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:

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

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

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

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:

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

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: $\mathbb{V}ar\big[\widehat{\Theta}\big] = \big[\mathbb{E}\frac{\partial \boldsymbol{\rho}(\Theta)}{\partial \Theta'}\big]^{-1}\mathbb{E}[\boldsymbol{\rho}(\Theta)\boldsymbol{\rho}(\Theta)']\big[\mathbb{E}\frac{\partial \boldsymbol{\rho}(\Theta)}{\partial \Theta}\big]^{-1}$. 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 $\{\widehat{E}^1, \ldots, \widehat{E}^B\}$ 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,

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

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	$n = 100$ RMSE	$n = 100$ MAE	$n = 200$ Mean	$n = 200$ RMSE	$n = 200$ MAE	$n = 400$ Mean	$n = 400$ RMSE	$n = 400$ 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	$n = 100$		$n = 200$		$n = 400$	
	Value	Mean	RMSE	Mean	RMSE	Mean	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

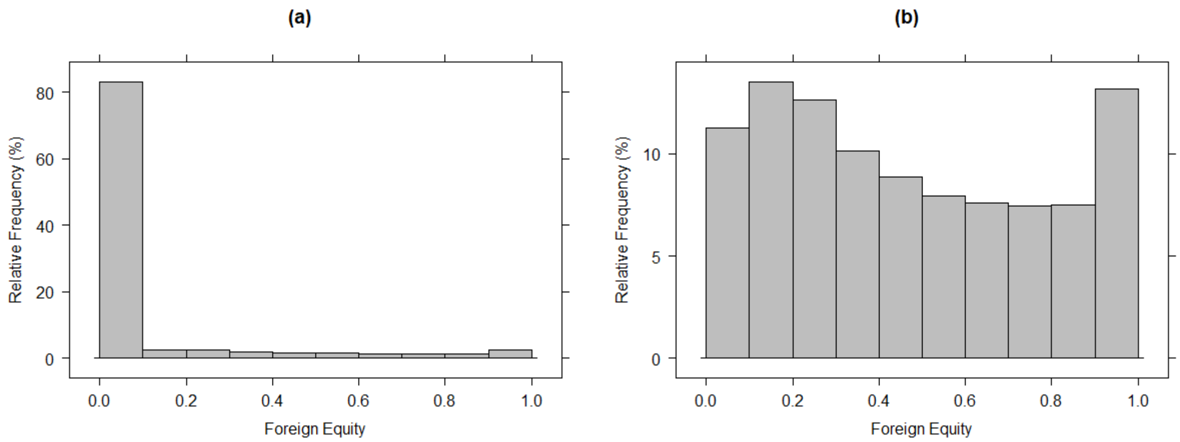

Figure H.1. Empirical Distribution of the Foreign Equity Share:
(a) All Firms; (b) Foreign-Invested Firms Only

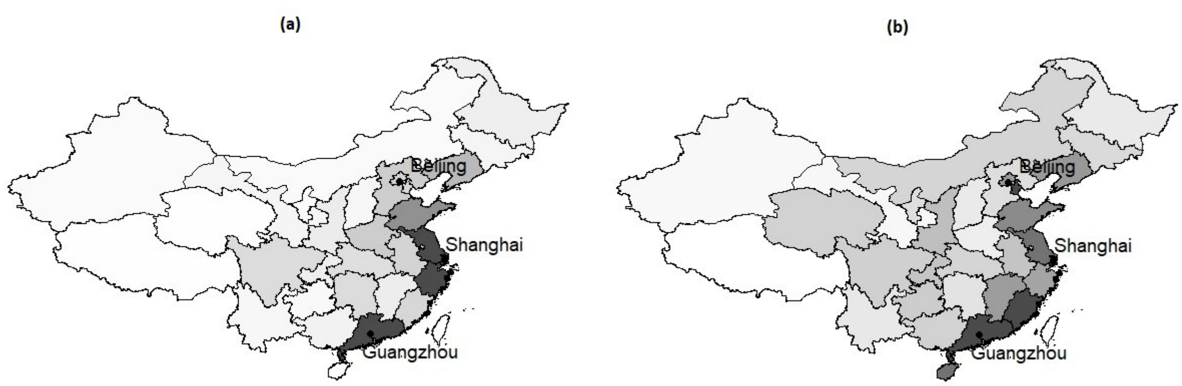

Figure H.2. Geographic Distribution of Firms (a) and the Foreign Equity Share (b)

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-

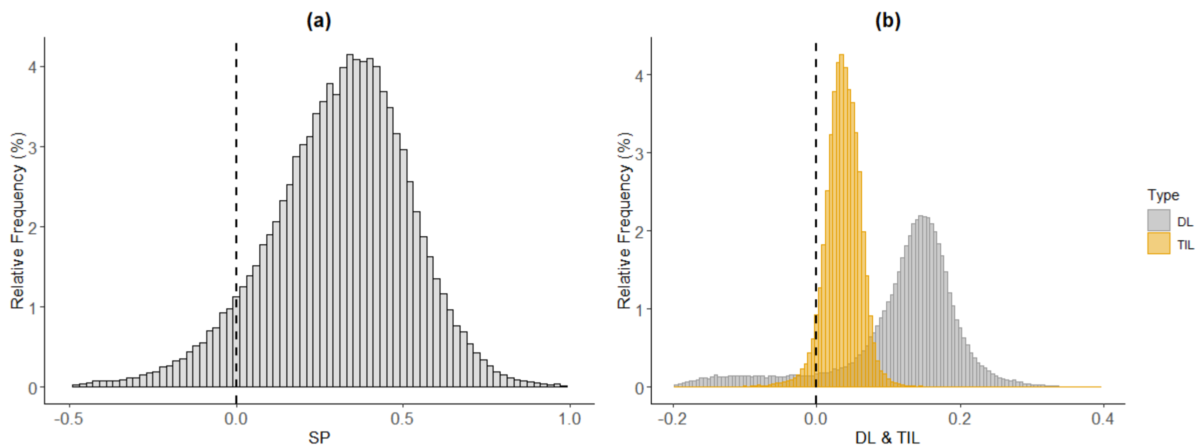

Figure I.3. Distributions of the Productivity Effects: (a) SP, (b) DL and TIL

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

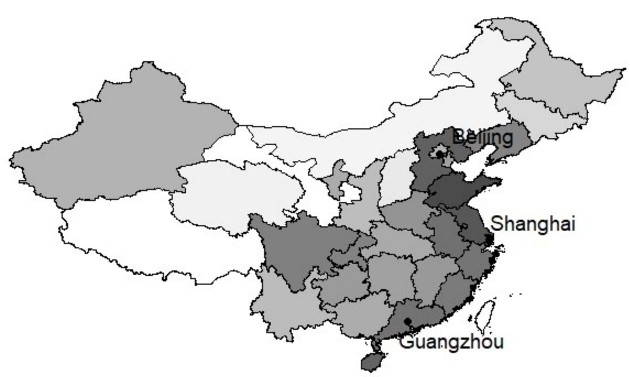

Figure I.4. Spatial Distribution of the Productivity Spillover Effects

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 $\text{Beta}(\alpha,\beta)$ 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 $SP^0_{it} = \partial\mathbb{E}[\omega_{it}|\cdot]/\partial\sum_{j(\neq i)} s^0_{ij,t-1}\omega_{j,t-1}$ and (ii) foreign-invested peers $SP^1_{it} = \partial\mathbb{E}[\omega_{it}|\cdot]/\partial\sum_{j(\neq i)} s^1_{ij,t-1}\omega_{j,t-1}$.

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 uni-dimensional 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_{it}^0 and SP_{it}^1 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

	Point Estimates			*Statistically* > 0
Estimand	1st Qu.	Median	3rd Qu.	(% 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.