Identifying Productivity Spillovers Using the Structure of Production Networks*

Samuel Bazzi
Boston University

Amalavoyal Chari
University of Sussex

Shanthi Nataraj
RAND Corporation

Alexander D. Rothenberg†
RAND Corporation

August 2017

Abstract

Despite the importance of agglomeration externalities in theoretical work, evidence for their nature, scale, and scope remains elusive, particularly in developing countries. Identification of productivity spillovers between firms is a challenging task, and estimation typically requires, at a minimum, panel data, which are often not available in developing country contexts. In this paper, we develop an identification strategy that uses information on the network structure of producer relationships to provide estimates of the size of productivity spillovers. Our strategy builds on that proposed by Bramoullé et al. (2009) for estimating peer effects, but we improve upon this network structure identification strategy by using panel data and validate it with exchange-rate induced trade shocks that provide additional identifying variation. We apply this strategy to a long panel dataset of manufacturers in Indonesia to provide new estimates of the scale and size of productivity spillovers. Our results suggest positive productivity spillovers between manufacturers in Indonesia, but estimates of TFP spillovers are considerably smaller than similar estimates based on firm-level data from the U.S. and Europe, and they are only observed in a few industries.

*Acknowledgement: We are grateful for financial support from Private Enterprise Development for Low-Income Countries (PEDL), a joint research initiative of the Centre for Economic Policy Research (CEPR) and the Department For International Development (DFID). Kun Gu provided excellent research assistance. All errors remain our own.

†Corresponding author: 1200 South Hayes St., Arlington, VA 22202-5050. Email: arothenb@rand.org.

1 Introduction

Despite the importance of agglomeration externalities in theoretical work, evidence for their nature, scale, and scope remains elusive. This is particularly a concern for developing countries, where high quality data are scarce, and where the potential scope for agglomeration externalities may be largest.[1] Among the key sources of agglomeration externalities are productivity spillovers between firms. Marshall (1890) describes several mechanisms through which such productivity spillovers may occur, including (1) technological spillovers, (2) labor market pooling, and (3) intermediate input linkeages.[2] Beyond theoretical concerns, the presence and magnitude of productivity spillovers feature prominently in evaluations of place-based policies. If spillovers are large enough, subsidies to locate in certain regions may ignite a virtuous circle of development and growth that enhances both local and national welfare.[3]

In this paper, we develop a strategy for identifying productivity spillovers between firms and apply our methodology to high quality firm-level panel data on Indonesian manufacturers. Identifying productivity spillovers is a challenging task.[4] First, measuring firm-level productivity is challenging, and estimating production functions requires addressing the fact that inputs and outputs are simultaneously determined by productivity, which is typically observed by the firm but not the econometrician (Olley and Pakes, 1996; Levinsohn and Petrin, 2003). Second, it is difficult to distinguish the effects of local spillovers from other unobserved local factors that may lower production costs or raise productivity (Ellison and Glaeser, 1999). Another identification problem comes from the difficulty of disentangling the impact of one firm's productivity on the productivity of other firms. This is due to the well-known reflection problem, which creates difficulties for identifying peer effects (Manski, 1993).

High quality panel data can help to resolve certain omitted variables problems (Henderson, 2003). In our empirical application, we use plant-level panel data from Indonesia's Manufacturing Survey (*Survei Industri*), which allow us to control for fixed factors specific to individual firms, such as their managerial capacity or technological sophistication, that may be correlated with input choices or the strength of the industry in a particular location. Panel data also allow us to control for the impact of any location-specific unobservables that are time-invariant, such as the regulatory climate or favourable geography, that may impact both the decisions of firms to locate in a region and how productive they are when they get there. Importantly, firm-level panel data enable us to estimate production functions using control function and GMM approaches, which rely on proxies for time-varying unobservables that are correlated with input choices (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015).

After estimating firm-level productivity, we specify a linear-in-means model that relates a firm's own productivity to the average productivity of firms to which it is connected. This model suffers from the reflection problem, and to address it, we employ a technique from Bramoullé et al. (2009), which leverages the fact that firms in certain industries are related to each other due to supply chains networks,

[1] Chauvin et al. (2016) describe how human capital externalities and agglomeration elasticities seem to be larger in Brazil, India, and China than in developed countries like the United States. They hypothesize that in developing countries, barriers to migration may create large spatial arbitrage opportunities.

[2] Duranton and Puga (2004) consider the sources of agglomeration spillovers as coming from (1) sharing, (2) matching, and (3) learning.

[3] See Kline and Moretti (2014) for more discussion. In a companion paper evaluating a place-based policy in Indonesia, Rothenberg et al. (2016b) argue that the welfare justification for regional policies depends critically on the shape of the agglomeration function.

[4] See Duranton et al. (2015) for a review of approaches to this problem.

and these networks have systematic patterns that can be measured through input-output tables. As long as these networks contain intransitive triads (i.e. industries that are not directly related to each other, but only related through a common third industry), identification can be achieved.

To measure firm-level connections, we begin by using Indonesian data on the products that industries produce and use as raw materials to construct a network of forward and backward linkeages. We describe the resulting network structure in the data, using descriptive statistics from graph theory, and compare the Indonesian industrial network to other networks. We find both that industrial relationships contain many intransitive triads, and that the network structure has many "small worlds" properties that are similar to other networks, including the network of U.S. industries, the structure of the internet, and gene networks (Acemoglu et al., 2016; Carvalho, 2014).

Next, we create firm-level networks by assuming that firms are connected to one another if they are in the same industry, or if their industries are related to each other through forward and backward linkeages, or if they are located in close physical proximity. This results in a family of network structures that provide us with significant identifying variation through firms in related industries that are located in multiple places. Intuitively, shocks to a firm's neighbors-of-neighbors should affect the productivity of neighbors to that firm, but they should not directly affect that firm's productivity. This exclusion restriction provides us with the identification we need.

We combine the network structure for identification with panel data, and we make use of exchange rate shocks that provide additional identifying variation. Implementing this strategy requires calculating network-level averages of variables, at different orders of connections (such as first-order connections, second-order connections, etc.). With roughly 20,000 firms each year, the size of the network is quite large, and computations require sparse matrix routines to increase speed.

Implementing this identification strategy with Indonesian manufacturing data, we find that exchange rate shocks have a meaningful first-stage relationship, meaning that neighbors-of-neighbors depreciations increase neighors' average productivity. Using this network-based instrument on firm-level panel data with individual firm fixed effects, we estimate positive average productivity spillovers between firms, but our estimates are substantially smaller than those found in the literature on U.S. and European firms. Moreover, the productivity spillovers we observe are driven by only a small number of industries. These relatively small estimates of TFP spillovers echo other work on Indonesia's agglomerations. For instance, Amiti and Cameron (2007) find small impacts of labor market pooling on wages in Indonesia, while a companion paper does not find a consistent correlation between an industry's use of highly skilled workers and spatial concentration (Rothenberg et al., 2016a). This suggests that one of the most important drivers of agglomeration externalities, knowledge spillovers, may not be operating well in Indonesian cities.

This paper contributes to several strands of literature. The first is the literature on estimating agglomeration externalities, which is now well developed. Most work infers agglomeration effects from wages, but increasingly, researchers have used firm-level data to estimate TFP spillovers (Duranton et al., 2015). Some authors instrument the determinants of agglomeration forces with long lagged values of historical variables, such as historical population density (Ciccone and Hall, 1996; Combes et al., 2008) or geographic variables that influence construction (Rosenthal and Strange, 2008; Combes et al., 2010). Particularly with firm-level data on TFP, many authors use GMM to estimate specifications in first differences, using lags of variables as instruments. These can be useful for uncovering agglomeration effects in static or dynamic specifications of TFP (Henderson, 2003; Mion, 2004; Graham et al., 2010; Martin et al., 2011). More recently, several authors have turned to natural experiments to estimate productivity spillovers (Hanson, 1997; Redding and Sturm, 2008; Greenstone et al., 2010).

We add to this literature by making use of a novel network structure identification strategy. This strategy is related to a series of papers that uses spatial lags to estimate productivity spillovers, but our work considers identification carefully and rectifies some of the identification shortcomings (Gibbons and Overman, 2012). Although exploiting the structure of the network in which agents interact has been used to estimate individual peer effects in other settings (e.g. De Giorgi et al., 2010), to our knowledge, this represents one of the first uses of this technique for estimating productivity spillovers between firms.[5]

Finally, there is a relatively new literature in macroeconomics that views the input-output network as a mechanism for propogating shocks (Acemoglu et al., 2016; Carvalho, 2014). We build upon these ideas by expanding the industrial network to encompass firms and physical proximity. However, our urban focus on productivity spillovers and agglomeration externalities is somewhat distinct from the macro focus of this literature.

The rest of this paper is organized as follows. Section 2 describes both the industry-level and firm-level networks that we work with to estimate productivity spillovers. Section 3 describes how we estimate production functions and implement the network structure identification strategy to estimate productivity spillovers. Section 4 discusses results, and Section 5 concludes.

2 The Social Network of Firms

In this section, we present a new method for using input-output tables to construct different measures of connections between medium and large manufacturing firms in Indonesia. We first describe our measures of industrial proximity, and we present descriptive statistics on the network of industries generated by these measures. Next, we explore how to combine data on industrial connections with information on the physical distances between firms to generate a measure of firm connections. Finally, we present summary statistics on the connected network of firms.

2.1 Industrial Proximity

Let $\mathcal{S}$ denote the set of (5-digit) industries (sectors), where two industries $A, B \in \mathcal{S}$ represent typical elements of that set. Firms in each industry produce one or more products, where $\mathcal{J}$ denotes the universe of products. Let $\mathcal{J}_A$ denote the set of products produced by firms in industry A, so that $\mathcal{J} = \bigcup_{A \in \mathcal{S}} \mathcal{J}_A$.

To measure the vertical relationships between industries, we focus on both upstream connections (or forward linkages), where industry A supplies raw materials to industry B, and downstream connections (or backward linkages), where industry A is supplied by products that industry C produces. To illustrate, Figure 1 depicts a production line, with industry A in the center. In this figure, the gray arrows are drawn outward from the source of production. Relative to industry A, industry B is an upstream connection,

[5] Serpa and Krishnan (2017) follow a similar approach using U.S. data.

because industry A produces products that B uses as raw materials. Industry A also has a downstream connection to industry C, because industry A uses products that C produces.

To better operationalize the connectedness of industries, we focus first on upstream connections. Let $\sigma^U_{A,B} \in [0,1]$ denote the share of the total value of industry B's raw materials that are produced by industry A. In writing $\sigma^U_{A,B}$, we adopt a convention that the first subscript denotes the *producing* industry, while the second subscript denotes the *consuming* industry. We can write $\sigma^U_{A,B}$ as follows:

$$\sigma^U_{A,B} \equiv \frac{RM_{A,B}}{RM_B} = \frac{\sum_{j \in \mathcal{J}_A} RM^B_j}{\sum_{j \in \mathcal{J}} RM^B_j}$$

where RM^B_j denotes the total raw material value of product j used by industry B. As $\sigma_{A,B}$ increases to 1, industry A produces a larger share of the products consumed by industry B as raw materials, increasing the intensity of the upstream connection. The definition of $\sigma^U_{A,B}$ naturally creates a *weighted directed network* of industries.

Downstream connections (or backward linkages) are defined similarly. Let $\sigma^D_{A,C} \in [0,1]$ denote the share of industry A's raw materials that are produced by industry C. We can write $\sigma^D_{A,C}$ as follows:

$$\sigma^D_{A,C} \equiv \frac{RM_{C,A}}{RM_A} = \frac{\sum_{j \in \mathcal{J}_C} RM^A_j}{\sum_{j \in \mathcal{J}} RM^A_j}$$

As $\sigma^D_{A,C}$ increases to 1, industry A uses a larger share of products produced by industry C as raw materials, increasing the downstream connection. The definition of $\sigma^D_{A,C}$ also creates a *weighted directed network* of industries.

Using $\sigma^U_{A,B}$ and $\sigma^D_{A,C}$, we can use matrix notation to define two separate families of *unweighted* directed networks measuring upstream and downstream connections:

$$\begin{aligned} \mathbf{g}^U(s) &= \left[g^U_{A,B} \right] \quad \text{where} \quad g^U_{A,B} = \mathbb{1}\left\{ \sigma^U_{A,B} \geq s \right\} \\ \mathbf{g}^D(s) &= \left[g^D_{A,C} \right] \quad \text{where} \quad g^D_{A,B} = \mathbb{1}\left\{ \sigma^D_{A,C} \geq s \right\} \end{aligned} \tag{1}$$

where s indicates the threshold input intensity level. Apart from direction, a major difference between these two networks is the normalization of the respective σ's; $\sigma^U_{A,B}$ is normalized by the total value of industry B's raw materials, while $\sigma^D_{A,B}$ is normalized by the total value of industry A's raw materials. Note also that these definitions generate some natural equivalences; using a weak threshold of any upstream or downstream connection, the upstream network is the just transpose of the downstream network; i.e. $\mathbf{g}^U(0) = \mathbf{g}^D(0)^T$.

In our empirical analysis, we will look for separate impacts of upstream and downstream connections. However, it is also instructive to focus on industrial networks where industries are connected if they display any upstream or downstream connection:

$$\mathbf{g}^A(s) = \left[g^A_{A,B} \right] \quad \text{where} \quad g^A_{A,B} = \mathbb{1}\left\{ \sigma^U_{A,B} \geq s \ \text{ or } \ \sigma^D_{A,B} \geq s \right\} \tag{2}$$

To measure the forward and backward linkages between industries, we use data on products produced and raw materials used by each 5-digit industry from the year 2000. These data are produced by Indonesia's Central Statistical Agency, or *Badan Pusat Statistik* (BPS). The "products produced" data contain a list of the total quantity and value of products produced by each 5-digit industry, with product code identifiers, while the "raw materials" data contain a list of the total quantity and value of products used as raw materials by each 5-digit industry, also with product-code identifiers. We aggregated both datasets to the industry-by-product level and merged the data to measure industrial connections.[6]

2.2 Descriptive Statistics on the Network of Industries

Figure 2 depicts a visualization of $\mathbf{g}^A$ $(s = 0.01)$, the network of relationships between 5-digit industries where industries are connected if industry A produced at least 1 percent of industry B's total value of raw materials, or at least 1 percent of industry A's raw materials are produced by industry B.[7] The number of industries (nodes) in this figure is 261, and there are 8,763 one-way connections (edges) between these industries, so that the network is only 12.9 percent complete, a measure indicating the fraction of possible edges that are actual edges.

It is easy to see the mass of highly connected industries in the center right portion of the figure. The industries with only a few connections line the exterior of the graph. To get a better sense of the upstream and downstream relationships, we computed the degree distributions of $\mathbf{g}^U$ $(s = 0.01)$, the network of upstream connections. Figure 3 presents histograms of the degree distributions of the nodes in $\mathbf{g}^U$ $(s = 0.01)$. For this network, an industry's *in-degree* is the number of other industries that supply it raw materials (downstream relationships), while an industry's *out-degree* is the number of other industries to which it supplies raw materials (upstream relationships). Mathematically, for node A, the in and out-degree terms are given by:

$$\text{in-degree}_A = \sum_{C \in \mathcal{S}} g^U_{C,A}$$
$$\text{out-degree}_A = \sum_{B \in \mathcal{S}} g^U_{A,B}$$

The median industry had 13 downstream relationships and 7 upstream relationships. However, Figure 3 shows that the distribution of downstream connections is much more uniform than the distribution of upstream connections, which is skewed. Over 10 percent of industries had no upstream connections; these industries tended to focus on producing finished consumer products.

Table 1 presents information on the five industries with the most in-degree connections (downstream relationships, in Panel A) and the five industries with the most out-degree connections (upstream relationships, in Panel B). Industries with more upstream relationships tended to be more basic, like wire producers (38194) and manufacturers of basic organic chemicals (35118) which are used in a wide variety of different manufacturing processes. Industries like the motor vehicle industry (38443) or wood working machinery (38232) have many subcomponents that rely on other industries, so they have more

[6] Both datasets were digitized from two scanned PDFs of BPS reports, titled *Statistik Industri Besar dan Sedang: Bagian II, 2000* and *Statistik Industri Besar dan Sedang: Bagian III, 2000*. Unfortunately, we were unable to access this data at the firm level, unlike Amiti and Konings (2007). Further details can be found in Appendix **??**.

[7] This network was drawn by the force-placement algorithm of Fruchterman and Reingold (1991), as implemented in Gephi. For simplicity, we do not draw the direction of the connections between firms.

downstream relationships.

Panel C presents information on the most connected industries, measured by betweenness centrality, which captures the fraction of times that an industry lies on the shortest paths that connect other nodes. Industries like the motor vehicle industry (38433), the footwear industry (32411), and synthetic resins (35131) have lots of upstream and downstream connections to other highly connected industries, so they tend to be much more central in the network.

In Table 2, we return to $\mathbf{g}^{A}(s)$ but present overall statistics on the network after varying s, the threshold determining whether or not industries are connected. In the network $\mathbf{g}^{A}(s=0)$, industries are connected if they have any upstream or downstream connections, and we increase the s threshold from 0.01 to 0.05, reporting statistics on the resulting networks in each respective columns. As s increases, this threshold governing industrial connections becomes more difficult to cross, so we would expect fewer connections to be drawn. As expected, when s increases, the number of edges falls (row 2), and the completeness of the network (row 3) also falls.

Row 4 reports the *transitivity* of the network, which measures the fraction of triples of nodes (with at least one connected pair) that are completely connected triangles. In Section 3, when we discuss how to use the structure of production networks to identify spillovers, our identification strategy relies on intransitive triads, or incomplete triangles. We find that as s increases, the transitivity of the network falls substantially. At $s=0$, 56.3 percent of triples with at least one connected pair are fully connected triangles, while at $s=0.05$, this measure falls to 28.2 percent.

Rows 5 and 6 of Table 2 highlight the "small worlds" feature of the Indonesian manufacturing network, a feature that is common to other industrial networks (Acemoglu et al., 2016) and is also seen in other contexts, such as social networks, the architecture of the internet (Brin and Page, 1998), and gene networks (Carvalho, 2014). Most industries are not directly connected to each other, as evidenced by the relatively low network completeness. However, nodes can be reached by a small number of connections. Across the distribution of s, the networks tend to have a fairly small diameter (at least for their largest connected component), and the shortest path length between nodes is also quite small.

2.3 Incorporating Physical Distance

Beyond industrial proximity, another important feature of firm networks lies in physical proximity. We assume that all firms within a certain physical distance or a certain industrial distance of one another interact, while firms that are too far away do not. We use the SI data on each firm's district of operation to enrich our measure of interfirm relationships.[8]

Formally, let $d(i)$ denote the district of firm i, and let $\delta_{ij} \equiv \delta_{d(i),d(j)}$ denote the distance between the centroids of districts $d(i)$ and $d(j)$, measured as the crow flies, in kilometers. For ease of notation, also let $\sigma^{U}_{ij} \equiv \sigma^{U}_{I(i),I(j)}$ denote the upstream industrial proximity between firm i's industry, $I(i)$, and firm j's industry, $I(j)$, and let $\sigma^{D}_{ij} \equiv \sigma^{U}_{I(i),I(j)}$ denote the downstream industrial proximity between firm i and j's industries.

[8] There are 301 districts in our data with a median land area of 1,886 km^2, which is slightly larger than the median U.S. county with a median area of 1,595 km^2. In a companion paper, we use address-level data to geocode the locations of Indonesian manufacturers, but this address-level data cannot be directly linked to the SI (Rothenberg et al., 2016a).

Using σ_{ij}^U, σ_{ij}^D, and δ_{ij}, we define several families of firm-level networks:

$$\begin{aligned} \mathbf{G}^U(s,d) &= \left[g_{ij}^U(s,d)\right] \quad \text{where} \quad g_{ij}^U(s,d) = \mathbb{1}\left\{\sigma_{ij}^U \geq s \text{ or } \delta_{ij} \leq d\right\} \\ \mathbf{G}^D(s,d) &= \left[g_{ij}^D(s,d)\right] \quad \text{where} \quad g_{ij}^D(s,d) = \mathbb{1}\left\{\sigma_{ij}^D \geq s \text{ or } \delta_{ij} \leq d\right\} \end{aligned} \tag{3}$$

Thus, in the firm-level network $\mathbf{G}^U(s,d)$, firms i and j are connected if they are either located in districts within d kilometers of one another, or their industrial proximity measure, σ_{ij}^U, is greater than some value s. Note that we assume that firms are connected if they belong to the same industry, but we also follow the convention that they are not connected to themselves: for all firms i, $g_{ii}^U(s,d) = 0$ and $g_{ii}^D(s,d) = 0$.

This measure of firm-level connections captures several important ideas in the literature on agglomeration economies and productivity spillovers. Our networks capture localization economies, because all firms in the same district are connected. Within-industry technological spillovers are reflected in the fact that firms in the same industry are also directly connected. Moreover, we provide direct measures of intermediate-input linkages through our measures of industrial proximity. Varying s and d allows these forces to evolve smoothly within industrial–physical proximity space.

2.4 Descriptive Statistics on the Network of Firms

Figure 4 presents a visualization of $\mathbf{G}^U(s = 0.01, d = 0)$, using data from the year 2000. In 2000, there are 21,834 firms and slightly more than 1 million connections drawn between firms. This is a typical year of data, and in our empirical analysis below, we will use similar data on roughly 20,000 firms observed every year, from 1990 to 2012. Because of the large size of $\mathbf{G}^U(s = 0.01, d = 0)$, and the need to repeatedly take local averages of variables on this large network, we use sparse matrix routines, implemented by Python's scientific computing environment, SciPy, for almost all of the network calculations in the paper.

The image of these connections that appears in Figure 4 depicts many clusters of firms connected together, and some clusters connected to one another through a relatively smaller number of links. The clusters in this image reflect the distance-based connections; all firms within the same district are connected to one another. Across clusters, the connections reflect industrial linkages.

In Table 3, we present statistics on the networks in the family of $\mathbf{G}^U(s,d)$. This table is organized similarly to Table 2, where s varies along the column dimension, and different network statistics appear in different rows. We vary s across columns and d across panels. Panel A fixes $d = 0$ while Panel B fixes $d = 10$. Most districts in Indonesia are farther apart than $d = 10$, but the special capital cities of Jakarta and Yogyakarta are divided into multiple districts, so increasing $d = 10$ allows for more connections between firms in those cities.

As in Table 3, when we increase s in Panels A and B, the number of edges decreases (row 2), and the percentage completeness falls (row 3). However, the network transitivity behaves slightly differently. In $\mathbf{G}^U(s = 0, d = 0)$, 88.1 percent of triples with at least one connected pair are fully connected. This large number reflects the fact that because of physical distance connections, many of the triples with at least one connected pair already have full connectivity. As we increase s, the threshold of industrial connections increases, and there are fewer connections that span multiple districts, where intransitive triads occur. This tends to increase the transitivity measure.

3 Identifying and Estimating Productivity Spillovers

In this section, we describe how to use the networks developed in Section 2 to identify and estimate productivity spillovers between firms. Our goal is to isolate the causal impact of changes in other firms' productivities on changes in firm i productivity. To do so, we propose an approach that builds on a large and growing literature on the identification of peer effects. Two major identification challenges are *endogeneity*, due to common group effects, and *reflection*, a particular form of simultaneity bias (Manski, 1993). Recent literatures shows how data on the structure of social networks can help with identification (e.g. Bramoullé et al., 2009; De Giorgi et al., 2010; Goldsmith-Pinkham and Imbens, 2013). Social networks with intransitive triads break the reflection problem, and the exogenous characteristics of neighbors-of-neighbors (second-degree connections in industrial–physical space) that are not directly peers with each other provide instrumental variables. However, instead of using the characteristics of neighbors-of-neighbors, we use externally determined variables to which neighbors-of-neighbors are exposed but which do not directly impact the firm itself.

We start by specifying a log-linear production function for firm i:

$$y_{it} = \alpha + \mathbf{w}_{it}'\beta + \mathbf{x}_{it}'\gamma + \underbrace{u_{it}}_{v_{it}+e_{it}}, \qquad t = 1, ..., T \;, \tag{4}$$

where y_{it} denotes firm i's log value added at time t, $\mathbf{w}_{it}$ is a vector of "adjustable" inputs, which include the logs of the number of production and non-production workers, and $\mathbf{x}_{it}$ is a vector of state variables, including log capital.[9] The term u_{it} is the firm's total factor productivity, which can be decomposed into a portion of productivity that is observed by the firm before making input choices, v_{it}, and an idiosyncratic, unobserved component, e_{it}.

Because v_{it} is observed by the firm when it makes input choices, estimating (4) using ordinary least squares will lead to biased parameter estimates (Marshak and Andrews, 1944). There is now a large, well-developed literature on how to use control functions, such as an investment function (Olley and Pakes, 1996) or an intermediate-input demand function (Levinsohn and Petrin, 2003), to control for the simultaneous correlation between input choices and unobserved productivity and to provide consistent estimates of the production function parameters, $\theta \equiv (\alpha, \beta', \gamma')$.

In our empirical results below, we primarily focus on TFP residuals derived from a production function that is estimated using the investment control function approach developed by Olley and Pakes (1996).[10] However, we show that our results are robust to estimates of TFP based on other techniques, including the intermediate inputs control function (Levinsohn and Petrin, 2003), the Wooldridge (2009) GMM approach, the Ackerberg et al. (2015) conditional control function approach, and the index number approach of Aw et al. (2001).

After estimating production function parameters, θ, we can recover consistent estimates of ω_{it}, the firm's observed portion of log productivity plus the constant mean efficiency term:

$$\widehat{\omega_{it}} = \widehat{\alpha} + \widehat{v_{it}}$$

[9] This notation and the accompanying discussion borrows heavily from Wooldridge (2009).

[10] This choice follows Amiti and Konings (2007), who use the investment control function approach in estimating TFP using the same SI data that we use.

$$= \left(y_{it} - \mathbf{w}_{it}'\widehat{\beta} - \mathbf{x}_{it}'\widehat{\gamma} \right)$$

To model productivity spillovers, we assume that firm i's productivity at time t depends on that firm's own characteristics, $\mathbf{z}_{it}$, the average productivity of firms connected to i, $\overline{\omega_{(i)t}}$, and the average characteristics of connected firms, $\overline{\mathbf{z}_{(i)t}}$. We assume this relationship can be expressed as a linear-in-means model (Manski, 1993):

$$\omega_{it} = \theta_0 + \mathbf{z}_{it}'\theta_z + \theta_1\overline{\omega_{(i)t}} + \overline{\mathbf{z}_{(i)t}}'\theta_{\overline{z}} + \eta_{it} \tag{5}$$

In this notation, the expression $\overline{x}_{(i)t}$ refers to the average value of x, where that average is taken over individuals who are directly connected to i. Focusing on the graph $\mathbf{G} = \mathbf{G}(s, d)$, note that we can write this local average as follows:

$$\overline{x}_{(i)t} = \frac{\sum_j g_{ij} x_{jt}}{\sum_j g_{ij}}$$

If we let $M_i = \sum_j g_{ij}$ denote the number of direct connections that firm i has to other firms in network $\mathbf{G}$, we can express this local average in matrix notation:

$$\overline{\mathbf{x}} = \mathbf{H}\mathbf{x}$$

where $\mathbf{H}$ is just the matrix $\mathbf{G}$ with rows normalized by M_i:

$$\mathbf{H} = \left[\frac{g_{ij}}{M_i} \right]$$

Using matrix notation, we can rewrite equation (5) as:

$$\omega = \theta_0\iota + \mathbf{z}'\theta_z + \theta_1\mathbf{H}\omega + \theta_{\overline{z}}\mathbf{H}\mathbf{z} + \eta \tag{6}$$

This linear-in-means model, where agents interact in networks, is identical to the model studied by Bramoullé et al. (2009). The parameter θ_z measures the "own effects" of increasing $\mathbf{z}$ on productivity, while $\theta_{\overline{z}}$ measures "exogenous effects" of increasing the average characteristics of the local network on the firm's own productivity. In the peer effects and learning literature, these exogenous, or contextual, social interactions take place when student achievement varies with the socioeconomic composition of peer groups (Manski, 2000).

The parameter of interest is θ_1, which measures how a firm's own productivity varies structurally with the productivity of connected firms. To think about what this parameter measures, imagine using an RCT similar to one conducted by Bloom et al. (2013) on management and productivity in Indian textile manufacturers. If we hired McKinsey to exogenously increase the productivity of firms that are connected to your firm, by how much does this increase your firm's productivity?

Bramoullé et al. (2009) argue that the average exogenous characteristics of neighbors-of-neighbors, which can be calculated as $\mathbf{H}^2\mathbf{z}$, can be used as instruments for $\mathbf{H}\omega$ in (6). The system of equations actually generates a family of instruments, where other instruments include averages of $\mathbf{z}$ can be taken over third-order neighbors ($\mathbf{H}^3\mathbf{z}$), fourth-order neighbors ($\mathbf{H}^4\mathbf{z}$), . . . , but because of computational considerations, we only consider second-order and third-order averages. In the firm productivity context, the question becomes which variables to use as characteristics, $\mathbf{z}$.

In a paper similar in spirit to this one, but using data from the U.S., Serpa and Krishnan (2017) estimate TFP spillovers using a variety of firm-level variables as exogenous characteristics. These include the firms' financial leverage, liquidity, turnover, capital labor ratios, and measures of firm sizes, among others. One concern with these variables is that they are *choices* made by firms and are potentially correlated with other group, network-level, or individual-level unobservables that are correlated with average peers' productivity and influence individual productivity directly.

In this paper, we use industry-level exchange rate shocks as our source of exogenous variation in productivity to estimate TFP spillovers. To illustrate the identification strategy, consider Figure 5. In this diagram, firms are represented as nodes, and lines are drawn connecting them. Firms A, B, C, D, E, and F form a fully connected sub-graph because they are all located in the same city, Jakarta. Firms G, H, and I also form a fully connected sub-graph, being located in a different city, Bandung. Firm nodes are also colored to reflect the different industries to which they belong; firms B and C belong to industry 1, firm A belongs to industry 2, firms D and G belong to industry 3, firms F, E, and I belong to industry 4, and firm H belongs to industry 5.

Notice that firms D and G form a cross-city connection through their industrial relationship, as do firms F, E, and I. Because we ignore inter-industry connections in this diagram, this network corresponds to $\mathbf{G}\,(s, d = 0)$ if we let s approach ∞. If each industry had been entirely contained in a single location, and if we had let $g_{ii} = 1$, this would correspond exactly to Manski (1993)'s setting, where identification is impossible because of the reflection problem. However, in Figure 5, we highlight how observing firms from the same industry who appear in different locations creates intransitive triads that are useful for identification.

Focusing on firm H, the set of neighbors of H are the other firms located in Bandung ($\mathcal{N}(H) = \{G, I\}$), while the second and third-order neighbor sets are given by $\mathcal{N}_2(H) = \{D, E, F\}$ and $\mathcal{N}_3(H) = \{A, B, C\}$. Notice that firm H appears in several intransitive triads, such as H-I-F and H-G-D; these triads span cities and are formed through industry relationships. Bramoullé et al. (2009) characterize the set of networks in which identification can be achieved; in such networks, it must be possible to find intransitive triads like those involving firm H.

The Z's that appear in Figure 5 represent industry-specific exchange-rate shocks. To find an instrument that is correlated with H's neighbors' productivity ($\overline{\omega_{H(i)}} = 0.5\omega_I + 0.5\omega_G$) but only affects firm H through its effect on H's neighbors' productivity, consider the following:

$$Z^{\overline{\omega}}_{\mathcal{N}_2}(H) = \frac{1}{3}Z_3 + \frac{2}{3}Z_4$$
$$Z^{\overline{\omega}}_{\mathcal{N}_3}(H) = \frac{2}{3}Z_1 + \frac{1}{3}Z_2$$

These two variables are appropriately weighted averages of industry-level Z's. The differences in weights reflect differences in the numbers of firms in each industry who are in different neighbors' sets. Because we have several possible instruments, we use GMM to appropriately weigh the instruments and account for the natural heteroskedasticity and clustering in the data.[11]

[11]Note that Bramoullé et al. (2009) propose to use a generalized 2SLS procedure for estimating $\widehat{\theta}$, that uses moments from both the structural and reduced form equations. Such an approach may be pursued in future work, but ultimately, our standard

In our empirical work, we can vary s and d to look for different spillovers at different levels of connectedness, but only up to a point. As discussed in Section 2, when s increases in the firm network, this reduces the number of industrial connections, and this leads to greater transitivity and fewer intrasitive triads. However, as long as firms in different industries appear in different locations, identification can still be achieved. On the other hand, if we were to increase d to ∞, so that all firms in Indonesia interacted together, all firms would be directly connected and there would be no possibility for identifying spillovers between them.

3.1 Exchange Rate Shocks

We measure exchange rate shocks for industry i in year t as follows:

$$ER_{it} = \frac{\sum_c w_{ict} ER_{ct}}{\sum_c w_{ict}} \tag{7}$$

This is a weighted average of country-specific exchange rates, measured as the local currency units per Indonesian Rupiah (IDR), where the weights w_{ict} measure the share of industry i's exports from Indonesia that go to country c in year t. The country-industry weights are calculated from the United Nations' COMTRADE database and are based on 6-digit products mapped to the 5-digit industry codes we work with in the SI data. The exchange rate data are from the International Monetary Fund's *International Financial Statistics*. As an industry's trade-weighted exchange rate falls, the currency depreciates and exports become cheaper. This generates an increase in export demand and can lead to increased productivity through standard terms-of-trade and trade opening effects (Melitz, 2003).

Figure 6 plots ER_{it} in logs across years; each line drawn represents a separate industry exchange-rate path. The variation across industries and time is substantial, and the peak in 2008, reflecting the global recession, is easy to discern. Note that only 45 percent of the variation in these data can be explained by industry and year fixed effects.

4 Results

We begin our investigation of productivity spillovers by focusing on the network $\mathbf{G}^A$ $(s = 0.01, d = 0)$. As discussed in Section 3, we estimate productivity by first estimating firm-level production functions using the Olley and Pakes (1996) investment control function approach. In Table 4, we report parameter estimates from a regression of neighbors' average TFP residuals, $\mathbf{H}\omega$, on second-order and third-order neighbors' average exchange rate shocks. This is the first-stage relationship in our investigation of productivity spillovers.

In order to identify productivity spillovers, we need a strong relationship between second and third-order neighbors' average ER shocks and neighbors' average TFP. Columns 1 and 2 use district and industry fixed effects, in addition to year fixed effects, while columns 3 and 4 focus on firm and year fixed effects, thereby restricting the identification to incumbent firms. These regressions show a strong, negative, and statistically significant relationship between the ER shocks of neighbors-of-neighbors and

errors need to be corrected to account for the fact that ω_{it} and its local averages are generated in a first-step procedure.

neighbors' average productivity. As the trade-weighted exchange rates of neighbors-of-neighbors depreciate, average productivity of neighbors increases. Although we find significant impacts of the third-order ER shock, the F-stat of the regressions fall somewhat, leading us to just use the second-order ER shock as the primary instrumental variables specification.

Table 5 reports the baseline spillover regression estimates using firm-level panel data. Here, we are estimating (6), and columns 1 and 2 report fixed-effects least squares regressions, while columns 3 and 4 report IV-GMM regressions, where $\mathbf{H}\omega$ is instrumented using $\mathbf{H}^2\mathbf{z}$. All regressions include firm and year fixed effects. In Columns 1 and 3, we just report regressions of productivity on the purely exogenous ER shocks, while in columns 2 and 4, we include a few selected time-varying firm-level controls (and neighbors' averages of those controls) as exogenous characteristics. Note that we do not use second-order averages of these firm-level characteristics as instruments. Robust standard errors, clustered at the district-by-industry level, are reported in parentheses.[12]

Overall, the results point to significant productivity spillovers albeit smaller than comparable estimates in the literature. In general, the fixed-effects least squares regressions show a positive, statistically significant relationship between own productivity and neighboring firms' average productivity. Once we instrument for average productivity using the neighbors-of-neighbors ER shock, the coefficient estimates increase slightly and remain statistically significant. The Kleibergen-Paap Wald Rank F-Stat and other identification tests also suggest a well-identified first stage; not to mention, key insights hold up to weak instrument robust inference. However, the point estimates themselves are relatively smaller than similar estimates in the literature. From column 3, a one standard deviation increase in firms' average productivity ($\sigma \approx 0.85$) leads to a 1.2 percent increase in own productivity on average. This is much smaller than the approximately 10 percent increase in own productivity reported by Greenstone et al. (2010) and is an order of magnitude smaller than the estimate reported by Serpa and Krishnan (2017), both of whom focus on U.S. data.

The results in Table 5 provide suggestive evidence that TFP spillovers in Indonesia, and possibly in other developing countries, are much smaller than productivity spillovers in the U.S. and Europe. Our findings also stand in contrast to Chauvin et al. (2016), who argue that compared to the U.S., human capital spillovers and the relationship between wages and density are even stronger in China and India. Nevertheless, more research is needed to determine whether these differences with other studies are due to the different sources of identification or different settings or both.

Robustness. In Table 8, we report estimates of θ_1 after varying the network definition. As in Table 7, each entry reports a separate estimate of θ_1 from a different regression, but here, we fix the any connections network and vary s along the columns and d along the rows. Looking across all specifications of the network structure, θ_1 appears to be positively estimated, but the results vary in terms of their precision. When $s = 0$, there are probably too many industrial connections, so the point estimate, while positive, is only marginally significant. However, as s increases, the point estimates tend to fall slightly, but the confidence intervals overlap across all of these different specifications.

In Table 6, we explore the robustness of these main regression results to different TFP estimates.

[12] In Appendix Table B.1, we show that these main results are robust to two-way clustering at the district and industry-level, but the significance falls slightly.

Row 1 reports the main estimates of θ_1 from Table 5, Column 3, where the TFP residual was estimated using the Olley and Pakes (1996) investment control function approach. Rows 2-4 use the intermediate-inputs control function approach of (Levinsohn and Petrin, 2003) but vary the proxy variable used as the intermediate input. Row 5 uses the Wooldridge (2009) GMM approach, row 6 uses the Ackerberg et al. (2015) conditional control function approach, and row 7 uses the index number approach of Aw et al. (2001). With the exception of the Wooldridge (2009) TFP residual, all spillover estimates are positive, statistically significant, and similar in magnitude as those reported in Table 5.

Mechanisms. In order to shed light on the mechanisms driving our results above, we separately estimate productivity spillovers, θ_1, by network type (upstream, downstream, or all connections), and 2-digit industry. In Table 7, each entry is a separate estimate of θ_1 from a different IV-GMM regression with firm fixed effects. In the first row and first column, we again report the main estimate of θ_1 with the network given by $\mathbf{G}^A\,(s = 0.01, d = 0)$. Column 2 uses the network of upstream connections, $\mathbf{G}^U\,(s = 0.01, d = 0)$, to construct all network variables in the regression, while column 3 uses the network of downstream connections, $\mathbf{G}^D\,(s = 0.01, d = 0)$.

In row 1, we find that the overall productivity spillovers are driven by both upstream and downstream connections. However, we find more precisely estimated spillovers form downstream connections suggesting that as your network of suppliers grows more productive, so does your productivity. The slightly more limited spillovers from upstream productivity improvements suggests that forward linkages may be slightly weaker than backward ones in Indonesian manufacturing.

Rows 2-10 report separate estimates of θ_1 for each of broad 2-digit industries. We find the largest and most significant spillover impacts for furniture and wood products (ISIC 33) and for finished metal, machines, and electronics (ISIC 38); the other industries do not seem to have significant TFP spillovers. The importance of downstream spillovers for the finished metal, machines, and electronics is consistent with this industry having some of the strongest backward linkages as seen in Table 1. Moreover, this industry is particularly susceptible to competitive effects of trade openness (Amiti and Konings, 2007), which is precisely the source of variation identifying changes in neighbors' productivity.

In summary, we find that productivity spillovers between Indonesian manufacturers are positive and significant, but substantially smaller than similar estimates from firm-level data in the U.S. and Europe. We also find that the overall TFP spillovers tend to be driven by spillovers that come from downstream connections, and we find that finished metals, machines, and electronics as well as furniture and wood producers tend to be the most amenable to productivity spillovers.

5 Conclusion

In this paper, we have developed and implemented a new identification strategy that uses information on the network structure of producer relationships to provide estimates of the size and scale of productivity spillovers. Our strategy builds on that proposed by Bramoullé et al. (2009) for estimating peer effects, and is one of the first applications of this idea to the estimation of productivity spillovers. We clarify the identification arguments required for implementing this strategy, use panel data, and validate it using external, trade-weighted exchange-rate shocks.

Our results, which are based on data from Indonesia, suggest positive productivity spillovers between manufacturers. However, our estimates of TFP spillovers in Indonesia are considerably smaller than similar estimates based on firm-level data from the U.S. and Europe, and positive effects are only observed in a few industries. These relatively small estimates of TFP spillovers are consistent with other work on Indonesia's agglomerations. For instance, Amiti and Cameron (2007) find relatively small impacts of labor market pooling on wages, at least when compared to effects estimated using U.S. and European data. In a companion paper, we also find that the spatial concentration of Indonesian manufacturing firms is not consistently related to an industry's use of highly skilled workers (Rothenberg et al., 2016a). This may suggest that one of the most important drivers of agglomeration externalities, knowledge spillovers, may not be operating well in Indonesian cities.

One advantage to our identification approach is that it is broadly applicable to other contexts. Researchers with access to both input-output data, physical distance data, and firm-level panel data would be able to use this strategy in other contexts. More research is needed to determine whether the TFP spillovers estimated from this approach are similar to other TFP spillover estimates; implementing this approach with firm-level panel data in many countries would allow us to cleanly study differences between TFP spillovers in developed and developing countries.

References

ACEMOGLU, D., U. AKCIGIT, AND W. KERR (2016): "Networks and the Macroeconomy: An Empirical Exploration," *NBER Macroeconomics Annual*, 30, 273–335.

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 86, 2411–2451.

AMITI, M. AND L. CAMERON (2007): "Economic Geography and Wages," *Review of Economics and Statistics*, 89, 15–29.

AMITI, M. AND J. KONINGS (2007): "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *American Economic Review*, 97, 1611–1638.

AW, B. Y., X. CHEN, AND M. J. ROBERTS (2001): "Firm-level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing," *Journal of Development Economics*, 66, 51–86.

BLOOM, N., B. EIFERT, A. MAHAJAN, D. MCKENZIE, AND J. ROBERTS (2013): "Does Management Matter? Evidence from India," *Quarterly Journal of Economics*, 128, 203–224.

BRAMOULLÉ, Y., H. DJEBBARI, AND B. FORTIN (2009): "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 150, 41–55.

BRIN, S. AND L. PAGE (1998): "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks*, 30, 107–117.

CARVALHO, V. M. (2014): "From Micro to Macro via Production Networks," *Journal of Economic Perspectives*, 28, 23–48.

CHAMBERLAIN, G. (1992): "Comment: Sequential Moment Restrictions in Panel Data," *Journal of Business & Economic Statistics*, 10, 20–26.

CHAUVIN, J. P., E. GLAESER, Y. MA, AND K. TOBIO (2016): "What is Different About Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States," NBER Working Paper No. 22002.

CICCONE, A. AND R. E. HALL (1996): "Productivity and the Density of Economic Activity," *American Economic Review*, 86, 54–70.

COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2008): "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics*, 63, 723–742.

COMBES, P.-P., G. DURANTON, L. GOBILLON, AND S. ROUX (2010): "Estimating Agglomeration Effects with History, Geology, and Worker Fixed-Effects," in *Agglomeration Economics*, ed. by E. L. Glaeser, Chicago University Press, 15–65.

DE GIORGI, G., M. PELLIZZARI, AND S. REDAELLI (2010): "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2, 241–275.

DURANTON, G., J. V. HENDERSON, AND W. C. STRANGE (2015): "Chapter 5: The Empirics of Agglomeration Economies," in *Handbook of Regional and Urban Economics, Volume 5A*, ed. by P.-P. Combes and L. Gobillon, Elsevier, 247–348.

DURANTON, G. AND D. PUGA (2004): "Micro-foundations of Urban Agglomeration Economies," in *Handbook of Regional and Urban Economics, Vol. 4*, ed. by J. V. Henderson and J. F. Thisse, Elsevier, 2063–2117.

ELLISON, G. AND E. L. GLAESER (1999): "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review*, 89, 311–316.

FRUCHTERMAN, T. M. J. AND E. M. REINGOLD (1991): "Graph Drawing by Force-Directed Placement," *Software: Practice and Experience*, 21, 1129–1164.

GIBBONS, S. AND H. G. OVERMAN (2012): "Mostly Pointless Spatial Econometrics?" *Journal of Regional Science*, 52, 172–191.

GOLDSMITH-PINKHAM, P. AND G. W. IMBENS (2013): "Social Networks and the Identification of Peer Effects," *Journal of Business and Economic Statistics*, 31, 253–264.

GRAHAM, D. J., P. S. MELO, P. JIWATTANAKULPAISARN, AND R. B. NOLAND (2010): "Testing for Causality Between Productivity and Agglomeration Economies," *Journal of Regional Science*, 50, 935–951.

GREENSTONE, M., E. MORETTI, AND R. HORNBECK (2010): "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 118, 536–598.

HANSON, G. H. (1997): "Increasing Returns, Trade, and the Regional Structure of Wages," *The Economic Journal*, 107, 113–133.

HENDERSON, J. V. (2003): "Marshall's Scale Economies," *Journal of Urban Economics*, 53, 1–28.

KLINE, P. AND E. MORETTI (2014): "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development," *Annual Review of Economics*, 6, 629–662.

LEVINSOHN, J. AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–342.

MANSKI, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542.

——— (2000): "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14, 115–136.

MARSHAK, J. AND W. H. ANDREWS (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12, 143–205.

MARSHALL, A. (1890): *Principles of Economics*, London: Macmillan and Co.

MARTIN, P., T. MAYER, AND F. MAYNERIS (2011): "Spatial Concentration and Plant-Level Productivity in France," *Journal of Urban Economics*, 69, 182–195.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

MION, G. (2004): "Spatial Externalities and Empirical Analysis: the Case of Italy," *Journal of Urban Economics*, 56, 97–118.

OLLEY, G. S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

REDDING, S. AND D. STURM (2008): "The Costs of Remoteness: Evidence from German Division and Reunification," *American Economic Review*, 98, 1766–1797.

ROSENTHAL, S. S. AND W. C. STRANGE (2008): "The Attenuation of Human Capital Spillovers," *Journal of Urban Economics*, 64, 373–389.

ROTHENBERG, A. D., S. BAZZI, A. CHARI, AND S. NATARAJ (2016a): "Assessing the Spatial Concentration of Indonesia's Manufacturing Sector: Evidence from Three Decades," Working Paper.

——— (2016b): "When Regional Policies Fail: An Evaluation of Indonesia's Integrated Economic Development Zones," Working Paper.

SERPA, J. C. AND H. KRISHNAN (2017): "The Impact of Supply Chains on Firm-Level Productivity," *Management Science*, Forthcoming, XXX–XXX.

WOOLDRIDGE, J. M. (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 104, 112–114.

Table 1: Most Connected and Clustered Industries, $\mathbf{g}^U(s = 0.01)$

Panel A: Top-5 In Degree (Downstream Connections)	In Degree	Out Degree	Between Cent.
38120. Manufacture of furniture and fixtures primarily made of metal	51	31	0.028
38433. Manufacture of motor vehicle component and apparatus	41	55	0.068
38232. Manufacture of wood working machineries	39	1	0.000
38113. Manufacture of kitchen ware made of aluminium	38	19	0.032
38514. Manufacture of instruments for practicum purposes	37	6	0.004
Panel B: Top 5 Out Degree (Upstream Connections)			
38194. Manufacture of wire	18	73	0.021
35118. Manufacture of basic organic chemicals resulting special chemicals	20	71	0.028
35606. Manufacture of plastics bags, containers	25	70	0.029
35210. Manufacture of paints, varnishes and lacquers	10	67	0.009
35131. Manufacture of synthetic resins	17	65	0.036
Panel C: Top 5 Betweenness Centrality			
38433. Manufacture of motor vehicle component and apparatus	41	55	0.068
32411. Manufacture of footwear for daily use	32	43	0.044
35131. Manufacture of synthetic resins	17	65	0.036
33211. Manufacture of furniture and fixtures mainly made of wood	34	39	0.033
38247. Alteration and repair of special industrial machineries	33	26	0.033

Notes: Authors' calculations. Network connections are based on data from the 2000 products produced and raw materials used datasets.

Table 2: Industrial Connections: Network Statistics, $\mathbf{g}^A(s)$

	$s = 0$	$s = 0.01$	$s = 0.02$	$s = 0.03$	$s = 0.04$	$s = 0.05$
Number of Nodes	261	261	261	261	261	261
Number of Edges	19368	8763	7366	6561	5865	5384
Completeness	0.285	0.129	0.109	0.097	0.086	0.079
Transitivity (Percentage of All Possible 3-Way Triangles)	0.536	0.357	0.333	0.314	0.294	0.282
Diameter of Largest Connected Component	4	4	4	5	5	5
Average Shortest Path Length of Largest Connected Component	1.760	1.910	1.954	1.991	2.035	2.065

Notes: Authors' calculations. We report statistics on the networks defined by $\mathbf{g}^A(s)$, where this family of networks is defined in (2). Different network statistics appear in separate rows, and we vary s along the columns. Network connections are based on data from the 2000 products produced and raw materials used datasets.

Table 3: Firm Connections: Network Statistics, $\mathbf{G}^U(s, d)$

Panel A: $\mathbf{G}^U(s, d = 0)$	$s = 0$	$s = 0.01$	$s = 0.02$	$s = 0.03$	$s = 0.04$	$s = 0.05$
Number of Nodes	21834	21834	21834	21834	21834	21834
Number of Edges	2068307	1087461	955728	904535	870895	830650
Completeness	0.004	0.002	0.002	0.002	0.002	0.002
Transitivity (Percentage of All Possible 3-Way Triangles)	0.881	0.884	0.914	0.928	0.938	0.950
Diameter of Largest Connected Component	3	4	5	6	6	6
Average Shortest Path Length of Largest Connected Component	1.286	1.810	2.062	2.110	2.175	2.353
Panel B: $\mathbf{G}^U(s, d = 10)$	$s = 0$	$s = 0.01$	$s = 0.02$	$s = 0.03$	$s = 0.04$	$s = 0.05$
Number of Nodes	21834	21834	21834	21834	21834	21834
Number of Edges	3304546	1584033	1349266	1271461	1221146	1147757
Completeness	0.007	0.003	0.003	0.003	0.003	0.002
Transitivity (Percentage of All Possible 3-Way Triangles)	0.769	0.764	0.816	0.838	0.852	0.873
Diameter of Largest Connected Component	4	5	5	6	6	6
Average Shortest Path Length of Largest Connected Component	1.656	2.047	2.216	2.284	2.345	2.477

Notes: Authors' calculations. We report statistics on the networks defined by $\mathbf{G}^U(s, d)$, where this family of firm-level networks is defined in (3). Different network statistics appear in separate rows, and we vary s along the columns. Industrial relationships are based on data from the 2000 products produced and raw materials used datasets, while the physical relationships are drawn using district information based on the 1990 census.

Table 4: First Stage Regressions; Dependent Variable: $\mathbf{H}\omega$

	(1)	(2)	(3)	(4)
2ND-ORDER NEIGHBORS ER SHOCK ($\mathbf{G}^A$ $(s = 0.01, d = 0)$)	-0.108	-0.072	-0.109	-0.072
	(0.021)***	(0.009)***	(0.018)***	(0.008)***
3RD-ORDER NEIGHBORS ER SHOCK ($\mathbf{G}^A$ $(s = 0.01, d = 0)$)		-0.060		-0.065
		(0.029)**		(0.028)**
N	347672	347593	341238	341154
F STAT	98.64	72.92	89.81	66.19
ADJUSTED R^2	0.361	0.362	0.482	0.483
ADJUSTED R^2 (WITHIN)	0.041	0.042	0.036	0.037
YEAR FE	YES	YES	YES	YES
DISTRICT FE	YES	YES	.	.
INDUSTRY FE	YES	YES	.	.
FIRM FE	.	.	YES	YES

Notes: This table reports first-stage regression results of firms' average TFP, $\mathbf{H}\omega$, on neighbors-of-neighbors and 3rd-order neighbors average ER shocks. Own ER shocks and neighbors ER shocks are also included in the specification, but coefficient estimates are not shown. Columns 1 and 2 include district and industry fixed effects, in addition to year fixed effects, while columns 3 and 4 include firm and year fixed effects. Robust standard errors, clustered at the district-by-industry level, are reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 5: Spillover Regressions

	FELS		GMM	
	(1)	(2)	(3)	(4)
NEIGHBORS AVG TFP (OP, $\mathbf{G}^A$ $(s = 0.01, d = 0)$)	0.074	0.071	0.120	0.127
	(0.009)***	(0.009)***	(0.040)***	(0.052)**
ER SHOCK	0.006	0.007	0.010	0.010
	(0.004)*	(0.004)*	(0.005)**	(0.004)**
NEIGHBORS ER SHOCK ($\mathbf{G}^A$ $(s = 0.01, d = 0)$)	-0.002	0.000	-0.008	-0.006
	(0.004)	(0.004)	(0.008)	(0.007)
TOTAL WORKERS (PAID AND UNPAID)		-0.059		-0.059
		(0.009)***		(0.009)***
EXPORTER (0 1)		0.057		0.058
		(0.011)***		(0.011)***
FOREIGN OWNED (0 1)		0.138		0.137
		(0.028)***		(0.027)***
N	198593	189322	197281	188017
ADJUSTED R^2	0.632	0.643	0.632	0.643
ADJUSTED R^2 (WITHIN)	0.004	0.005	0.002	0.003
KLEIBERGEN-PAAP WALD RANK F STAT			65.534	52.090
UNDER ID. TEST (KP RANK LM STAT)			99.763	80.550
P-VALUE			0.000	0.000
ANDERSON-RUBIN WALD TEST (WEAK IV ROBUST INF.)			8.230	5.664
P-VALUE			0.004	0.017
FIRM FE	YES	YES	YES	YES
YEAR FE	YES	YES	YES	YES
FIRM CONTROLS	.	YES	.	YES

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column). Robust standard errors, clustered at the district-by-industry level, are reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 6: Spillover Regressions: Robustness to Different TFP Measures

	(1)
Olley-Pakes (1996) (Investment Proxy)	0.120 (0.040)***
Levinson-Petrin (2003) (Electricity as Proxy Variable)	0.141 (0.050)***
Levinson-Petrin (2003) (Raw Materials as Proxy Variable)	0.113 (0.042)***
Levinson-Petrin (2003) (Electricity and Raw Materials as Proxy Variables)	0.117 (0.045)***
Wooldridge (2009)	-5.041 (11.510)
Ackerberg, Caves, and Frazer (2015)	0.075 (0.030)**
Aw, Chen, and Roberts (1991) Index-Number Method	0.175 (0.077)**
Year FE	Yes
District FE	.
Industry FE	.
Firm FE	Yes

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column). Robust standard errors, clustered at the district-by-industry level, are reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 7: Spillover Regressions: By Industry and Network Type

	All	Up	Down
	(1)	(2)	(3)
All Firms	0.120 (0.040)***	0.076 (0.051)	0.093 (0.039)**
31 - Food Processing	0.100 (0.072)	0.154 (0.100)	-0.011 (0.073)
32 - Textiles and Garments	-0.021 (0.065)	-0.056 (0.077)	0.010 (0.075)
33 - Furniture and Wood Products	0.340 (0.067)***	0.333 (0.071)***	0.346 (0.077)***
34 - Paper Products	-0.017 (0.076)	-0.053 (0.075)	0.006 (0.068)
35 - Chemical Products	0.107 (0.075)	0.078 (0.088)	0.043 (0.067)
36 - Ceramics, Glass, Cement, and Clay Products	0.084 (0.156)	0.135 (0.361)	0.010 (0.145)
37 - Iron and Steel	-0.037 (0.368)	0.632 (0.632)	-0.039 (0.282)
38 - Finished Metal, Machines, and Electronics	0.166 (0.072)**	0.179 (0.082)**	0.167 (0.072)**
39 - Other Manufacturing	0.112 (0.139)	0.042 (0.133)	0.766 (0.587)
Year FE	Yes	Yes	Yes
District FE	.	.	.
Industry FE	.	.	.
Firm FE	Yes	Yes	Yes

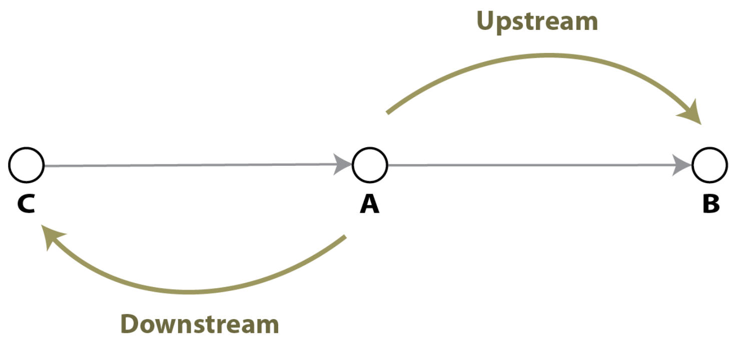
Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column). Robust standard errors, clustered at the district-by-industry level, are reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Table 8: Spillover Regressions: By Network Definition

	GMM					
	$s = 0$	$s = 0.01$	$s = 0.02$	$s = 0.03$	$s = 0.04$	$s = 0.05$
	(1)	(2)	(3)	(4)	(5)	(6)
Neighbors' Average TFP (d = 0)	0.092 (0.052)*	0.120 (0.040)***	0.078 (0.044)*	0.083 (0.045)*	0.082 (0.044)*	0.073 (0.042)*
Neighbors' Average TFP (d = 10)	0.092 (0.049)*	0.110 (0.038)***	0.078 (0.041)*	0.083 (0.043)*	0.087 (0.042)**	0.074 (0.042)*
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes

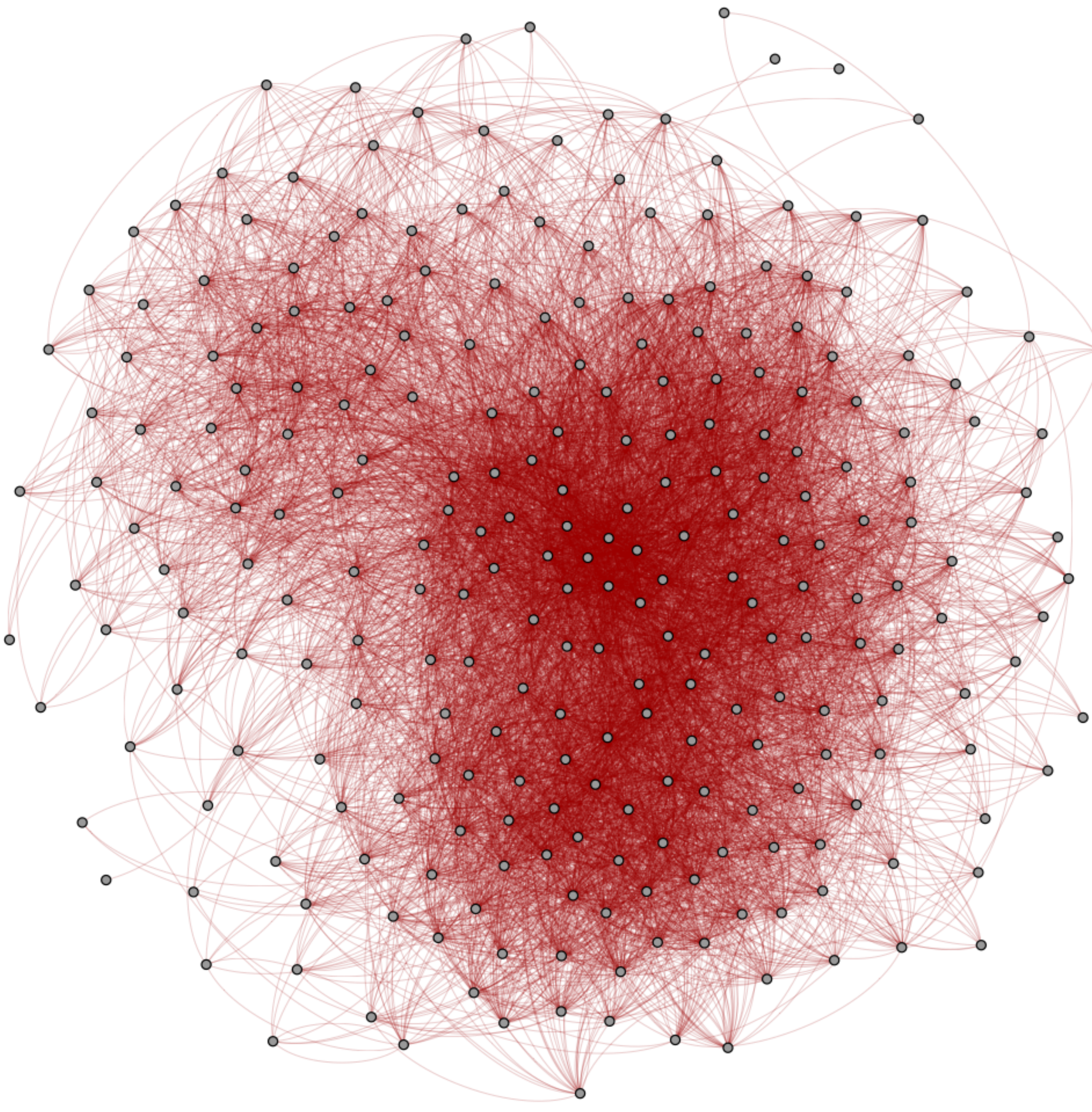
Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column). Robust standard errors, clustered at the district-by-industry level, are reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.

Figure 1: Definitions: Upstream and Downstream Connections



Notes: This image depicts the upstream and downstream connections between industries A, B, and C. Relative to industry A, industry B is an upstream connection, while industry C is a downstream connection.

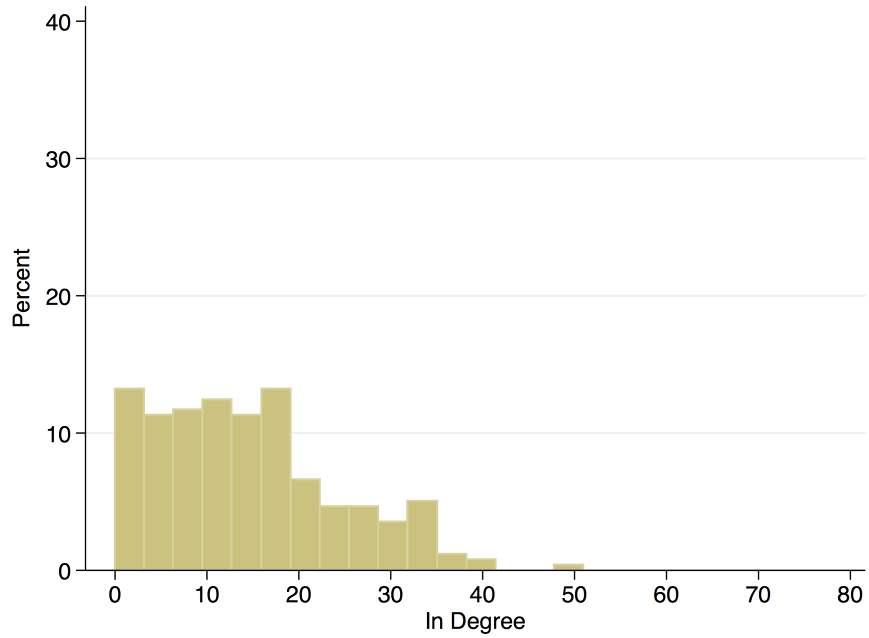
Figure 2: Network of 5-Digit Industries, Any Upstream or Downstream Connection; $\mathbf{g}^A\,(s = 0.01)$



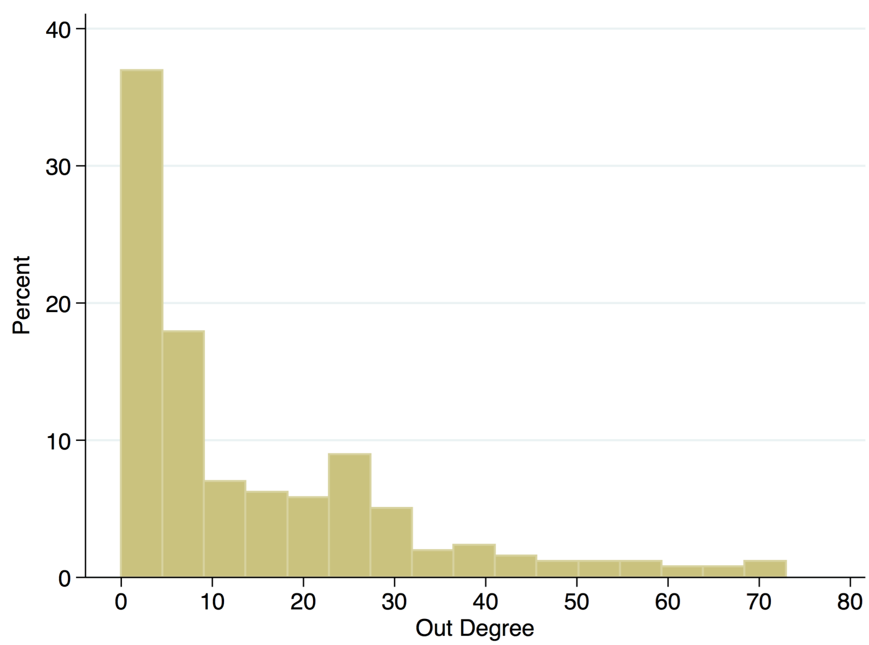
Notes: Authors' calculations. Network connections are based on data from the 2000 products produced and raw materials used datasets. Visualization uses the force-placement algorithm of Fruchterman and Reingold (1991). To simplify the image, we do not draw the direction of the connections between firms.

Figure 3: Network of 5-Digit Industries: Degree Distributions of g^U $(s = 0.01)$

(A) In Degree Distribution (Number of Downstream Connections)

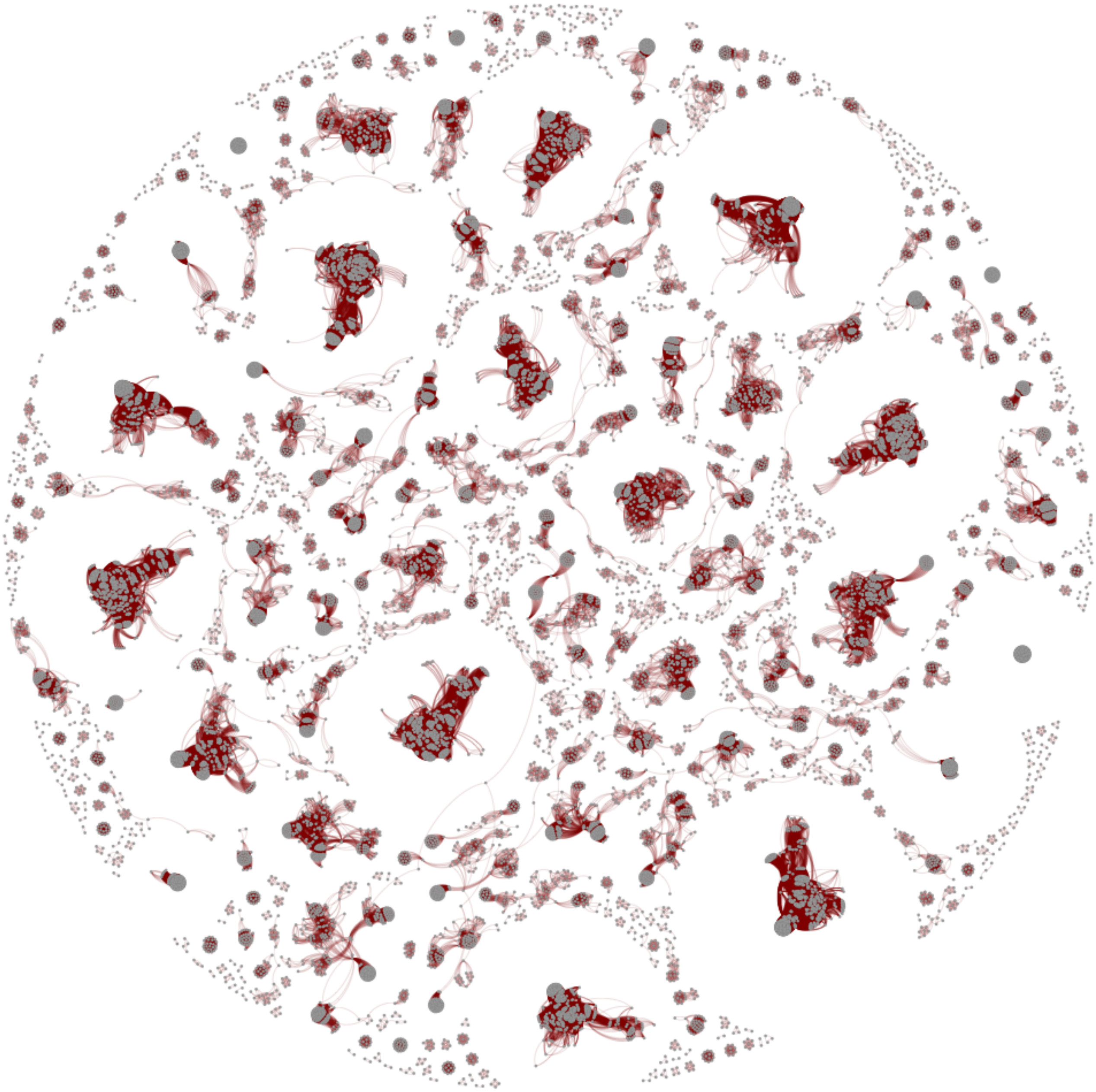


(B) Out Degree Distribution (Number of Upstream Connections)



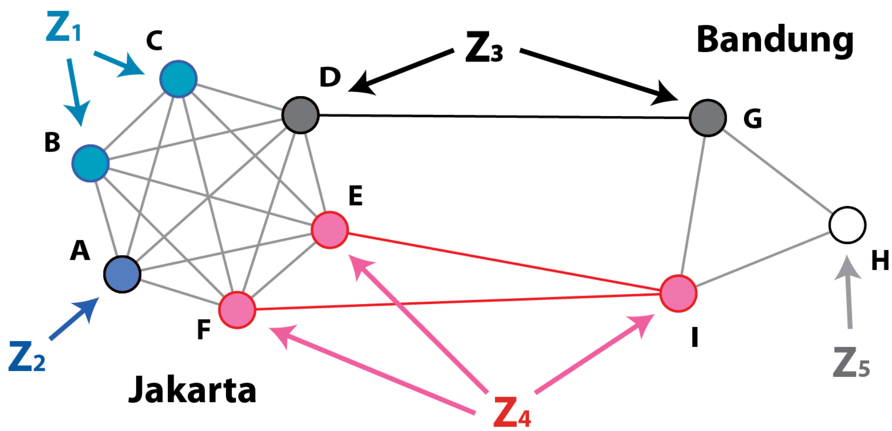
Notes: Authors' calculations. Network connections are based on data from the 2000 products produced and raw materials used datasets.

Figure 4: Network of Firms ($\mathbf{G}^U(s = 0.01, d = 0)$)



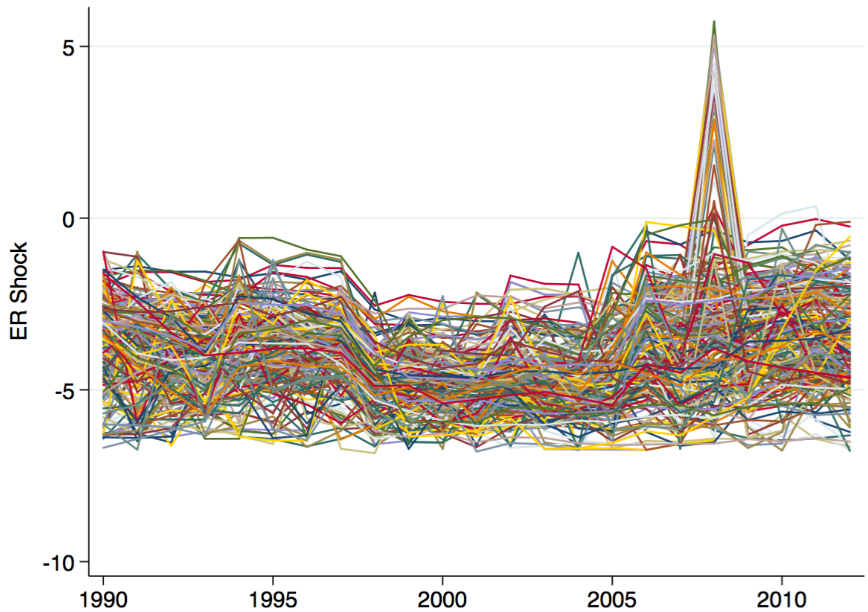
Source: Authors' calculations. Visualization uses the force-placement algorithm of Fruchterman and Reingold (1991).

Figure 5: Identification Diagram



Notes: This image depicts two cities, Jakarta and Bandung, and 9 firms belonging to 5 different industries. Firms are depicted as nodes in the network, and lines drawn between nodes form network connections. There are distance-based connections (in gray) and industry-based connections that span cities (in red and black).

Figure 6: Variation in Exchange Rates



Notes: This image depicts ER_{it} across industries i and years t, where ER_{it} is defined in (7). A separate line is drawn for each industry history.

A Production Function Estimation

Wooldridge (2009) shows how to recast the control function approach for estimating production functions (e.g. Ackerberg et al., 2015; Levinsohn and Petrin, 2003; Olley and Pakes, 1996) in a generalized method of moments (GMM) framework. Following his notation, we write the firm-level production function as follows:

$$y_{it} = \alpha + \mathbf{w}_{it}'\beta + \mathbf{x}_{it}'\gamma + v_{it} + e_{it}, \qquad t = 1, ..., T \tag{8}$$

where we define:

- y_{it}: Log Value Added
- $\mathbf{w}_{it}$: a $(J \times 1)$ vector of "adjustable" inputs, including:
  - LP_{it}: Log Total Production Workers
  - LNP_{it}: Log Total Non-Production Workers
- $\mathbf{x}_{it}$: a $(K \times 1)$ vector of state variables, including:
  - k_{it}: Log Capital (book value, but estimated if missing)
- v_{it}: transmitted component of the error term. This is a state variable, observed by the firm before its decisions about labor and capital are made.
- e_{it}: an i.i.d. error, not observed by the firm before decisions are made.

A major concern with estimating (8) directly, either with OLS or fixed effects, is that because the firm observes v_{it} but we do not, v_{it} is an omitted variable correlated with choices of labor and capital. This will lead to biased estimates of the production function parameters, and by implication, biased estimates of firm-level productivity residuals.

To resolve this, Levinsohn and Petrin (2003) use a $(M \times 1)$ vector of proxy variables, $\mathbf{m}_{it}$, to control for the correlation between input levels and unobserved productivity. They assume that intermediate input demand is given by:

$$\mathbf{m}_{it} = m\left(\mathbf{x}_{it}, v_{it}\right), \qquad t = 1, ..., T$$

and that this vector of functions is monotonic in v_{it}. Monotonicity allows you to invert this function and solve for v_{it}:

$$v_{it} = g\left(\mathbf{x}_{it}, \mathbf{m}_{it}\right), \qquad t = 1, ..., T$$

This gives us our first equation for identifying the production function parameters:

$$y_{it} = \alpha + \mathbf{w}_{it}'\beta + \mathbf{x}_{it}'\gamma + g\left(\mathbf{x}_{it}, \mathbf{m}_{it}\right) + e_{it}, \qquad t = 1, ..., T \tag{9}$$

where Wooldridge (2009) assumes that e_{it} is sequentially exogenous (Chamberlain, 1992), mean zero conditional on the history of $\mathbf{w}$, $\mathbf{x}$, and $\mathbf{m}$ up to this point:

$$\mathbb{E}\left[e_{it} \,\middle|\, \left(\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}\right), \left(\mathbf{w}_{it-1}, \mathbf{x}_{it-1}, \mathbf{m}_{it-1}\right), ..., \left(\mathbf{w}_{i1}, \mathbf{x}_{i1}, \mathbf{m}_{i1}\right)\right] = 0, \qquad t = 1, ..., T \tag{10}$$

According to Olley and Pakes (1996) and Levinsohn and Petrin (2003), (9) and (10) identifies β, but Ackerberg et al. (2015) argues that this equation has no identifying information if intermediate inputs and labor are chosen at the same time.

To identify γ (or, if you believe the critique of Ackerberg et al. (2015), β and γ together), we need to make stronger assumptions on the dynamics of the productivity process. Define the current period's innovation in productivity as:

$$a_{it} \equiv v_{it} - \mathbb{E}\left[\, v_{it} \,|\, v_{it-1} \,\right]$$

where we have:

$$\begin{aligned}\mathbb{E}\left[\, v_{it} \,|\, v_{it-1} \,\right] &= f\left[v_{it-1}\right] \\ &= f\left[g\left(\mathbf{x}_{it-1}, \mathbf{m}_{it-1}\right)\right]\end{aligned}$$

Adding and subtracting this term in (9), we write:

$$y_{it} = \alpha + \mathbf{w}_{it}'\beta + \mathbf{x}_{it}'\gamma + f\left[g\left(\mathbf{x}_{it-1}, \mathbf{m}_{it-1}\right)\right] + u_{it}, \qquad t = 1, ..., T \tag{11}$$

where $u_{it} \equiv a_{it} + e_{it}$. Wooldridge (2009) argues that to identify α, β, and γ, we need to assume:

$$\mathbb{E}\left[\, u_{it} \,\middle|\, \mathbf{x}_{it}, \left(\mathbf{w}_{it-1}, \mathbf{x}_{it-1}, \mathbf{m}_{it-1}\right), ..., \left(\mathbf{w}_{i1}, \mathbf{x}_{i1}, \mathbf{m}_{i1}\right)\right] = 0, \qquad t = 1, ..., T \tag{12}$$

Equations (9) and (11), together with their respective moment restrictions (10) and (12), identify α, β and γ. Wooldridge (2009) proposes a GMM approach to estimate α, β, and γ that involves approximating f and g with flexible polynomials and stacking sample versions of the moment restrictions 10 and 12 in a GMM objective function.

B Robustness Tables and Figures

Table B.1: Spillover Regressions (2-Way Clustering)

	FELS		GMM	
	(1)	**(2)**	**(3)**	**(4)**
Neighbors Avg TFP (OP, $\mathbf{G}^A$ $(s = 0.01, d = 0)$)	0.074	0.071	0.120	0.127
	(0.009)***	(0.009)***	(0.040)***	(0.052)**
ER Shock	0.006	0.007	0.010	0.010
	(0.004)*	(0.004)*	(0.005)**	(0.004)**
Neighbors ER Shock ($\mathbf{G}^A$ $(s = 0.01, d = 0)$)	-0.002	0.000	-0.008	-0.006
	(0.004)	(0.004)	(0.008)	(0.007)
Total Workers (Paid and Unpaid)		-0.059		-0.059
		(0.009)***		(0.009)***
Exporter (0 1)		0.057		0.058
		(0.011)***		(0.011)***
Foreign Owned (0 1)		0.138		0.137
		(0.028)***		(0.027)***
N	198593	189322	197281	188017
Adjusted R^2	0.632	0.643	0.632	0.643
Adjusted R^2 (Within)	0.004	0.005	0.002	0.003
Kleibergen-Paap Wald Rank F Stat			65.534	52.090
Under Id. Test (KP Rank LM Stat)			99.763	80.550
p-Value			0.000	0.000
Anderson-Rubin Wald Test (Weak IV Robust Inf.)			8.230	5.664
p-Value			0.004	0.017
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm Controls	.	Yes	.	Yes

Notes: Each cell reports the coefficient from a regression of the given dependent variable (listed in the left-most column). Robust standard errors, two-way clustered by district and industry, are reported in parentheses. */**/*** denotes significant at the 10% / 5% / 1% levels.