08 \end{pmatrix},$$

²⁹In Appendix Figure C1, I construct type proxies as in Figure 2. I group inventors according to the quartiles of the citations of their sole-authored patents, and restrict the sample to inventors with at least 5 patents on their own, so the sample is small, with 1554 inventors. The figure suggests the presence of heterogeneity and sorting, although complementarity is less salient than for economists, and sorting seems more evenly spread out along the diagonal (compare with Figure 2).

Table 6: Nonlinear production, patents and inventors

	$K = 2$		$K = 4$		$K = 5$		$K = 6$	
	$n = 1$	$n = 2$	$n = 1$	$n = 2$	$n = 1$	$n = 2$	$n = 1$	$n = 2$
Total variance	1263.5	1231.7	1263.5	1231.7	1263.5	1231.7	1263.5	1231.7
Heterogeneity	237.2	302.9	324.6	366.9	375.2	412.8	390.1	384.4
Sorting	-	46.6	-	80.1	-	72.7	-	109.9
Nonlinearities	-	42.2	-	60.9	-	96.7	-	70.2
Other factors	1026.4	840.0	938.9	723.7	888.3	649.5	873.4	667.1

Notes: Estimates of variance components for different team sizes n in the nonlinear model with K types. Descriptive statistics on the sample are given in panel (b) of Table 4.

and in 2-inventor teams:

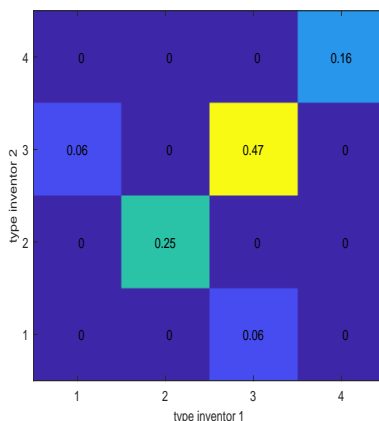
$$\begin{pmatrix} 7.76 & 6.65 & 23.36 & 13.97 \\ 6.65 & 10.71 & 18.56 & 29.70 \\ 23.36 & 18.56 & 27.30 & 53.32 \\ 13.97 & 29.70 & 53.32 & 112.99 \end{pmatrix}.$$

In Figure 6, I report similar quantities as in Figure 3 to illustrate sorting, heterogeneity, and complementarity in the sample of inventors. Panel (a) shows some evidence of sorting towards the top, however the pattern is less concentrated than in the sample of economists (compare with panel (a) of Figure 3). Panel (b) of Figure 6 suggests the presence of complementarity, since the return to two high-type inventors working together is 2 standard deviations above the return to two low or middle types working together. This suggests gains from complementarity in patent production, with somewhat less sorting than in the case of economists.

In Table 6, I show variance decomposition results based on the nonlinear model, for $K = 2, 4, 5, 6$. Focusing on the results for $K = 4, 5, 6$, which are broadly comparable, inventor heterogeneity explains approximately 30% of output variance, and sorting explains between 6% and 9%. Nonlinearities explain a larger share than in the economists' sample, between 5% and 8% of the variance.

In addition, in Figure 7 I report the allocation that maximizes total output according to (17). As in the case of economists, the allocation has perfect assortative matching for the top type but not for the bottom types. This optimal allocation differs quite substantially from the estimated allocation shown in panel (a) of Figure 6.

Figure 7: Optimal allocation, patents and inventors ($K = 4$)



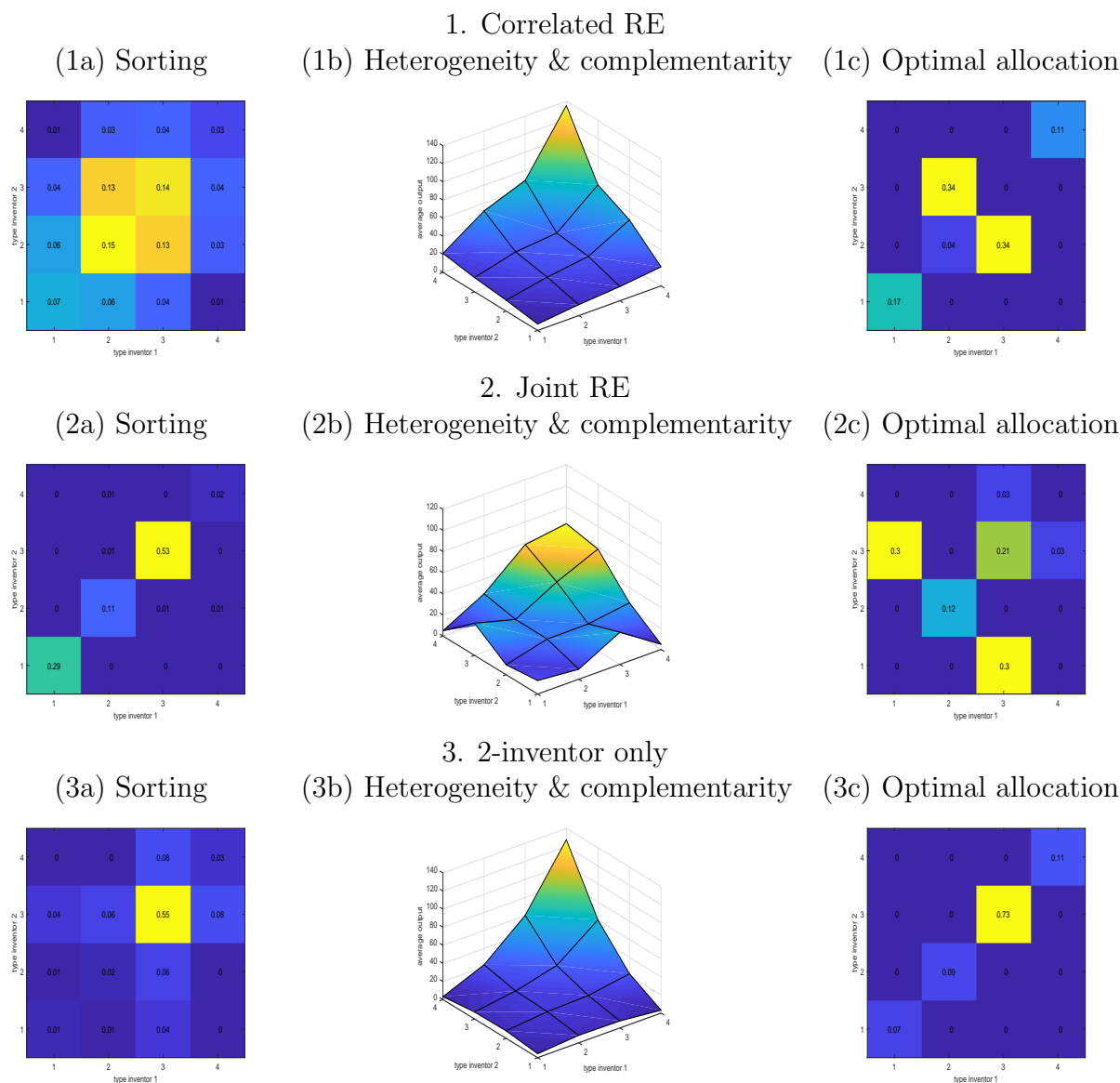
Notes: Proportions of types for inventors producing together in a 2-inventor team, in the allocation that solves (17). Descriptive statistics on the sample are given in panel (b) of Table 4.

Other specifications. In Figure 8, I show estimates based on three alternative specifications: a correlated random-effects estimator, a joint random-effects estimator where 1-inventor and 2-inventor collaborations follow independent Poisson distributions with type-specific parameters, and an estimator that is solely based on 2-inventor collaborations. In these specifications I net out multiplicative inventor-age effects from the output, in addition to year effects. I show the variance decomposition results in Appendix Table C5. As in the case of economists, I find that the correlated RE estimates do not substantially differ from the baseline. The optimal allocation in panel (1c) of Figure 8 does differ somewhat from the baseline (compare with Figure 7), yet both allocations exhibit perfect assortative matching at the top.

In addition, the estimates based on 2-inventor teams are also broadly similar to the ones based on both 1-inventor and 2-inventor teams in this sample. One difference is that the group shares are more unbalanced when using only 2-inventor collaborations. Another difference is that the implied optimal allocation in panel (3c) of Figure 8 exhibits perfect assortative matching along the entire distribution, and not only at the top.

However, the joint RE estimates show important differences compared to the baseline. Note that, as in the corresponding specification in the economists' sample, here inventor types do not only reflect heterogeneity in the quality of innovation through patent citations, but also in the quantity of patents produced by an inventor. In 2-inventor teams, heterogeneity,

Figure 8: Nonlinear model estimates and optimal allocation, patents and inventors, other specifications ($K = 4$)



Notes: Random-effects estimates of a finite mixture model with $K = 4$ types. In panels (a) I show the proportions of types for inventors producing together in a 2-inventor team. In panels (b) I show average output (i.e., average truncation-adjusted forward citations) for different combinations of the types. In panels (c) I show the proportions of types for inventors producing together in a 2-inventor team, in the allocation that solves (17). “Correlated RE” comes from a model where the types follow a multinomial logit distribution given the numbers of 1-inventor and 2-inventor collaborations at the extensive and intensive margins. “Joint RE” comes from a model where collaborations follow independent type-specific Poisson distributions. “2-inventor only” only uses information from 2-inventor teams. Descriptive statistics on the sample are given in panel (b) of Table 4. For these results I net out multiplicative year and inventor-age effects from the output.

sorting and nonlinearity account for 9%, 8%, and less than 1% of the variance of output, respectively, compared to 30%, 7%, and 5% in the baseline (see Appendix Table C5). While the estimates in panels (2a) and (2b) of Figure 8 show a high degree of sorting, they exhibit less evidence of complementarity compared to the other specifications. The implied optimal allocation in panel (2c) is also quite different in this case. This suggests that the modeling of inventor types and how types affect team formation is important to accurately assess the contributions of heterogeneity, sorting and complementarity in this sample.³⁰

7 Conclusion

In this chapter I have outlined a measurement framework to assess the contributions of individuals to team output. While an additive production specification leads to a tractable estimator, it rules out complementarity that appears to be a feature of the samples of economists and inventors that I study. In the applications, a natural next step will be to relate the latent types to characteristics of authors and inventors, such as their national origin or education background.

The discrete-type approach that I have implemented is promising, yet it needs to be studied more. In particular, it will be important to provide formal consistency arguments and to derive asymptotic distributions for variational estimators in this setting. Spectral clustering methods (e.g., Lei and Rinaldo, 2015) could be possible alternatives to the non-linear random-effects methods that I have described. It will also be important to augment the framework to account for the productive effects of team-specific factors, both observed and latent.

The literature on matching and sorting has made substantial progress in one-to-one settings (Chiappori and Salanié, 2016, Chade *et al.*, 2017). However, less is known about many-to-one and many-to-many sorting environments. Eeckhout and Kircher (2018) model sorting in large firms, while abstracting from complementarities between workers. Chade and Eeckhout (2018) propose a model of information aggregation in teams that has tight implications for sorting.³¹ Moreover, it is well-known that models with general complemen-

³⁰In the appendix, I probe the predictive performance of inventor types out of sample, using the period 2000-2005. I find that sorting patterns are stable, and that out-of-sample estimates are consistent with the persistence of inventor heterogeneity out of sample. However, the effect of worker types on output is less strong than in the period previous to 1999.

³¹A related strand of the literature proposes models of co-authorship networks, see among others Goyal *et al.* (2004) and Gans and Murray (2014), and the co-author model in Jackson and Wolinsky (1996).

tarities may not have equilibria (Kelso and Crawford, 1982). Addressing these challenges, and combining the quantitative framework of team production that I have introduced, with economic models of team formation and effort allocation (such as those recently proposed by Hsieh *et al.*, 2020, and Anderson and Richards-Shubik, 2019), is an interesting avenue for future work.

References

- [1] Abowd, J. M., and F. Kramarz (1999): “The Analysis of Labor Markets Using Matched Employer-Employee Data,” *Handbook of Labor Economics*, 3, 2629–2710.
- [2] Abowd, J., F. Kramarz, and D. Margolis (1999): “High Wage Workers and High Wage Firms”, *Econometrica*, 67(2), 251–333.
- [3] Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008): “Mixed Membership Stochastic Blockmodels,” *Journal of Machine Learning Research*, 9(Sep), 1981–2014.
- [4] Akcigit, U., S. Baslandze, and S. Stantcheva (2016): “Taxation and the International Mobility of Inventors,” *American Economic Review*, 106(10), 2930–81.
- [5] Akcigit, U., J. Grigsby, and T. Nicholas (2017): “The Rise of American Ingenuity: Innovation and Inventors of the Golden Age,” (No. w23047). National Bureau of Economic Research.
- [6] Ahmadpoor, M., and B. F. Jones (2019): “Decoding Team and Individual Impact in Science and Invention,” *Proceedings of the National Academy of Sciences*, 116(28), 13885–13890.
- [7] Allman, E. S., C. Matias, and J. A. Rhodes (2009): “Identifiability of Parameters in Latent Structure Models with Many Observed Variables,” *Annals of Statistics*, 3099–3132.
- [8] Allman, E. S., C. Matias, and J. A. Rhodes (2011): “Parameter Identifiability in a Class of Random Graph Mixture Models,” *Journal of Statistical Planning and Inference*, 141(5), 1719–1736.
- [9] Anderson, K. A., and S. Richards-Shubik (2019): “Collaborative Production in Science: An Empirical Analysis of Coauthorships in Economics,” CMU and Lehigh University Working Paper.
- [10] Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008): “High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?” *Journal of the Royal Statistical Society: Series A*, 171(3), 673–697.
- [11] Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012): “Estimating Spillovers Using Panel Data, with an Application to the Classroom,” *Quantitative Economics*, 3(3), 421–470.
- [12] Arcidiacono, P., J. Kinsler, and J. Price (2017): “Productivity Spillovers in Team Production: Evidence from Professional Basketball,” *Journal of Labor Economics*, 35(1), 191–225.
- [13] Arellano, M., and S. Bonhomme (2009): “Robust Priors in Nonlinear Panel Data Models”, *Econometrica*, 77, 489–536.
- [14] Becker, G. S. (1973): “A Theory of Marriage: Part I,” *Journal of Political Economy*, 81(4), 813–846.
- [15] Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019): “Who Becomes an Inventor in America? The Importance of Exposure to Innovation,” *The Quarterly Journal of Economics*, 134(2), 647–713.
- [16] Bertrand, M., and A. Schoar (2003): “Managing with Style: The Effect of Managers on Firm Policies,” *The Quarterly Journal of Economics*, 118(4), 1169–1208.

- [17] Bickel, P., D. Choi, X. Chang, and H. Zhang (2013): “Asymptotic Normality of Maximum Likelihood and its Variational Approximation for Stochastic Blockmodels,” *Annals of Statistics*, 41(4), 1922–1943.
- [18] Bishop, C. M. (2006): *Pattern Recognition and Machine Learning*. springer.
- [19] Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017): “Variational Inference: A Review for Statisticians,” *Journal of the American statistical Association*, 112(518), 859–877.
- [20] Bonhomme, S. (2017): “Econometric Analysis of Bipartite Networks,” to appear in *The Analysis of Network Data*, B. Graham and A. de Paula (eds).
- [21] Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2020): “How Much Should we Trust Estimates of Firm Effects and Worker Sorting?” (No. w27368). National Bureau of Economic Research.
- [22] Bonhomme, S., K. Jochmans, and J.M. Robin (2016): “Nonparametric Estimation of Finite Mixtures from Repeated Measurements,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1), 211–229.
- [23] Bonhomme, S., T. Lamadon, and E. Manresa (2019): “A Distributional Framework for Matched Employer-Employee Data,” *Econometrica*, 87(3), 699-739.
- [24] Bonhomme, S., T. Lamadon, and E. Manresa (2021): “Discretizing Unobserved Heterogeneity,” unpublished manuscript.
- [25] Bonhomme, S., and E. Manresa (2015): “Grouped Patterns of Heterogeneity in Panel Data,” *Econometrica*, 83(3), 1147–1184.
- [26] Bramoullé, Y., H. Djebbari, and B. Fortin (2009): “Identification of Peer Effects through Social Networks,” *Journal of Econometrics*, 150(1), 41–55.
- [27] Carlier, G., and I. Ekeland (2010): “Matching for Teams,” *Economic Theory*, 42(2), 397–418.
- [28] Celisse, A., J. J. Daudin, and L. Pierre (2012): “Consistency of Maximum-Likelihood and Variational Estimators in the Stochastic Block Model,” *Electronic Journal of Statistics*, 6, 1847–1899.
- [29] Chade, H., and J. Eeckhout (2018): “Matching Information,” *Theoretical Economics*, 13(1), 377–414.
- [30] Chade, H., J. Eeckhout, and L. Smith (2017): “Sorting Through Search and Matching Models in Economics,” *Journal of Economic Literature*, 55(2), 493–544.
- [31] Chamberlain, G. (1984): “Panel data”, in *Handbook of Econometrics*, 2, 1247–1318.
- [32] Chamberlain, G. (1992): “Efficiency Bounds for Semiparametric Regression”, *Econometrica*, 60, 567–596.
- [33] Chiappori, P. A., and B. Salanié (2016): “The Econometrics of Matching Models,” *Journal of Economic Literature*, 54(3), 832–61.
- [34] Chiappori, P. A., B. Salanié, and A. Galichon (2019): “On Human Capital and Team Stability,” *Journal of Human Capital*, 13(2), 236–259.

- [35] Daudin, J. J., F. Picard, and S. Robin (2008): “A Mixture Model for Random Graphs,” *Statistics and Computing*, 18(2), 173–183.
- [36] Devereux, K. (2018): “Identifying the Value of Teamwork: Application to Professional Tennis,” No. 14, Working Paper Series.
- [37] Ductor, L., M. Fafchamps, S. Goyal, and M. J. van der Leij (2014): “Social Networks and Research Output,” *Review of Economics and Statistics*, 96(5), 936–948.
- [38] Eeckhout, J., and P. Kircher (2011): “Identifying Sorting – In Theory,” *The Review of Economic Studies*, 78(3), 872–906.
- [39] Eeckhout, J., and P. Kircher (2018): “Assortative Matching with Large Firms,” *Econometrica*, 86(1), 85–132.
- [40] Fafchamps, M., M. J. Van der Leij, and S. Goyal (2010): “Matching and Network Effects,” *Journal of the European Economic Association*, 8(1), 203–231.
- [41] Furman, J. L., and P. Gaule (2013): “A Review of Economic Perspectives on Collaboration in Science,” In *Workshop on Institutional & Organizational Supports for Team Science*.
- [42] Gans, J., and F. Murray (2014): “Markets for Scientific Attribution,” (No. w20677). National Bureau of Economic Research.
- [43] Garicano, L., and E. Rossi-Hansberg (2006): “Organization and Inequality in a Knowledge Economy,” *The Quarterly Journal of Economics*, 121(4), 1383–1435.
- [44] Gaure, S. (2014): “Correlation Bias Correction in Two-Way Fixed-Effects Linear Regression,” *Stat*, 3, 379–390.
- [45] Goyal, S., M. J. Van Der Leij, and J. L. Moraga-González (2004): “Economics: An Emerging Small World,” Tinbergen Institute Discussion Paper, No. 04-001/1.
- [46] Goyal, S., M. J. Van Der Leij, and J. L. Moraga-González (2006): “Economics: An Emerging Small World,” *Journal of Political Economy*, 114(2), 403–412.
- [47] Hahn, J., and H. Moon (2010): “Panel Data Models with Finite Number of Multiple Equilibria,” *Econometric Theory*, 26(3), 863–881.
- [48] Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001): “The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools,” (No. w8498). National Bureau of Economic Research.
- [49] Hall, P., and X. H. Zhou (2003): “Nonparametric Estimation of Component Distributions in a Multivariate Mixture,” *Annals of Statistics*, 201–224.
- [50] Herkenhoff, K., J. Lise, G. Menzio, and G. M. Phillips (2018): “Production and Learning in Teams,” (No. w25179). National Bureau of Economic Research.
- [51] Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988): “Estimating Vector Autoregressions with Panel Data,” *Econometrica*, 1371–1395.
- [52] Hsieh, C. S., M. D. Konig, X. Liu, and C. Zimmermann (2020): “Collaboration in Bipartite Networks,” unpublished manuscript.

- [53] Hu, Y. (2008): “Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution,” *Journal of Econometrics*, 144(1), 27–61.
- [54] Hvattum, L. M. (2019): “A Comprehensive Review of Plus-Minus Ratings for Evaluating Individual Players in Team Sports,” *International Journal of Computer Science in Sport*, 18(1), 1–23.
- [55] Jackson, M. O., and A. Wolinsky (1996): “A Strategic Model of Social and Economic Networks,” *Journal of Economic Theory*, 71(1), 44–74.
- [56] Jaravel, X., N. Petkova, and A. Bell (2018): “Team-Specific Capital and Innovation,” *American Economic Review*, 108(4-5), 1034–73.
- [57] Jochmans, K., and M. Weidner (2020): “Fixed-Effect Regressions on Network Data,” to appear in *Econometrica*.
- [58] Kelso Jr, A. S., and V. P. Crawford (1982): “Job Matching, Coalition Formation, and Gross Substitutes,” *Econometrica*, 1483–1504.
- [59] Kline, P., R. Saggio, and M. Solvsten (2020): “Leave-out Estimation of Variance Components,” to appear in *Econometrica*.
- [60] Kodrzycki, Y. K., and P. Yu (2006): “New Approaches to Ranking Economics Journals,” *The BE Journal of Economic Analysis & Policy*, 5(1).
- [61] Lei, J., and A. Rinaldo (2015): “Consistency of Spectral Clustering in Stochastic Block Models,” *The Annals of Statistics*, 43(1), 215–237.
- [62] Lerner, J., M. Tranmer, J. Mowbray, and M. G. Hancean (2019): “REM Beyond Dyads: Relational Hyperevent Models for Multi-Actor Interaction Networks,” arXiv preprint arXiv:1912.07403.
- [63] Li, G. C., R. Lai, A. D’Amour, D. M. Doolin, Y. Sun, V. I. Torvik, A. Z. Yu, and L. Fleming (2014): “Disambiguation and Co-Authorship Networks of the US Patent Inventor Database (1975-2010),” *Research Policy*, 43(6), 941–955.
- [64] Mariadassou, M., S. Robin, and C. Vacher (2010): “Uncovering Latent Structure in Valued Graphs: a Variational Approach,” *The Annals of Applied Statistics*, 4(2), 715–742.
- [65] Pearce, J. (2019): “Firm-Inventor Links and the Composition of Innovation,” unpublished manuscript.
- [66] Snijders, T. A., and K. Nowicki (1997): “Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure,” *Journal of Classification*, 14(1), 75–100.
- [67] Talman, D., and Z. Yang (2011): “A Model of Partnership Formation,” *Journal of Mathematical Economics*, 47(2), 206–212.
- [68] Turnbull, K., S. Lunagómez, C. Nemeth, and E. Airolidi (2019): “Latent Space Representations of Hypergraphs,” arXiv preprint arXiv:1909.00472.
- [69] Weidmann, B., and D. J. Deming (2020): “Team Players: How Social Skills Improve Group Performance” (No. w27071), National Bureau of Economic Research.
- [70] Woodcock, S. D. (2008): “Wage Differentials in the Presence of Unobserved Worker, Firm, and Match Heterogeneity,” *Labour Economics*, 15(4), 771–793.

APPENDIX

A Monte Carlo simulations

To probe the accuracy of the variational estimator in finite sample, I run Monte Carlo simulations. I generate models based on two sets of collaborations, both drawn from the data on economic researchers who work either on their own or with one co-author (see Section 5). For the smaller sample I select articles that were published in 1999. Since I impose that all authors have at least 5 articles, this gives a small number of authors and teams (about 150 and 900, respectively). For the larger sample I select articles that were published between 1998 and 1999. This gives a sample of about 900 authors and 5000 articles.

I first consider a log-normal model with $K_0 = 2$ types, and use $K = 2$ types in estimation. The true parameters, as well as means and 2.5% and 97.5% quantiles of the variational estimator across 500 simulations, are given in panel (a) of Table A1. The results suggest that both the parameters of sole-authored productions and the type proportions are well recovered in both samples. However, the parameters of 2-author teams are not as well recovered in the smaller sample. In particular, while the true mean output for two high-type workers is 4, the mean Monte Carlo estimate is 3.71. This suggests that the variational approximation, which only matters for 2-author collaborations, is imperfectly accurate in this case. In contrast, estimates in the larger sample are close to unbiased. For example, the mean output for two high-type workers is indistinguishable from the truth. In addition, Monte Carlo estimates are quite tightly concentrated around true values. This suggests that the variational approximation is accurate in the larger sample.

I next consider a log-normal model with $K_0 = 4$ types, and use $K = 4$ types in estimation. In this case, I set the true parameter values to the parameters of the model estimated on the 1995-2000 sample; see the column labeled “True” in panel (b) of Table A1. While this second simulation design is closer to the data, it is also more challenging for the variational estimator. Panel (b) of Table A1 shows that, in the smaller sample, both the parameters of sole-authored productions and those of 2-author teams are biased and imprecise. In the larger sample, biases are quite small in general, and estimates are more precise. In the empirical analysis in Section 5, I focus on a much larger sample of articles produced between 1995 and 1999. Since I also impose that every author produces at least 5 articles, I expect the variational method to be accurate in this case. Developing methods to assess parameter uncertainty in the variational approach is an important avenue for future work.

B How do inventor types perform out of sample?

I study the predictive performance of inventor types out of sample. To do so, I use the parameters of the baseline random-effects model with $K = 4$ types, estimated on the 1995-1999 period. Using the estimated variational posterior type probabilities, I then calculate average patent output produced by inventors of particular types between 2000 and 2005, and the type proportions in those collaborations. Figure C2 shows that sorting patterns and productivity differences between types persist out of sample, and that the higher returns to collaborations between high inventor types relative to other type combinations persist as well. However, the out-of-sample estimates show less inventor heterogeneity than in the baseline (compare with panel (b) of Figure 6). There may be several explanations for this: posterior type probabilities are only based on five years of observations, truncation issues with patent citations may make the outcome less informative towards the end of the sample (Akcigit *et al.*, 2016), and over a 10-year period inventor quality may change.

Table A1: Monte Carlo simulation

	True	Smaller sample			Larger sample		
		Mean	p2.5%	p97.5%	Mean	p2.5%	p97.5%
(a) $K_0 = 2, K = 2$							
Mean type 1	0.00	0.00	-0.05	0.06	0.00	-0.03	0.02
Mean type 2	2.00	2.00	1.92	2.07	2.00	1.97	2.03
Var. type 1	0.50	0.50	0.43	0.56	0.50	0.47	0.53
Var. type 2	0.50	0.50	0.42	0.57	0.50	0.47	0.53
Mean type (1,1)	0.00	0.05	-0.49	0.74	0.01	-0.10	0.12
Mean type (1,2)	1.00	1.01	0.59	1.42	1.01	0.92	1.10
Mean type (2,2)	4.00	3.71	0.78	4.72	4.00	3.85	4.16
Var. type (1,1)	0.50	0.46	0.11	0.93	0.50	0.40	0.62
Var. type (1,2)	0.50	0.46	0.16	0.84	0.50	0.42	0.59
Var. type (2,2)	0.50	0.38	0.00	1.24	0.49	0.35	0.69
Prop. type 1	0.60	0.60	0.53	0.67	0.60	0.57	0.63
(b) $K_0 = 4, K = 4$							
Mean type 1	-0.52	-0.52	-0.54	-0.50	-0.52	-0.53	-0.52
Mean type 2	-0.47	-0.42	-0.56	-0.22	-0.47	-0.52	-0.41
Mean type 3	0.07	0.21	-0.09	0.71	0.07	-0.03	0.18
Mean type 4	1.62	1.68	1.39	2.01	1.64	1.53	1.75
Var. type 1	0.01	0.02	0.01	0.08	0.01	0.01	0.01
Var. type 2	0.48	0.64	0.38	1.15	0.48	0.41	0.55
Var. type 3	1.74	1.88	1.42	2.44	1.74	1.59	1.89
Var. type 4	2.16	2.10	1.65	2.59	2.15	1.96	2.31
Mean type (1,1)	-0.53	-0.03	-1.48	2.90	-0.52	-0.60	-0.34
Mean type (1,2)	-0.39	-0.28	-1.83	1.58	-0.38	-0.73	0.06
Mean type (2,2)	-0.53	-0.01	-1.73	2.17	-0.50	-0.63	-0.08
Mean type (1,3)	0.01	0.26	-2.13	2.94	0.02	-0.43	0.52
Mean type (2,3)	-0.11	0.25	-1.50	2.25	-0.07	-0.42	0.26
Mean type (3,3)	0.35	0.83	-1.28	3.23	0.42	-0.13	0.95
Mean type (1,4)	1.38	1.07	-0.68	3.34	1.37	0.75	1.95
Mean type (2,4)	0.66	0.94	-1.01	3.01	0.67	0.12	1.17
Mean type (3,4)	1.36	1.46	-0.59	3.53	1.35	0.96	1.71
Mean type (4,4)	2.35	2.02	-0.47	4.19	2.24	1.71	2.75
Var. type (1,1)	0.01	0.58	0.00	2.99	0.02	0.00	0.24
Var. type (1,2)	0.43	0.55	0.00	2.89	0.47	0.11	1.18
Var. type (2,2)	0.01	0.61	0.00	2.89	0.13	0.00	0.98
Var. type (1,3)	1.76	0.79	0.00	4.03	1.68	0.89	2.80
Var. type (2,3)	1.18	0.83	0.00	3.42	1.25	0.76	1.85
Var. type (3,3)	1.98	1.06	0.00	4.84	1.92	1.02	2.94
Var. type (1,4)	2.09	1.02	0.00	4.31	1.98	1.13	3.11
Var. type (2,4)	1.48	0.88	0.00	3.27	1.43	0.70	2.49
Var. type (3,4)	1.91	1.02	0.00	3.60	1.87	1.24	2.60
Var. type (4,4)	1.61	0.94	0.00	3.80	1.65	0.77	2.58
Prop. type 1	0.15	0.15	0.10	0.22	0.14	0.12	0.17
Prop. type 2	0.20	0.26	0.15	0.42	0.20	0.17	0.24
Prop. type 3	0.36	0.31	0.13	0.45	0.36	0.31	0.41

Notes: Estimates of the random-effects model with K groups, in data generated according to that model with K_0 groups. 500 simulations. The smaller sample has 156 workers and 896 teams, the larger sample has 921 workers and 5447 teams. In panel (a) $K_0 = K = 2$, in panel (b) $K_0 = K = 4$.

C Additional tables and figures

Table C2: Additive production, economic researchers, robustness

	$n = 1$	$n = 2$	$n = 3$
(a) Log-journal quality as dependent variable			
Total variance	2.00	2.41	2.30
Heterogeneity	0.64	0.85	1.02
Heterogeneity (uncorrected)	0.80	1.33	1.76
Sorting	-	0.59	0.93
Sorting (uncorrected)	-	0.38	0.49
Other factors	1.15	1.29	1.19
Other factors (uncorrected)	1.20	1.12	1.20
Team shift μ_n	0.00	-0.20	-0.61
(b) Ranks as dependent variables			
Total variance	0.089	0.090	0.086
Heterogeneity	0.034	0.019	0.013
Heterogeneity (uncorrected)	0.043	0.029	0.023
Sorting	-	0.015	0.016
Sorting (uncorrected)	-	0.012	0.012
Other factors	0.055	0.055	0.054
Other factors (uncorrected)	0.046	0.049	0.052
Team scale λ_n	1.00	0.55	0.38
(c) ≥ 2 articles per author			
Total variance	65.05	127.78	155.92
Heterogeneity	21.76	23.53	29.29
Heterogeneity (uncorrected)	36.63	69.89	78.47
Sorting	-	24.04	35.41
Sorting (uncorrected)	-	2.81	8.65
Other factors	41.11	73.13	65.96
Other factors (uncorrected)	28.42	49.85	63.53
Team scale λ_n	1.00	0.67	0.49
(d) ≥ 1 articles per author			
Total variance	53.68	115.17	145.22
Heterogeneity	6.36	0.40	29.92
Heterogeneity (uncorrected)	32.72	95.30	139.92
Sorting	-	30.58	31.67
Sorting (uncorrected)	-	-18.58	-34.85
Other factors	47.89	34.63	70.54
Other factors (uncorrected)	20.96	37.97	63.53
Team scale λ_n	1.00	0.68	0.50

Notes: Estimates of variance components for different team sizes n . In panels (a) and (b) I use log-journal quality and year-specific ranks of journal quality, respectively, as dependent variables in the baseline sample. In panels (c) and (d) I enlarge the sample to include at most 2 and 1 articles per author, compared to at most 5 in the baseline sample, using journal quality as the dependent variable.

Table C3: Nonlinear production, economic researchers, other specifications ($K = 4$)

	Correlated RE		Joint RE		2-author only
	$n = 1$	$n = 2$	$n = 1$	$n = 2$	$n = 2$
Total variance	97.81	162.64	97.81	162.64	162.64
Heterogeneity	36.49	46.71	32.86	35.03	51.97
Sorting	-	17.04	-	30.90	7.84
Nonlinearities	-	4.29	-	1.15	2.82
Other factors	61.32	94.61	64.96	95.56	100.01

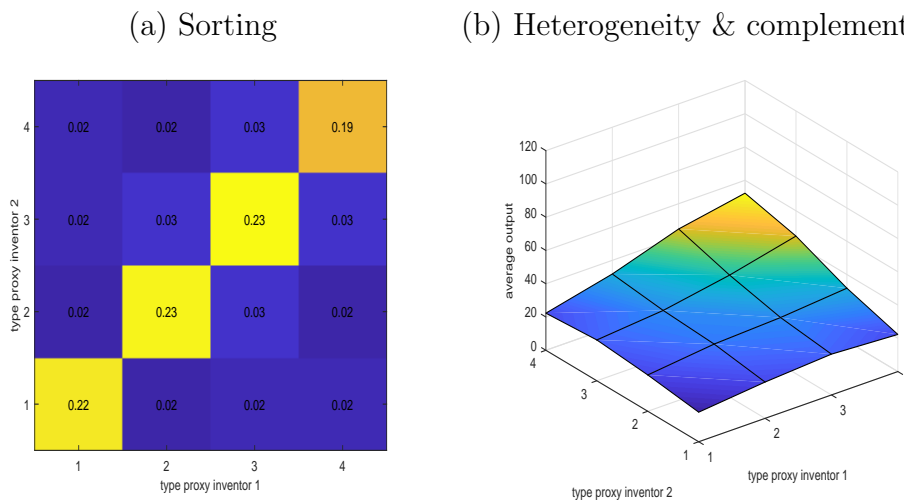
Notes: Estimates of variance components for different team sizes n in the nonlinear model with K types, for three specifications. “Correlated RE” comes from a model where the types follow a multinomial logit distribution given the numbers of 1-author and 2-author collaborations at the extensive and intensive margins. “Joint RE” comes from a model where collaborations follow independent type-specific Poisson distributions. “2-author only” only uses information from 2-author teams. Descriptive statistics on the sample are given in panel (b) of Table 1.

Table C4: Additive production, patents and inventors, robustness

	$n = 1$	$n = 2$	$n = 3$
(a) ≥ 2 patents per inventor			
Total variance	1443.5	1277.1	1597.1
Heterogeneity	591.5	699.7	886.7
Heterogeneity (uncorrected)	958.9	1368.6	2029.5
Sorting	-	-97.4	-179.5
Sorting (uncorrected)	-	-515.3	-1000.4
Other factors	872.6	596.6	756.0
Other factors (uncorrected)	484.6	428.4	497.2
Team scale λ_n	1.00	0.54	0.40
(b) ≥ 1 patents per inventor			
Total variance	1446.9	1265.8	1546.5
Heterogeneity	538.7	1079.3	698.1
Heterogeneity (uncorrected)	1098.6	1989.1	2241.2
Sorting	-	-479.8	-72.9
Sorting (uncorrected)	-	-1021.9	-1135.2
Other factors	918.1	637.8	854.0
Other factors (uncorrected)	348.3	311.8	381.4
Team scale λ_n	1.00	0.54	0.39

Notes: Estimates of variance components for different team sizes n . In panels (a) and (b) I enlarge the sample to include at most 2 and 1 patents per inventor, compared to at most 5 in the baseline sample. For these results I net out multiplicative year and inventor-age effects from the output.

Figure C1: Preliminary evidence based on type proxy, patents and inventors



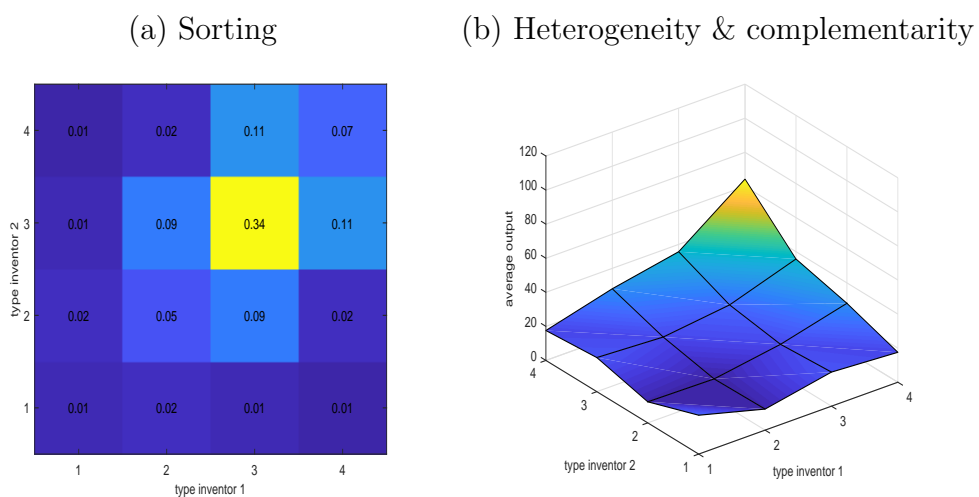
Notes: Subsample of inventors who produce at least 5 patents on their own. I compute type proxies as quartiles of average sole-authored output. In panel (a) I show the proportions of type proxies for inventors producing together in a 2-inventor team. In panel (b) I show average output (i.e., average truncation-adjusted forward citations) for different combinations of the type proxies.

Table C5: Nonlinear production, patents and inventors, other specifications ($K = 4$)

	Correlated RE		Joint RE		2-inventor only
	$n = 1$	$n = 2$	$n = 1$	$n = 2$	$n = 2$
Total variance	1270.4	1202.0	1270.4	1202.0	1202.0
Heterogeneity	355.9	394.8	197.0	103.0	401.3
Sorting	-	79.7	-	92.5	86.2
Nonlinearities	-	52.9	-	6.9	47.1
Other factors	914.5	674.5	1073.4	999.6	667.3

Notes: Estimates of variance components for different team sizes n in the nonlinear model with K types, for three specifications. “Correlated RE” comes from a model where the types follow a multinomial logit distribution given the numbers of 1-inventor and 2-inventor collaborations at the extensive and intensive margins. “Joint RE” comes from a model where collaborations follow independent type-specific Poisson distributions. “2-inventor only” only uses information from 2-inventor teams. Descriptive statistics on the sample are given in panel (b) of Table 4. For these results I net out multiplicative year and inventor-age effects from the output.

Figure C2: Heterogeneity, sorting and complementarity out of sample (2000 to 2005, $K = 4$)



Notes: The random-effects model with $K = 4$ inventor types is estimated on the 1995-1999 period. In panel (a) I show the proportions of types, and in panel (b) I show average output (i.e., average truncation-adjusted forward citations) for different combinations of the types, for inventors producing together in 2-worker teams between 2000 and 2005.