AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

On the Estimation of Cross-Firm Productivity Spillovers with an Application to FDI

Emir Malikov, University of Nevada, Las Vegas, emir.malikov@unlv.edu

Invited Paper prepared for presentation at the **2022 AEA/ASSA Annual Meeting VIRTUAL, January 7-9, 2022**

*Copyright 2022 by **Emir Malikov and Shunan Zhao**. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.*

On the Estimation of Cross-Firm Productivity Spillovers with an Application to FDI*

EMIR MALIKOV[1] SHUNAN ZHAO[2]

[1]University of Nevada, Las Vegas
[2]Oakland University

June 20, 2021

Abstract

We develop a novel methodology for the proxy variable identification of firm productivity in the presence of productivity-modifying learning and spillovers which facilitates a unified "internally consistent" analysis of the spillover effects between firms. Contrary to the popular two-step empirical approach, ours does not postulate contradictory assumptions about firm productivity across the estimation steps. Instead, we explicitly accommodate cross-sectional dependence in productivity induced by spillovers which facilitates identification of both the productivity and spillover effects therein simultaneously. We apply our model to study cross-firm spillovers in China's electric machinery manufacturing, with a particular focus on productivity effects of inbound FDI.

Keywords: productivity spillovers, production function, proxy variable, FDI spillovers

JEL Classification: C14, C23, D24, F21, L20, O30

*Email: emir.malikov@unlv.edu (Malikov) and shunanzhao@oakland.edu (Zhao).

1 Introduction

Since its popularization by Marshall (1890), the concept of cross-firm knowledge, or technology, spillovers has increasingly become a central fixture in many economic theories, including of long-run growth, spatial agglomeration, research and innovation, international trade and more. The idea is simple: firms improve their productivity by learning from one another, with the most commonly conjectured drivers of these knowledge exchanges (technology transfers) being human interaction along with spatial and industrial/technological proximity. These productivity spillovers can also propel significant positive externalities in many productivity-enhancing activities such as research and development (R&D), foreign direct investment (FDI) or exporting. In this paper, we develop a new methodology for the proxy variable structural identification of firm productivity in the presence of productivity-modifying learning and spillovers which facilitates a unified "internally consistent" analysis of the spillover effects between peer firms.

Although productivity is straightforward in concept, its measurement is not trivial for a multitude of reasons among which is the inherent latency of firm productivity/efficiency. Naturally, the identification of productivity spillovers across firms is even more challenging a task because, as Krugman (1991, p.53) points out, "knowledge flows ... are invisible; they leave no paper trail by which they may be measured and tracked." On this account, most empirical work on cross-firm technology spillovers abstracts away from pinpointing specific mechanisms by which such spillovers occur[1] and instead focuses on a simpler but more feasible objective of testing for the presence of cross-firm spillovers in general. The most common frameworks either (i) focus squarely on "productivity spillovers" by seeking to identify how a firm's productivity is affected by that of its peers—the "endogenous effect" in the Manski (1993) nomenclature—or (ii) take a more reduced-form approach centered only on measuring "contextual" spillover effects of

[1]Nuanced empirical studies of spillovers are rare and require detailed matched (and usually proprietary or confidential) datasets and, by design, have a limited identifying ability restricted to particular channels/mechanisms of knowledge diffusion such as labor turnover (e.g., Balsvik, 2011; Poole, 2013; Stoyanov & Zubanov, 2012) or coauthorship networks (e.g., Zacchia, 2020). Others resort to limiting the scope of analyzed spillovers: e.g., Jiang et al. (2019) restrict FDI spillovers to direct links between joint ventures and their domestic partners, whereas Newman et al. (2015) focus on productivity spillovers along vertical supply chains.

various productivity-modifying activities (FDI, R&D, exporting, etc.) facilitated by cross-firm spillovers in productivity. Recent examples of the first include Bazzi et al. (2017) and Serpa & Krishnan (2018) who study vertical productivity spillovers along supply chains and material-product connections. As it happens, the literature embracing the second framework is more predominant and has a longer history: e.g., see Alvarez & López (2008) on spillovers from exporting; Javorcik (2004), Javorcik & Spatareanu (2008), Haskel et al. (2007), Blalock & Gertler (2008), Keller & Yeaple (2009), Barrios et al. (2011), Lu et al. (2017) on FDI spillovers; Branstetter (2001), Griffith et al. (2006), Bloom et al. (2013), Zacchia (2020) on technology spillovers from R&D; and Acharya & Keller (2008) on productivity spillover effects of imports.

Both empirical frameworks are usually operationalized in two steps, whereby one first recovers firm productivity from the production function estimates and then examines spillovers in the second step by (linearly) regressing these productivity estimates on various peer-group averages capturing firms' exposure to potential spillovers. Owing to its popularity and ease of implementation, most studies estimate firm productivity in the first step via the proxy variable approach à la Olley & Pakes (1996) and Levinsohn & Petrin (2003) that typically assumes that each firm's productivity process is an independent (over firms) exogenous Markov chain. However, if present, spillovers would generate cross-sectional dependence among firms, which is nonetheless being overlooked in the first-step estimation of productivity. Not only does this raise reservations about the identification of production function (and hence, productivity) econometrically, but more importantly, such a two-step procedure suffers from the conceptual "internal inconsistency" because the second-step regressions, in effect, contradictorily postulate the existence of spillover-induced cross-firm dependence in productivity which is at odds with the identifying assumptions used in the first step. As such, conclusions about spillovers based on a two-step procedure may be spurious.

With the above in mind, we provide a novel (semiparametric) methodology for the estimation of productivity spillovers. In line with the existence of cross-firm spillovers, in building our model, we explicitly accommodate cross-sectional peer dependence in firm-specific (latent)

productivity that such spillovers induce. This is fundamentally different from the aforementioned traditional two-step approach.[2]

To keep our methodology amenable to a wide range of contexts, we conceptualize peer dependence in firm performance via spatiotemporal spillovers in latent productivity itself. We generalize the conventional setup of firm production assumed in the literature to introduce the dependence of each firm's productivity on its (geographically and industrially proximate) peer-group average productivity. To that end, we dispense with the standard assumption of independent (over i) exogenous Markov process for latent productivity (e.g., Olley & Pakes, 1996; Levinsohn & Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020, and others) in favor of a controlled productivity process with explicitly incorporated cross-sectional dependence. This permits the firm to improve its productivity by learning not only directly from its own productivity-modifying activities but also indirectly from the activities of its peers.

Explicit modeling of cross-sectional dependence in firm productivity directly affecting its evolution (along with a structural timing assumption about learning process) enables us to build upon the popular proxy variable technique to develop a unified identification scheme for both the latent productivity and spillover effects therein *simultaneously* that is also robust to Ackerberg et al.'s (2015) and Gandhi et al.'s (2020) critiques. In fact, as we show in the paper, estimating the firm production function or productivity using traditional proxy methods while ignoring the spillover-induced cross-sectional dependence, as customarily done in the literature, likely leads to misspecification and omitted variable bias. This underscores the key practical advantage of our proposed methodology. Also, by virtue of a nonparametric formulation of the firm productivity process, we transcend restrictive additively linear specifications favored in the spillovers literature which lets us accommodate heterogeneous spillover effects.

Because our methodology can be easily adapted to admit various spillover origins, it is fit to investigate productivity spillovers in many contexts, including spatial agglomeration, R&D

[2]We should note that a two-step framework is not universal across empirical studies of productivity spillovers. The exceptions are predominantly from the literature on R&D-borne spillovers that favors the estimation of "augmented production functions." We discuss the benefits of our methodology over the latter in Appendix A.

externalities, learning from exporters, and others. In our paper, for example, we consider an application to the FDI inflows.

We also contribute to the literature on proxy-based identification of production functions more broadly, by providing a practical, easy-to-implement semiparametric adaptation of Gandhi et al.'s (2020) estimator. Our point of departure is a parametric assumption about the functional form of production function, which is the predominant modeling strategy in productivity literature with the Cobb-Douglas specification being the most popular among researchers. Along the lines of Doraszelski & Jaumandreu (2013), our modeling approach fully embraces the assumed parametric specification of the production function by explicitly utilizing a known functional form of the static first-order condition for materials and the inverse conditional input demand function that it implies. By doing so, we circumvent the need to integrate the estimated material elasticity function at *each* observation in order to recover the unknown production function required by Gandhi et al.'s (2020) more computationally demanding, albeit admittedly less restrictive, nonparametric methodology. In contrast, our parametric inversion of the material demand yields a much simpler semiparametric estimator. We also show how to extend our methodology to more flexible specifications of the firm's production function such as translog.

Besides the empirical application of our methodology, we also demonstrate its ability to successfully identify firm productivity and cross-firm spillovers therein in a set of Monte Carlo experiments. The results are encouraging and show that our approach recovers the true parameters well, thereby lending strong support to the validity of our identification strategy. We also use the simulations to show how estimating spillovers via the popular but internally inconsistent two-step procedure can lead to spurious and misleading results.

The rest of the paper unfolds as follows. Section 2 provides context for our application centered on FDI spillovers. Section 3 describes a generic model of firm production with productivity-modifying learning and spillovers. We discuss identification and estimation in Sections 4 and 5, respectively. Section 6 reports simulation results. We present our empirical application in Section 7 and conclude in Section 8.

2 Application to FDI

Cross-firm productivity spillovers can propel significant positive externalities in many productivity-enhancing activities, which are especially important from a policy perspective. Take, for instance, inbound foreign direct investment.

Public policies aimed at attracting FDI are commonplace both in developing and developed economies. Besides immediate returns in the form of capital inflows and employment gains, the primary justification for the FDI-promoting government incentives is mostly centered around gaining access to intangible productive "knowledge" assets from abroad such as new technologies, proprietary know-hows, more efficient and innovative marketing and management practices, established relational networks, reputation, etc., which can boost productivity of domestic firms. More crucially, productivity-enhancing effects of inbound FDI are widely believed to realize broadly beyond immediate recipients, who benefit from direct learning of foreign knowledge, by also benefiting many other domestic firms via productivity spillovers. These spillovers may occur via informal contacts (e.g., attendance of trade shows, exposure to affiliate and/or competitor products and marketing, learning by imitation, customer-supplier discussions), more formal reverse engineering, or labor turnover, eventually yielding large within- and/or cross-industry productivity gains. Measuring the extent and significance of these "social returns" of FDI is therefore imperative for the design of effective industrial policy.

To empirically showcase our estimator, we apply it to study horizontal productivity spillovers in China's electric machinery manufacturing industry in 1998–2007, with a particular focus on the technology-transfer effects of inbound FDI on productivity via domestic firms' learning of more advanced/efficient foreign knowledge to which they may gain access directly through their *own* foreign investors and indirectly through spillovers from their foreign-invested *peers*. Among the world's top destinations for foreign investment, China presents a natural environment for the analysis of broad productivity effects of FDI on domestic firms especially because of its "open door" policies aimed at promoting foreign investment (e.g., special economic zones

with regulatory environments favorable to foreign capital) and its fairly recent accession to the World Trade Organization in 2001. Focus on the electric machinery industry in particular is motivated by its being historically one of China's most fundamental manufacturing sectors and among the largest FDI recipients (see Appendix B for more on the choice of the industry).

The empirical literature on FDI spillovers has generally produced mixed findings, especially for the long-sought-after horizontal productivity spillovers (see Keller, 2008, 2010, for excellent surveys). Few earlier studies that have analyzed external productivity spillovers from inbound FDI in China (see Jiang et al., 2019, and the references therein) have done so using the two-step approach or by "augmenting" the firm's production function with the associated methodological issues as discussed earlier. The results have been mixed, further heightening the appeal of our study based on a new methodology. The reanalysis of FDI-borne technology spillovers in China is also timely and relevant in light of the ongoing trade disputes between the U.S. and China fostered, among other things, by grievances of the former against China's "unfair technology transfer regime" for foreign companies. Investigating the extent of external spillovers can therefore provide an informative context for a more holistic understanding of the FDI environment in the country and the implications of its technology-transfer rules and regulations.

To briefly preview our key results, we find that at least 87% of manufacturers of electric machinery in China enjoy significant productivity-boosting effects of inbound FDI, both directly and indirectly. At the median, an increase of the foreign share in all firms' equity by 10 percentage points, in the short run, improves each firm's productivity by 1.4% via direct learning and by 0.4% via external effects. The latter indirect effect of FDI is facilitated by substantial cross-firm productivity spillovers in the industry, with the median spillover elasticity estimated at 0.33. These productivity spillovers are significantly positive for about 84% of firms in the industry.

3 Production with Learning and Spillovers

We now describe a generic paradigm of production in the presence of productivity-modifying learning and spillovers. Consider the production process of a firm $i = 1, \ldots, n$ in time period

$t = 1, \ldots, T$ in which physical capital K_{it}, labor L_{it} and an intermediate input such as materials M_{it} are transformed into the output Y_{it} via production function, given the log-additive Hicks-neutral firm productivity. Following the popular convention in the literature (e.g., Olley & Pakes, 1996; Levinsohn & Petrin, 2003; Topalova & Khandelwal, 2011; Doraszelski & Jaumandreu, 2013), we assume that the firm's stochastic production process is Cobb-Douglas:

$$Y_{it} = K_{it}^{\beta_K} L_{it}^{\beta_L} M_{it}^{\beta_M} \exp\left\{\omega_{it} + \eta_{it}\right\}, \tag{3.1}$$

where the exponent $\omega_{it} + \eta_{it}$ is the latent "composite" productivity residual consisting of (i) the firm i's persistent productivity ω_{it} and (ii) a random transitory productivity shock η_{it}. Our methodology can also adopt more flexible specifications of the firm's production function such as the log-quadratic translog specification. See Appendix C for this extension of our model.

We assume that K_{it} and L_{it} are subject to adjustment frictions (e.g., time-to-install, hiring and training costs) and thus are quasi-fixed, whereas M_{it} is freely varying. That is, M_{it} is chosen in period t, whereas K_{it} and L_{it} are determined in period $t-1$. Both K_{it} and L_{it} are the state variables with dynamic implications and follow their respective deterministic laws of motion:

$$K_{it} = I_{it-1} + (1-\delta)K_{it-1} \quad \text{and} \quad L_{it} = H_{it-1} + L_{it-1}, \tag{3.2}$$

where I_{it}, H_{it} and δ are the gross investment, net hiring and the depreciation rate, respectively. The firm maximizes a discounted stream of expected life-time profits in perfectly competitive output and factor markets. Also, for convenience, let $\mathscr{I}_{it}$ denote the information set available to the firm i for making period t production decisions.

Our main objective is to study the role of learning and spillovers in the evolution of firm productivity. To that end, we need to dispense with the standard assumption of *exogenous* Markov process for ω_{it} in favor of a *controlled* productivity process and, more importantly, to explicitly recognize the potential for cross-sectional dependence therein. For generality sake, we denote the productivity-modifying "controls" via a (vector of) generic variable(s) G_{it}. This variable may measure the firm's deliberate activities aimed at improving its productivity such as R&D expenditures (Doraszelski & Jaumandreu, 2013) or some other aspects of its behavior in the market-

place that have productivity implications such as exporting (De Loecker, 2013). Depending on the application of interest, G_{it} may also admit measures of the firm's exposure to technological innovations from investors or partners—the focus of our empirical illustration—or its access to public subsidies and other forms of favorable treatment from the government owing to political connections, etc. In the end, no matter the choice of G_{it}, the rationale for its inclusion in the firm productivity evolution is to capture within-firm "learning" facilitated by the firm's own productivity-modifying activities or characteristics.

Next, we permit the firm i to improve its productivity by learning not only from its *own* activities but also from its *peer* firms. We do so by relaxing the usual assumption of firm productivity being an independent (over i) Markov chain to allow for cross-sectional dependence. More concretely, we assume that the ith firm's productivity ω_{it} evolves according to the following controlled first-order process:

$$\omega_{it} = \mathbb{E}\left[\omega_{it} \middle| \omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}\right] + \zeta_{it}, \tag{3.3}$$

where $\{s_{ij,t-1};\ j(\neq i) = 1, \ldots, n\}$ are the peer-identifying weights (from the perspective of firm i), and ζ_{it} is a mean-independent unanticipated random innovation in persistent productivity normalized to have a zero mean: $\mathbb{E}[\zeta_{it}|\mathscr{I}_{i,t-1}] = \mathbb{E}[\zeta_{it}|\omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}] = 0$.

While the exact choice of how to construct peer weights $\{s_{ijt}\}$ depends on the empirical context, for a general baseline case here, we let the peers be identified based on their spatial vicinity and industrial similarity to firm i. Thus, letting $\mathscr{L}(i,t)$ represent a set of spatially proximate "neighbors" of the firm i in time period t that also operate in the same industry, peer weights $\{s_{ijt}\}$ are constructed for each (i,t) as follows:

$$s_{ijt} = \frac{\mathbb{1}\{(j,t) \in \mathscr{L}(i,t)\}}{\sum_{k(\neq i)=1}^{n} \mathbb{1}\{(k,t) \in \mathscr{L}(i,t)\}}, \tag{3.4}$$

where the normalization in the denominator yields a convenient interpretation of $\sum_{j(\neq i)} s_{ijt}\omega_{jt}$ as the average peer productivity. Focusing on geographically proximate peers within the industry fits a broader narrative in regional and urban economics about the scopes of agglomeration economies and the localized cross-firm productivity spillovers (due to technology and knowl-

edge diffusion, labor market interactions, etc.) being one of the main sources of such externalities (e.g., see Duranton & Puga, 2004; Rosenthal & Strange, 2004). By restricting the scope to the same industry we effectively focus on intra-industry horizontal productivity spillovers.

Remark 1. The weighing scheme in (3.4) treats cross-firm spillovers symmetrically in that all members of a peer group affect each other's productivity. Not only is this a standard approach to measuring "peer effects," but in doing so we also wish to remain as agnostic about spillovers as possible and avoid imposing priors about the directionality of peer dependence. But should one choose to regulate the direction of spillovers by restricting them to occur, say, from more productive to less productive firms only, our framework can be modified to accommodate that too. For more discussion, see Appendix D.

Remark 2. In (3.4), a uniform weighting is applied across all peers of the firm i that are located in its spatial proximity. This implicitly assumes that within boundaries of the firm's spatial "neighborhood" the distance gradient is of second-order importance for knowledge spillovers. The main benefit of postulating such a feature of peer networks is that it does not require granular geographic data about individual firms and can be operationalized using coarse location information such as ZIP code, census track, city, region. The degree to which this is a reasonable weighting scheme obviously depends on the selected "level" of neighborhoods as well as the application-specific institutional context. If desired and feasible, peer weights $\{s_{ijt}\}$ can be appended to incorporate a (decay) function of the distance between i and its peers $\{j\}$.

The innovativeness of our model in the context of a broader literature on the structural proxy variable estimation of production functions is as follows. The "controlled" formulation in (3.3) is more general than the most commonly assumed exogenous Markov process à la Olley & Pakes (1996) whereby $\omega_{it} = \mathbb{E}\left[\omega_{it}|\omega_{i,t-1}\right] + \zeta_{it}$ because it enables the firm to influence the evolution of its productivity via its own productivity-enhancing activities/characteristics as well as by interacting with other local firms in the industry as captured by $G_{i,t-1}$ and $\sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}$, respectively. While controlled Markov processes for ω_{it} are not novel to the literature (e.g., Do-

raszelski & Jaumandreu, 2013; De Loecker, 2013), all such studies have focused exclusively on an *independently* (over i) distributed ω_{it} having the latter depend on the firm's own productivity and productivity-modifying variables. Our important generalization is that we permit cross-sectional dependence in firm productivity within peer networks due to agglomeration.

Consider the spatiotemporal autoregressive conditional mean of ω_{it} in (3.3) that represents the ith firm's expected one-period-ahead productivity at time $t-1$. First, by letting it depend on the firm's *own* productivity modifier $G_{i,t-1}$, we are able to account for (internal) *direct* learning taking place within the firm, with the corresponding estimand of interest being

$$DL_{it} = \frac{\partial \mathbb{E}[\omega_{it}|\cdot]}{\partial G_{i,t-1}}. \tag{3.5}$$

Second, in including the spatial average of other firms' productivities $\sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}$, not only can we accommodate potential agglomeration externalities facilitated by productivity *spillovers* across firms, but we are also able to capture the (external) *indirect* learning whereby the productivity-modifying activities may have secondary effects on firms beyond their immediate (and intended) beneficiary. Concretely, defining the cross-firm productivity spillovers as

$$SP_{it} = \frac{\partial \mathbb{E}[\omega_{it}|\cdot]}{\partial \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}}, \tag{3.6}$$

the measure of firm i's indirect learning from firm j's productivity-modifying activities is

$$IL_{ijt} = \frac{\partial \mathbb{E}[\omega_{it}|\cdot]}{\partial G_{j,t-2}} = \frac{\partial \mathbb{E}[\omega_{it}|\cdot]}{\partial \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}} \left(s_{ij,t-1} \frac{\partial \omega_{j,t-1}}{\partial G_{j,t-2}} \right) = SP_{it} \times s_{ij,t-1} \times DL_{j,t-1}. \tag{3.7}$$

The IL_{ijt} effect in (3.7) is defined for an (i, j) pair of firms, and we can aggregate it to the total indirect learning of firm i from all of its peers as

$$TIL_{it} = \sum_{j(\neq i)} IL_{ijt} = SP_{it} \times \sum_{j(\neq i)} s_{ij,t-1} DL_{j,t-1}. \tag{3.8}$$

As defined in (3.5)–(3.8), the learning and spillover effects on firm productivity are "short-run," but they accumulate and diffuse over time owing to a persistent nature of the firm's productivity evolution. Indirectly, this dynamic feature permits two peer firms that are separated temporally to continue to affect one another with the effect size attenuating over time. Such time-separated interactions characterize the temporal scope of productivity spillovers which

helps propel *dynamic* agglomeration economies. The underlying idea here is that the knowledge acquired either through internal learning or from peers takes time to accumulate. Together with the geographic and industrial scopes of spillovers embedded in the definition of peer weights $\{s_{ijt}\}$, the autoregressiveness of productivity specification covers the three main dimensions of external economies (see Rosenthal & Strange, 2004).

Remark 3. The total indirect learning effect in (3.8) is, effectively, a measure of spillovers *specifically* in G. In this, our conceptualization of external effects of G as operating through the firm's exposure to the aggregate of its peers' unobservable productivities—that is, via "productivity spillovers" more broadly—fundamentally differs from the conventional approach to measuring spillovers in productivity-modifying activities (think, FDI, R&D or export spillovers) that relies on observable industry aggregates of G. That is, we measure the ith firm's exposure to the *external* knowledge using an aggregate of $\{\omega_{jt};\ j(\neq i) = 1, \ldots, n\}$ as opposed to an aggregate of $\{G_{jt};\ j(\neq i) = 1, \ldots, n\}$. Our formulation is more flexible because it does not restrict the origins of cross-firm productivity spillovers to G alone. It is also more realistic and conceptually congruous because it incorporates secondary information about the peer firms' own direct/internal learning facilitated by the productivity-modifying activities they undertake: namely, to learn from one's peers' G, peers themselves should learn from their own G first.

The productivity evolution process in (3.3) characterizes the peer interaction between firms through their productivity. Each firm i has a "reference group" of spatially proximate peers from the same industry $\mathscr{L}(i,t)$ with which it interacts. The identification of such cross-peer relations in networks is a notoriously challenging problem (e.g., see Manski, 1993, 2000; Moffitt, 2001; Blume et al., 2011). The potential obstacles include (i) the perfect functional dependence between the average outcome of the group and its mean characteristics due to the so-called "reflection problem" which may leave no exogenous variation excluded to instrument the endogenous peer behavior when there is more than one channel for the peer effects, (ii) the confounding presence of unobserved "correlated" group effects, and (iii) the endogenous group membership (or network structure). In our case, the additional layer of complexity is the la-

tency of firm productivity. This aspect is addressed in the proxy variable framework by making a full use of the behavioral model of firm production, and we discuss this in detail in Section 4. We now consider the issues pertaining to the measurement of peer effects between firms.

The identification of learning and spillover effects on firm productivity in our model is based on several structural assumptions about the timing as well as the underlying form of peer interactions and network organization. To begin with, the productivity process in (3.3) is a dynamic analogue of a "pure endogenous-effects model" in Manski's nomenclature. It postulates that the cross-firm peer interactions occur only through the outcomes (i.e., ω) whereby each firm's productivity is affected by the mean productivity of the peers in its reference group. As noted in Remark 3, we effectively assume away the "contextual effects" of the peers' productivity modifiers and, in doing so, address the first of two Manski's (1993) unidentification results about the indistinguishability of endogenous and exogenous peer effects.[3] The latter issue becomes moot because in the absence of contextual effects of $\{G_{j,t-1};\ j(\neq i) = 1,\ldots,n\}$ on ω_{it} our model postulates a single channel of cross-peer dependence. Appendix E discusses how our setup may be augmented to allow for such contextual effects.

The evolution process in (3.3) also implicitly assumes that learning occurs with a delay which is why the dependence of ω_{it} on both its own productivity-modifying controls and peers' productivity is lagged, implying that the improvements in firm productivity take a period to materialize. Furthermore, in $\mathbb{E}[\zeta_{it}|\mathscr{I}_{i,t-1}] = 0$ we assume that firms do not experience changes in their location and/or productivity modifiers in light of expected *future* innovations in productivity. This timing assumption about the arrival of ζ_{it} renders both the lagged $G_{i,t-1}$ and a set of spatially proximate peers $\mathscr{L}(i,t-1)$ at time $t-1$ that defines the peer weights $\{s_{ij,t-1};\ j(\neq i) = 1,\ldots,n\}$ predetermined (weakly exogenous) with respect to the firm i's productivity innovation at time t, which helps identify both the learning and spillover effects on firm productivity.

When it comes to internal learning effects (via own $G_{i,t-1}$), such a timing assumption is quite

[3] His second result is about the difficulty to distinguish "real" peer interactions through observables from the unobservable "correlated effects;" more on this later.

common in the productivity literature (e.g., see Van Biesebroeck, 2005; De Loecker, 2013; Doraszelski & Jaumandreu, 2013; Malikov et al., 2020). More specifically, $\mathbb{E}[\zeta_{it}|\mathscr{I}_{i,t-1}] = 0$ rules out the firm's ability to systematically predict future productivity shocks. Instead, the Markovian process in (3.3) states that the firm anticipates the effect of its G productivity modifier on ω_{it} in period t when adjusting the former in period $t-1$, and the conditional mean $\mathbb{E}\left[\omega_{it}|\omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}\right]$ is what captures that *expected* productivity. But the *actual* firm productivity at time t also includes a random innovation ζ_{it}. Essentially, the conditional-expectation-function error ζ_{it} represents unpredictable uncertainty that is naturally associated with productivity-modifying activities (new R&D investments, entering export markets or attracting new foreign investors) such as chance in discovery, success in implementation, etc. This productivity innovation ζ_{it} is realized after $G_{i,t-1}$ is fully determined.[4]

In our paper, we also extend this timing assumption to external cross-firm learning via spillovers, which yields mean-orthogonality of the spatiotemporal "lag" of peers' productivities $\sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}$ and the innovation ζ_{it}. That is, the assumed is weak exogeneity of the location-dependent peer weights $\{s_{ij,t-1}\}$, according to which firms do not relocate in anticipation of *future* productivity shocks because such shocks are purely random. This rules out endogeneity of the firm's peer network in period $t-1$ with respect to the productivity shock ζ_{it} it experiences at time t. The plausibility of this is further buttressed by the fact that firm *re*location in most industries (e.g., agriculture, manufacturing, utilities) is highly, if not prohibitively, costly. In fact, our assumption about the weak exogeneity of group membership is not as strong as the standard assumption of fixed (non-random) networks commonly made in the (empirical) social-effects or spatial literature.

Note that our timing assumption about learning and spillover effects does *not* rule out a *contemporaneous* correlation between firm productivity and its productivity modifiers or even the location. That is, we do not assume that $\mathbb{E}[\zeta_{it}|\mathscr{I}_{it}] = 0$. Consequently, firms are permitted to endogenously update their G_{it} as well as to change their locations based on the (observable by

[4]Depending on the source of learning, it may be possible to reasonably relax the assumption of a delayed learning effect of G_{it} on firm productivity. See discussion in Appendix E.

firms) period t level of their productivity ω_{it}. For instance, in the presence of inbound FDI opportunities that can help improve a domestic firm's productivity (when G_{it} measures the firm's exposure to investors from abroad), the more productive firms are more likely to be attractive for investors, and the corresponding non-zero $\text{Cov}[G_{it}, \omega_{it}]$ is well within our framework.

An important implication of our structural assumption about $\mathbb{E}[\zeta_{it} | \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}] = 0$ is that the innovation in the productivity evolution process (3.3) does *not* contain any unobservable "correlated effects" at the reference group level—to borrow Manski's terminology—the presence of which can complicate, if not hinder, the identification of peer effects occurring through the group mean productivity $\sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}$. Effectively, we attribute all cross-firm dependence in productivity to the within-group dependence of each firm's underlying productivity on that of its peers as opposed to the tendency of all group firm-members to see their productivities evolve in a similar fashion due to the influence of common group unobservables such as shared locational/institutional environments. This is an admittedly strong but fairly common working assumption in the literature, given the well-known challenges in tackling group-level unobservables in network models (for an excellent review, see Blume et al., 2011). Our no-group-effects assumption echoes the existing studies of R&D/FDI/export spillovers and the productivity literature more broadly, and we maintain it to maximize comparability with the commonly used methodologies. Having said that, this assumption can be relaxed—we do so in our robustness checks—if we restrict the group-level unobservables to be time-invariant à la Graham & Hahn (2005) and Bramoullé et al. (2009).

Fundamentally, the potential threats to identification of the spillover effects posed by the correlated group effects can otherwise be cast as a spatial selection/sorting problem, whereby more productive firms may be *ex ante* sorting into the what-then-become high productivity locations. Under this scenario, when we compare the firm to its spatial peers, we may mistakenly attribute any future productivity improvements to spillovers from the peers (i.e., agglomeration), while in actuality it merely reflects the underlying propensity of *all* firms in this location to be more productive and, consequently, more apt at improving their productivity.

While there has recently been notable progress in formalizing and understanding these coincident phenomena theoretically (e.g., Behrens et al., 2014; Gaubert, 2018), disentangling firm sorting and agglomeration remains a non-trivial task empirically.[5] However, by including the firm's *own* lagged productivity in the autoregressive ω_{it} process in (3.3), we are able (at least to some extent) to account for this potential self-sorting because sorting into locations is heavily influenced by the firm's own productivity (oftentimes stylized as the "talent" or "efficiency" in theoretical models). That is, the spillover effect SP_{it} on future firm productivity in our model is measured after partialling out the contribution of its own productivity. Incidentally, De Loecker (2013) argues the same in the context of export-based learning and self-selection of exporters.

We maintain the *i.i.d.* assumption about the random transitory shock η_{it}, from where it follows that $\mathbb{E}[\eta_{it}|\mathscr{I}_{it}] = \mathbb{E}[\eta_{it}] = 0$ with the mean normalized to zero. The latter implies that the shock η_{it} is observable to firms in period t only *ex post* after all production decisions.

4 A System Approach to Identification via Proxy Variables

Logging the production function in (3.1) and making use of the Markovian nature of ω_{it} from (3.3), we obtain

$$y_{it} = \beta_K k_{it} + \beta_L l_{it} + \beta_M m_{it} + h\left(\omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}\right) + \zeta_{it} + \eta_{it}, \tag{4.1}$$

where the lower-case variables denote the logs of the respective upper-case variables, and $h[\cdot] \equiv \mathbb{E}[\omega_{it}|\cdot]$ is some unknown function. Under our structural assumptions about firm behavior, all right-hand-side covariates in (4.1) are predetermined and thus mean-independent of $\zeta_{it} + \eta_{it}$, except for the freely varying input m_{it} that the firm chooses in time period t conditional on ω_{it}

[5] Urban economics literature also distinguishes the third endogenous process usually referred to as the "selection" which differs from sorting in that it occurs *ex post* after the firms had self-sorted into locations and which determines their continuing survival. We abstract away from this low-productivity-driven attrition issue in the light of the growing empirical evidence suggesting that it explains none of spatial productivity differences which, in contrast, are mainly driven by agglomeration economies (see Combes et al., 2012). Relatedly, the firm attrition out of the sample has also become commonly accepted as a *practical* non-issue in the productivity literature so long as the data are kept unbalanced. For instance, Levinsohn & Petrin (2003, p.324) write: "The original work by Olley and Pakes devoted significant effort to highlighting the importance of not using an artificially balanced sample (and the selection issues that arise with the balanced sample). They also show once they move to the unbalanced panel, their selection correction does not change their results."

(among other state variables including quasi-fixed inputs) thereby making it a function of ζ_{it}. That is, the materials variable is endogenous.

To consistently estimate (4.1), we first need to address the latency of firm productivity ω_{it}. A widely popular solution to this problem in the literature is to adopt a proxy variable approach à la Levinsohn & Petrin (2003) whereby unobservable ω_{it} is proxied by inverting the firm's conditional demand for an observable static input m_{it}. However, Gandhi et al. (2020) show that identification generally fails under such a standard estimation procedure due to the lack of a valid instrument for the endogenous m_{it} despite the abundance of predetermined higher-order lags of inputs. Therefore, the production function remains unidentified in the flexible input. To solve this problem, they suggest exploiting a structural link between the production function and the firm's (static) optimality condition. In what follows, we build on this idea which we adapt in the spirit of Doraszelski & Jaumandreu (2013), whereby we explicitly make use of the assumed functional form of the production function to streamline identification of the material elasticity and to ease computational burden of estimation (also see Remark 4).

We first focus on the identification of the production function in its flexible input M_{it}. Specifically, given the Cobb-Douglas form, we seek to identify the material elasticity parameter β_M. To do so, we consider an equation for the firm's first-order condition with respect to M_{it}. Since it is a static input, the firm's optimal choice of M_{it} can be modeled as the restricted expected profit-maximization problem[6] subject to the (already) optimal allocation of quasi-fixed inputs:

$$\max_{M_{it}} P_t^Y K_{it}^{\beta_K} L_{it}^{\beta_L} M_{it}^{\beta_M} \exp\{\omega_{it}\}\theta - P_t^M M_{it}, \tag{4.2}$$

where P_t^Y and P_t^M are respectively the output and material prices that, under the commonly invoked assumption of perfect competition, need not vary across firms; and $\theta \equiv \mathbb{E}[\exp\{\eta_{it}\} | \mathscr{I}_{it}]$. The first-order condition is given by

$$\beta_M P_t^Y K_{it}^{\beta_K} L_{it}^{\beta_L} M_{it}^{\beta_M - 1} \exp\{\omega_{it}\}\theta = P_t^M, \tag{4.3}$$

which can be transformed via dividing it by the production function in (3.1) to obtain the fol-

[6]Under the risk neutrality of firms.

lowing stochastic material share equation (in logs):

$$\ln V_{it} = \ln(\beta_M \theta) - \eta_{it}, \tag{4.4}$$

where $V_{it} \equiv P_t^M M_{it} / (P_t^Y Y_{it})$ is the nominal share of material costs in total revenue. This share is readily observable in the data, and the construction thereof does not require firm-level prices.

Intuitively, equation (4.4) says that *unobservable* material elasticity of the production function β_M can be identified from *observable* material share V_{it} because the two must be equal on average (in logs) to maximize profits. Specifically, it identifies $\beta_M \times \theta$ (and the random productivity residual η_{it}) based on $\mathbb{E}[\eta_{it}] = 0$:

$$\ln(\beta_M \theta) = \mathbb{E}[\ln V_{it}]. \tag{4.5}$$

To identify the material elasticity β_M net of constant θ, recognize that θ can be identified via $\theta = \mathbb{E}[\exp\{\eta_{it}\}] = \mathbb{E}\left[\exp\{\ln(\beta_M\theta) - \ln V_{it}\}\right] = \mathbb{E}\left[\exp\{\mathbb{E}[\ln V_{it}] - \ln V_{it}\}\right]$. Then, we have that

$$\beta_M = \exp\{\mathbb{E}[\ln V_{it}]\} \big/ \mathbb{E}\left[\exp\{\mathbb{E}[\ln V_{it}] - \ln V_{it}\}\right]. \tag{4.6}$$

With β_M identified from (4.6), we have thus identified the production function in the dimension of its endogenous freely varying input M_{it} thereby effectively circumventing Gandhi et al.'s (2017) critique. To see the latter, we rewrite (4.1) as follows:

$$y_{it}^* = \beta_K k_{it} + \beta_L l_{it} + h\left(\omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}\right) + \zeta_{it} + \eta_{it}, \tag{4.7}$$

where $y_{it}^* \equiv y_{it} - \beta_M m_{it}$ is already identified and thus observable, and our model in (4.7) no longer contains endogenous variables needing instrumentation.

To identify the remaining parameters of the production function $(\beta_K, \beta_L)'$ as well as latent firm productivity ω_{it} in (4.7), we make use of the *known* parametric form of the conditional material demand function $M_{it} = \mathbb{M}(\omega_{it}, K_{it}, L_{it}, P_t^Y, P_t^M)$ implied by the first-order condition in (4.3) which we invert for ω_{it}. Under our standard assumptions about firm behavior and regularity conditions on the production function, $\mathbb{M}(\cdot)|M_{it} > 0$ must be strictly monotonic in ω_{it} for any given $(K_{it}, L_{it}, P_t^Y, P_t^M)$, and hence we can invert $\mathbb{M}(\cdot)$ to control for unobserved persistent productivity via $\omega_{it} = \mathbb{M}^{-1}(M_{it}, K_{it}, L_{it}, P_t^Y, P_t^M)$. Specifically, substituting for $\omega_{i,t-1}$ and $\omega_{j,t-1}$

using the inverted material function derived analytically from (4.3), from (4.7) we get

$$y_{it}^{*} = \beta_K k_{it} + \beta_L l_{it} + h\left(\omega_{i,t-1}^{*}\left(\beta_K, \beta_L\right), G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}^{*}\left(\beta_K, \beta_L\right)\right) + \zeta_{it} + \eta_{it}, \tag{4.8}$$

where

$$\omega_{it}^{*}\left(\beta_K, \beta_L\right) = \left[(1-\beta_M)m_{it} - \ln(\beta_M\theta) - \ln(P_t^Y / P_t^M)\right] - \beta_K k_{it} - \beta_L l_{it} \quad \forall i, t \tag{4.9}$$

is the inverted material demand function in which the bracketed component is already observable and therefore the only remaining unknown parameters in it are $(\beta_K, \beta_L)'$. All right-hand-side covariates in the semiparametric model (4.8) are weakly exogenous and thus self-instrument. The model is thus identified on the basis of

$$\mathbb{E}\left[\zeta_{it} + \eta_{it} \middle| k_{it}, l_{it}, k_{i,t-1}, l_{i,t-1}, m_{i,t-1}, G_{i,t-1}, \sum_{j\neq i} s_{ij,t-1}k_{j,t-1}, \sum_{j\neq i} s_{ij,t-1}l_{j,t-1}, \sum_{j\neq i} s_{ij,t-1}m_{j,t-1}\right] = 0, \tag{4.10}$$

where we have made explicit use of the variables entering the proxy function $\omega_{it}^{*}\left(\beta_K, \beta_L\right)$.

The appearance of group averages of the peers' predetermined inputs in (4.10) is akin to the idea of instrumenting the endogenous group mean of an outcome with the exogenous group mean characteristics, which is a common identification strategy in both the social-effects and spatial econometrics literature (e.g., see LeSage & Pace, 2009; Bramoullé et al., 2009). The critical distinction here is that, in our case, the "group mean of an outcome" $\sum_{j\neq i} s_{ij,t-1}\omega_{j,t-1}$ is *not* endogenous with respect to $\zeta_{it} + \eta_{it}$ and therefore needs no instrumentation. In contrast, our use of the "group mean characteristics" $\left(\sum_{j\neq i} s_{ij,t-1}k_{j,t-1}, \sum_{j\neq i} s_{ij,t-1}l_{j,t-1}, \sum_{j\neq i} s_{ij,t-1}m_{j,t-1}\right)'$ is effectively in their proxy-variable capacity given latency of $\omega_{j,t-1}$.

Remark 4. Following the steps of Doraszelski & Jaumandreu (2013), our approach fully embraces the assumed parametric specification of the firm's production function by explicitly utilizing the known functional form of the first-order condition for materials and the inverse conditional input demand function that it implies. By doing so, we circumvent the need to integrate the estimated material elasticity function at *each* observation in order to recover the unknown production function required by Gandhi et al.'s (2017) nonparametric methodology. Importantly, by relying on parameter restrictions between the production function and inverted ma-

terial demand function in (4.8), we do not have to rely on nonparametric methods to estimate the unknown proxy function for ω that appears inside the also unknown $h(\cdot)$ function. Otherwise, identification of (4.8) would have been complicated by the presence of a nonparametric $\mathbb{M}^{-1}(\cdot)$ function (evaluated at multiple data points[7]) inside another nonparametric function $h(\cdot)$. Our parametric inversion of $\mathbb{M}^{-1}(\cdot)$ yields a much simpler semiparametric estimator.

Remark 5. Our model is also robust to Ackerberg et al.'s (2015) critique that focuses on the potential inability of structural proxy variable estimators to separably identify the production function and productivity proxy. This issue arises in the wake of perfect functional dependence between variable inputs appearing both inside the unknown production function and productivity proxy function. Our second-stage equation (4.8) does not suffer from such a problem because it contains no (endogenous) variable input on the right-hand side, the corresponding parameter of which has already been identified from the share equation in the first stage.

Lastly, with all parameters of the production function $(\beta_K, \beta_L, \beta_m)'$ and the transitory productivity shock η_{it} successfully identified in the two stages, we readily identify latent firm productivity ω_{it} from the production function in logs: $\omega_{it} = y_{it} - \beta_K k_{it} - \beta_L l_{it} - \beta_M m_{it} - \eta_{it}$.

5 Estimation Procedure

We implement our identification strategy in two stages. In the first stage, we estimate the material elasticity of the production function. Based on (4.6), the consistent estimator of β_M is

$$\widehat{\beta}_M = \exp\left\{\frac{1}{N}\sum_i\sum_t \ln V_{it}\right\} \Big/ \left[\frac{1}{N}\sum_i\sum_t \exp\left\{\left(\frac{1}{N}\sum_i\sum_t \ln V_{it}\right) - \ln V_{it}\right\}\right], \tag{5.1}$$

where N is the total number of observations, which equals nT in the case of a balanced panel.

We then estimate y^*_{it} via $\widehat{y}^*_{it} = y_{it} - \widehat{\beta}_M m_{it}$ and also construct "partial" estimates of the pro-

[7] That is, evaluated at $(m_{i,t-1}, k_{i,t-1}, l_{i,t-1})$ to proxy for $\omega_{i,t-1}$ as well as at $(m_{j,t-1}, k_{j,t-1}, l_{j,t-1})$ $\forall$ j to proxy for $\omega_{j,t-1}$ entering the spillover-capturing peer group average.

ductivity proxy function $\omega^*_{it}(\beta_K, \beta_L)$ in (4.9) as

$$\widehat{\omega}^*_{it}(\beta_K, \beta_L) = \underbrace{\left[(1-\widehat{\beta}_M)m_{it} - \ln(\widehat{\beta_M\theta}) - \ln(P^Y_t/P^M_t)\right]}_{\widehat{\varkappa}_{it}} - \beta_K k_{it} - \beta_L l_{it},$$

where $\ln(\widehat{\beta_M\theta}) = \frac{1}{N}\sum_i \sum_t \ln V_{it}$ on the basis of (4.5). Note that $\widehat{\omega}^*_{it}(\beta_K, \beta_L)$ still contains two unknowns which enter the function linearly: $(\beta_K, \beta_L)'$. For convenience, let the already identified portion of productivity be denoted by $\widehat{\varkappa}_{it} = (1-\widehat{\beta}_M)m_{it} - \ln(\widehat{\beta_M\theta}) - \ln(P^Y_t/P^M_t)$.

With $\widehat{y}^*_{it}$ and $\widehat{\omega}^*_{it}(\beta_K, \beta_L)$ from the first stage in hand, we proceed to the second-stage estimation of (4.8), where we approximate unknown function $h(\cdot)$ using polynomial sieves. Specifically, recognize that $\widehat{\omega}^*_{it}(\beta_K, \beta_L) = \widehat{\varkappa}_{it} - \beta_K k_{it} - \beta_L l_{it}$ and let $\widehat{\mathbf{z}}_{i,t-1}(\boldsymbol{\beta}) = ([\widehat{\varkappa}_{i,t-1} - \beta_K k_{i,t-1} - \beta_L l_{i,t-1}],\ G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}[\widehat{\varkappa}_{j,t-1} - \beta_K k_{j,t-1} - \beta_L l_{j,t-1}])'$ be a 3×1 vector with $\boldsymbol{\beta} = (\beta_K, \beta_L)'$. Then, for each $\widehat{\boldsymbol{z}}(\boldsymbol{\beta})$, we approximate the unknown function $h(\widehat{\boldsymbol{z}}(\boldsymbol{\beta}))$ in (4.8) as follows:

$$h(\widehat{\boldsymbol{z}}(\boldsymbol{\beta})) \approx \mathscr{A}_{L_n}(\widehat{\boldsymbol{z}}(\boldsymbol{\beta}))'\boldsymbol{\gamma}, \tag{5.2}$$

where $\mathscr{A}_{L_n}(\cdot) = (A_1(\cdot), \ldots, A_{L_n}(\cdot))'$ is an $L_n \times 1$ vector of known basis functions of $\widehat{\boldsymbol{z}}(\boldsymbol{\beta})$ including a vector of ones, $\boldsymbol{\gamma}$ is a conformable vector of parameters, and $L_n \to \infty$ slowly with n.

Given the orthogonality conditions in (4.10), we estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ via nonparametric nonlinear least squares. Letting $\mathbf{x}_{it} = (k_{it}, l_{it})'$, the parameter estimators are given by

$$(\widehat{\beta}_K, \widehat{\beta}_L, \widehat{\boldsymbol{\gamma}}')' = \underset{\boldsymbol{\beta},\boldsymbol{\gamma}}{\operatorname{argmin}} \sum_i \sum_t \left[\widehat{y}^*_{it} - \mathbf{x}'_{it}\boldsymbol{\beta} - \mathscr{A}_{L_n}(\widehat{\mathbf{z}}_{i,t-1}(\boldsymbol{\beta}))'\boldsymbol{\gamma}\right]^2, \tag{5.3}$$

where the minimand is the sum of squared errors corresponding to our sieve estimator of (4.8).

Using the already estimated $(\widehat{\beta}_K, \widehat{\beta}_L, \widehat{\beta}_M)'$ and $\widehat{\boldsymbol{\gamma}}$, we then readily have the estimators for our primary estimands of interest: $\widehat{DL}_{it} \equiv \partial\widehat{h}_{it}(\cdot)/\partial G_{i,t-1}$, $\widehat{SP}_{it} \equiv \partial\widehat{h}_{it}(\cdot)/\partial\sum_{j(\neq i)} s_{ij,t-1}\widehat{\omega}_{j,t-1}$ and $TIL_{it} = \widehat{SP}_{it} \times \sum_{j(\neq i)} s_{ij,t-1}\widehat{DL}_{j,t-1}$ respectively measuring the direct learning, cross-firm spillover and total indirect learning effects, where $\widehat{h}_{it}(\cdot) = \mathscr{A}_{L_n}(\widehat{\mathbf{z}}_{it}(\widehat{\boldsymbol{\beta}}))'\widehat{\boldsymbol{\gamma}}$ and $\widehat{\omega}_{j,t-1} = \widehat{\varkappa}_{j,t-1} - \widehat{\beta}_K k_{j,t-1} - \widehat{\beta}_L l_{j,t-1}$. Lastly, the estimator of latent firm productivity is $\widehat{\omega}_{it} = y_{it} - \widehat{\beta}_K k_{it} - \widehat{\beta}_L l_{it} - \widehat{\beta}_M m_{it} - \widehat{\eta}_{it}$, where $\widehat{\eta}_{it} = \ln(\widehat{\beta_M\theta}) - \ln V_{it}$ from the first stage.

For the limit results, our sequential estimation methodology can be recast as a moment-

based semiparametric sieve M-estimation problem. Specifically, letting

$$\mathbf{z}_{i,t-1}\left(\beta_M,\beta_K,\beta_L,\theta\right)=\begin{bmatrix}(1-\beta_M)m_{it-1}-\ln(\beta_M\theta)-\ln(P^Y_{t-1}/P^M_{t-1})-\beta_K k_{it-1}-\beta_L l_{it-1}\\ G_{it-1}\\ \sum_{j(\neq i)} s_{ijt-1}\left[(1-\beta_M)m_{jt-1}-\ln(\beta_M\theta)-\ln(P^Y_{t-1}/P^M_{t-1})-\beta_K k_{jt-1}-\beta_L l_{jt-1}\right]\end{bmatrix}$$

and $r_{it}\left(\beta_M,\beta_K,\beta_L,\theta\right)=y_{it}-\beta_M m_{it}-\beta_K k_{it}-\beta_L l_{it}-\mathscr{A}_{L_n}\left(\mathbf{z}_{i,t-1}\left(\beta_M,\beta_K,\beta_L,\theta\right)\right)'\boldsymbol{\gamma}$, we can rewrite our two estimation stages in the form of the following multiple-equation moment restrictions:

$$\mathbb{E}\begin{bmatrix}\ln V_{it}-\ln(\beta_M\theta)\\ \exp\left\{\ln(\beta_M\theta)-\ln V_{it}\right\}-\theta\\ r_{it}\left(\beta_M,\beta_K,\beta_L,\theta\right)\begin{bmatrix}\partial r_{it}\left(\beta_M,\beta_K,\beta_L,\theta\right)/\partial\left(\beta_K,\beta_L\right)'\\ -\mathscr{A}_{L_n}\left(\mathbf{z}_{i,t-1}\left(\beta_M,\beta_K,\beta_L,\theta\right)\right)\end{bmatrix}\end{bmatrix}=\mathbf{0}_{4+L_n}, \tag{5.4}$$

consisting of two blocks, where the first two moments correspond to the estimator of the material elasticity (first block) and the remaining orthogonality conditions correspond to the nonlinear sieve least-squares estimation of the proxied production function and productivity in (5.3).

In the above, $\mathscr{A}_{L_n}(\cdot)$ is a sieve approximation of the unknown infinite-dimensional nonparametric function $h(\cdot)$, and $\left(\beta_M,\beta_K,\beta_L,\theta\right)'$ are the unknown parameters of fixed dimension. Thus, our estimator falls within Ai & Chen's (2003) general framework for the minimum distance estimation based on the conditional moment restrictions of a generic form $\mathbb{E}[\rho(X,\delta_0,g_0(\cdot))|Z]=0$, where X and Z are data and $\rho(\cdot)$ is a vector of "residual functions" with finite-dimensional unknown parameters δ and infinite-dimensional unknown functions g. The large-sample limit results from Ai & Chen (2003) and Chen et al. (2003) therefore extend to our two-step estimator. Inference can be asymptotic or via bootstrap; we discuss both in detail in Appendix F.

6 Simulations

We conduct a set of Monte Carlo experiments. Our data generating process draws from those used by Grieco et al. (2016) and Gandhi et al. (2020). Specifically, we consider a balanced panel of $n=\{100,200,400\}$ firms operating during $T=10$ periods.[8] Each panel is simulated 1,000

[8]We have also experimented with 5 and 50 periods. The results are qualitatively unchanged.

times. To simplify matters, we dispense with labor and consider the production process only with the quasi-fixed dynamic K_{it} and freely-varying static M_{it}. The production technology is

$$Y_{it} = K_{it}^{\beta_K} M_{it}^{\beta_M} \exp\{\omega_{it} + \eta_{it}\}, \tag{6.1}$$

where we set $\beta_K = 0.25$ and $\beta_M = 0.65$, and the noise $\eta_{it} \sim$ i.i.d. $\mathbb{N}(0, \sigma_\eta^2)$ with $\sigma_\eta = \sqrt{0.07}$.

The firm's capital is set to evolve according to $K_{it} = I_{i,t-1} + (1-\delta_i)K_{i,t-1}$, with the firm-specific depreciation rates $\delta_i \in \{0.05, 0.075, 0.10, 0.125, 0.15\}$ distributed uniformly across i. The initial levels of capital K_{i0} is drawn from *i.i.d.* $\mathbb{U}(10, 200)$. The investment function takes the following form: $I_{i,t-1} = K_{i,t-1}^{\alpha_1} \exp\{\alpha_2 \omega_{it-1}\}$, where $\alpha_1 = 0.8$ and $\alpha_2 = 0.1$.

The materials M_{it} series is generated solving the firm's restricted expected profit maximization problem along the lines of (4.2). The conditional demand for M_{it} is given by

$$M_{it} = \underset{\mathscr{M}_{it}}{\operatorname{argmax}} \left\{ P_t^Y K_{it}^{\beta_K} \mathscr{M}_{it}^{\beta_M} \exp\{\omega_{it}\}\theta - P_t^M \mathscr{M}_{it} \right\} = \left(\beta_M K_{it}^{\beta_K} \exp\{\omega_{it}\} \right)^{1/(1-\beta_M)}, \tag{6.2}$$

where, in the second equality, we have normalized $P_t^M = \theta \ \forall \ t$ and have assumed no temporal variation in output prices: $P_t^Y = 1$ for all t.

We assume that the firm's productivity modifier G_{it} is autoregressively persistent. More specifically, we consider two laws of motion for G_{it}: (a) an exogenous autoregressive process whereby $G_{it} = \gamma_0 + \gamma_1 G_{i,t-1} + \epsilon_{it}$ and (b) a controlled autoregressive process, contemporaneously conditional on the firm's latent productivity: $G_{it} = \gamma_0 + \gamma_1 G_{i,t-1} + \gamma_2 \omega_{it} + \epsilon_{it}$, where $\gamma_0 = 0.01$, $\gamma_1 = 0.6$, $\gamma_2 = 0.3$ and $\epsilon_{it} \sim$ i.i.d. $\mathbb{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 0.1$. Of the two processes, the second one assumes that more productive firms engage in higher levels of the productivity-modifying activities. The process (b) permits firms to endogenously update their G_{it} based on the (observable by them) period t level of their productivity ω_{it}. For example, if G_{it} measures the firm's exposure to investors from abroad, this accommodates the scenario when foreign investors choose to invest in more productive domestic firms in the first place.

We let firm productivity ω_{it} be a linear spatiotemporal first-order autoregressive process:

$$\omega_{it} = \rho_0 + \rho_1 \omega_{i,t-1} + \rho_2 \sum_{j(\neq i)} s_{ij,t-1} \omega_{j,t-1} + \rho_3 G_{i,t-1} + \zeta_{it}, \tag{6.3}$$

where, unless stated otherwise, we set $\rho_0 = 0.2$, $\rho_1 = 0.55$, $\rho_2 = 0.4$ and $\rho_3 = 0.5$. The innovation

is generated as $\zeta_{it} \sim$ i.i.d. $\mathbb{N}(0, \sigma_\zeta^2)$ with $\sigma_\zeta = 0.2$. The initial level of productivity $\omega_{i0} \sim$ i.i.d. $\mathbb{U}(1,3)$ over i. In Appendix G, we also present the results for a *nonlinear* specification for ω_{it}.

To keep matters simple, we consider one common spatial region for all firms and assume that all firms belong to the same industry. Hence, cardinality of the set $\mathscr{L}(i,t)$ is the same across all i and equals $n-1$. The peer weights $\{s_{ijt};\ j(\neq i)\}$ are constructed according to (3.4) and, given the setup, are equal to $1/(n-1)$ for all firms and time periods.

Proposed Methodology. First, we evaluate the performance of our proposed estimator with the focus on its ability to successfully identify productivity spillovers across firms. For each combination of the G and ω processes, we consider the following three DGP scenarios: (**i**) a general case scenario in which firm productivity is modified via both the direct learning ($DL_{it} \neq 0$) and cross-firm spillovers ($SP_{it} \neq 0$); (**ii**) a special case scenario in which we assume no direct learning ($DL_{it} = 0$ globally) in order to focus our attention exclusively on the agglomeration-driven learning via spillovers ($SP_{it} \neq 0$); (**iii**) an even more special case scenario in which firm productivity evolves exogenously (both $DL_{it} = 0$ and $SP_{it} = 0$ globally) as traditionally assumed in the proxy variable production function estimation literature. The special case scenarios are implemented by setting the appropriate coefficients in the productivity process (6.3) to zero.

We estimate the model via the two-stage estimation algorithm outlined in Section 5, where we approximate unknown $h(\cdot)$ using second-degree polynomial sieves. Table 1 reports the simulation results for our proposed estimator, when G_{it} evolves exogenously [top panel] and following an ω_{it}-dependent controlled process [bottom panel]. Each of these two panels includes the results from the three different scenarios. Reported are the mean, root mean squared error (RMSE) and mean absolute error (MAE) of the fixed-parameter β_K estimates[9] and the averages (across simulation iterations) of these metrics corresponding for each iteration computed using observation-specific nonparametric estimates of the autoregressive gradient $AR_{it} = \partial h(\cdot)/\partial \omega_{i,t-1}$,

[9]We omit the results corresponding to the material elasticity β_M from the first stage because the estimator yields very precise estimates of β_M via (5.1) with the MSE and MAE being at least as small as 10^{-10} owing to the small sampling error induced by η_{it} in our DGPs. We also experimented with much larger values of σ_η with no significant changes to the qualitative results.

$DL_{it} = \partial h(\cdot)/\partial G_{i,t-1}$, $SP_{it} = \partial h(\cdot)/\partial \sum_{j\neq i} \omega_{j,t-1}$ and $TIL_{it} = SP_{it} \times \sum_{j(\neq i)} s_{ij,t-1} DL_{j,t-1}$.

The results in Table 1 are encouraging and show that our methodology recovers the true parameters remarkably well, thereby lending strong support to the validity of our identification strategy. As expected of a consistent estimator, the estimation becomes more stable as n grows. Same is the case when the productivity DGP in nonlinear (see Table G.1 in Appendix G).

Alternative Procedures. Next, to demonstrate the advantage of our internally consistent methodology, we inspect the performance of a widely used alternative procedure for estimating spillovers via a two-step approach. In this case, the unobserved firm productivity ω_{it} is first estimated via the standard proxy variable estimator (which assumes that the productivity process is an exogenous Markov chain and thus ignores spillovers) and then linearly regressed on some measure of the firm's exposure to its peers in the second step. As already discussed at length, such second-step regressions are inconsistent with the assumptions made in the first step because they contradictorily postulate the existence of cross-peer dependence which was assumed away when recovering firm productivity in the first place. Consequently, the productivity estimates (by means of the production function) obtained via such an approach are prone to biases due to the endogeneity-inducing misspecification of the productivity proxy. The empirical evidence of spillovers can thus be spurious. This is unsurprising because the unaccounted cross-sectional dependence is a hindrance to identification in general (see Pesaran, 2006; Bai, 2009).

The second-step regressions used in spillovers literature have numerous variations but can be by and large categorized into two distinct types: those that measure the firm's exposure to spillovers from peers using the group means of characteristics which are said to facilitate such spillovers (FDI, R&D, exports, etc.), and those that measure the firm's exposure to spillovers using the peer group mean of an outcome (that is, firm productivity). Essentially, the first type of regressions focuses on the "contextual effects" while the second type models cross-peer dependence via "endogenous effects." Rarely do researchers allow for both effects at the same time. The first type is arguably the predominant choice in spillovers literature. Such studies overwhelmingly estimate linear specifications, and virtually all omit the temporal lag of the firm's

own productivity in the second-step analysis.

We consider alternative methodologies with the second-step regressions of both these types. To facilitate a level-playing-field comparison between these and our models in the ability to identify spillovers, we specify the second-step regressions in lags. This is to ensure the maximal compatibility of the second-step regressions with the fashion in which learning and spillovers occur in the DGP. For concreteness, we run the following two second-step regressions:

$$\text{[ALT1]} \qquad \widehat{\omega}_{it} = \alpha_{11} + \alpha_{12} \sum_{j(\neq i)} s_{ij,t-1} G_{j,t-2} + \alpha_{13} G_{i,t-1} + e_{1,it}, \tag{6.4}$$

$$\text{[ALT2]} \qquad \widehat{\omega}_{it} = \alpha_{21} + \alpha_{22} \sum_{j(\neq i)} s_{ij,t-1} \widehat{\omega}_{j,t-1} + \alpha_{23} G_{i,t-1} + e_{2,it}, \tag{6.5}$$

using $\widehat{\omega}_{it}$ recovered in the first step via our semiparametric production function estimator but assuming an exogenous Markov process for productivity $\omega_{it} = \mathbb{E}[\omega_{it}|\omega_{i,t-1}] + \zeta_{it}$.[10] Here we also permit the DL effects as oftentimes done in this literature. Because regressors in both alternative procedures in (6.4)–(6.5) are all weakly exogenous per our DGP and the assumptions, these second-step regressions are estimates via least squares.

To be able to meaningfully analyze the performance of alternative models as well as to fairly compare them to our methodology (especially, in case of the popular ALT1 specification), we focus on the estimands that match in terms of their *qualitative* interpretations. Instead of looking at specific parameters that may not always be directly comparable across the models and with the DGP, we consider the derived measures of DL, SP and TIL as appropriate/available. For instance, of the two alternative methodologies, only the ALT2 specification postulates cross-firm spillovers via the mean peer productivity as in our proposed conceptualization in Section 3 and the DGP. Therefore, α_{22} is essentially comparable to ρ_2 in the DGP: both measure the SP effect. This is however not the case with the ALT1 specification which only models the contextual effect. Hence, we cannot contrast α_{12} to the true ρ_2 value in the DGP. Having said that, α_{12} measuring the (twice lagged) total indirect effect of the peers' G can indeed be meaningfully compared to the similarly interpretable $TIL = SP \times DL = \rho_2 \times \rho_3$ effect derived from the

[10] Essentially, firm productivity here is estimated via the semiparametric adaptation of the original Gandhi et al. (2020) procedure modified to take advantage of the "known" parametric form of the production technology.

DGP that also occurs over two periods. When it comes to direct learning, both α_{13} and α_{23} are comparable to the true $DL = \rho_3$ from the DGP in (6.3). Tables 2–3 summarize these results.

To examine the ability of alternative models to identify firm productivity, we first study if they can consistently estimate the production function coefficients (here β_K) because $\widehat{\omega}_{it}$ is a direct construct of these parameters: $\widehat{\omega}_{it} = y_{it} - \widehat{\beta}_K k_{it} - \widehat{\beta}_M m_{it} - \widehat{\eta}_{it}$.[11] The corresponding estimates of β_K are reported in Table G.2 in Appendix G. These first-step results apply to both the ALT1 and ALT2 models and are obtained assuming that ω_{it} is an exogenous first-order Markov process. As expected, the estimation of production-function parameters (and hence, firm productivity) becomes biased with no tangible improvement following the growth in n as soon as we deviate from the exogenous productivity process [scenarios (i) and (ii)]. In the latter case, biases originate from misspecification of the productivity proxy function that is missing relevant controls pertaining to productivity-modifying learning and/or spillovers.

As seen in Tables 2–3, the misestimation of productivity feeds into the second-step regressions. Across all experiments, both the ALT1 and ALT2 models exhibit non-vanishing biases in the estimation of spillovers. The same is also generally the case for estimation of within-firm direct learning, with the exception of the ALT2 estimator in the least probable scenarios when G is an irrelevant uncorrelated covariate (i.e., when $DL = 0$ *and* G evolves exogenously in population). Notably and perhaps more importantly, the alternative estimators fail at identifying (zero) cross-firm spillovers even when exogeneity of the Markov productivity process assumed in the first step is true [scenario (iii)]. This is because the second-step regressions remain misspecified due to their omission of the lagged productivity as customarily done in spillovers studies. Thus, if the first-step assumption of exogenous productivity in such analyses is indeed correct, the "evidence" of cross-firm spillovers uncovered in the second step is likely spurious and effectively driven by the missing *auto*regressive dynamics in productivity *within* the firm. This is not just a feature of specifications in (6.4)–(6.5). In Appendix G, we consider their multiple variants drawn from the literature. Those results provide further evidence of the potential for spurious

[11] The estimates of $\widehat{\beta}_M$ and $\widehat{\eta}_{it}$ are obtained from the material revenue share regression which does not depend on the Markovian assumption about ω_{it}. Hence, they are exactly the same as those in our methodology.

findings of spillovers using the popular two-step analysis procedure.

7 Empirical Application

We apply our methodology to study cross-firm spillovers with a particular focus on the productivity effects of inbound FDI via the domestic firms' learning of more advanced/efficient foreign knowledge. We proxy the firm's exposure to foreign knowledge using information on the share of foreign capital in its equity. This is a standard measure of foreign knowledge exposure in the literature. Thus, the foreign equity share $G_{it} \in [0,1]$ is our productivity modifier of interest.

Our objective is to study two potential channels—direct and indirect—of the productivity-boosting effects of inbound FDI. First, domestic firms may boost their productivity levels via "importing" better/new technology and learning more efficient management and marketing practices from abroad that they gain *direct* access to through foreign investors; these are direct technology transfers. A second mechanism by which domestic firms may *indirectly* improve their productivity is by learning from other spatially proximate foreign-invested/owned firms in the industry and then adopting their superior practices already imported into the country. The latter channel is indirect and works through cross-firm peer effects. To model these indirect productivity effects of FDI, we need to explicitly recognize the potential for cross-sectional dependence in productivity which would permit FDI spillovers capable of influencing the domestic firms' productivity levels (and hence their output) beyond the immediate recipients. Our proposed model in Section 3 readily provides an empirical framework for this analysis. It allows identification of both the direct/internal (DL) and indirect/external (TIL) effects of inbound FDI in the presence of non-zero productivity spillovers (SP) among peer firms. In line with Remark 3, we model "FDI spillovers" as operating through the firm's exposure to the average peer productivity, i.e., via "productivity spillovers" due to agglomeration externalities more broadly.

Data. Our data come from the Chinese Industrial Enterprises Database survey conducted by China's National Bureau of Statistics. We focus on the electric machinery and equipment manufacturing industry, SIC 2-digit code 39. The rationale behind the choice of this industry is

discussed in Appendix B. Our sample period runs from 1998 to 2007, and the operational sample is an unbalanced panel of 23,720 firms with a total of 73,095 observations. In Appendix H, we provide the details of variable construction and describe the data.

We use postal codes to identify spatial neighbors included in each firm's peer group $\mathscr{L}(i,t)$. Peers are defined at the city level and at the level of the upper administrative division (provinces, autonomous regions, municipalities under the direct rule of government and special administrative regions) to allow for a broader geographical extent of spillovers while also respecting regulatory, administrative and cultural heterogeneity across regions. For the baseline results, the industrial scope of peer effects is defined at the level of the whole 2-digit industry. We consider a more granular definition of industrial similarity at the 4-digit level in robustness checks.

7.1 Results

Owing to the nonparametric specification of the firm productivity process, we obtain observation-specific heterogeneous estimates of SP_{it}, DL_{it} and TIL_{it}. We estimate the unknown $h(\cdot)$ via sieve methods using the popular second-degree polynomial series.[12] All estimations include time effects (the quadratic time trend yields qualitatively similar results). Also note that, because ω_{it} is the log-productivity, SP_{it} is an elasticity measured in percents per unit percent of firm productivity, whereas both the DL_{it} and TIL_{it} are semi-elasticities measured in percents per unit percentage *point* change in the firm's foreign equity share. The measured learning effects on productivity are short-run and partial (i.e., for one given firm only). They do not capture mutual peer effects of an FDI injection across the network of firms, and neither do they account for dynamic effects over time. Obviously, owing to the persistence and cross-peer dependence of productivity, the cumulative implications of FDI for domestic firms' productivity in the long run equilibrium will be more sizable due to accumulation and diffusion over time and space.

Table 4 summarizes semiparametric point estimates of cross-firm productivity spillovers along with the direct and indirect effects of FDI on the productivity of domestic firms from our

[12] For instance, see De Loecker et al. (2016) or Gandhi et al. (2020). We have also experimented with higher-order polynomials, and the results are very similar except somewhat noisier, as expected.

baseline specification,[13] in which each firm's peer group is restricted to the same province and the industrial scope of spillovers is defined at the level of the entire 2-digit industry. All reported estimates are accompanied by the two-tailed 95% bootstrap intervals. In addition, we formally test for significantly *positive* productivity effects at each observation using the *one*-sided 95% bootstrap lower bounds. Throughout, we use accelerated bias-corrected bootstrap percentile confidence intervals (see Appendix F). The share of firms for which the estimates statistically exceed zero are reported in the last column of Table 4. In Appendix I, we also summarize these productivity effect estimates graphically via empirical distributions across firms.

The estimated median DL effect of own FDI is 0.14, whereby an increase of the foreign share in the median firm's equity by 10 percentage points boosts its productivity next year by 1.4%. Expectedly, the TIL effect of peers' FDI is smaller in magnitude—0.04 at the median—so a 10 percentage point increase in the peer group average of the foreign equity share boosts the firm's productivity by only 0.4%. Overall, at least 87% of firms enjoy significant productivity-boosting effects of inbound FDI, both directly and indirectly.

The non-zero external/indirect learning effect of FDI is facilitated by the presence of substantial and positive cross-firm productivity spillovers in the industry, with the median spillover elasticity SP estimated at 0.33 along with the corresponding interquartile range of 0.18–0.45. Thus, a 10% improvement in the average productivity of the firm's peers is estimated to increase the median firm's own productivity by about 3.3%. These productivity spillovers are significantly positive for roughly 84% of firms in the industry. We examine their geographic distribution in Appendix I.

Heterogeneity and Nonlinearity. Even within a given industry, firms are highly heterogeneous across many dimensions including their productivity and the extent of their exposure to foreign investors, both direct and through their peers. These characteristics can influence the effect size of spillovers and learning. Conveniently, our model readily facilitates testing of that.

[13] The associated production function parameter estimates are $\widehat{\beta}_M = 0.74$, $\widehat{\beta}_K = 0.05$ and $\widehat{\beta}_L = 0.12$ with the implied scale elasticity of 0.91. These are in line with the Cobb-Douglas estimates for Chinese manufacturing reported in the literature (e.g., see Brandt et al., 2017) and suggest that the industry exhibits the decreasing returns to scale.

Recall that we estimate the productivity effects of interest via $\widehat{SP}_{it} = \partial \widehat{h}_{it}(\cdot)/\partial \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}$ and $\widehat{DL}_{it} = \partial \widehat{h}_{it}(\cdot)/\partial G_{i,t-1}$, where we estimate $h(\cdot)$ using the second-order polynomial sieve approximation. Thus, by analytical derivation, the estimated $\widehat{SP}_{it}$ and $\widehat{DL}_{it}$ are the linear functions of the "determinants" of firm productivity $(\omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1})'$. Table 5 reports the estimates of sieve coefficients on these three variables in the SP and DL functions.

Consider the spillovers first. The coefficient on the firm's own productivity is negative, indicating that the spillover effects decline in magnitude as firms themselves become more productive. Thus, less productive manufacturers have a greater potential to benefit from positive peer effects. Also consistent with economic intuition, the effect size of spillovers increases with the average productivity of peers: there is more to learn from highly productive neighbors in the industry. We also find a negative relationship between the firm's foreign equity share and the effect size of spillovers. This suggests that the domestic firms experiencing larger productivity improvements via indirect learning from their foreign-invested peers—thanks to positive productivity spillovers—are those with limited direct access to foreign knowledge through their own investors (i.e., low-foreign-equity-share firms).

In the case of FDI effects on productivity, results in the far right column of Table 5 suggest that the direct learning effects diminish as the firm's productivity rises, implying that the more productive firms have less absorptive capacity to learn. The foreign share in a firm's equity negatively affects the learning effect size, which basically indicates the diminishing productivity returns to receiving FDI. Lastly, there is evidence that the higher the average of peer productivity, the lesser the productivity boosts from FDI. Thus, positive spillovers from highly productive neighbors essentially diminish the importance of direct FDI effects.

Robustness Analysis. We first assess robustness of our empirical findings of significantly positive productivity spillovers and learning effects of FDI to the following modeling choices: (i) the inclusion of peer group effects to control for unobservable "correlated effects;" (ii) the composition of a reference peer group $\mathscr{L}(i,t)$; and (iii) the peer-weighting scheme $\{s_{ijt}\}$.

As discussed in Section 3, to structurally identify productivity spillovers SP, we rule out

unobservable "correlated effects" at the peer group level. However, we can replace this no-group-effects assumption with a much milder assumption allowing for network unobservables but having them be time-invariant. In this case, we can control for the potential network confounders using group-level fixed effects (see Graham & Hahn, 2005; Bramoullé et al., 2009). We consider group effects across both the spatial and industrial dimensions. Specifically, we re-estimate our baseline specification by adding fixed effects at the level of the entire peer group as well as more granular subgroups.[14] The corresponding results are summarized in columns F1–F4 of Table 6 (see table notes for the details on group fixed effects). While predictably there is no dramatic change in the DL estimates of within-firm learning, the median effect size of cross-firm spillovers SP increases notably when we rely solely on the within-group variation over time to estimate the productivity peer effects.[15] The latter is especially true when the correlated group effects are defined narrowly at the 4-digit sub-industry level. In this case, the median spillover effect is 0.61 (against the baseline estimate of 0.33) and the effect is statistically positive for almost all firms (99%). The increase in the effect size of spillovers is indicative of the substantial between-group heterogeneity in (peer) firm productivity, which is consistent with the well-documented differential in productivity levels across regions in China. Thus, when omitting group-specific effects, the measure of the strength of peer dependence across firms gets "diluted" in the baseline model due to the variation across groups.[16]

In columns P1–P3 of Table 6, we estimate productivity spillovers under three alternative definitions of who the firm's relevant peers are. Each one presumes a much smaller reference group than the baseline. Namely, we consider narrowing the scope of local spillovers to the level of city and/or the 4-digit sub-industry. The direct effect of the firm's own FDI expectedly continues to stay largely unchanged, but the estimates of productivity spillovers diminish in size significantly. The latter indirectly corroborates the rationale of our baseline specification in that the agglomeration effects have broad geographical and industrial scopes. By restricting

[14] These peer group effects are included in the second-stage estimation that models the productivity process.
[15] Larger magnitudes of TIL are a direct result of the increased SP estimates.
[16] This is the case even if the strength of within-group spillovers is the same for all groups.

the extent of spillovers to the local city and/or the firm's sub-industry only, we also restrict the reach of cross-firm externalities in productivity. Intuitively, when restricting the firm's learning opportunities to a narrower group of neighbors, a 10% improvement in the average peer productivity is estimated to help boost the firm's own productivity by only about 0.6–0.9% at the median. In contrast, if the relevant peer reference group is actually larger in scope, the same 10% improvement across all peers (as in the baseline specification) implies a bigger industry-wide aggregate effect and, consequently, a larger estimated spillover effect on the firm of 3.3%.

In Table 6, we also consider an alternative way of weighting peers, whereby bigger neighbors get assigned larger relative weights (column W1). The spillover effects only modestly decline in size. Overall, the cross-model variation in the spillover estimates we observe in Table 6 is unsurprising and, in fact, expected because each model treats peer interactions a bit differently and/or utilizes different variation in data to identify productivity spillovers. Having said that, the *SP* point estimates across all models are highly positively correlated, with the rank correlation coefficient being 0.81 on average. Consistently across all specifications, we continue to find that the overwhelming majority of the electric machinery manufacturers in China enjoy positive and significant productivity spillovers, in general, and FDI spillovers, specifically.

Appendix I contains additional robustness checks, including to the potential violation of the weak exogeneity of the *lagged* foreign equity share. Controlling for potential endogeneity of the FDI exposure, we continue to find strong evidence in support of significantly positive productivity spillovers for 80% of firms or more, with our findings remaining qualitatively unchanged. In the same appendix, we also explore potential heterogeneity in the external productivity spillovers from the peers *conditional* on their FDI status, which gives rise to *bi*dimensional spillovers. We find evidence of heterogeneity in the strength of spillovers from wholly-domestic versus foreign-invested peers but, overall, our main findings stay the same: productivity spillovers are positive and significant for most firms. For the details, see Appendix I.

8 Conclusion

This paper develops a novel methodology for the proxy variable structural identification of (latent) firm productivity in the presence of learning and cross-firm spillovers which allows a unified one-step analysis of the knowledge-transfer effects *between* peer firms. Our framework is fundamentally different from the popular empirical approach traditionally implemented in two steps, whereby one first recovers firm productivity using the available standard proxy variable estimators and then tests for spillovers in the second step by regressing these productivity estimates on various peer-group averages capturing firms' exposure to potential spillovers. Contrary to such an approach, our methodology is "internally consistent" in that it does not postulate contradictory assumptions. In building our model, we explicitly accommodate cross-sectional dependence in firm productivity induced by spillovers. We also show that estimating the firm production function or productivity using traditional proxy methods while ignoring the spillover-induced cross-sectional dependence, as customarily done in the literature, likely leads to misspecification and endogeneity-generating omitted variable bias. Because our methodology can be easily adapted to admit various spillover origins such as spatial agglomeration, R&D, FDI, exporting, etc., it is fit to investigate cross-firm productivity spillovers in many contexts.

References

Acharya, R. C. & Keller, W. (2008). Estimating the productivity selection and technology spillover effects of imports. NBER Working Paper 14079.

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, *83*, 2411—2451.

Ai, C. & Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, *71*, 1795–1843.

Alvarez, R. & López, R. A. (2008). Is exporting a source of productivity spillovers? *Review of World Economics*, *144*(4), 723–749.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, *77*, 1229–1279.

Balsvik, R. (2011). Is labor mobility a channel for spillovers from multinationals? Evidence from Norwegian manufacturing. *Review of Economics and Statistics*, *93*, 285–297.

Barrios, S., Görg, H., & Strobl, E. (2011). Spillovers through backward linkages from multinationals: Mea-

surement matters! *European Economic Review*, *55*, 862–875.

Bazzi, S., Chari, A. V., Nataraj, S., & Rothenberg, A. D. (2017). Identifying productivity spillovers using the structure of production networks. *RAND Working Paper* WR-1128.

Behrens, K., Duranton, G., & Robert-Nicoud, F. (2014). Productive cities: Sorting, seleciton, and agglomeration. *Journal of Political Economy*, *122*, 507–553.

Blalock, G. & Gertler, P. J. (2008). Welfare gains from foreign direct investment through technology transfer to local suppliers. *Journal of International Economics*, *74*, 402–421.

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

Blume, L. E., Brock, W. A., Durlauf, S. N., & Ioannides, Y. M. (2011). Identification of social interactions. In J. Benhabib, A. Bisin, & M. O. Jackson (Eds.), *Handbook of Social Economics*, volume 1B. North-Holland, Amsterdam.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*, 41–55.

Brandt, L., Van Biesebroeck, J., Wang, L., & Zhang, Y. (2017). WTO accession and performance of Chinese manufacturing firms. *American Economic Review*, *107*, 2784–2820.

Branstetter, L. G. (2001). Are knowledge spillovers international of intranational in scope? Microeconometric evidence from the U.S. and Japan. *Journal of International Economics*, *53*, 53–79.

Chen, X., Linton, O., & Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, *71*, 1591–1608.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Dinstinguishing agglomeration from firm selection. *Econometrica*, *80*, 2543–2594.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, *5*, 1–21.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., & Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, *84*, 445–510.

De Loecker, J. & Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, *102*, 2437–2471.

Deloitte China Manufacturing Industry Group (2013). A new stage for overseas expansion for China's equipment manufacturing industry. Report by Deloitte China Research and Insight Centre.

Doraszelski, U. & Jaumandreu, J. (2013). R&D and productivity: Estimating endogenous productivity. *Review of Economic Studies*, *80*, 1338–1383.

Doraszelski, U. & Jaumandreu, J. (2018). Measuring the bias of technological change. *Journal of Political Economy*, *126*, 1027–1084.

Duranton, G. & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In J. V. Henderson & J. Thisse (Eds.), *Handbook of Regional and Urban Economics*, volume 4. Elsevier B.V.

Efron, B. (1987). Better bootstrap confidence interval. *Journal of American Statistical Association*, *82*, 171–200.

Eichengreen, B. & Tong, H. (2007). Is China's FDI coming at the expense of other countries? *Journal of the Japanese and International Economies*, *21*(2), 153–172.

Euro Exim Bank (2020). Export of electrical machinery from China. Euro Exim Bank Global Finance Blog; October 30, 2020.

Gandhi, A., Navarro, S., & Rivers, D. (2020). On the identification of gross output production functions. *Journal of Political Economy*, *128*, 2973–3016.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, *108*, 3117–3153.

Graham, B. S. & Hahn, J. (2005). Identification and estimation of the linear-in-means model of social interactions. *Economics Letters*, *88*, 1–6.

Grieco, P. L. E., Li, S., & Zhang, H. (2016). Production function estimation with unobserved input price dispersion. *International Economic Review*, *57*, 665–689.

Griffith, R., Harrison, R., & Van Reenen, J. (2006). How special is the special relationship? Using the impact of U.S. R&D spillovers on U.K. firms as a test of technology sourcing. *American Economic Review*, *96*, 1859–1875.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics*, *10*, 92–116.

Griliches, Z. & Mairesse, J. (1998). Production functions: The search for identification. In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium* (pp. 169–203). Cambridge University Press.

Hahn, J., Liao, Z., & Ridder, G. (2018). Nonparametric two-step sieve M estimation and inference. *Econometric Theory*. forthcoming.

Haskel, J. E., Pereira, S. C., & Slaughter, M. J. (2007). Does inward foreign direct investment boost the productivity of domestic firms? *Review of Economics and Statistics*, *89*, 482–496.

Hernán, M. A. & Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall.

Hirano, K. & Imbens, G. W. (2004). The propensity score with continuous treatments. In W. A. Shewhart & S. S. Wilks (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*. New York: Wiley & Sons, Ltd.

Horowitz, J. L. (2001). The bootstrap. In J. J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics* (5 ed.). chapter 52, (pp. 3159–3228). Elsevier Science B.V.

Ihrcke, J. & Becker, K. (2006). Study on the future opportunities and challenges of EU-China trade and investment relations. Study 1 of 12: Machinery. Report commissioned and financed by the Commission of the European Communities.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*, 5–86.

Javorcik, B. S. (2004). Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review*, *94*, 605–627.

Javorcik, B. S. & Spatareanu, M. (2008). To share or not to share: Does local participation matter for spillovers from foreign direct investment? *Journal of Development Economics*, *85*, 194–217.

Jiang, K., Keller, W., Qiu, L. D., & Ridley, W. (2019). International joint ventures and internal vs. external technology transfer: Evidence from China. NBER Working Paper 24455.

Kelejian, H. H. & Prucha, I. R. (1999). A generalized moment estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40*, 509–533.

Keller, W. (2008). Transfer of technology. In L. E. Blume & S. N. Durlauf (Eds.), *The New Palgrave Dictionary of Econometrics* (2 ed.). New York: Macmillan.

Keller, W. (2010). International trade, foreign direct investment, and technology spillovers. In B. Hall & N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, volume 2 (pp. 793–829). Amsterdam: North Holland.

Keller, W. & Yeaple, S. R. (2009). Multinational enterprises, international trade, and productivity growth: Firm-level evidence from the United States. *Review of Economics and Statistics*, *91*, 821–831.

Krugman, P. R. (1991). *Geography and Trade*. Cambridge, MA: MIT Press.

Kuersteiner, G. M. & Prucha, I. R. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, *88*, 2109–2146.

Lee, L.-f. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, *137*, 489–514.

LeSage, J. & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton: Taylor & Francis CRC Press.

Levinsohn, J. & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, *70*, 317–341.

Lewbel, A. (2012). Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics*, *30*, 67–80.

Lu, Y., Tai, Z., & Zhu, L. (2017). Identifying FDI spillovers. *Journal of International Economics*, *107*, 75–90.

Malikov, E., Zhao, S., & Kumbhakar, S. C. (2020). Estimation of firm-level productivity in the presence of exports: Evidence from China's manufacturing. *Journal of Applied Econometrics*, *81*, 457–480.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics*, *21*, 255–285.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*, 531–542.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, *14*, 115–136.

Marshall, A. (1890). *Principles of Economics* (1st ed.). London: Macmillan.

Moffitt, R. A. (2001). Policy interventions, low-level equilibria and social internations. In S. N. Durlauf & H. P. Young (Eds.), *Social Dynamics*. MIT Press.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters*, *14*(2), 201–206.

Newman, C., Rand, J., Talbot, T., & Tarp, F. (2015). Technology transfers, foreign investment and productivity spillovers. *European Economic Review*, *76*, 168–187.

Olley, G. S. & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, *64*, 1263–1297.

Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with a multifactor error structure. *Econometrica*, *74*, 967–1012.

Poole, J. P. (2013). Knowledge transfers from multinationals to domestic firms: Evidence from worker mobility. *Review of Economics and Statistics*, *95*, 393–406.

Rosenthal, S. S. & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In J. V. Henderson & J. Thisse (Eds.), *Handbook of Regional and Urban Economics*, volume 4. Elsevier B.V.

Serpa, J. C. & Krishnan, H. (2018). The impact of supply chains on firm-level productivity. *Management Science*, *64*, 511–532.

Shao, J. & Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag New York Inc.

Stoyanov, A. & Zubanov, N. (2012). Productivity spillovers across firm through worker mobility. *American Economic Journal: Applied Economics*, *4*, 168–198.

Topalova, P. & Khandelwal, A. (2011). Trade liberalization and firm productivity: The case of India. *Review of Economics and Statistics*, *93*, 995–1009.

Van Biesebroeck, J. (2005). Exporting raises productivity in sub-Saharan African manufacturing firms. *Journal of International Economics*, *67*, 373–391.

Zacchia, P. (2020). Knowledge spillovers through networks of scientists. *Review of Economic Studies*, *87*, 1989–2018.

Table 1. Simulation Results for Our Estimation Methodology

	True Value	$n = 100$ Mean	RMSE	MAE	$n = 200$ Mean	RMSE	MAE	$n = 400$ Mean	RMSE	MAE
DGP WITH G EVOLVING EXOGENOUSLY										
Scenario (i): $DL \neq 0$ **and** $SP \neq 0$										
β_K	0.25	0.249	0.055	0.043	0.250	0.038	0.030	0.252	0.027	0.021
AR	0.55	0.546	0.080	0.065	0.547	0.054	0.044	0.549	0.038	0.031
DL	0.50	0.500	0.164	0.134	0.502	0.113	0.091	0.501	0.079	0.064
SP	0.40	0.399	0.144	0.116	0.400	0.102	0.082	0.399	0.070	0.057
TIL	0.20	0.199	0.121	0.097	0.201	0.085	0.069	0.199	0.059	0.047
Scenario (ii): $DL = 0$ **and** $SP \neq 0$										
β_K	0.25	0.249	0.055	0.042	0.250	0.037	0.029	0.252	0.027	0.021
AR	0.55	0.546	0.085	0.070	0.546	0.059	0.048	0.549	0.042	0.034
DL	0	–0.002	0.154	0.124	0.001	0.106	0.087	0.000	0.075	0.060
SP	0.40	0.395	0.172	0.140	0.402	0.124	0.102	0.400	0.085	0.069
TIL	0	0.004	0.067	0.049	0.002	0.045	0.034	0.000	0.030	0.024
Scenario (iii): $DL = 0$ **and** $SP = 0$										
β_K	0.25	0.248	0.051	0.040	0.250	0.035	0.028	0.251	0.025	0.020
AR	0.55	0.546	0.083	0.067	0.546	0.058	0.047	0.549	0.040	0.033
DL	0	0.003	0.155	0.125	0.003	0.106	0.086	0.001	0.075	0.061
SP	0	–0.031	0.157	0.131	–0.009	0.108	0.091	–0.004	0.078	0.065
TIL	0	–0.004	0.028	0.018	–0.002	0.013	0.008	–0.001	0.006	0.004
DGP WITH G FOLLOWING AN ω-CONTROLLED PROCESS										
Scenario (i): $DL \neq 0$ **and** $SP \neq 0$										
β_K	0.25	0.249	0.060	0.047	0.251	0.040	0.032	0.252	0.029	0.023
AR	0.55	0.544	0.110	0.089	0.546	0.073	0.060	0.548	0.053	0.043
DL	0.50	0.503	0.148	0.120	0.502	0.099	0.081	0.502	0.070	0.057
SP	0.40	0.404	0.058	0.047	0.403	0.041	0.033	0.401	0.028	0.023
TIL	0.20	0.201	0.069	0.056	0.201	0.047	0.038	0.201	0.032	0.026
Scenario (ii): $DL = 0$ **and** $SP \neq 0$										
β_K	0.25	0.251	0.058	0.044	0.250	0.037	0.029	0.252	0.027	0.021
AR	0.55	0.544	0.105	0.087	0.545	0.070	0.058	0.548	0.051	0.042
DL	0	0.004	0.134	0.108	0.003	0.092	0.076	0.002	0.067	0.055
SP	0.40	0.387	0.224	0.183	0.400	0.158	0.130	0.399	0.116	0.094
TIL	0	–0.007	0.060	0.043	–0.003	0.038	0.028	–0.001	0.027	0.021
Scenario (iii): $DL = 0$ **and** $SP = 0$										
β_K	0.25	0.248	0.052	0.041	0.250	0.035	0.028	0.251	0.025	0.019
AR	0.55	0.546	0.101	0.083	0.546	0.069	0.056	0.548	0.050	0.040
DL	0	–0.001	0.134	0.109	0.000	0.092	0.075	0.001	0.068	0.054
SP	0	–0.031	0.155	0.128	–0.008	0.105	0.088	–0.004	0.075	0.063
TIL	0	0.000	0.022	0.015	0.000	0.011	0.007	0.000	0.005	0.003

NOTES: Owing to linearity of the productivity process in (6.3), the true values of AR, DL, SP and TIL are the fixed coefficients for all i and t, and $TIL = SP \times DL$ is derived indirectly. Throughout, $T = 10$.

Table 2. Simulation Results for the Two-Step Alternative Estimator of Spillovers ALT1

	True Value	$n = 100$ Mean	$n = 100$ RMSE	$n = 200$ Mean	$n = 200$ RMSE	$n = 400$ Mean	$n = 400$ RMSE
DGP WITH G EVOLVING EXOGENOUSLY							
Scenario (i): $DL \neq 0$ **and** $SP \neq 0$							
DL	0.50	0.167	0.355	0.171	0.340	0.173	0.333
TIL	0.20	–0.577	0.793	–0.578	0.787	–0.577	0.781
Scenario (ii): $DL = 0$ **and** $SP \neq 0$							
DL	0	–0.202	0.225	–0.200	0.211	–0.200	0.206
TIL	0	–0.348	0.372	–0.345	0.361	–0.352	0.357
Scenario (iii): $DL = 0$ **and** $SP = 0$							
DL	0	0.461	0.470	0.464	0.468	0.465	0.467
TIL	0	0.165	0.204	0.170	0.191	0.169	0.180
DGP WITH G FOLLOWING AN ω-CONTROLLED PROCESS							
Scenario (i): $DL \neq 0$ **and** $SP \neq 0$							
DL	0.50	1.651	1.152	1.650	1.151	1.652	1.152
TIL	0.20	0.582	0.387	0.585	0.387	0.586	0.387
Scenario (ii): $DL = 0$ **and** $SP \neq 0$							
DL	0	0.416	0.420	0.420	0.422	0.424	0.425
TIL	0	–0.112	0.119	–0.113	0.117	–0.115	0.117
Scenario (iii): $DL = 0$ **and** $SP = 0$							
DL	0	0.616	0.619	0.617	0.619	0.621	0.621
TIL	0	–0.362	0.365	–0.361	0.362	–0.362	0.363

NOTES: Reported are the results from the second-step regression in (6.4) estimated with $\widehat{\omega}_{it}$ obtained in the first step using the standard proxy estimator under the assumption of exogenous Markov productivity process. DL and TIL are respectively measured by α_{13} and α_{12}, with the latter capturing spillovers. Throughout, $T = 10$.

Table 3. Simulation Results for the Two-Step Alternative Estimator of Spillovers ALT2

	True Value	$n = 100$ Mean	$n = 100$ RMSE	$n = 200$ Mean	$n = 200$ RMSE	$n = 400$ Mean	$n = 400$ RMSE
DGP WITH G EVOLVING EXOGENOUSLY							
Scenario (i): $DL \neq 0$ **and** $SP \neq 0$							
DL	0.50	0.800	0.313	0.814	0.320	0.817	0.320
SP	0.40	1.074	0.676	1.083	0.684	1.090	0.690
Scenario (ii): $DL = 0$ **and** $SP \neq 0$							
DL	0	–0.036	0.093	–0.015	0.063	–0.008	0.044
SP	0.40	0.751	0.518	0.874	0.502	0.924	0.527
Scenario (iii): $DL = 0$ **and** $SP = 0$							
DL	0	0.005	0.084	0.004	0.059	0.003	0.042
SP	0	0.539	0.541	0.545	0.546	0.548	0.548
DGP WITH G FOLLOWING AN ω-CONTROLLED PROCESS							
Scenario (i): $DL \neq 0$ **and** $SP \neq 0$							
DL	0.50	1.149	0.650	1.152	0.652	1.156	0.656
SP	0.40	0.580	0.182	0.579	0.180	0.577	0.178
Scenario (ii): $DL = 0$ **and** $SP \neq 0$							
DL	0	0.369	0.376	0.376	0.379	0.375	0.377
SP	0.40	–0.708	1.304	–0.635	1.121	–0.583	1.020
Scenario (iii): $DL = 0$ **and** $SP = 0$							
DL	0	0.406	0.406	0.405	0.406	0.407	0.407
SP	0	0.390	0.393	0.395	0.396	0.397	0.397

NOTES: Reported are the results from the second-step regression in (6.5) estimated with $\widehat{\omega}_{it}$ obtained in the first step using the standard proxy estimator under the assumption of exogenous Markov productivity process. DL and SP are respectively measured by α_{23} and $SP = \alpha_{22}$. Throughout, $T = 10$.

Table 4. Estimates of the Productivity Effects

	Point Estimates			*Statistically* > 0
Estimand	1st Qu.	Median	3rd Qu.	(% Obs.)
SP	0.182	0.327	0.452	83.84
	(0.108, 0.240)	(0.211, 0.399)	(0.302, 0.543)	
DL	0.096	0.138	0.168	89.13
	(0.076, 0.118)	(0.118, 0.164)	(0.146, 0.197)	
TIL	0.022	0.037	0.053	86.71
	(0.011, 0.032)	(0.022, 0.050)	(0.034, 0.068)	

NOTES: Reported are the results based on our baseline specification of the productivity process in (3.3). The left panel summarizes point estimates of SP_{it}, DL_{it} and TIL_{it} with the corresponding two-sided 95% bootstrap percentile confidence intervals in parentheses. The last column reports the share of observations for which the point estimates are statistically positive at the 5% significance level using one-sided bootstrap confidence intervals.

Table 5. Heterogeneity and Nonlinearity in the Productivity Effects

	SP	DL
$\omega_{i,t-1}$	–0.735	–0.164
	(–0.890, –0.550)	(–0.203, –0.126)
$G_{i,t-1}$	–0.198	–0.155
	(–0.403, –0.103)	(–0.184, –0.127)
$\sum_j s_{ij,t-1}\omega_{j,t-1}$	1.249	–0.198
	(0.423, 1.527)	(–0.403, –0.103)

NOTES: Reported are the parameter estimates for the SP and DL functions derived from the polynomial approximation of the conditional mean of ω_{it} in the productivity process formulation in (3.3). Two-sided 95% bootstrap percentile confidence intervals in parentheses. These correspond to our baseline specification.

Table 6. Estimates of the Productivity Effects: Robustness Analysis

Estimand	F1	F2	F3	F4	W1	P1	P2	P3
				—Median Estimates—				
SP	0.499	0.410	0.607	0.605	0.222	0.076	0.085	0.064
	(0.386, 0.535)	(0.313, 0.428)	(0.540, 0.614)	(0.524, 0.647)	(0.186, 0.254)	(0.048, 0.100)	(0.049, 0.122)	(0.051, 0.077)
DL	0.167	0.176	0.171	0.177	0.144	0.146	0.145	0.149
	(0.145, 0.194)	(0.155, 0.205)	(0.149, 0.198)	(0.156, 0.205)	(0.123, 0.171)	(0.123, 0.172)	(0.121, 0.171)	(0.128, 0.174)
TIL	0.071	0.060	0.089	0.092	0.009	0.009	0.010	0.008
	(0.050, 0.079)	(0.046, 0.067)	(0.074, 0.100)	(0.076, 0.106)	(0.006, 0.014)	(0.006, 0.013)	(0.006, 0.016)	(0.006, 0.011)
				—Statistically > 0 **(% Obs.)—**				
SP	95.98	94.19	98.78	99.15	77.96	92.25	89.69	88.89
DL	90.24	90.39	90.34	90.43	89.11	89.43	89.69	89.57
TIL	96.31	94.72	99.07	99.31	72.46	93.11	89.34	94.97

NOTES: Reported are the results for the productivity process formulation in (3.3). The two-sided 95% bootstrap percentile confidence intervals for the median point estimates are in parentheses. Statistical positiveness is at the 5% significance level, using one-sided bootstrap percentile confidence intervals. In the baseline specification: (i) each firm's peers are restricted to the firms located in the same province and the industrial scope of spillovers defined at the level of the entire 2-digit industry, (ii) the technical change is flexibly controlled for using a series of year effects. The other model specifications are based on the baseline except for the following alternate features. F1: includes location fixed effects at the level of peer groups (here, province); F2: includes location fixed effects at the level of peer subgroups (here, city); F3: includes location-industry fixed effects at the level of peer subgroups (here, province and 4-digit sub-industry); F4: includes spatial and industry fixed effects at the level of peer subgroups (here, city and 4-digit sub-industry); W1: peers are size-weighted, i.e., $s_{ijt} = L_{jt}\mathbb{1}\{(j,t) \in \mathscr{L}(i,t)\}/\sum_k L_{kt}\mathbb{1}\{(k,t) \in \mathscr{L}(i,t)\}$; P1: peers in $\mathscr{L}(i,t)$ are restricted to the 4-digit sub-industry; P2: peers in $\mathscr{L}(i,t)$ are restricted to the same city; P3: peers in $\mathscr{L}(i,t)$ are restricted to the 4-digit sub-industry *and* the same city.

Online Appendix to

On the Estimation of Cross-Firm Productivity Spillovers with an Application to FDI

EMIR MALIKOV[1] SHUNAN ZHAO[2]

[1]University of Nevada, Las Vegas

[2]Oakland University

A Relation to the Augmented Production Function Approach

A two-step framework, which we seek to improve upon in this paper, is not universal across empirical studies of productivity spillovers. The exceptions are predominantly from the literature on R&D-borne productivity spillovers, where some studies instead adopt a singe-step methodology centered on the estimation of the Griliches (1979)-style "augmented production function" which, besides the conventional inputs, also explicitly admits the firm's own and external knowledge capital stock. Seemingly, such a model readily provides estimates of the "contextual" spillover effects of R&D on firm production in one step. However, this framework is rather unique to the studies of spillovers in R&D, because this productivity-enhancing activity is the most input-accumulation-like in that it is an investment into the knowledge capital. Augmenting the firm's production function to include the FDI/exports/imports variables and their respective spillover pool measures is however not as conceptually unambiguous. Specifically, this is problematic on at least two fronts. First, the spillover effects on firm productivity in such a setup is essentially assumed to be deterministic, whereby the impact on productivity is improbably the same for all firms without a possibility of the varying degree of success (say, due to random luck or misfortune). Second, including internal and external measures of productivity modifiers directly into the production function effectively implies substitutability of the firm's inputs with not only its own productivity-enhancing activities such FDI or exporting but also—and perhaps more eyebrow-raising—with those of its peers. This remark equally applies to the case of R&D spillovers[1] and is along the lines of De Loecker's (2013) critique in the context of estimating the learning-by-exporting effects on firm productivity. Not least importantly, identification of productivity spillovers in prior studies (including those via R&D) may also be seriously hindered by the well-known econometric problems with standard proxy-based or (dynamic panel) fixed-effects production function estimators.[2] This further highlights the practical

[1] More recently, the literature has been gravitating towards embedding the firm's R&D behavior into the productivity process and taking it out of the production function itself; e.g., see Doraszelski & Jaumandreu (2013, 2018).

[2] See Griliches & Mairesse (1998), Ackerberg et al. (2015) and Gandhi et al. (2020) for the discussion of various production-function estimators and identification challenges associated with them.

usefulness of our proposed methodology.

B China's Electric Machinery Manufacturing

Our empirical analysis focuses on China's electric machinery and equipment manufacturing industry which includes manufacturing of generators and motors, power transmission and distribution equipment, wires and cables, batteries, household electric and non-electric appliances, lighting appliances, etc. We select this industry because it is histrionically one of the country's most fundamental manufacturing sectors. The development of this industry has been closely related to the growth of GDP and the ever-expanding demand of electricity. By its very nature, the industry has thus been crucial for promoting the overall industrialization in China. Besides that, electric machinery and equipment is also China's most exported product (Euro Exim Bank, 2020), and the industry amounts to over a quarter of global sales (Deloitte China Manufacturing Industry Group, 2013). It is also one of the manufacturing industries that receive most of FDI. For instance, in 2005 alone (near the end of our sample period) the machinery and equipment industry in China attracted $4 billion in foreign investment, which was about 10% of FDI inflows to Chinese manufacturing that year (Ihrcke & Becker, 2006).

Foreign-invested firms are the dominant players in this industry. Arguably, this is mainly due to China's lack of domestic innovation capabilities, excessive failure rates of R&D and, consequently, high dependence on new technologies imported from abroad, particularly during the first decade following renewed privatization efforts in the late 1990s (the period of our analysis). For example, according to the Xiamen Bureau of Statistics, in Xiamen (a large and important port-city on the East coast) just 48 large foreign-invested firms owned 82% of fixed assets and produced 79% of the output value in the local electric machinery industry in 2005.

More generally, the Chinese electric machinery and equipment manufacturing industry is characterized by a high degree of spatial clustering and industrial agglomeration [mainly on the coast; see Figure H.2(a) in Appendix H] typical for such technology- and skill-intensive industries. Along with the government's emphasis on innovations and new technologies as a means

for sustainable development of this industry, this makes it an interesting application for studying productivity effects of inbound FDI and the associated spillovers across firms.

C Translog Production Function

Our methodology can adapt more flexible specifications of the firm's production function. The log-quadratic translog specification provides a natural extension of the log-linear Cobb-Douglas form that we have assumed in (3.1). The former is more flexible and implies input and scale elasticities that vary both over time and across firms thereby being more robust to firm heterogeneity. For instance, see De Loecker & Warzynski (2012) and De Loecker et al. (2016) for recent applications of the translog production functions in the structural proxy estimation.

Let the firm's stochastic production function takes the following form *in logs*:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_K k_{it} + \tfrac{1}{2}\beta_{KK} k_{it}^2 + \beta_L l_{it} + \tfrac{1}{2}\beta_{LL} l_{it}^2 + \beta_M m_{it} + \tfrac{1}{2}\beta_{MM} m_{it}^2 + \\ &\quad \beta_{KL} k_{it} l_{it} + \beta_{KM} k_{it} m_{it} + \beta_{LM} l_{it} m_{it} + \omega_{it} + \eta_{it} \\ &\equiv T(k_{it}, l_{it}, m_{it}) + \omega_{it} + \eta_{it}, \end{aligned} \tag{C.1}$$

where $T(k_{it}, l_{it}, m_{it})$ is a shorthand for the translog expansion of inputs. All the remaining assumptions about the market environment, productivity processes, timing of production decisions and learning, etc. stay unchanged.

The firm's static optimization problem with respect to materials now is

$$\max_{M_{it}} P_t^Y \exp\{T(k_{it}, l_{it}, m_{it})\} \exp\{\omega_{it}\}\theta - P_t^M M_{it}, \tag{C.2}$$

with the corresponding first-order condition given by

$$P_t^Y \exp\{T(k_{it}, l_{it}, m_{it})\} \frac{\beta_M + \beta_{MM} m_{it} + \beta_{KM} k_{it} + \beta_{LM} l_{it}}{M_{it}} \exp\{\omega_{it}\}\theta = P_t^M. \tag{C.3}$$

Dividing (C.3) by the translog production function expressed in levels and then taking logs of both sides, we obtain the following material share equation:

$$\ln V_{it} = \ln(\left[\beta_M + \beta_{MM} m_{it} + \beta_{KM} k_{it} + \beta_{LM} l_{it}\right]\theta) - \eta_{it}, \tag{C.4}$$

where $\beta_M + \beta_{MM} m_{it} + \beta_{KM} k_{it} + \beta_{LM} l_{it}$ is the material elasticity function. Analogous to the

discussion in Section 4, the above share equation identifies the material-related production-function parameters $(\beta_M, \beta_{MM}, \beta_{KM}, \beta_{LM})'$ as well as the mean of exponentiated shocks $\theta = \mathbb{E}[\exp\{\eta_{it}\}]$ based on the mean-orthogonality condition $\mathbb{E}[\eta_{it}|\,\mathscr{I}_{it}] = \mathbb{E}[\eta_{it}] = 0$. These parameters are to be estimated in the first stage via nonlinear least squares on (C.4).

Having identified the production function in the dimension of its endogenous static input m_{it}, we focus on the remaining production-function parameters as well as the nonparametric evolution process for ω_{it}. With the already identified $y^*_{it} \equiv y_{it} - \beta_M m_{it} - \frac{1}{2}\beta_{MM} m^2_{it} - \beta_{KM} k_{it} m_{it} - \beta_{LM} l_{it} m_{it}$ and using the Markovian process for productivity, we now have the analogue of (4.7):

$$y^*_{it} = \beta_K k_{it} + \tfrac{1}{2}\beta_{KK} k^2_{it} + \beta_L l_{it} + \tfrac{1}{2}\beta_{LL} l^2_{it} + \beta_{KL} k_{it} l_{it} + h\left(\omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}\right) + \zeta_{it} + \eta_{it} \tag{C.5}$$

that contains no endogenous variables on the right-hand side. Next, proxying for $\omega_{i,t-1}$ and $\omega_{j,t-1}$ via the inverted material function derived from (C.3), we obtain

$$\begin{aligned} y^*_{it} = {} & \beta_K k_{it} + \tfrac{1}{2}\beta_{KK} k^2_{it} + \beta_L l_{it} + \tfrac{1}{2}\beta_{LL} l^2_{it} + \beta_{KL} k_{it} l_{it} + \\ & h\left(\omega^*_{i,t-1}\left(\beta_K, \beta_L, \beta_{KK}, \beta_{LL}, \beta_{KL}\right), G_{i,t-1}, \sum_{j(\neq i)} s_{ij,t-1}\omega^*_{j,t-1}\left(\beta_K, \beta_L, \beta_{KK}, \beta_{LL}, \beta_{KL}\right)\right) + \zeta_{it} + \eta_{it}, \end{aligned} \tag{C.6}$$

where the productivity proxy function is given by

$$\omega^*_{it}\left(\beta_K, \beta_L, \beta_{KK}, \beta_{LL}, \beta_{KL}\right) = \varkappa^*_{it} - \beta_K k_{it} - \tfrac{1}{2}\beta_{KK} k^2_{it} - \beta_L l_{it} - \tfrac{1}{2}\beta_{LL} l^2_{it} - \beta_{KL} k_{it} l_{it} \quad \forall i, t, \tag{C.7}$$

with

$$\begin{aligned} \varkappa^*_{it} = {} & \ln(P^M_t / P^Y_t) - \ln([\beta_M + \beta_{MM} m_{it} + \beta_{KM} k_{it} + \beta_{LM} l_{it}]\,\theta) - \\ & (1 - \beta_M) m_{it} - \tfrac{1}{2}\beta_{MM} m^2_{it} - \beta_{KM} k_{it} m_{it} - \beta_{LM} l_{it} m_{it} \end{aligned}$$

being a function of the parameters that have already been identified in the first stage.

A semiparametric model in (C.6) is then identified based on the same moment restriction as in (4.10), with all right-hand-side covariates being weakly exogenous and thus self-instrumenting. Approximating the unknown $h(\cdot)$ via linear sieves, (C.6) is to be estimated in the second stage via semiparametric nonlinear least-squares. The remaining aspects closely

follow the estimation procedure outlined in Section 5.

D Asymmetric Productivity Spillovers

Our baseline peer weighing scheme in (3.4) treats cross-firm spillovers symmetrically in that all members of a peer group affect each other's productivity. That is, each ith firm's productivity is influenced by the average productivity of all its peers: those that are more *and* those that are less productive than the firm i itself. Given that we have no prior beliefs about the directionality of productivity spillovers in China's electric machinery manufacturing that we study in our empirical application, we opt for a symmetric specification. But should one choose to regulate the direction of productivity spillovers by restricting them to occur *from* more productive *to* less productive firms, our framework can be modified to accommodate that too.

The latter case however implies a somewhat different conceptualization of cross-firm dependence in which firms are said to learn exclusively from (relative) productivity "leaders." The identification of such *a*symmetric spillovers, which are conditional on the firm's own productivity relative to that of its peers, generally requires additional structural/timing assumptions.

To model productivity spillovers between firms asymmetrically, we can redefine peer weights $\{s_{ijt}\}$ as follows:

$$s^*_{ijt} = \frac{\mathbb{1}\{(j,t) \in \mathscr{L}(i,t) \text{ and } \omega_{j,t-1} > \omega_{i,t-1}\}}{\sum_{k(\neq i)=1}^{n} \mathbb{1}\{(k,t) \in \mathscr{L}(i,t) \text{ and } \omega_{j,t-1} > \omega_{i,t-1}\}}, \tag{D.1}$$

so that only the neighbors who are more productive than the firm i are identified as its peers for external cross-firm learning. Note that, in the above, the relevant peers at time t are selected based on their relative productivity superiority in the *previous* period $t-1$. Without this, we would not be able to separate the cross-firm spillover effect from the firm's own autoregressive effect, conflating the two. The latter becomes obvious when we substitute (D.1) into the Markov

productivity process (3.3) that describes the evolution of firm i's productivity over time:

$$\omega_{it} = \mathbb{E}\left[\omega_{it} \middle| \omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} \underbrace{\frac{\mathbb{1}\{(j,t-1) \in \mathscr{L}(i,t-1) \text{ and } \omega_{j,t-2} > \omega_{i,t-2}\}}{\sum_{k(\neq i)=1}^{n} \mathbb{1}\{(k,t-1) \in \mathscr{L}(i,t-1) \text{ and } \omega_{j,t-2} > \omega_{i,t-2}\}}}_{s^*_{ij,t-1}} \omega_{j,t-1}\right] + \zeta_{it}. \tag{D.2}$$

By making the asymmetry in external learning be a function of the *twice*-lagged pair-wise productivity differentials between the firm and its peers, we avoid the appearance of $\omega_{i,t-1}$ in two places thereby allowing us to partial out the cross-firm spillovers from the *auto*regressive persistence in productivity. Thus, to separably identify asymmetric spillovers in productivity, in addition to assuming that (both the internal and external) learning occurs with a delay, one also requires an assumption that the firm takes an additional period to identify more productive peers. However, we do not need this additional timing assumption in our baseline analysis (with symmetric interactions).

E Additional Modeling Considerations

Contextual Effects. Because we assume delayed cross-firm peer interactions, as noted by Manski (1993), the *dynamic* nature of productivity model (3.3) can potentially provide an additional avenue to circumvent the unidentification problems and separate different types of peer effects, should one be interested in also modeling the "contextual effects" on productivity via $\sum_{j(\neq i)} s_{ij,t-1} G_{j,t-1}$. In such a setup, per the results in Bramoullé et al. (2009), the separable identification of "endogenous" and "contextual" effects can also be achieved by relying on variation in the size of peer reference groups, so long as the firm i is excluded when computing group means, as is in our case. Alternatively, (co)variance-based quadratic moment conditions may be used to aid identification (Kelejian & Prucha, 1999; Lee, 2007; Kuersteiner & Prucha, 2020).

Contemporaneous Effects. Depending on the particular source of learning, it may sometimes be possible to reasonably relax the timing assumption that the learning effect of G_{it} on firm productivity be with a delay. Take, for example, the firm's export status in the context of "learning by exporting." Consistent with much theoretical and empirical work in international trade, the

decision to start exporting is usually associated with large sunk entry costs, which would impede firms from adjusting their export status *immediately* after experiencing an improvement in their productivity. Analogous arguments can be made about costliness of swift geographic relocations. If so, it may be feasible to replace weak exogeneity of lagged $G_{i,t-1}$ and $\{s_{ij,t-1}\}$ with a stronger assumption of weak exogeneity of G_{it} and $\{s_{ijt}\}$. The productivity process in (3.3) can then be modified as follows: $\omega_{it} = \mathbb{E}\left[\omega_{it} | \, \omega_{i,t-1}, G_{it}, \sum_{j(\neq i)} s_{ijt}\omega_{j,t-1}\right] + \zeta_{it}$, where the implied mean-orthogonality of ζ_{it} and $(G_{it}, \sum_{j(\neq i)} s_{ijt}\omega_{j,t-1})'$ is effectively paramount to assuming that, due to adjustment costs, both the G_{it} and firm location in period t are determined in period $t-1$ based on $\omega_{i,t-1}$ just like the dynamic inputs are.

F Inference

Asymptotic Inference. Let the moment vector in (5.4) be concisely written as $\mathbb{E}[\boldsymbol{\rho}(\Theta)] = \mathbf{0}$, where Θ is a collection of both the finite-dimensional coefficients $\left(\beta_M, \beta_K, \beta_L, \theta\right)'$ and nonparametric sieve "parameters" $\boldsymbol{\gamma}$. Given the just-identification of the model and so long as we use linear sieves for $\mathscr{A}_{L_n}(\cdot)$ such as polynomial or B-spline series, we can make use of the numerical equivalence (see Hahn et al., 2018) between the consistent estimator of the asymptotic variance of *semiparametric* "parameters" $\widehat{\Theta}$ and a consistent estimator of the asymptotic variance derived for these parameter estimators as if the estimated model were of a *parametric* form specified in (5.1) and (5.3). Thus, in practice, one can use the variance formula for a parametric two-step estimator to consistently estimate the variance of a semiparametric sieve estimator.[3] The asymptotic variance for such a parametric two-step estimator can in turn be derived following Newey's (1984) suggestion by making use of the optimal GMM covariance formula: $\mathbb{V}ar\left[\widehat{\Theta}\right] = \left[\mathbb{E}\frac{\partial \boldsymbol{\rho}(\Theta)}{\partial \Theta'}\right]^{-1} \mathbb{E}[\boldsymbol{\rho}(\Theta)\boldsymbol{\rho}(\Theta)'] \left[\mathbb{E}\frac{\partial \boldsymbol{\rho}(\Theta)}{\partial \Theta}\right]^{-1}$. This streamlines asymptotic inference.

[3] Note that this equivalence applies to finite samples only because, asymptotically, the number of sieve "parameters" will diverge to infinity with the sample size whereas the number of parameters in a parametric specification will stay a finite constant. Furthermore, the numerical equivalence holds more generally for fully *non*parametric two-step sieve estimators. In our case, the estimator is *semi*parametric, with the first step implemented using the known parametric form. Since ours is a special case of the nonparametric setup studied by Hahn et al. (2018), their results continue to apply.

Bias-Corrected Bootstrap Inference. However, because asymptotic inference for semi- and nonparametric estimators is well-known to perform unreliably due to finite-sample biases as well as the first-order asymptotic theory's poor ability to approximate the distribution of estimators in finite samples (Horowitz, 2001), for hypothesis testing, we therefore rely on Efron's (1987) accelerated bias-corrected bootstrap percentile confidence intervals, which are second-order accurate and provide means not only to correct for the estimator's finite-sample bias but also to account for higher-order moments (particularly, skewness) in the sampling distribution.

We approximate sampling distributions of the estimator via wild residual block bootstrap that takes into account a panel structure of the data, with both stages resampled jointly owing to a sequential nature of our estimation procedure. More specifically, when constructing wild bootstrap residuals, we work with the *joint* distribution of firm-specific time series of $\{\widehat{\eta}_{it}\}$ and $\{\widehat{\zeta}_{it}\}$, with the auxiliary random variable drawn from the Mammen (1993) two-point distribution independently over i. Note that this independence over i is consistent with our model's assumption about random productivity shocks. We set the number of bootstrap replications to $B = 400$. Having first obtained bootstrap parameter estimates $\{(\widehat{\beta}_K^b, \widehat{\beta}_L^b, \widehat{\beta}_M^b)'; \ b = 1, \ldots, B\}$ and $\{\widehat{\boldsymbol{\gamma}}^b; \ b = 1, \ldots, B\}$, we then obtain bootstrap values for our main estimands of interest: $\widehat{DL}_{it}^b$, $\widehat{SP}_{it}^b$ and $\widehat{TIL}_{it}^b$ for $b = 1, \ldots, B$ (at each observation). Next, we use the accelerated bias-correction method to make inference about DL, SP and TIL.

To make matters concrete, let the (observation-specific) estimand of focus be denoted by $\widehat{E}$. We use the empirical distribution of B bootstrap estimates $\left\{\widehat{E}^1, \ldots, \widehat{E}^B\right\}$ to estimate $(1-a) \times 100\%$ confidence bounds for $\widehat{E}$ as intervals between the $[a_1 \times 100]$th and $[a_2 \times 100]$th percentiles of its bootstrap distribution with

$$a_1 = \Phi\left(\widehat{\phi}_0 + \frac{\widehat{\phi}_0 + \phi_{a/2}}{1 - \widehat{c}\left(\widehat{\phi}_0 + \phi_{a/2}\right)}\right) \quad \text{and} \quad a_2 = \Phi\left(\widehat{\phi}_0 + \frac{\widehat{\phi}_0 + \phi_{(1-a/2)}}{1 - \widehat{c}\left(\widehat{\phi}_0 + \phi_{(1-a/2)}\right)}\right), \tag{F.1}$$

where $\Phi(\cdot)$ is the standard normal cdf, ϕ_α is the $(\alpha \times 100)$th percentile of the standard normal distribution,

$$\widehat{\phi}_0 = \Phi^{-1}\left(\#\left\{\widehat{E}^b < \widehat{E}\right\} / B\right) \tag{F.2}$$

is a bias-correction factor, and $\widehat{c}$ is an acceleration parameter which, following the literature, is estimated via jackknife as follows (e.g., see Shao & Tu, 1995):

$$\widehat{c} = \frac{\sum_{j=1}^{J}\left(\sum_{s=1}^{J}\widehat{E}^{s} - \widehat{E}^{j}\right)^{3}}{6\left[\sum_{j=1}^{J}\left(\sum_{s=1}^{J}\widehat{E}^{s} - \widehat{E}^{j}\right)^{2}\right]^{3/2}}, \tag{F.3}$$

where $\widehat{E}^{j}$ is the $j(=1,\ldots,J)$th jackknife estimate of E.[4]

Note that both the acceleration and bias-correction factors are different for each estimator, denoted here generically by $\widehat{E}$. That is, the bias-correction procedure is not only estimand-specific but may also be observation-specific as is in our case. Also, the estimated confidence intervals may not contain the original estimates if the finite-sample bias is large.

G Additional Simulation Results

Nonlinear Productivity Process. Table G.1 presents the results for our proposed estimator when the productivity DGP is nonlinear. Specifically, we consider the following nonlinear productivity process:

$$\begin{aligned}\omega_{it} = {} & \rho_0 + \rho_{11}\omega_{i,t-1} + \rho_{12}\omega_{i,t-1}^{2} + \rho_{21}\sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1} + \rho_{22}\left(\sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}\right)^{2} + \\ & \rho_{31}G_{i,t-1} + \rho_{32}G_{i,t-1}^{2} + \varrho_{12}\omega_{i,t-1}\left(\sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}\right) + \varrho_{13}\omega_{i,t-1}G_{i,t-1} + \\ & \lambda_{23}G_{i,t-1}\left(\sum_{j(\neq i)} s_{ij,t-1}\omega_{j,t-1}\right) + \zeta_{it},\end{aligned} \tag{G.1}$$

where $\rho_0 = 0.2$, $\rho_{11} = 0.65$, $\rho_{12} = -0.015$, $\rho_{21} = 0.18$, $\rho_{22} = 0.025$, $\rho_{31} = 0.37$, $\rho_{32} = 0.12$, $\varrho_{12} = 0.006$, $\varrho_{13} = -0.06$ and $\lambda_{23} = 0.07$. The rest of the DGP is kept unchanged (see Section 6).

Table G.1 essentially replicates Table 1 using this new productivity process. The simulation results remain encouraging and show that our estimation methodology is consistent and recovers the true parameters well.

[4]We have tried different versions of jackknife with similar results. We settle on a delete-50T jackknife (i.e., leave-50-cross-sections-out) which respects the panel structure of our data while yielding a reasonable number of subsamples the estimation on which is not computationally prohibitive.

Table G.1. Simulation Results for Our Estimator under the Nonlinear Productivity Process

	Mean True Value	$n = 100$ Mean	RMSE	MAE	$n = 200$ Mean	RMSE	MAE	$n = 400$ Mean	RMSE	MAE
DGP WITH G EVOLVING EXOGENOUSLY										
Scenario (i): $DL \neq 0$ **and** $SP \neq 0$										
β_K	0.250	0.237	0.077	0.059	0.247	0.046	0.035	0.251	0.031	0.024
AR	0.596	0.593	0.081	0.066	0.592	0.054	0.044	0.594	0.038	0.031
DL	0.412	0.416	0.157	0.128	0.418	0.106	0.087	0.414	0.075	0.061
SP	0.301	0.274	0.313	0.257	0.303	0.200	0.166	0.302	0.138	0.115
TIL	0.124	0.102	0.144	0.112	0.119	0.085	0.069	0.120	0.059	0.048
Scenario (ii): $DL = 0$ **and** $SP \neq 0$										
β_K	0.250	0.233	0.082	0.063	0.247	0.049	0.037	0.251	0.033	0.025
AR	0.609	0.606	0.081	0.067	0.606	0.055	0.045	0.608	0.039	0.032
DL	0	0.003	0.158	0.127	0.003	0.107	0.087	0.001	0.076	0.061
SP	0.274	0.088	0.301	0.246	0.222	0.198	0.162	0.257	0.130	0.107
TIL	0	–0.011	0.075	0.049	–0.005	0.038	0.026	–0.002	0.023	0.017
Scenario (iii): $DL = 0$ **and** $SP = 0$										
β_K	0.250	0.246	0.064	0.051	0.249	0.044	0.035	0.251	0.032	0.025
AR	0.672	0.623	0.078	0.064	0.623	0.054	0.044	0.626	0.038	0.031
DL	0	0.003	0.155	0.126	0.003	0.106	0.087	0.001	0.075	0.061
SP	0	–0.034	0.152	0.126	–0.010	0.106	0.086	–0.004	0.073	0.060
TIL	0	–0.004	0.027	0.017	–0.002	0.012	0.008	–0.001	0.006	0.004
DGP WITH G FOLLOWING AN ω-CONTROLLED PROCESS										
Scenario (i): $DL \neq 0$ **and** $SP \neq 0$										
β_K	0.250	0.249	0.028	0.022	0.250	0.020	0.015	0.251	0.014	0.011
AR	0.338	0.328	0.078	0.060	0.334	0.054	0.043	0.338	0.038	0.030
DL	1.148	1.157	0.119	0.093	0.152	0.083	0.064	0.148	0.058	0.046
SP	0.695	0.697	0.036	0.029	0.697	0.025	0.021	0.695	0.017	0.014
TIL	0.798	0.808	0.294	0.166	0.803	0.207	0.119	0.798	0.146	0.082
Scenario (ii): $DL = 0$ **and** $SP \neq 0$										
β_K	0.250	0.233	0.086	0.065	0.245	0.053	0.039	0.251	0.033	0.025
AR	0.609	0.609	0.102	0.084	0.607	0.068	0.055	0.608	0.050	0.040
DL	0	–0.006	0.135	0.109	–0.002	0.092	0.075	0.000	0.066	0.054
SP	0.274	0.085	0.341	0.277	0.207	0.215	0.175	0.259	0.152	0.124
TIL	0	0.023	0.069	0.045	0.005	0.034	0.023	0.003	0.021	0.015
Scenario (iii): $DL = 0$ **and** $SP = 0$										
β_K	0.250	0.246	0.068	0.053	0.249	0.044	0.035	0.251	0.031	0.025
AR	0.627	0.624	0.100	0.082	0.623	0.067	0.054	0.626	0.049	0.040
DL	0	–0.001	0.134	0.108	0.001	0.092	0.075	0.001	0.066	0.054
SP	0	–0.035	0.147	0.122	–0.009	0.101	0.084	–0.004	0.071	0.060
TIL	0	0.002	0.021	0.014	0.000	0.010	0.007	0.000	0.005	0.003

NOTES: Owing to nonlinearity of the productivity process in (G.1), AR, DL, SP and TIL are all observation-specific. With the sole exception for the fixed parameter $\beta_K = 0.25$, the mean true values are the averages (across simulation repetitions) of the mean simulated values over i and t. Throughout, $T = 10$.

Table G.2. Simulation Results for the Alternative Estimator of β_K

	True Value	$n = 100$ Mean	$n = 100$ RMSE	$n = 200$ Mean	$n = 200$ RMSE	$n = 400$ Mean	$n = 400$ RMSE
		DGP WITH G EVOLVING EXOGENOUSLY					
Scenario (i): $DL \neq 0$ and $SP \neq 0$	0.25	0.376	0.136	0.373	0.128	0.374	0.126
Scenario (ii): $DL = 0$ and $SP \neq 0$	0.25	0.438	0.193	0.435	0.188	0.436	0.187
Scenario (iii): $DL = 0$ and $SP = 0$	0.25	0.251	0.040	0.250	0.029	0.251	0.020
		DGP WITH G FOLLOWING AN ω-CONTROLLED PROCESS					
Scenario (i): $DL \neq 0$ and $SP \neq 0$	0.25	0.118	0.152	0.113	0.148	0.109	0.146

NOTES: Reported are the first-step results for $\widehat{\beta}_K$ from the alternative estimators which proxy for latent productivity under the assumption of exogenous Markov process for ω_{it}. The results corresponding to scenarios (ii) and (iii) of the second DGP [bottom panel] are omitted because they are identical to those for the first DGP [top panel]. This is because not only does G not enter the alternative estimator but it also does not affect the evolution of firm productivity by design ($DL = 0$) in these two scenarios. Throughout, $T = 10$.

First-Step Estimates of β_K from the Two-Step Estimator. To examine the ability of alternative models to identify firm productivity, we first study if these two-step estimators can consistently estimate the production function coefficients (here β_K) because $\widehat{\omega}_{it}$ is a direct construct of these parameters. The corresponding estimates of β_K are reported in Table G.2. These first-step results apply to both the ALT1 and ALT2 models and are obtained assuming that ω_{it} is an exogenous first-order Markov process.

Alternative Two-Step Estimators of Spillovers. Table G.3 reports the results for the "spillovers" estimator (defined as either SP or TIL) from different variants of the second-step regressions in (6.4)–(6.5) estimated with $\widehat{\omega}_{it}$ obtained in the first step using the standard proxy estimator under the assumption of exogenous Markov productivity process. The data are simulated assuming a linear productivity process under scenario (iii) with the true $DL = 0$ and $SP = 0$. Thus, the first-step estimation of productivity is correctly specified and consistent. For the estimation of second-step regressions, the G series is generated as an ω-controlled process (b). The specifications containing contemporaneously endogenous regressors are estimated using their respective first lags as predetermined instruments.

Examining the results in Table G.3, we find that, across all specifications, the second-step

Table G.3. Simulation Results for Different Variants of the Two-Step Alternative Estimator

	Results for α_{12} (TIL) in eq. (6.4)							Results for α_{22} (SP) in eq. (6.5)			
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Mean Estimate											
$n = 100$	1.154	0.558	1.057	0.620	–0.359	0.378	–0.362	0.990	0.591	0.540	0.390
$n = 200$	1.157	0.556	1.064	0.624	–0.355	0.386	–0.361	0.995	0.597	0.546	0.395
$n = 400$	1.167	0.560	1.073	0.628	–0.356	0.387	–0.362	0.998	0.599	0.548	0.397
Root Mean Squared Error											
$n = 100$	1.171	0.596	1.072	0.647	0.359	0.416	0.365	0.987	0.595	0.541	0.393
$n = 200$	1.166	0.575	1.072	0.638	0.356	0.405	0.362	0.995	0.599	0.547	0.396
$n = 400$	1.170	0.570	1.077	0.635	0.357	0.398	0.363	0.998	0.600	0.549	0.397
Rejection Frequency for H_0 : Spillover Parameter = 0											
$n = 100$	1.00	0.97	1.00	0.99	1.00	0.90	1.00	1.00	0.99	1.00	1.00
$n = 200$	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00
$n = 400$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Variables in the Second-Step Regression											
Spillover Variable	$\overline{G}_{it}$	$\overline{G}_{it}$	$\overline{G}_{it-1}$	$\overline{G}_{it-1}$	$\overline{G}_{it-2}$	$\overline{G}_{it-2}$	$\overline{G}_{it-2}$	$\overline{\omega}_{it}$	$\overline{\omega}_{it}$	$\overline{\omega}_{it-1}$	$\overline{\omega}_{it-1}$
DL Variable	–	G_{it}	–	G_{it-1}	–	G_{it-2}	G_{it-1}	–	G_{it}	–	G_{it-1}

NOTES: Reported are the results for "spillovers" (defined as either SP or TIL) from different variants of the second-step regressions in (6.4)–(6.5) estimated with $\widehat{\omega}_{it}$ obtained in the first step using the standard proxy estimator under the assumption of exogenous Markov productivity. The data are simulated assuming linear productivity process under scenario (iii) with true $DL = 0$ and $SP = 0$. Thus, the true spillovers are zero. For the estimation of second-step regressions, the G series is generated as an ω-controlled process. The specifications containing contemporaneously endogenous regressors are estimated using first lags as instruments. $\overline{G}_{it} \equiv \sum_j s_{ijt} G_{jt}$ and $\overline{\omega}_{it} \equiv \sum_j s_{ijt} \omega_{jt}$. Throughout, $T = 10$.

estimator exhibits non-vanishing biases in the estimation of spillovers. All models spuriously fail at identifying *zero* cross-firm spillovers. Here we also report the rejection frequencies (over simulation repetitions) for the asymptotic z-test of the null that the coefficient of a spillover variable in the model is zero at the 95% confidence level. Had the second-step been correctly specified and consistent, we were to expect these rejection frequencies to be all around 0.05. Consistent with our expectations, the results in the table indicate drastic size distortions due to misspecification of the two-step approach. The results are qualitatively the same when we also control for firm and/or time fixed effects.

Table H.4. Data Summary Statistics

Variable	Mean	1st Qu.	Median	3rd Qu.
Y	80,031	9,165	19,577	51,320
K	14,330	1,039	2,957	8,929
L	3,627	533	1,128	2,707
M	55,761	6,372	13,640	35,878
G	0.09	0	0	0
$\mathbb{1}\{G > 0\}$	0.19			

NOTES: Y – gross output; K – capital stock; L – labor; M – materials; G – foreign equity share; $\mathbb{1}\{G > 0\}$ – binary indicator signifying foreign-invested firms. Y, K, L and M are measured in thousands of RMB; G is a unit-free proportion.

H Data

Our data are drawn from the Chinese Industrial Enterprises Database survey conducted by China's National Bureau of Statistics (NBS). This database covers all firms with sales above 5 million yuan (about 0.6 million in U.S. dollar) and includes most industries including mining, manufacturing and public utilities. We focus on the electric machinery and equipment manufacturing industry, SIC 2-digit code 39.

The production variables are as follows. The firm's capital stock (K_{it}) is the net fixed assets deflated by the price index of investment into fixed assets. Labor (L_{it}) is measured as the total wage bill plus benefits deflated by the GDP deflator. Materials (M_{it}) are the total intermediate inputs, including raw materials and other production-related inputs, deflated by the purchasing price index for industrial inputs. The output (Y_{it}) is defined as the gross industrial output value deflated by the producer price index. The price indices are obtained from NBS and the World Bank. The four variables are measured in thousands of real RMB. In addition, the foreign equity share (G_{it}) is a bounded proportion that lies between zero and one, by construction.

We exclude observations with missing values for these variables as well as a small number of likely erroneous observations with the foreign equity share values outside the unit interval. With the sample period running from 1998 to 2007, the operational sample is an unbalanced

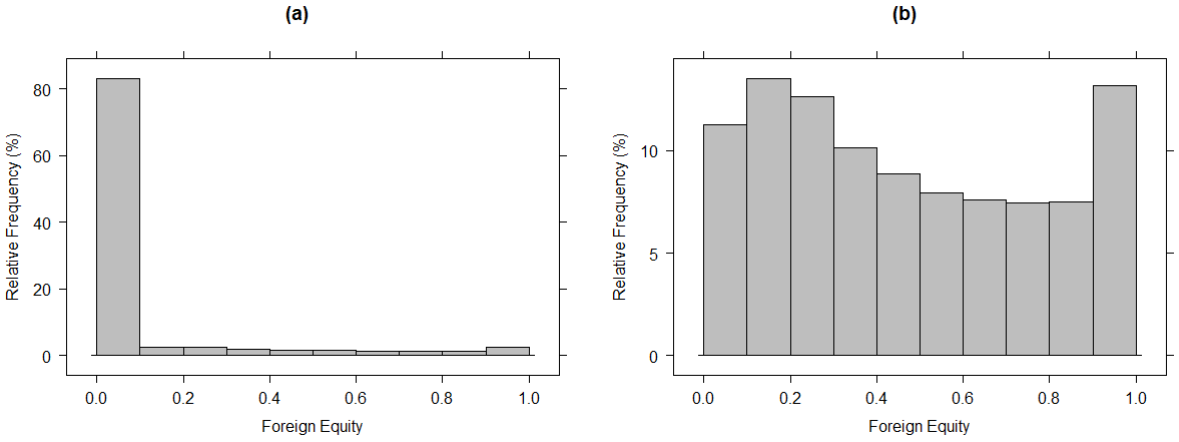

Figure H.1. Empirical Distribution of the Foreign Equity Share:
(a) All Firms; (b) Foreign-Invested Firms Only

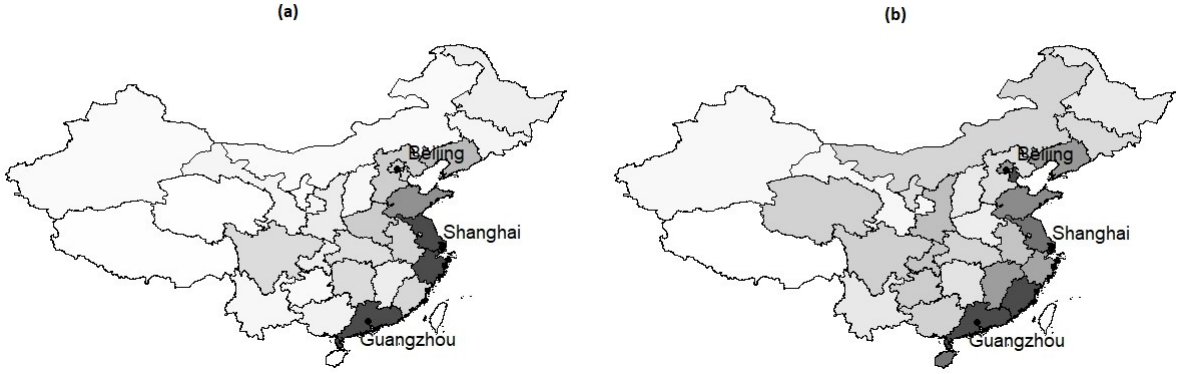

Figure H.2. Geographic Distribution of Firms (a) and the Foreign Equity Share (b)

NOTE: Shown are the number of firms in the sample from each province (a) and the sample mean at the province level (b), with the darker areas corresponding to higher values.

panel of 23,720 firms with a total of 73,095 observations. Table H.4 reports summary statistics for these data.

Figure H.1(a) plots a histogram of G_{it} across firms which expectedly has a zero mode because the manufacturing sector in China is dominated by wholly domestic firms. Figure H.1(b) on the right plots the distribution of $G_{it}|G_{it} > 0$, i.e., for foreign-invested firms only. Overall, 81% firms in our sample are wholly domestically-owned, 1% are pure foreign multinationals, with the remaining 18% being (partially) foreign-invested domestic firms. The map in Figure H.2(b) shows the spatial distribution of the (average) foreign equity share across regions where, consistent with one's priors, we see the heightened concentration of FDI along the coast.

I Additional Empirical Results

Baseline Results. Figure I.3 plots empirical histograms of the point estimates of productivity effects under the baseline specification. These estimates are the same as those summarized in Table 4. Subfigure I.3(a) shows the distribution of productivity spillover elasticities SP; the direct/internal and indirect/external learning effects of FDI (DL and TIL, respectively) are presented in subfigure I.3(b).

We also examine the geographic distribution of productivity spillovers in Figure I.4. The map shows by-province median estimates of productivity spillovers. We observe that productivity spillovers are stronger in the highly industrialized, fast-growing provinces in the Southeast and near the coast. Interestingly, comparing Figure I.4 with the spatial distribution of the industry in Figure H.2(a) in Appendix H, we find that productivity spillovers are comparable in strength (little, if any, shade gradient) across most Southeastern and coastal provinces and thus extend beyond the Shanghai and Guangzhou areas where the majority of the electric machinery manufacturing industry is concentrated. Therefore, the evidence of productivity spillovers that we find is not just a "mechanical" function of the spatial density of data.

Endogenous Exposure to FDI. The structural identification of our model only requires that the *lagged* foreign equity share $G_{i,t-1}$ be weakly exogenous with respect to the future produc-

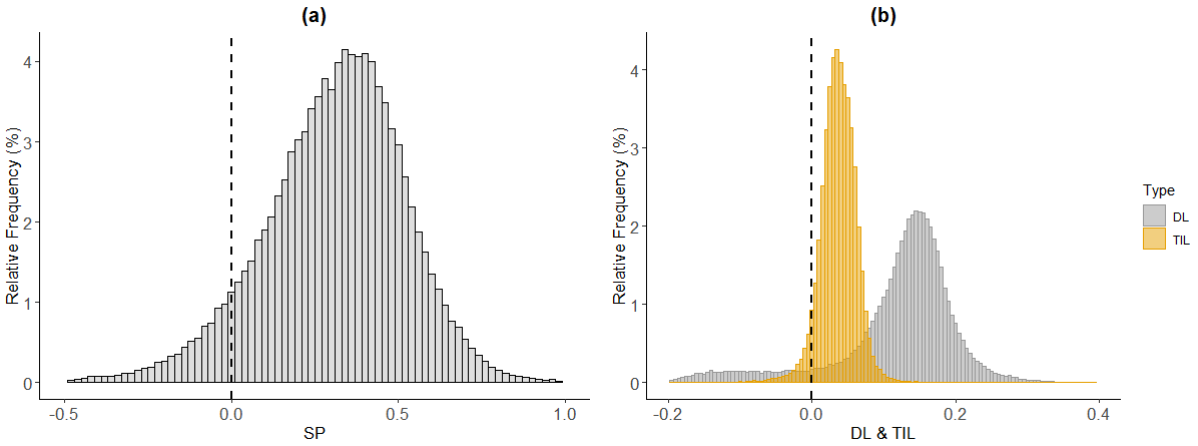

Figure I.3. Distributions of the Productivity Effects: (a) SP, (b) DL and TIL

NOTE: Plotted are the point estimates based on our baseline specification of the productivity process in (3.3).

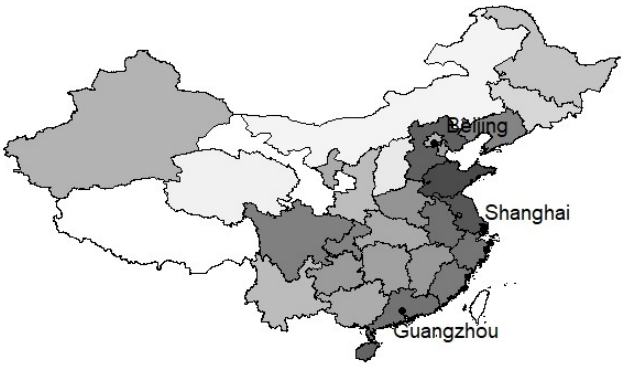

Figure I.4. Spatial Distribution of the Productivity Spillover Effects

NOTE: Plotted are the sample medians at the province level, with the darker areas corresponding to higher values.

Table I.5. Estimates of the Productivity Effects: Endogeneity of Lagged FDI

	Exogenous	IPW	IV-External	IV-Lewbel
		—Median Estimates—		
SP	0.327	0.477	0.466	0.414
	(0.211, 0.399)	(0.413, 0.512)	(0.411, 0.585)	(0.276, 0.515)
DL	0.138	0.128	0.171	0.191
	(0.118, 0.164)	(0.105, 0.158)	(0.138, 0.241)	(0.053, 0.341)
TIL	0.037	0.055	0.067	0.066
	(0.022, 0.050)	(0.044, 0.069)	(0.056, 0.100)	(0.019, 0.126)
		—Statistically > 0 **(% Obs.)—**		
SP	83.84	98.55	97.17	79.56
DL	89.13	91.40	78.48	82.30
TIL	86.71	98.99	97.64	83.48

NOTES: Reported are the results for the productivity process in (3.3) under baseline specification. The two-sided 95% bootstrap percentile confidence intervals for the median point estimates are in parentheses. Statistical positiveness is at the 5% significance level, using one-sided bootstrap confidence intervals. The "Exogenous" column corresponds to our proposed estimation procedure under the structural assumption of weak exogeneity of $G_{i,t-1}$, with the second-stage eq. (4.8) estimated via least squares. The three models on the right allow for the violation of $\mathbb{E}[\zeta_{it} + \eta_{it}|G_{i,t-1}] = 0$ and address the potential endogeneity of lagged FDI in (4.8) via (i) inverse probability weighting [IPW], (ii) instrumentation using external IVs including coastal province dummy, province-level openness measure and their interactions with predetermined $L_{i,t-1}$ [IV-External], (iii) instrumentation with the heteroskedasticity-based internal IV à la Lewbel (2012) [IV-Lewbel].

tivity innovation ζ_{it}. Therefore, firms in our framework may experience endogenous updates to their exposure to foreign knowledge G_{it} and even relocate based on the contemporaneous productivity ω_{it}. Nonetheless, however mild an assumption, predeterminedness of $G_{i,t-1}$ may still be violated in case the firms—or their foreign investors—can forecast their future productivity (shocks). As a robustness check to this potential violation of the weak exogeneity of lagged foreign equity share, we re-estimate our baseline model using the inverse probability weighting (IPW) procedure and instrumentation. In our case, we need to weight only the second-stage regression since the first-stage share equation contains neither the FDI variable nor a productivity innovation. Analogously, instrumenting the lagged foreign equity share affects only the second stage. The results are summarized in Table I.5.

By means of IPW, we seek to account for the potential selection (on observables) of firms by

their foreign investors based on their *future* productivity. We use the stabilized IPWs which are typically more numerically stable and produce narrow confidence bounds (Hernán & Robins, 2020). Also, note that we deal with a continuous "treatment" G_{it} which is why our approach is different from the more conventional propensity score estimation suitable for binary treatments (e.g., Imbens & Wooldridge, 2009). The IPWs for a continuous treatment G_{it} are given by $f_G(G_{it})/f_{G|\mathbf{d}}(G_{it}|\mathbf{d}_{i,t+1})$, where $f(\cdot)$ is a *pdf*. [For more, also see Hirano & Imbens (2004) and Hernán & Robins (2020).] Note that the vector of observables $\mathbf{d}_{i,t+1}$ includes firm characteristics reflective of its performance next period, because of concern is the selection into treatment based on the future productivity. We include the following correlates of the firm productivity that may influence its foreign exposure: size proxied by the logged labor, age, state equity share, government subsidy receipts, export intensity, normalized profits, the return on assets, leverage, logged total assets as well as the East coast dummy and the time trend.

To avoid the curse of dimensionality as well as the problem of near-zero extreme values associated with nonparametric estimation of densities, we employ a parametric maximum-likelihood approach to estimate f_G and $f_{G|\mathbf{d}}$. Given the bounded nature of a fractional variable $G_{it} \in [0,1]$, we assume it is Beta-distributed. Because beta distribution is not trivial to estimate, we impose few data-motivated restrictions on it. More specifically, we let $G_{it} \sim \text{Beta}(\alpha,\beta)$ and $G_{it}|\mathbf{d}_{i,t+1} \sim \text{Beta}(\alpha,\beta(\mathbf{d}_{i,t+1}))$ where, to match the data, we restrict the first shape parameter to a unit value ($\alpha = 1$) and the second parameter/function $\beta(\cdot)$ to be greater than 1 so that the distribution of G_{it} is unimodal with a zero mode in both instances.[5] The densities are estimated via maximum likelihood (ML), although the method of moments provides an alternative route.

Since we seek to address the potential endogeneity of lagged $G_{i,t-1}$ in the second-stage least squares regression, we also lag the estimated IPWs to have them match the time period of "treatment." In other words, we weight each observation in the second stage by $\widehat{f}_G(G_{i,t-1})/\widehat{f}_{G|\mathbf{d}}(G_{i,t-1}|\mathbf{d}_{it})$.[6]

[5] Recall that the mode of the Beta(α,β) distribution is 0 for $0 < \alpha \leq 1$ and $\beta > 1$.

[6] More concretely, the IPW function is

$$\frac{f_G(G_{i,t-1})}{f_{G|\mathbf{d}}(G_{i,t-1}|\mathbf{d}_{it})} = \frac{(1-G_{i,t-1})^{\beta_0-1}\Gamma(1+\beta_0)}{\Gamma(\beta_0)} \times \left[\frac{(1-G_{i,t-1})^{\beta(\mathbf{d}_{it})-1}\Gamma(1+\beta(\mathbf{d}_{it}))}{\Gamma(\beta(\mathbf{d}_{it}))}\right]^{-1},$$

The results are summarized in the IPW column of Table I.5.

Table I.5 also reports the estimates of productivity effects from the second stage estimated via generalized method of moments using external instruments (the IV-External column) and Lewbel's (2012) heteroskedasticity-based internal instruments (the IV-Lewbel column) for $G_{i,t-1}$. The external instruments include the East coast dummy, a province-level measure of openness (defined as the ratio of the sum of imports and exports to the gross domestic product) and their interactions with the firm-level lagged labor input. The two external instruments are motivated by the previous studies, such as Eichengreen & Tong (2007) and Keller & Yeaple (2009), and are selected to proxy friendly regional policies towards foreign capital, shipping costs and the overall ease of engaging in international trade and finance. We interact these instruments with the predetermined labor at the firm level to gain variation. For identification of the firm's productivity process based on heteroskedasticity, adapting Lewbel (2012) we first estimate an auxiliary equation for the endogenous regressor by regressing $G_{i,t-1}$ on all the other exogenous variables in the ω_{it} process, namely, $\omega_{i,t-1}$ and $\sum_{j\neq i} s_{ij,t-1}\omega_{j,t-1}$. The residuals from this auxiliary regression are then interacted with the demeaned $\omega_{i,t-1}$ and used to instrument for $G_{i,t-1}$. Here, we use the firm's predetermined lagged productivity $\omega_{i,t-1}$ as a "Z" variable that is uncorrelated with the *product* of productivity innovation ζ_{it} (with which $G_{i,t-1}$ is suspected to be correlated) and the error in the auxiliary equation for $G_{i,t-1}$. For more details on instrumentation via heteroskedasticity, see Lewbel (2012).

Controlling for potential endogeneity of the FDI exposure, we continue to find strong empirical evidence in support of significantly positive productivity spillovers for 80% of firms or more. At least 78% of manufacturers benefit from significant productivity boosts associated with receiving FDI. The changes in the effect sizes are not out of the ordinary either, and the rank correlation coefficient of the point estimate of productivity effects across these estimators is at least as high as 0.71. All in all, our findings remain qualitatively unchanged.

where β_0 is a scalar shape parameter estimated via ML using $G_{it} \sim \text{Beta}(1, \beta_0)$; $\beta(\mathbf{d}_{it})$ is a scalar function estimated via ML using $G_{it}|\mathbf{d}_{i,t+1} \sim \text{Beta}(1, \beta(\mathbf{d}_{i,t+1}))$ with $\beta(\mathbf{d}_{it})$ parameterized using the exponential function of a single index; and $\Gamma(\cdot)$ is the Gamma function.

Bidimensional Spillovers. In our main analysis, per the productivity process (3.3), spillovers from all spatially proximate peers in the industry have the potential to affect the recipient firm's productivity in the same manner, no matter the exposure of these peers to foreign knowledge. Given the documented impact of FDI on firm productivity, it may also be of interest to allow for heterogeneity in the external productivity spillovers from the peers *conditional* on their FDI status. To this end, we adapt our methodology to allow a more general evolution process for productivity that permits bidimensional spillovers by means of two spatiotemporal lags. Thus, we now consider the following evolution process of firm productivity:

$$\omega_{it} = \mathbb{E}\left[\omega_{it} \middle| \omega_{i,t-1}, G_{i,t-1}, \sum_{j(\neq i)} s^0_{ij,t-1}\omega_{j,t-1}, \sum_{j(\neq i)} s^1_{ij,t-1}\omega_{j,t-1}\right] + \zeta_{it}, \tag{I.1}$$

with the distinction between peer weights $\{s^0_{ijt}\}$ and $\{s^1_{ijt}\}$ based on the peers' FDI status:

$$s^0_{ijt} = \frac{\mathbb{1}\{G_{jt} = 0 \text{ and } (j,t) \in \mathscr{L}(i,t)\}}{\sum_{k(\neq i)=1}^{n} \mathbb{1}\{G_{kt} = 0 \text{ and } (k,t) \in \mathscr{L}(i,t)\}}, \tag{I.2}$$

$$s^1_{ijt} = \frac{\mathbb{1}\{G_{jt} > 0 \text{ and } (j,t) \in \mathscr{L}(i,t)\}}{\sum_{k(\neq i)=1}^{n} \mathbb{1}\{G_{kt} > 0 \text{ and } (k,t) \in \mathscr{L}(i,t)\}}. \tag{I.3}$$

The cross-firm productivity spillovers are now bidimensional. The null intersection of the two sets of peers ensures separable identifiability of heterogeneous spillovers from (i) the wholly domestically owned peers $SP^0_{it} = \partial\mathbb{E}[\omega_{it}|\cdot]\big/\partial\sum_{j(\neq i)} s^0_{ij,t-1}\omega_{j,t-1}$ and (ii) foreign-invested peers $SP^1_{it} = \partial\mathbb{E}[\omega_{it}|\cdot]\big/\partial\sum_{j(\neq i)} s^1_{ij,t-1}\omega_{j,t-1}$.

Table I.6 summarizes point estimates of bidimensional productivity spillovers under our baseline specification, whereby the firm's peer group $\mathscr{L}(i,t)$ is defined at the level of the same province and the entire 2-digit industry. Just like in the case of our main model with the unidimensional cross-firm dependence, we continue to find substantial and positive productivity spillovers in the industry. However, by disentangling the peer effects of wholly domestically owned and foreign-invested neighbors, we find that the former group has a significantly larger effect on its neighbors. The median spillover elasticity from fully domestic firms SP^0 is 0.30, whereas the counterpart estimate SP^1 from the foreign-invested peers is 0.17 only. Regardless of the peer group, the productivity spillovers are statistically positive for the overwhelming ma-

Table I.6. Estimates of Bidimensional Productivity Spillovers

Estimand	1st Qu.	*Point Estimates* Median	3rd Qu.	*Statistically* > 0 (% Obs.)
SP^0	0.153	0.300	0.428	82.58
	(0.090, 0.180)	(0.198, 0.349)	(0.302, 0.504)	
SP^1	0.158	0.172	0.185	98.89
	(0.119, 0.188)	(0.129, 0.204)	(0.137, 0.216)	

NOTES: Reported are the semiparametric estimates of bidimensional productivity spillovers from (I.1) under our baseline specification. The left panel summarizes point estimates of SP^0_{it} and SP^1_{it} with the corresponding two-sided 95% bootstrap percentile confidence intervals in parentheses. The last column reports the share of observations for which the point estimates are statistically positive at the 5% significance level using one-sided bootstrap percentile confidence intervals.

jority of firms in the industry (83% or more). This is on par with the extent of spillovers that we have found in our main model with unidimensional spillovers.

The documented heterogeneity in the magnitudes of spillovers across the two types of peers suggests that it is relatively "easier" for Chinese manufacturers to learn from other domestic firms that are *not* recipients of FDI. This may be because foreign-invested firms are more protective of their newly adopted foreign technologies/knowledge, which makes it more difficult to learn from them. At the same time, it also may be that learning from these foreign-invested firms is simply more difficult because their practices are too advanced and biased towards more productive/efficient firms in the first place. If so, firms that are already foreign-invested—and, hence, are more productivity due to direct learning effects of FDI—are to enjoy larger spillovers from their peers who have also received foreign investments. This is corroborated by the evidence in Table I.7: the coefficient on the firms' *own* FDI exposure is negative for SP^0 and positive for SP^1.

The results in Table I.7 continue to indicate that the more productivity firms have less absorptive capacity to learn from their peers. For both the SP^0 and SP^1 spillovers, the effect size increases with the average productivity of peers from whom spillovers originate. But on the other hand, the strength of spillovers from one peer group declines with the average productivity of the other group (note the negative coefficient on the cross-peer productivity). Taken to-

Table I.7. Heterogeneity in Bidimensional Productivity Spillovers

	SP^0	SP^1
$\omega_{i,t-1}$	–0.720	–0.070
	(–0.882, –0.571)	(–0.125, –0.019)
$G_{i,t-1}$	–0.185	0.036
	(–0.355, –0.073)	(–0.151, 0.130)
$\sum_j s^0_{ij,t-1}\omega_{j,t-1}$	1.206	–0.147
	(0.591, 1.459)	(–0.276, –0.023)
$\sum_j s^1_{ij,t-1}\omega_{j,t-1}$	–0.147	0.144
	(–0.276, –0.023)	(0.101, 0.178)

NOTES: Reported are the parameter estimates for the SP^0 and SP^1 functions derived from the polynomial approximation of the conditional mean of ω_{it} in the productivity process formulation with bidimensional spillovers in (I.1). Two-sided 95% bootstrap percentile confidence intervals in parentheses. These correspond to our baseline specification, with (i) each firm's peers restricted to the firms located in the same province and the industrial scope of spillovers defined at the level of the entire 2-digit industry, (ii) the technical change flexibly controlled for using a series of year effects.

gether, these two findings suggest some substitutability between learning from the two groups along with the recipient firm's finite capacity to absorb such spillovers from the peers in a given period. Thus, if the foreign-invested peers improve their productivity, the firm starts learning more from them and less from the non-foreign-invested peers, and vice versa.

To conclude, although we find evidence of heterogeneity in the strength of spillovers from wholly-domestic versus foreign-invested peers (with those from the former being relatively stronger), in the grand scheme of things, our main findings stay the same: productivity spillovers are positive and significant for most firms in the industry.

Asymmetric Spillovers. As we explain in Appendix D, the peer weighing scheme that we use in our analysis treats cross-firm spillovers symmetrically in that all members of a peer group affect each other's productivity. That is, each ith firm's productivity is influenced by the average productivity of all its peers: those that are more *and* those that are less productive that the firm i itself. But should one choose to regulate the direction of productivity spillovers by restricting

Table I.8. Estimates of the Productivity Effects with Asymmetric Spillovers

Estimand	1st Qu.	*Point Estimates* Median	3rd Qu.	*Statistically* > 0 (% Obs.)
SP	0.301	0.331	0.362	99.54
	(0.277, 0.328)	(0.306, 0.361)	(0.335, 0.400)	
DL	0.083	0.133	0.169	87.01
	(0.058, 0.106)	(0.104, 0.156)	(0.135, 0.192)	

NOTES: Reported are the results based on the productivity process formulation with asymmetric productivity spillovers in (D.2) using our baseline specification of $\mathscr{L}(i,t)$. The left panel summarizes point estimates of SP_{it} and DL_{it} with the corresponding two-sided 95% bootstrap percentile confidence intervals in parentheses. The last column reports the share of observations for which the point estimates are statistically positive at the 5% significance level using one-sided bootstrap percentile confidence intervals.

them to occur *from* more productive *to* less productive firms, our framework can be modified to accommodate that, albeit with additional timing assumptions.

We estimate such an asymmetric specification given in (D.2). Table I.8 summarizes the corresponding estimates of productivity effects. Comparing these results with our main estimates in Table 4, we see that the estimates of direct learning stay by and large unchanged, as expected. While comparable at the median, the asymmetric spillover effect estimates exhibit a much smaller variation in effect size (perhaps, because the peer pool is more homogeneous now) but are *as* prevalent as they are when we model them symmetrically.

References

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83, 2411–2451.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150, 41–55.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, 5, 1–21.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., & Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, 84, 445–510.

De Loecker, J. & Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102, 2437–2471.

Deloitte China Manufacturing Industry Group (2013). A new stage for overseas expansion for China's equipment manufacturing industry. Report by Deloitte China Research and Insight Centre.

Doraszelski, U. & Jaumandreu, J. (2013). R&D and productivity: Estimating endogenous productivity. *Review of Economic Studies*, 80, 1338–1383.

Doraszelski, U. & Jaumandreu, J. (2018). Measuring the bias of technological change. *Journal of Political Economy*, 126, 1027–1084.

Euro Exim Bank (2020). Export of electrical machinery from China. Euro Exim Bank Global Finance Blog; October 30, 2020.

Efron, B. (1987). Better bootstrap confidence interval. *Journal of American Statistical Association*, 82, 171–200.

Eichengreen, B. & Tong, H. (2007). Is China's FDI coming at the expense of other countries? *Journal of the Japanese and International Economies*, 21(2), 153–172.

Gandhi, A., Navarro, S., & Rivers, D. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128, 2973–3016.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics*, 10, 92–116.

Griliches, Z. & Mairesse, J. (1998). Production functions: The search for identification. In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium* (pp. 169–203). Cambridge University Press.

Hahn, J., Liao, Z., & Ridder, G. (2018). Nonparametric two-step sieve M estimation and inference. *Econometric Theory*. forthcoming.

Hernán, M. A. & Robins, J.M. (2020). *Causal Inference: What If.* Boca Raton: Chapman & Hall.

Hirano, K. & Imbens, G.W. (2004). The propensity score with continuous treatments. In W. A. Shewhart & S. S.Wilks (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family.* New York: Wiley & Sons, Ltd.

Horowitz, J. L. (2001). The bootstrap. In J. J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics* (5 ed.). chapter 52, (pp. 3159–3228). Elsevier Science B.V.

Ihrcke, J. & Becker, K. (2006). Study on the future opportunities and challenges of EU-China trade and investment relations. Study 1 of 12: Machinery. Report commissioned and financed by the Commission of the European Communities.

Imbens, G.W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47, 5–86.

Kelejian, H. H. & Prucha, I. R. (1999). A generalized moment estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40, 509–533.

Keller, W. & Yeaple, S. R. (2009). Multinational enterprises, international trade, and productivity growth: Firm-level evidence from the United States. *Review of Economics and Statistics*, 91, 821–831.

Kuersteiner, G. M. & Prucha, I. R. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, 88, 2109–2146.

Lee, L.-f. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137, 489–514.

Lewbel, A. (2012). Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics*, 30, 67–80.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics*, 21, 255–285.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60, 531–542.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters*, 14(2), 201–206.

Shao, J. & Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag New York Inc.