Journal of Econometrics 190 (2016) 222–232

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Endogenous network production functions with selectivity

William C. Horrace[a,*], Xiaodong Liu[b], Eleonora Patacchini[c,d]

[a] Department of Economics, 110 Eggers Hall, Syracuse University, Syracuse, NY, 13244, United States
[b] Department of Economics, University of Colorado, Boulder, CO, 80309, United States
[c] Department of Economics, Cornell University, Ithaca, NY, 14850, United States
[d] Sapienza University and EIEF, Rome, Italy

ARTICLE INFO

Article history:
Available online 29 June 2015

JEL classification:
C31
C44
D24

Keywords:
Stochastic frontier model
Spatial autoregressive model
Peer effects
Endogenous network formation
Selectivity

ABSTRACT

We consider a production function that transforms inputs into outputs through peer effect networks. The distinguishing features of this model are that the network is formal and observable through worker scheduling, and selection into the network is done by a manager. We discuss identification and suggest several estimation techniques. We tackle endogeneity arising from selection into groups and exposure to common group factors by employing a polychotomous Heckman-type selection correction. We illustrate our method using data from the Syracuse University Men's Basketball team, where at any time the coach selects a lineup and players interact strategically to win games.

1. Introduction

Endogeneity in production function estimation is not a new issue. Endogeneity of inputs can arise for a variety of reasons: input measurement error, simultaneity of unobservables and inputs, and endogeneity of "explanatory" outputs in multiple-output distance function analysis (to name a few). In service industries, these problems are exacerbated in obvious ways. However, one could imagine that the main challenge in estimating a service production function is the specification of the function itself. In particular, the way that labor is transformed into output may be unclear. Production in a service industry is typically not "serial" as it might be on a manufacturing assembly line, where productivity of worker *A* may only affect the productivity of worker *B*, who (in turn) only affects worker *C*.[1] Service industries may be characterized by teams of workers whose individual productivities are interrelated in complex ways and (in particular) through networks. Consider an architectural firm which simultaneously produces design plans for a variety of projects with teams of architects and draftsmen, who may work across multiple projects in a given workday. In this setting worker interrelatedness may be determined by networks established by a single manager, who assigns workers to teams based on both observable and unobservable characteristics of workers. This implies formal and measurable time-varying networks which may be endogenous due to selectivity.[2] Understanding network effects in production may be important for worker scheduling and design of worker incentive schemes.

The purpose of this paper is to specify an econometric model that incorporates peer effects on worker productivity (output).[3] That is, a worker's productivity is a function of the productivities of the co-workers on her team, where teams are assigned by managers. Individual team members interact through time-varying interaction schemes which serve as proxies for the managerial decision and which function as the mechanism for group formation and individual interrelatedness. In most econometric network models, selection into groups is as much an individual choice as is

* Corresponding author. Tel.: +1 315 443 9061; fax: +1 315 4431081.
E-mail address: whorrace@maxwell.syr.edu (W.C. Horrace).

1 This is not to suggest that a manufacturing process could not be more complicated, but the traditional assembly line process possesses this feature.

2 There may also be informal networks, but they are not the focus here. Informal networks may arise through a principal–agent problem of imperfect monitoring. A manager may order a worker to split her time evenly on the two projects, but she may not, in practice. An alternative way to conceptualize this phenomenon is that the formal network is measured with error.

3 Peer effects have been indicated as one of the main empirical determinants of several important social phenomena (see Jackson and Zenou, 2013, part III, for a collection of recent studies).

http://dx.doi.org/10.1016/j.jeconom.2015.06.005
0304-4076/

the behavior that stems from a given network structure. In this setting endogeneity problems may arise if the model does not account for unobserved individual characteristics driving both network formation and behavior over networks.

We consider the unique situation where a manager selects workers into teams (networks) to produce output, and we call this model a *Network Production Function Model*. In the model, network connections are captured by a binary adjacency matrix, where adjacency is specified as a binary indicator of team membership. The salient feature of this model is that team membership is perfectly observable.[4] In this model, the manager's selection decisions depend on the combination of individual characteristics at the team level, rather than individual-level characteristics. Such team-level factors contribute to the so called "correlated effects" (Manski, 1993), which could be confounded with peer effects and lead to identification problems.

We use a polychotomous Heckman-type correction to address this problem in the context of production networks. In team projects, the probability of selecting a worker for the project is not independent across workers. We exploit this interdependency for the identification and estimation of peer effects in network production functions. This is the main contribution of the paper.

More specifically, we consider productivity of a single project, involving a two-stage process. First, the manager chooses a team (lineup) of m workers (m is predetermined) from a population of n workers to work on the project of interest. Residual workers are assigned to other projects.[5] Next, workers work on the project to produce output for a given time period. For the population of n workers, the $n \times n$ adjacency matrix across all projects is potentially endogenous. By focusing on a single project of interest, we have an $m \times m$ submatrix of the adjacency matrix which is exogenous conditional on selection into the specific project. Thus, the network endogeneity is reduced to a selectivity bias, which can be corrected using a fixed effect estimator or a polychotomous Heckman-type bias correction procedure due to Lee (1983) and Dahl (2002).[6]

The resulting selectivity bias term is an inverse mills ratio (in the case of Lee's parametric estimate) or a single index (in the case of Dahl's semi-parametric estimate), varies across lineups and time, and can be interpreted in two interesting ways. First, it can be thought of as a fixed effect that represents the correlated effect, "wherein individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments" (Manski, 1993, page 533). In this case the group is the observed lineup, and the "institutional environment" is the manager's selection of the lineup into the project of interest. In this sense we use Heckman (1979) to solve Manski's correlated effects problem. In fact, in terms of estimation, we employ a fixed effect estimator in the style of Lee (2007) that differences out the correlated effect. Second, the selectivity bias term is loosely interpretable as managerial competence or efficiency. That is, all things being equal and averaging out luck, it is the manager's lineup selection that produces any unobserved team effect and, hence, variability of worker output. This is similar to the notion of inefficiency in the stochastic frontier literature (Aigner et al., 1977; Meeusen and van den Broeck, 1977), so our selectivity bias term can be thought of as efficiency if it increases output and as inefficiency if it lowers it. Also, insofar as our bias term may be estimated from a first-stage selection equation, it is interpretable as x-efficiency in the stochastic frontier literature Alvarez et al. (2006).[7]

Our empirical example is the network production function for college basketball. While this may only loosely represent a service industry production process, it is sufficient for the purpose of illustration. In this setting there are n players on a team engaged in two projects at any given period of time: five players interact to produce offense and defense, and $n-5$ players sit on the bench to produce rest (which is inversely correlated with fatigue).[8] Our measure of active player productivity is *player efficiency,* which aggregates time-averaged performance statistics on points, rebounds, blocks, steals, misses, assists, and other measures of offensive and defensive activity for each player. We include a measure of lagged fatigue as an explanatory variable to control for the productivity of benched players. Our data are all player substitutions during the regular 2011–2012 season of the Syracuse University men's college basketball team. We find statistically significant positive production spillovers across players in the same category (guards or forwards), but insignificant effects across players in different categories. When selectivity bias is taken into account, our estimate of peer effects in productivity is 0.0534. That is, a one unit increase in the average efficiency of the other active guards (forwards) induces a 0.0534 increase in the efficiency of an individual guard (forward) once selectivity bias is taken into consideration.

The rest of the paper is organized as follows. The next section reviews the related literature, while highlighting the contribution of our paper. Section 3 introduces the econometric specification of a network production model, while Section 4 considers the specification and estimation of a network production model with selectivity. Section 5 provides an empirical example, using data from the 2011–2012 Syracuse University Men's basketball team. Section 6 concludes.

2. Related literature

Our paper lies at the intersection of different literatures. We briefly review them below, while highlighting our contribution.

2.1. Econometric network models

A number of papers have dealt with the identification and estimation of peer effects with network data (see Blume et al., 2011 for an excellent survey). There are two main methodological approaches.

(i) The network is assumed exogenous once potential unobserved factors responsible for network endogeneity are treated by network fixed effects. Identification relies on network topology and estimation is performed using 2SLS or GMM (see, e.g., Lee, 2007; Bramoullé et al., 2009; Calvó-Armengol et al., 2009; Davezies et al., 2009; Lee et al., 2010; Liu and Lee, 2010). Network fixed effects can be interpreted as originating from a two-step link formation process, where individuals self-select *into* different networks in the first step based on network-specific characteristics and, then, in the second step, link formation takes place *within*

4 It is also possible for adjacency to be measured as cumulative time that individuals worked together on a project. This would be directly measurable from time-cards, but we do not explore it here.

5 We note that, in any period the $n - m$ residual workers are assigned to other projects, and lags of the output from these projects (as well as the project of interest) are treated as explanatory variables in the output and selection equations. In this sense our specification is not unlike the multiple-output distance function (Färe and Primont, 1990) where a single output is modeled as functions of the remaining outputs.

6 It is also interesting to note that the word "lineup" evokes an image of workers standing in a line. Our notion of lineup allows us to abstract from the complicated endogenous network for all the workers to a simple, fixed and complete network of workers in a project.

7 More generally, it is interpretable as another source of heterogeneity. However, it is still interesting to speculate on the ways it may embody (in)efficiency.

8 We take the managerial decisions and performance of the opposing team as exogenous. In this sense our notion of strategic equilibrium is only partial.