Estimation of Firm-Level Productivity in the Presence of Exports: Evidence from China's Manufacturing

Malikov, Emir and Zhao, Shunan and Kumbhakar, Subal C.

University of Nevada Las Vegas, Oakland University, Binghamton University

2020

Estimation of Firm-Level Productivity in the Presence of Exports: Evidence from China's Manufacturing*

Emir Malikov[1] Shunan Zhao[2] Subal C. Kumbhakar[3]

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

January 8, 2020

Abstract

Motivated by the longstanding interest of economists in understanding the nexus between firm productivity and export behavior, this paper develops a novel structural framework for control-function-based nonparametric identification of the gross production function and latent firm productivity in the presence of endogenous export opportunities that is robust to recent unidentification critiques of proxy estimators. We provide a workable identification strategy, whereby the firm's degree of export orientation provides the needed (excluded) relevant independent exogenous variation in endogenous freely varying inputs, thus allowing us to identify the production function. We estimate our fully nonparametric IV model using the Landweber-Fridman regularization with the unknown functions approximated via artificial neural network sieves with a sigmoid activation function which are known for their superior performance relative to other popular sieve approximators, including the polynomial series favored in the literature. Using our methodology, we obtain robust productivity estimates for manufacturing firms from twenty eight industries in China during the 1999–2006 period to take a close look at China's exporter productivity puzzle, whereby exporters are found to exhibit lower productivity levels than non-exports.

Keywords: ANN sieves, control function, export, nonparametric, productivity, proxy, regularized estimation, TFP

JEL Classification: D24, F10, L10

*Email: emir.malikov@unlv.edu (Malikov), shunanzhao@oakland.edu (Zhao), kkar@binghamton.edu (Kumbhakar).

We would like to thank the co-editor, Edward Vytlacil, and three anonymous referees for many insightful comments and suggestions that greatly helped improve this paper. We are also thankful to Yiguo Sun, Nic Ziebarth, seminar participants at Auburn and Oakland as well as conference participants at the 2016 North American Productivity Workshop in Quebec City for helpful comments. Any remaining errors are our own.

1 Introduction

The importance of production function formalizing the transformation of inputs into the output needs no justification for economists. The concept lies at the very core of the theory of the firm. It is thus unsurprising that the issues surrounding the consistent estimation of the production function have been among the most studied in empirical economics. The primary hurdle to the identification of the firm's production function is latent productivity which, while unobserved by an econometrician, is one of the key determinants of the firm's endogenous input allocation. If not accounted for explicitly, the presence of the firm's productivity leads to the simultaneity (endogeneity) problem as first pointed out by Marschak & Andrews (1944).

In light of rather unsatisfactory performance of conventional approaches to tackling endogeneity in the production function context, such as fixed effects estimation or instrumenting for inputs using prices (see Griliches & Mairesse, 1998; Ackerberg et al., 2007, 2015; Gandhi et al., 2017), the recently developed alternative control-function-based approach to the *structural* identification of production functions by Olley & Pakes (1996) and Levinsohn & Petrin (2003) has gained wide popularity among practitioners. The method relies on the use of lagged inputs as a source of exogenous variation (under some assumptions about the firm's economic environment) to achieve identification. More recently, many such proxy-based production function estimators have been subject to serious critiques for the lack of identification due to perfect functional dependence (Ackerberg et al., 2015) and/or violation of the "rank condition" (Gandhi et al., 2017). Nonetheless, the estimators remain, if not have become more, dominant in the applied work on productivity.

In this paper, we develop a novel structural framework for control-function-based nonparametric identification of the gross production function and latent firm productivity that is robust to both Ackerberg et al.'s (2015) and, more importantly, Gandhi et al.'s (2017) critiques and provide an alternative estimation methodology. Namely, we consider the firm's production decisions in the presence of export opportunities.

Our motivation is largely rooted in the longstanding interest of economists in studying the nexus between firm productivity and export behavior, empirical research on which has proliferated greatly since Bernard & Jensen's (1995) seminal publication (e.g., Aw & Hwang, 1995; Baldwin & Gu, 2003; Bernard et al., 2003; Wagner, 2007, to name a few). Exporting firms are generally believed to be more productive than their domestically oriented counterparts. Not only are the exporters usually more productive *a priori* (Clerides et al., 1998; Delgado et al., 2002) but, more importantly, they are also more likely to enjoy productivity gains due to learning by exporting, quality and variety effects, exposure to tougher competition as well as the absorption of new technologies from abroad (Pavcnik, 2002; Amiti & Konings, 2007; De Loecker, 2007, 2013). In fact, the latter aspects are often invoked in support of active export promotion in developing countries. However, no such regularities seem to apply to the Chinese manufacturing sector. Puzzlingly, exporting manufacturers in China have been found to exhibit lower productivity levels than non-exporting firms (Lu, 2010; Lu et al., 2010; Yu, 2014; Dai et al., 2016). In drawing their conclusions, most these studies either use crude measures of average productivity or rely on standard proxy estimators that are prone to the above-referenced unidentification issues as well as assume restrictive parametric assumptions. In order to take a fresh look at China's exporter productivity puzzle, we therefore seek to build the framework for a robust measurement of firm productivity.

In our conceptual model of the firm's production decisions, we pay close attention to modeling of its export behavior, particularly the firm's endogenous determination of the degree of its export orientation (in addition to input choices). Failure to explicitly accommodate such endogenous export decisions and the productivity implications thereof in the production function estimation may lead to the model misspecification and the omission of relevant export variables from the proxy for latent productivity which, owing to their correlation with inputs, results in the endogeneity-inducing omitted variable bias. Perhaps even more importantly, the conventional structural framework does *not* allow the identification of both the production function and productivity in line with Gandhi et al.'s (2017) critique. Therefore, our contribution to the literature is as follows. We generalize the standard behavioral assumptions about the firm-level production customarily assumed in the literature to explicitly formalize firms' endogenous export decisions along with their potential learning-by-exporting effects on productivity. We show that, unlike under the conventional assumptions, our structural framework provides a workable identification strategy, whereby the firm's degree of export orientation provides the needed (excluded) *relevant* independent exogenous variation in endogenous freely varying inputs, thus allowing us to tackle Gandhi et al.'s (2017) under-identification critique and to successfully identify the production function and latent firm productivity. Our ability to achieve structural identification stems from both the underlying features of our conceptual framework, whereby the firm's export decisions are said to exhibit adjustment frictions implying their quasi-fixity and to depend on some external exogenous cost shifter (e.g., a locational or institutional cost differential) which need *not* be observable, and the fact that we use the information about firms' exports in the form of a continuous measure such as export intensity as opposed to a discrete export status as commonly done in the productivity literature. Thus, our methodology constitutes a useful addition to the toolkit of practitioners interested in a robust empirical modeling of the nexus between the firm's export behavior and its productivity.

In order to avoid a likely possibility of misspecification due to the widely-imposed restrictive Cobb-Douglas parametric assumption, our methodology employs the nonparametric formulation not only for the control function but also for the production process itself. However, owing to its fully nonparametric nature, the estimation of our IV model becomes non-trivial due an "ill-posed inverse" problem. To bypass the ill-posedness, we regularize our estimator (see Carrasco et al., 2007, for an excellent review). Specifically, we adopt the Landweber-Fridman regularization technique with its primary appeal being that it is iterative thereby not requiring direct inversion of large-dimensional matrices, which is of great practical importance when working with large panels of firms (as in our case). Given the well-documented poor finite-sample performance of polynomial series favored in the productivity literature (e.g., Gandhi et al., 2017), in this paper, we approximate unknown functions via the single-layer feed-forward artificial neural network (ANN) sieves with a sigmoid activation function, which can approximate any function to any desired accuracy with the use of a sufficiently large number of hidden intermediate units regulating the degree of approximation complexity (Hornik et al., 1989; White, 1989). We choose the ANN sieves primarily due to their superior ability to accommodate nonlinearities and non-separabilities between the elements of an approximated function as well as because other approximators such as polynomial, Fourier or spline sieves usually exhibit slower convergence rates requiring a higher degree of approximation complexity to attain comparable accuracy (also see Kuan, 2008; Chen & Ludvigson, 2009).

We first study our methodology in a small set of Monte Carlo experiments where we compare its performance to that of the traditional estimator. The results are encouraging, and simulation experiments show that our approach recovers the true parameters well, thereby lending strong support to the validity of our identification strategy. As expected of a consistent estimator, the estimation also becomes more stable as the sample size grows. In contrast, consistent with Gandhi et al.'s (2017) critique, the under-identified traditional proxy estimator exhibits non-vanishing biases in the estimates of production function. We then apply our model to a firm-level panel for 28 manufacturing industries in China during the 1999–2006 period to obtain estimates of the (total factor) productivity differential between exporters and non-exporters. Despite of a more robust methodology employed, we too find a generally negative productivity premium to exporting, although the results indicate non-negligible heterogeneity across industries and firm characteristics. We fairly consistently find an increasingly negative productivity differential between non-exporters and exporter with the higher degree of export orientation. Pooling the whole manufacturing sector together, the (statistically significant) conditional exporter productivity premium estimates decline from –7.1% for the first decile to –20.7% for the tenth decile of the export intensity distribution among exporters. At the disaggregated industry level, the evidence is a bit less clear-cut, with few premium estimates being statistically insignificant for low-export-intensity exporters. Although we largely fail to find evidence of these more domestically oriented exporters exhibiting a positive productivity differential over non-exporters: at most, they may be as productive as their non-exporting counterparts.

The rest of the paper proceeds as follows. Section 2 describes the model of firm-level production in the presence of endogenous exports. We describe our identification and estimation strategy in Sections 3 and 4, respectively. Section 5 describes the data. The results are discussed in Section 6. Section 7 concludes.

2 Production with Endogenous Exports

Consider the production process of a firm i ($i = 1, \dots, n$) in the time period t ($t = 1, \dots, T$) in which physical capital K_{it}, labor L_{it} and materials M_{it} (intermediate input) are being transformed into output Y_{it} via production function $F(\cdot)$ given Hicks-neutral productivity.[1] The firm's stochastic production process can be formalized as

$$Y_{it} = F(K_{it}, L_{it}, M_{it}) \exp\{\omega_{it} + \eta_{it}\}, \tag{2.1}$$

where the exponent $(\omega_{it} + \eta_{it})$ is the composite productivity term consisting of *(i)* the persistent first-order Markovian productivity ω_{it} and *(ii)* a random *i.i.d.* transitory productivity shock η_{it}.[2]

The $F(\cdot)$ function is said to satisfy the standard neo-classical assumptions, including differentiability, positive monotonicity and concavity. We assume that, unlike freely varying materials, capital and labor are subject to adjustment frictions (e.g., time-to-install, hiring costs) and thus are quasi-fixed.[3] That is, M_{it} is said to be statically determined by the firm in period t, whereas K_{it} and L_{it} are determined in period $t-1$ via dynamic optimization. Thus, both K_{it} and L_{it} are the state variables with dynamic implications and follow their respective deterministic laws of motion:

$$K_{it} = I_{it-1} + (1-\delta)K_{it-1} \quad \text{and} \quad L_{it} = H_{it-1} + L_{it-1}, \tag{2.2}$$

where I_{it}, H_{it} and δ are the gross investment, net hiring and the depreciation rate, respectively.

In this paper, we pay particularly close attention to modeling the firm's production decisions in the presence of export opportunities: exports are an outcome of the endogenous decision-making. On one hand, firms may self-select into foreign markets (Clerides et al., 1998; Delgado et al., 2002). The decision whether to start/halt exporting abroad or adjust the degree of its export orientation is likely to be correlated with the firm's quasi-fixed inputs and productivity in the production function (2.1), which, if not accounted for, will produce inconsistent and biased estimates of $F(\cdot)$ and ω_{it}. On the other hand, exporting firms are also more likely (than non-exporters) to enjoy productivity gains due to learning by exporting, quality and variety effects, exposure to tougher competition as well as the absorption of new technologies from abroad (Pavcnik, 2002; Amiti & Konings, 2007; De Loecker, 2007, 2013). Taking these productivity implications of the firm's export experiences for granted may thus lead to the misspecification of its productivity evolution process. As it is to be evident later in the paper, in either case, the failure to explicitly accommodate endogenous export decisions effectively leads to the omission of a relevant export variable from the proxy for latent productivity which, owing to its correlation with inputs, results in the endogeneity-inducing

[1]It is easy to let the production technology be time-varying, in which case $F(\cdot)$ is to be replaced with $F_t(\cdot)$. We opt against it to avoid notational clutter.

[2]Sometimes the latter shock is alternatively interpreted as a measurement error in the output.

[3]The timing assumption about L_{it} ensures that Gandhi et al.'s (2017) under-identification critique applies to a *single* freely varying input only. The latter is important because, had we assumed that L_{it} was freely varying too, our identification scheme would have fallen flat due to unavailability of the *second* source of excluded relevant exogenous variation necessary to instrument for L_{it} (also see the discussion in Appendix A). Also, our assumption about labor seems reasonable given the microeconomic evidence in support of quasi-fixity of labor in the wake of nonlinear adjustment costs (e.g., Caballero et al., 1997; Cooper & Willis, 2003). The latter may be particularly relevant in case of the Chinese labor market (given our empirical application) where, due to institutional/traditional reasons, firms oftentimes face difficulties in firing workers, especially during economic downturns. It is particularly well-documented for state-owned and -invested firms.

omitted variable bias.

We formalize firms' endogenous export decisions along the lines of Van Biesebroeck (2005) and Amiti & Konings (2007). For this, we generalize the standard behavioral framework customarily assumed in the literature. Specifically, we assume that the firm maximizes the discounted present value of the stream of its future profits subject to its own state covariates and expectations about the market structure variables including the input and output prices that are said to be common to all incumbent firms. The factor markets are assumed to be perfectly competitive. Unlike in the standard framework, the firm is now allowed to sell in domestic and/or foreign markets, both of which are perfectly competitive with potentially different market prices. The cross-market price differential may be attributed to different production cost fundamentals across countries that may persist in the presence of institutional barriers — both formal, such as export/import permits, unique packaging/labeling requirements or imperfect factor mobility, and informal like language — separating the two markets. The price differential may also be reflective of underlying differences in products themselves (horizontally or vertically) depending on the intended market of sale. Such qualitative differences across products may potentially be explained by the use of somewhat different technologies *within* the exporting firm (depending on the products' intended markets of sale), in which case the production function in (2.1) can be conceptualized as the firm-level "reduced-form" production technology with the left-hand-side output measuring the firm's total real sales in both markets. For instance, De Loecker (2011) essentially pursues a similar approach whereby representing the firm's multi-product technology (which may consist of heterogeneous product-specific technologies) via a single firm-level reduced-form/aggregate production function. Such an approach, while not ideal, allows to circumvent the pervasive lack of data on cross-product input allocations within the firm.

Under the first-order Markov process assumption both for the firm's persistent productivity and factor prices, we can adapt Ericson & Pakes's (1995) result about the existence of a Markov perfect Nash equilibrium in the firm's time t production decisions, now also including those about the export engagement. Instead of specifying the firm's exporting behavior in a standard trichotomous fashion like the bulk of the literature (e.g., Van Biesebroeck, 2005; Amiti & Konings, 2007; De Loecker, 2013), whereby the firm chooses between engaging in exports, staying oriented on domestic customers or, if already an exporter, withdrawing from foreign markets, we model export decisions in a richer, continuous framework.[4] Specifically, the firm is said to be choosing not only its exporter status but also the degree of its export orientation. We do so by defining the export intensity of the firm's sales $X_{it} \in [0, 1]$, the boundary values of which correspond to a non-exporter and a completely export-oriented firm, respectively. For instance, in the framework where products somewhat differ based on the intended market of sale, the firm's choice of X_{it} can be conveniently interpreted as its decision about the (re)configuration of production lines within the firm. We also assume a decision about the degree of export orientation X_{it} in period t is made in period $t-1$. That

[4]Furthermore, modeling export decisions as continuous also helps us achieve the identification of the firm's production function and productivity. We discuss this in more detail in Section 3.

is, we assume that changes in the firm's export intensity are costly and thereby subject to delay. The latter is meant to capture adjustment costs associated with changes in the degree of firm's export orientation which may include time for and cost of finding new intermediaries/buyers abroad, contract (re)negotiations, obtaining new permits and, perhaps most importantly, reconfiguring the production technology if products intended for sale abroad are distinct from those sold domestically, etc. Incidentally, this implied quasi-fixed treatment of the firm's export decisions is along the lines of that implicitly assumed by De Loecker (2013).[5] In such a setup, the firm's export intensity plays a role of an additional state variable that evolves according to the following controlled process:

$$X_{it} = \mathcal{X}_{it-1} + X_{it-1}, \tag{2.3}$$

with $\mathcal{X}_{it}$ representing endogenous adjustment—a choice variable—in the firm's degree of export orientation.

Thus, the firm's state variables are $(K_{it}, L_{it}, X_{it}, \omega_{it})$, with the Bellman equation corresponding to its dynamic optimization given by

$$\begin{aligned}\mathbb{V}_t(K_{it}, L_{it}, X_{it}, \omega_{it}) = \sup_{I_{it}, H_{it}, \mathcal{X}_{it}} \Big\{ &\Pi_t(K_{it}, L_{it}, X_{it}, \omega_{it}) - \mathrm{C}_t^I(I_{it}, K_{it}) - \mathrm{C}_t^H(H_{it}, L_{it}) - \mathrm{C}_t^{\mathcal{X}}(\mathcal{X}_{it}, X_{it}, \mathcal{V}_{it}) + \\ &\beta \mathbb{E}[\mathbb{V}_{t+1}(K_{it+1}, L_{it+1}, X_{it+1}, \omega_{it+1}) |\, \Omega_{it}] \Big\},\end{aligned} \tag{2.4}$$

where Ω_{it} is the information available to the firm i for making period t decisions; β is the time discount factor; and we do not explicitly include prices beyond subscripting the functions with t since, under perfect competition, they do not vary across firms in a given time period. Here, $\Pi_t(\cdot)$ is the restricted (static) profit function with the freely varying M_{it} being the choice variable in which it is maximized conditional on the already optimized quasi-fixed factors,[6] and $\mathrm{C}_t^{\kappa}(\cdot)$ is the cost of $\kappa \in \{I_{it}, H_{it}, \mathcal{X}_{it}\}$, respectively. By conditioning the cost of investment, hiring and export degree adjustment on their respective stocks (i.e., capital, labor and export intensity), in line with the convention in the literature, we are able to model the adjustment costs associated with changes in each of these choice variables.

Notably, the cost function $\mathrm{C}_t^{\mathcal{X}}(\cdot)$ warrants a few additional remarks. Not only does it allow for the adjustment costs associated with changes in the firm's export orientation but it can also implicitly accommodate the existence of one-time fixed irrecoverable exporting costs,[7] where the latter would take the form of a constant locational shift in $\mathrm{C}_t^{\mathcal{X}}$ once a non-exporter firm ($X_{it} = 0$) chooses to start exporting abroad in the next period (i.e., chooses $\mathcal{X}_{it} > 0$). More importantly, we assume the cost of changes in the firm's export orientation depends on a firm-varying exogenous variable

[5] De Loecker (2013) does not explicitly discuss the details of the firm's decisions about its export status. However, from the proxy function used in his estimation procedure (see his footnote 14), one can infer that De Loecker (2013) treats the contemporaneous export status as a state variable affecting the demand for investment or an intermediate input, which implies that the firm's export status is assumed to be predetermined.

[6] This is the value function corresponding to a static profit maximization with respect to freely varying M_{it} in (2.5).

[7] For instance, Melitz (2003) assumes that firms wishing to enter foreign markets must pay a fixed, sunk export investment cost.

$\mathcal{V}_{it}$ which represents some external factors influencing firms' export decisions. For instance, it may reflect effects of the firm's location on its export costs. Namely, since firms have different distances to their nearest shipping center (a port, freight train station, etc), it is reasonable to assume that those close to the shipping centers also face lower exporting costs via lower transportation costs. Furthermore, in the economic geography and trade literature it is widely accepted that the activities of neighboring peers may also help reduce the firm's exporting costs, and $\mathcal{V}_{it}$ thus may be thought of as capturing such spillover effects within region and/or industry. Lastly but not least importantly, the Chinese manufacturing firms' exporting costs also likely depend on their location given the country's unique institutional environment whereby the authorities, including both the regional or local governments, have been actively promoting exports through tax and policy incentives by establishing economic and technological development zones, high-tech industrial zones and export-processing zones around the country (Wang & Wei, 2010). Besides the locational effects, $\mathcal{V}_{it}$ may also capture the effects of other relevant variables, such as credit constraints, on the firm's export costs. For example, Feenstra et al. (2014) have recently found that, given China's credit market imperfections which adversely affect exporters and hence influence trade patterns, credit constraints become more stringent as the firm's export intensity grows. As it is to become more apparent in Section 3, the presence of an exogenous factor $\mathcal{V}_{it}$ affecting the firm's export behavior from outside its production environment enables us to circumvent Gandhi et al.'s (2017) critique and successfully identify the firm's production function. Crucially, such a variable need *not* be observed for the estimation purposes when employing *freely varying* inputs to proxy for the firm's latent productivity like we do in this paper.

Solving (2.4) for $(I_{it}, H_{it}, \mathcal{X}_{it})$, which represent changes in dynamic state variables, produces the export-behavior adjusted gross investment function $I_{it} = \mathbb{I}_t(K_{it}, L_{it}, X_{it}, \mathcal{V}_{it}, \omega_{it})$, the net hiring function $H_{it} = \mathbb{H}_t(K_{it}, L_{it}, X_{it}, \mathcal{V}_{it}, \omega_{it})$ along with the export orientation adjustment function $\mathcal{X}_{it} = \mathbb{X}_t(K_{it}, L_{it}, X_{it}, \mathcal{V}_{it}, \omega_{it})$. Note that all three policy functions depend on $\mathcal{V}_{it}$ which, if the latter is unobservable (like in our application), implies that neither investment nor hiring are eligible proxies for ω_{it} due to apparent violation of the "scalar unobservable" condition necessary to ensure the invertability of these functions. Fortunately, as shown below, this condition however continues to hold for M_{it} which, owing to its freely varying nature, does not depend on $\mathcal{V}_{it}$ or any other cost determinants pertaining to the choice of dynamic inputs.

To obtain the conditional demand for freely varying materials M_{it}, we consider the firm's static optimization problem (under risk neutrality) embedded in the restricted profit function $\Pi_t(\cdot)$ inside the dynamic problem in (2.4):

$$\max_{M_{it}} \; P_t(X_{it})F(K_{it}, L_{it}, M_{it}) \exp\{\omega_{it}\}\mathcal{E} - \mathrm{C}_t^M(M_{it}), \tag{2.5}$$

where $P_t(X_{it}) \equiv [P_t^X X_{it} + P_t^D(1 - X_{it})]$ is the firm-specific composite output price that is the export-intensity-weighted average of domestic (P_t^D) and export (P_t^X) output prices; $\mathcal{E} \equiv \mathbb{E}[\exp\{\eta_{it}\}| \, \Omega_{it}]$ is some constant; and $\mathrm{C}_t^M(\cdot)$ is the cost of materials. Solving (2.5) for M_{it} and omitting the elements common to all firms yields the material demand function $M_{it} = \mathbb{M}_t(K_{it}, L_{it}, X_{it}, \omega_{it})$ with just one unobservable. Given the profit-maximizing behavior by firms and the cross-input regularity conditions on the production function, $\mathbb{M}_t(\cdot)|\ M_{it} > 0$ must be strictly monotonic in ω_{it} for any given (K_{it}, L_{it}, X_{it}). This condition is similar to that derived by Levinsohn & Petrin (2003). Hence, so long as $\mathbb{M}_t(\cdot)$ satisfies a scalar unobservability condition, it can be inverted to proxy for persistent productivity via $\omega_{it} = \mathbb{M}_t^{-1}(K_{it}, L_{it}, X_{it}, M_{it})$.[8]

Remark 1 The optimal policy function for intermediate inputs of the form $M_{it} = \mathbb{M}_t(K_{it}, L_{it}, X_{it}, \omega_{it})$, whereby the quantity demanded of materials is conditioned not only on quasi-fixed inputs and productivity but also on the firm's (predetermined) degree of export orientation, may also be justified without invoking differential prices in the domestic and foreign markets as we do in (2.5). For instance, one can reasonably postulate dependence of the optimal M_{it} on X_{it} if the firm is said to potentially use materials of higher quality in the production of output intended for sale abroad. In case of a developing country like China, the latter may be reasonably justified on the basis of higher standards in developed countries whereto the products are exported.

Next, we formalize the productivity effects of firm's export behavior. We effectively assume that exporting impacts the firm's output only *indirectly* (via its productivity) which is why no export variable explicitly enters production function (2.1).[9] Specifically, we allow the evolution of ω_{it} to be impacted by the firm's past export experiences to capture the "learning by exporting" effects by letting its persistent productivity follow a *controlled* first-order Markov process à la De Loecker (2013):[10]

$$\omega_{it} = \mathbb{E}[\omega_{it}|\ \omega_{it-1}, X_{it-1}] + \zeta_{it} \quad \text{with} \quad \mathbb{E}[\zeta_{it}|\ \Omega_{it-1}] = 0, \tag{2.6}$$

where it is convenient to interpret ζ_{it} as the innovation in persistent productivity, unobservable to the firm in period $t-1$. Note that firms do observe ω_{it}, consisting of $\mathbb{E}[\omega_{it}|\ \omega_{it-1}, X_{it-1}]$ and ζ_{it}, in period t when decisions concerning freely varying M_{it} are being made, i.e., $\omega_{it} \in \Omega_{it}$. With regards to a transitory productivity shock η_{it}, following the convention, we assume that

$$\mathbb{E}[\eta_{it}|\ \Omega_{it}] = \mathbb{E}[\eta_{it}] = 0, \tag{2.7}$$

with its mean normalized to zero. The above implies that the random shock η_{it} is observable to firms in period t only *ex post* after all production decisions (including those about M_{it}) take place.

[8] Note that, while our main covariate of interest X_{it} can take zero values, this does *not* pose a zero-value problem that served as the original motivation for Levinsohn & Petrin's (2003) estimator because we do not seek to invert the policy function for export intensity in order to proxy for ω_{it} in the estimation. We proxy using the conditional demand for M_{it}.

[9] That is, following the convention in the literature, inputs are said to be the only variables affecting the firm's output directly, with all other production-related variables working through the latent productivity term ω_{it}.

[10] Despite the similarity between our approach to modeling productivity effects of the firm's export behavior and that by De Loecker (2013), our papers pursue different identification strategies. Unlike us, De Loecker (2013) considers the estimation of the value-added, not gross, production function and therefore, in the face of Gandhi et al.'s (2017) critique, his identification scheme *cannot* be easily extended to the case when intermediate inputs enter the production function directly. Our paper however pays particular attention to the latter issue.

The evolution process in (2.6) implicitly assumes that learning is a costly process which is why the dependence of ω_{it} on the export variable is lagged implying that the export-driven improvements in firm productivity take a period to materialize. This is a common assumption in the "learning" literature: e.g., De Loecker (2013) in the case of exporting or Doraszelski & Jaumandreu (2013, 2018) in the case of R&D. Further, due to adjustment costs, firms do not experience immediate changes in their degree of export orientation in light of a productivity shock thereby $\mathbb{E}[\zeta_{it}|\ \Omega_{it-1}] = 0$ with $X_{it-1} \in \Omega_{it-1}$.

To summarize, our framework implies that, in period $t-1$, the firm adjusts (if at all) the split of its total sales across the two markets for the next period t. The actual level of output, sold both domestically and abroad, is however determined come next period, after the random transitory productivity shock has realized as well as the quantity of intermediate inputs has been decided on the basis of the newly updated persistent productivity. Also, by conditioning the investment and hiring functions on the firm's degree of export orientation X_{it}, we implicitly allow the use of quasi-fixed inputs within the firm to depend on the product designation (i.e., whether the product is for export or not). That is, if the exported products are distinct from those sold locally and are produced using a somewhat different technology (e.g., using somewhat different machinery or better quality labor), the latter implies that these inputs may not be perfectly substitutable across the product lines. Van Biesebroeck (2005) and Amiti & Konings (2007) pursue a similar approach. Thus, our structural assumptions about costly and lagged adjustments in the firm's degree of export orientation (subject to unobservable exogenous cost shifter) along with costly learning by exporting and the timing of arrival of ζ_{it} and η_{it} help ensure that X_{it} is weekly exogenous, relevant for the contemporaneous material inputs and excludable from the production function.

Note that we model the firm's endogenous export decisions in the "input-allocation" paradigm as opposed to explicitly specifying a full-fledged export decision rule. That is, by modeling the firm choosing the degree of its export orientation (X_{it+1}) along with its inputs, we essentially conceptualize the change in export intensity ($\mathcal{X}_{it}$) as another "investment". Van Biesebroeck (2005) builds on a similar framework, although his estimation methodology differs from ours on multiple fronts with perhaps the most important difference being in his use of the original Olley & Pakes's (1996) estimator which has been shown to suffer from unidentification issues. In contrast, our identification strategy is robust to various unidentification critiques of the material-based proxy estimators of production functions (see Section 3). Further, by means of our investment-like conceptualization of export decisions, we avoid making *a priori* presumptions about specific categories of firms (e.g., high-productivity firms) self-selecting into foreign markets, thus mitigating the risks of artificially pre-tailoring the model to produce desirable results. Essentially, our approach takes a more agnostic view of export decisions. Our framework also suggests that the previous results about the Nash equilibrium with the corresponding conditional demand functions can be extended to our export adjusted model. Since the modified model of the firm's production decisions differs from the standard one essentially only by including an additional firm-specific state variable (X_{it}), all integral components necessary to ensure that Ericson & Pakes's (1995) and Olley & Pakes's

(1996) results hold are still in place.

3 Identification

The estimation of the production function in (2.1) is not trivial, among other things, due to the latent nature of firm productivity ω_{it}. Omitting it from the regression is not an option, since this would give rise to the endogeneity problem given that ω_{it} is correlated with inputs. We tackle this problem by adopting a control function approach à la Levinsohn & Petrin (2003) whereby we proxy for unobservable ω_{it} via the observable freely varying input M_{it}.

Taking logs of both sides of (2.1) yields

$$y_{it} = f(K_{it}, L_{it}, M_{it}) + \omega_{it} + \eta_{it}, \tag{3.1}$$

where the lower-case variables/functions denote the logs of the respective variables/functions, e.g., $f_{it} \equiv \log F_{it}$. Making use of the Markovian nature of ω_{it} from (2.6), we then rewrite (3.1) as

$$y_{it} = f(K_{it}, L_{it}, M_{it}) + h[\omega_{it-1}, X_{it-1}] + \zeta_{it} + \eta_{it}, \tag{3.2}$$

where we let $\mathbb{E}[\omega_{it}|\cdot]$ be an unknown function $h[\cdot]$. As discussed in Section 2, under our structural assumptions, we can invert the material demand to control for latent persistent productivity via $\omega_{it} = \mathbb{M}_t^{-1}(K_{it}, L_{it}, X_{it}, M_{it})$. Specifically, substituting for ω_{it-1} using the proxy, from (3.2) we get

$$\begin{aligned} y_{it} &= f(K_{it}, L_{it}, M_{it}) + h\left[\mathbb{M}_{t-1}^{-1}(K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1}), X_{it-1}\right] + \zeta_{it} + \eta_{it} \\ &\equiv f(K_{it}, L_{it}, M_{it}) + \varphi(K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it} + \eta_{it}, \end{aligned} \tag{3.3}$$

where $\varphi(\cdot)$ is some unknown function.

The identification of production function $f(\cdot)$ in (3.3) requires that the endogenous M_{it} (due to its correlation with ζ_{it}) be instrumented. Following the bulk of the literature, one may think that the nonparametric model in (3.3) can be seemingly identified by utilizing abundant lagged covariates to instrument for M_{it} given that they are predetermined under assumptions about the evolution of firm's productivity and hence meet the (weak) exogeneity requirement for valid instruments. More specifically, under the assumptions embedded in (2.6) and (2.7), equation (3.3) satisfies the following orthogonality condition:

$$\mathbb{E}[\zeta_{it} + \eta_{it}|\ K_{it}, L_{it}, X_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1}, \ldots, K_{i1}, L_{i1}, X_{i1}, M_{i1}] = 0. \tag{3.4}$$

However, Gandhi et al. (2017) have recently shown that model (3.3) may still be *un*identified despite numerous moments in (3.4), and that the identification can only be achieved if the additional predetermined variables provide some excluded *relevant* (exogenous) variation for M_{it} after conditioning on the already included self-instrumenting variables. Intuitively, in addition to satisfying the order condition, these variables must also be relevant to meet the rank condition for model identification.

In what follows, we show that Gandhi et al.'s (2017) critique of traditional proxy estimators, however, does *not* apply to our model. Under our conceptual assumptions, the freely varying input M_{it} appearing inside the production function $f(\cdot)$ in (3.3) has a valid instrument, namely X_{it}, which provides *excluded relevant* exogenous variation, conditional on the already included self-instrumenting $(K_{it}, L_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$. Therefore, we are able to structurally identify the model.

The (structural) validity of X_{it} as an instrument rests on *(i)* its exogeneity whereby $\mathbb{E}[\zeta_{it} + \eta_{it}|\ X_{it}] = 0$, *(ii)* the containment of independent variation from outside the model, and *(iii)* its relevancy for the firm's choice of M_{it}. The first condition holds owing to the quasi-fixed nature of firm's export intensity due to the costly adjustment therein. To see how X_{it} meets the remaining two conditions, note that the freely varying input M_{it}

$$\begin{aligned} M_{it} &= \mathbb{M}_t(K_{it}, L_{it}, X_{it}, \omega_{it}) \\ &= \mathbb{M}_t(K_{it}, L_{it}, X_{it}, h_t(\omega_{it-1}, X_{it-1}) + \zeta_{it}) \\ &= \mathbb{M}_t(K_{it}, L_{it}, X_{it}, \varphi(K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it}), \end{aligned} \tag{3.5}$$

is a function of the following observables $(K_{it}, L_{it}, X_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$. Comparing this set of variables with those that enter the right-hand side of (3.3) and already self-instrument, it is evident that M_{it} has an extra source of variation coming from X_{it} (strictly speaking, $\mathcal{X}_{it-1}$) which contains exogenous independent variation from outside the production function. Using the firm's export orientation adjustment function along with its law of motion, we have that

$$\begin{aligned} X_{it} &= \mathcal{X}_{it-1} + X_{it-1} \\ &= \mathbb{X}_{t-1}(K_{it-1}, L_{it-1}, X_{it-1}, \mathcal{V}_{it-1}, \omega_{it-1}) + X_{it-1} \\ &= \mathbb{X}_{t-1}(K_{it-1}, L_{it-1}, X_{it-1}, \mathcal{V}_{it-1}, \mathbb{M}_{t-1}^{-1}(K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})) + X_{it-1}, \end{aligned} \tag{3.6}$$

from where it can be easily seen that, conditional on the self-instrumenting variables $(K_{it}, L_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$ in our model, X_{it} provides relevant independent (excluded) exogenous variation for M_{it} originating from the exogenous, albeit unobservable, $\mathcal{V}_{it-1}$. Therefore, we can use X_{it} as an excluded valid instrument for M_{it} which ultimately enables us to identify production function $f(\cdot)$ from model (3.3) on the basis of the following moment conditions

$$\mathbb{E}[\zeta_{it} + \eta_{it}|\ K_{it}, L_{it}, X_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1}] = 0, \tag{3.7}$$

where all covariates except the endogenous materials M_{it} (instrumented by X_{it}) appearing in (3.3) instrument for themselves.[11] In other words, being a function of seven arguments $(K_{it}, L_{it}, M_{it}, K_{it-1},$

[11] While, owing to their exogeneity, second- and higher-order lags can also be added to the conditioning set, these additional instruments are however irrelevant for predicting M_{it}, conditional on the already included instrument set

$L_{it-1}, X_{it-1}, M_{it-1})$, the conditional expectation $\mathbb{E}[y_{it}|\,\Omega_{it}]$ from (3.3) has seven full-rank observable sources of relevant exogenous variation $(K_{it}, L_{it}, X_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$ thereby meeting the "rank condition" for identification.

Remark 2 Incidentally, our identification arguments based on the use of unobservable exogenous variation in the endogenous regressor (originating from an export adjustment cost shifter) are, in essence, along the lines of Matzkin's (2004) "unobservable instrument" idea for the identification of nonseparable nonparametric functions with scalar unobservables. The latter idea has also recently been adapted in the context of the production function estimation by Ackerberg & Hahn (2015), from whom we differ in that we do not require the assumption of "exogenous" Markov process for ω_{it} in order to accommodate learning effects of exporting and allow for an additional unobservable shock η_{it}. We accomplish this by assuming that the nonparametric production function is additively separable from both ω_{it} and η_{it} in logs.

Technically, owing to its fully nonparametric nature, the identification of (3.3) requires additional restrictions on the distribution of data (as well as regularized estimation). First, we would not have been able to identify our model if we were to use the information on exporting in the form of a discrete export status dummy as commonly done in the productivity literature, since the IV identification in nonparametric models generally requires an instrument to be as complex as the endogenous variable it is meant to instrument (e.g., see Newey & Powell, 2003). In our case, the continuous variable for materials M_{it} is instrumented using the firm's export intensity X_{it} that also has a continuous codomain. The second restriction is more nuanced. For concreteness, we first rewrite our model of interest (3.3) ignoring its additivity as $y_{it} = \psi(W_{it}, M_{it}) + u_{it}$, where $\psi(\cdot) \equiv f(\cdot) + \varphi(\cdot)$, $u_{it} \equiv \zeta_{it} + \eta_{it}$, and $W_{it} = (K_{it}, L_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})'$ is a vector of self-instrumenting exogenous covariates, and focus on the identification of $\psi(\cdot)$. Our identification problem can then be viewed as that of the familiar nonparametric IV identification of the following model:

$$y_{it} = \psi(W_{it}, M_{it}) + u_{it}, \qquad \mathbb{E}[u_{it}|Z_{it}] = 0, \qquad Z_{it} = (W_{it}', X_{it})'. \tag{3.8}$$

Equivalently, taking the conditional expectation of equation (3.8), our function of interest $\psi(\cdot)$ corresponds to any solution to the following integral Fredholm equation of the first kind:

$$r(z) = \int \psi(w, m) G_{M|Z}(\mathrm{d}m|z), \tag{3.9}$$

with $r(z) \equiv \mathbb{E}[y|z]$ on the left-hand side being the reduced-form conditional mean function and the integral on the right-hand side of equation being equal to $\mathbb{E}[\psi(w, m)|z]$, and where $G_{M|Z}$ is the conditional cdf of M_{it} given Z_{it}. Clearly, both $r(z)$ and $G_{M|Z}$ are identifiable from the observable data $(y_{it}, M_{it}, Z_{it}')'$, and the identification of $\psi(\cdot)$ thus depends on the uniqueness of solution to the

$(K_{it}, L_{it}, X_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$. As can be seen from (3.5), conditional on these covariates, the endogenous M_{it} has only one source of "free" variation left which is the unobservable ζ_{it}.

functional equation (3.9). By Proposition 2.1 in Newey & Powell (2003), this uniqueness condition is satisfied, and hence $\psi(\cdot)$ is identified, if and only if the conditional expectation of M_{it} given Z_{it} is complete in X_{it} whereby $\mathbb{E}[\psi^*(w,m)|z] = 0$ a.s. implies $\psi^*(w,m) = 0$ a.s. for all $\psi^*(w,m)$. Such a completeness condition is essentially a nonparametric analogue of the conventional rank condition for linear models to ensure the association between M_{it} and X_{it} and has been widely used in the literature on nonparametric IV identification (for references, e.g., see Freyberger, 2017). Following Newey & Powell (2003), we can achieve identification of $\psi(\cdot)$ by relying on the L^1-completeness[12] property of the exponential family by assuming the conditional distribution of M_{it} given Z_{it} belong to this class of distributions (by their Theorem 2.2). Else, we can allow for a much broader class of distributions by assuming an L^2-completeness à la Hall & Horowitz (2005) and Darolles et al. (2011) which imposes a second-moment condition on the joint distribution of data instead of a weaker first-moment condition imposed by an L^1-completeness (also see Andrews, 2017).

To see the intuition underlying the identification result more clearly, consider the following. We assume y_{it} is square-integrable with the observable data $(y_{it}, M_{it}, Z'_{it})'$ being characterized by its joint cdf $G_{y,M,Z}$, dominated by the Lebesgue measure. For a given $G_{y,M,Z}$, define the function space $L^2_G(\chi)$ of real-valued square-integrable functions of $\chi \in \{y, M, Z\}$, and recall that $Z = (W', X)'$. Further, assume that $\mathbb{E}[y^2|Z = z] < \infty$ and $r(z) \in L^2_G(Z)$. Next, define the linear conditional-expectation operator $\mathbb{T}$:

$$\mathbb{T} : L^2_G(W, M) \to L^2_G(Z),\ \psi \to \mathbb{T}\psi\ = \mathbb{E}[\psi(W, M)|Z], \tag{3.10}$$

where $\mathbb{T}$ projects functions of (W, M) onto the space of functions of Z. Using this notation, the analogue of our model in (3.9) is given by

$$r(z) = \mathbb{T}\psi(w, m), \tag{3.11}$$

from where it is evident that the solution for $\psi(\cdot)$ is unique if and only if $\mathbb{T}$ is nonsingular or one-to-one, which the completeness condition is meant to ensure. The estimation of unknown $\psi(\cdot)$ from (3.11) [or, equivalently, from (3.9)] is however non-trivial due an "ill-posed inverse" problem, more on which in Section 4.

With function $\psi(\cdot)$ identified, by Lemma 1 in Gandhi et al. (2013), we can identify up to an additive constant its additive components $f(\cdot)$ and $\varphi(\cdot)$, with former of the two being the production function of interest. To see this intuitively, first note that, with $\psi(\cdot) = f(\cdot) + \varphi(\cdot)$ identified, the conditional expectation of $\psi(\cdot)$ given Z_{it} is unique almost surely, so that any other additive nonparametric function $f^*(\cdot)+\varphi^*(\cdot)$ satisfying our model's assumptions must be such that $\Pr[f^*(\cdot) + \varphi^*(\cdot) = f(\cdot) + \varphi(\cdot)] = 1$ (also see Newey et al., 1999). Owing to the differentiability of $\psi(\cdot)$ and $f(\cdot)$ and assuming the boundary of the support of $(K_{it}, L_{it}, M_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$ has a probability measure zero, $\psi(\cdot)$ and its partial derivatives are uniquely identified at each point in the interior of their support. Thus, differentiating $\psi(\cdot)$ with respect to the production

[12] Normally, referred to simply as "completeness" (for details, see Andrews, 2017).

function covariates, we have that $\partial\psi(\cdot)/\partial v_{it} = \partial f^*(\cdot)/\partial v_{it} = \partial f(\cdot)/\partial v_{it}$ for each $v_{it} \in \{K_{it}, L_{it}, M_{it}\}$ implying that $f^*(\cdot)$ and $f(\cdot)$ differ only by an additive constant with probability one. Hence, production function $f(\cdot)$ is identified up to a constant (for more details, see Gandhi et al., 2013).

In summary, our model would have remained under-identified as in the traditional setup critiqued by Gandhi et al. (2017) [see why in Appendix A] had we not brought in an additional source of exogenous relevant independent variation (related to the firm's export behavior) from outside the production function. Intuitively, the logic of proxying for the latent productivity using the information on M_{it}, which already enters the production function on its own, is "circular", and the identification therefore requires an excluded valid instrument. Our identification strategy works because of the underling structural model of the firm's behavior discussed in Section 2, whereby the firm decides whether to start/halt exporting abroad and/or to adjust the degree of its export orientation (along with its input allocation decisions) in the dynamic profit maximization framework summarized in Bellman equation (2.4). Such a conceptual framework guarantees that the conditional demand for materials M_{it} is a function of the firm's export intensity X_{it} which provides an additional source of independent exogenous variation from outside the production function that originates from $\mathcal{V}_{it}$ which, importantly, need *not* be observed for the estimation purposes.

Remark 3 In principal, the core idea of our approach to the structural identification of production function $f(\cdot)$ can also be extended to the setting in which firms engage in production-related decisions other than exporting, such as R&D investments studied by Doraszelski & Jaumandreu (2013), so long as such (continuously measured) decisions just like X_{it} in our paper: *(i)* are subject to adjustment frictions, *(ii)* contain independent exogenous variation, and *(iii)* are among the state variables affecting the firm's freely varying input choices.

We also caution the reader against potentially confusing the identification issue we discuss here from that pointed out by Ackerberg et al. (2015) whose focus is on *separable* identifiability of two additive unknown functions $f(\cdot)$ and $\varphi(\cdot)$ after conditioning the equation of interest (3.3) on instruments. Such a problem may arise in the wake of potentially perfect functional dependence[13] between an *endogenous* freely varying input M_{it} inside $f(\cdot)$ and the arguments of the proxy function $\varphi(\cdot)$. In our model, we are able to separably identify these two functions because, conditional on our instrument set $(K_{it}, L_{it}, X_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$, as can be seen from (3.5) the variable M_{it} still has a source, albeit unobservable, of *independent* variation, namely ζ_{it}. Thus, our model does not suffer from the perfect functional dependence problem à la Ackerberg et al. (2015) [for more details, see Appendix B].

Lastly, we identify the persistent productivity ω_{it}. Clearly, it is not identified from (3.3) since we cannot isolate ζ_{it} from the composite error $(\zeta_{it} + \eta_{it})$. Therefore, in order to identify ω_{it}, we consider the production process (3.1) where ω_{it} is proxied by the *contemporaneous* inverted demand

[13]The analogue of "perfect functional dependence" in a linear parametric framework is the "perfect collinearity."

for intermediate inputs, i.e.,

$$y_{it} = f(K_{it}, L_{it}, M_{it}) + g(K_{it}, L_{it}, X_{it}, M_{it}) + \eta_{it}, \tag{3.12}$$

with $g(\cdot) \equiv \mathbb{M}_t^{-1}(\cdot)$. Persistent productivity $\omega_{it} = \mathbb{M}_t^{-1}(K_{it}, L_{it}, X_{it}, M_{it})$ is then identified by subtracting the already identified $f(\cdot)$ from both sides of (3.12) yielding the standard nonparametric regression:

$$y_{it}^* \equiv y_{it} - f(K_{it}, L_{it}, M_{it}) = g(K_{it}, L_{it}, X_{it}, M_{it}) + \eta_{it}, \tag{3.13}$$

such that

$$\mathbb{E}[\eta_{it}|\ K_{it}, L_{it}, X_{it}, M_{it}] = 0. \tag{3.14}$$

4 Estimation Procedure

We implement our identification strategy via a three-step estimation procedure.

Step 1. The first step concerns the estimation of an additive nonparametric function $\psi(\cdot) = f(\cdot) + \varphi(\cdot)$ from its identifying integral equation (3.11) to which it is a unique solution under the completeness assumption. This estimation is however non-trivial due to discontinuities in the inverse mapping from r to ψ as eigenvalues of $\mathbb{T}\psi$ approach zero. This lack of continuity of the estimator of ψ in the reduced-form estimators of r and $\mathbb{T}$ implies that small inaccuracies in the latter may lead to large inaccuracies in the estimates of ψ. Consequently, consistency of $\widehat{\psi}$ does not immediately follow from consistency of well-identified $\widehat{r}$ and $\widehat{\mathbb{T}}$. Regularization methods are a common approach to circumvent this "ill-posed inverse" problem, which we also pursue here.

Most regularization methods use the dual operator of $\mathbb{T}$. Define such an operator as

$$\mathbb{T}^* : L^2_G(Z) \to L^2_G(W, M),\ \vartheta \to \mathbb{T}^*\vartheta\ = \mathbb{E}[\vartheta(Z)|W, M], \tag{4.1}$$

where $\mathbb{T}^*$ does the opposite of what $\mathbb{T}$ does by projecting functions of Z onto the space of square-integrable functions of (W, M). Applying this projector to (3.11) gives us

$$\mathbb{T}^*r = \mathbb{T}^*\mathbb{T}\psi, \tag{4.2}$$

from where it is apparent that the identification of ψ can equivalently be expressed in terms of the non-singularity of $\mathbb{T}^*\mathbb{T}$. Heuristically, the unique solution to (4.2) is $\psi = [\mathbb{T}^*\mathbb{T}]^{-1}\mathbb{T}^*r$. While this solution can be assumed to exist and be well-defined in the population, it is generally not the case in the sample. Specifically, ψ is identifiable if and only if 0 is not an eigenvalue of $\mathbb{T}^*\mathbb{T}$ (also see Corollary 2.1 in Darolles et al., 2011). However, the smallest eigenvalues of the matrix can get arbitrarily close to zero and therefore, in practice, the direct inversion of $\mathbb{T}^*\mathbb{T}$ may lead to an explosive, non-continuous solution. For more intuition, see Centorrino et al. (2017); Horowitz (2014) also provides an excellent review of ill-posedness and regularization in economics.

To bypass the ill-posedness, we regularize $\mathbb{T}^*\mathbb{T}$, which effectively entails choosing a regularization (tuning) parameter to make the problem be well-posed. In essence, the regularization procedure replaces $\mathbb{T}^*\mathbb{T}$ with its continuous transformation to rule out explosive solutions. Several regularization methods are available (see Carrasco et al., 2007, for review), with the Tikhonov (e.g., Hall & Horowitz, 2005; Darolles et al., 2011) and Petrov-Galerkin (e.g., Blundell et al., 2007; Horowitz, 2011) regularizations perhaps being the most popular in the literature. In this paper, we adopt an alternative Landweber-Fridman regularization technique with its primary appeal being that it is iterative thereby not requiring direct inversion of a large-dimensional matrix $\mathbb{T}^*\mathbb{T}$, which is of particular importance to us given large values that n takes in our application. Recent applications of the Landweber-Fridman regularization include Centorrino (2016), Centorrino et al. (2017), Florens et al. (2018) and Centorrino et al. (2019).

The sample analogue of the identifying equation in (4.2) is

$$\widehat{\mathbb{T}}^*\widehat{r} = \widehat{\mathbb{T}}^*\widehat{\mathbb{T}}\psi, \tag{4.3}$$

which defines the estimator of ψ as a solution of this large-dimensional system of equation. In light of the ill-posed inverse problem, system in (4.3) is expected to be almost singular in the finite sample, which is why we are to regularize it.

To obtain consistent estimates of r, $\mathbb{T}$ and $\mathbb{T}^*$ to be used in (4.3), we employ series regressions. We opt for linear sieves here mainly for computational ease, thereby assuming that functions in $L^2_G(Z)$ and $L^2_G(W,M)$ can be approximated by a finite sum of basis functions. More concretely, we use polynomials (in logs). Also, remember that $W_{it} = (K_{it}, L_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})'$ and $Z_{it} = (W'_{it}, X_{it})'$.

The projection of y_{it} on Z_{it} provides an estimator for the reduced-form $r(Z_{it}) \equiv \mathbb{E}[y_{it}|Z_{it}]$. Specifically, let $\{\phi_1(\cdot), \phi_2(\cdot), \dots\}$ be a sequence of polynomial series (or the tensor product thereof). Then, for each z, we approximate $\mathbb{E}[y|z]$ by $\boldsymbol{\phi}_{L_n}(z)'\boldsymbol{\pi}$, where, for any integer $\kappa > 0$, we denote a $\kappa \times 1$ vector of known basis functions $\boldsymbol{\phi}_\kappa(v) = [\phi_1(v), \dots, \phi_\kappa(v)]'$, and the unknown parameter vector $\boldsymbol{\pi}$ is of dimension L_n. L_n controls the complexity of approximation and slowly increases with the sample size n. The series regression

$$y_{it} = \boldsymbol{\phi}_{L_n}(Z_{it})'\boldsymbol{\pi} + e_{it}, \tag{4.4}$$

where e_{it} is a conditional-mean projection error, is estimated via least squares. This yields the estimator of r:

$$\widehat{r} = \boldsymbol{\mathcal{Z}}_{L_n}\widehat{\boldsymbol{\pi}} = \boldsymbol{\mathcal{Z}}_{L_n}(\boldsymbol{\mathcal{Z}}'_{L_n}\boldsymbol{\mathcal{Z}}_{L_n})^{-1}\boldsymbol{\mathcal{Z}}'_{L_n}\mathbf{y}, \tag{4.5}$$

where $\boldsymbol{\mathcal{Z}}_{L_n}$ is the matrix of basis functions constructed by stacking up $\{\boldsymbol{\phi}'_{L_n}(Z_{it})\}$ in the ascending order of index i first then index t; and so is the vector $\mathbf{y}$ but using y_{it}.

To estimate $\mathbb{T}$ and $\mathbb{T}^*$, recall that they are conditional expectation operators and, hence, can be estimated using linear smoothers. Following Centorrino et al. (2017), we approximate $\mathbb{T}$ via

$$\widehat{\mathbb{T}} = \boldsymbol{\mathcal{Z}}_{L_n}(\boldsymbol{\mathcal{Z}}'_{L_n}\boldsymbol{\mathcal{Z}}_{L_n})^{-1}\boldsymbol{\mathcal{Z}}'_{L_n}, \tag{4.6}$$

which is essentially a projection matrix onto the space of $\mathbf{Z} = [\mathbf{W}\ \mathbf{X}]$, once we fix L_n. Analogously, the approximation of $\mathbb{T}^*$ can be obtained from the polynomial series regression of $\widehat{r}(Z_{it})$ on (W_{it}, M_{it}), which yields

$$\widehat{\mathbb{T}}^* = \boldsymbol{\mathcal{M}}_{J_n}(\boldsymbol{\mathcal{M}}'_{J_n}\boldsymbol{\mathcal{M}}_{J_n})^{-1}\boldsymbol{\mathcal{M}}'_{J_n}, \tag{4.7}$$

where $\boldsymbol{\mathcal{M}}_{J_n}$ is the matrix of basis functions constructed by stacking up $\{\boldsymbol{\phi}'_{J_n}(W_{it}, M_{it})\}$ in the ascending order of index i first then index t, with $\boldsymbol{\phi}_{J_n}(W_{it}, M_{it})$ denoting a $J_n \times 1$ vector of the tensor product of known polynomial series of $(W'_{it}, M_{it})'$; and $J_n \to \infty$ slowly as $n \to \infty$. Essentially, for a fixed J_n, $\widehat{\mathbb{T}}^*$ is a projection matrix onto the space of $[\mathbf{W}\ \mathbf{M}]$.

We select both smoothing parameters L_n and J_n via generalized cross-validation of Craven & Wahba (1979):

$$\bar{L}_n = \underset{L_n}{\operatorname{argmin}} \frac{\frac{1}{n}||(\mathbf{I} - \mathbb{T}_{L_n})\mathbf{d}||^2}{\left[1 - \frac{1}{n}\mathrm{tr}\{\mathbb{T}_{L_n}\}\right]^2} \qquad \text{and} \qquad \bar{J}_n = \underset{J_n}{\operatorname{argmin}} \frac{\frac{1}{n}||(\mathbf{I} - \mathbb{T}^*_{J_n})\mathbf{d}||^2}{\left[1 - \frac{1}{n}\mathrm{tr}\{\mathbb{T}^*_{J_n}\}\right]^2}, \tag{4.8}$$

with $\mathbf{d}$ being the vector of left-hand-side variables.

Having obtained $\widehat{r}$, $\widehat{\mathbb{T}}$ and $\widehat{\mathbb{T}}^*$, we next proceed to the regularized estimation of unknown ψ. Pre-multiplying both sides of (4.3) by a tuning parameter c such that $c||\mathbb{T}^*\mathbb{T}|| < 1$, to ensure convergence of the iterative algorithm, we have

$$c\widehat{\mathbb{T}}^*\widehat{r} = c\widehat{\mathbb{T}}^*\widehat{\mathbb{T}}\psi, \tag{4.9}$$

from where, by adding and subtracting ψ, we obtain a recursive solution that provides a basis for the iterative algorithm:

$$\widehat{\psi}^{s+1} = \widehat{\psi}^s + c\widehat{\mathbb{T}}^*\left(\widehat{r} - \widehat{\mathbb{T}}\widehat{\psi}^s\right) \qquad \forall\ s = 0, 1, 2, \ldots. \tag{4.10}$$

The choice of c is inconsequential and, since the largest eigenvalue of $\mathbb{T}^*\mathbb{T}$ equals 1, any value of c less than 1 would ensure convergence (for details, see Centorrino et al., 2017). Effectively, c controls the size of an iterative step with the larger values resulting in fewer but coarser iterations and, in contrast, values close to 0 yielding small iterative updates which, however, may require an impractically long computational time. Following Florens et al. (2018), we set $c = 0.5$ to balance computational speed and precision. Also, to improve over the MSE of $\widehat{\psi}$ in the finite sample, we follow Centorrino et al.'s (2017) advice and update L_n and J_n for the estimation of operators $\mathbb{T}$ and $\mathbb{T}^*$ at each iteration. Specifically, the iterative algorithm is as follows:

1. Initiate the algorithm (for $s = 0$) with $\widehat{\psi}^0 = c\widehat{\mathbb{T}}^*\widehat{r}$ constructed using estimators in (4.5) and (4.7) with the parameters L_n and J_n cross-validated via (4.8) setting $\mathbf{d}$ to $\mathbf{y}$ and $\widehat{r}$, respectively.

2. For each $s = 0, 1, 2, \dots$, use the current estimate $\widehat{\psi}^s$ to update L_n and J_n to be used in the construction of $\widehat{\mathbb{T}}$ and $\widehat{\mathbb{T}}^*$ by cross-validating them again, but this time, setting $\mathbf{d}$ respectively equal to $\widehat{\psi}^s$ and $\left(\widehat{r} - \widehat{\mathbb{T}}\widehat{\psi}^s\right)$, as suggested by the recursive solution.
3. Update the estimate of ψ via $\widehat{\psi}^{s+1} = \widehat{\psi}^s + c\widehat{\mathbb{T}}^*\left(\widehat{r} - \widehat{\mathbb{T}}\widehat{\psi}^s\right)$ with $\widehat{\mathbb{T}}$ and $\widehat{\mathbb{T}}^*$ constructed using the newly cross-validated smoothing parameters.
4. Repeat steps (2)–(3) until the following residual-based convergence criterion is minimized in s $(s = 1, 2, \dots)$:
$$RSS(s) = s\left(\widehat{r} - \widehat{\mathbb{T}}\widehat{\psi}^s\right)'\left(\widehat{r} - \widehat{\mathbb{T}}\widehat{\psi}^s\right), \tag{4.11}$$
where we stop when it starts increasing.

Under relatively mild regularity conditions, the regularized estimator $\widehat{\psi}$ is consistent as $n \to \infty$ (e.g., see Johannes et al., 2013, Proposition 3.2) which helps ensure consistency of the least-squares sieve estimators in the second and third steps.

Step 2. As an important byproduct of estimating an additive function $\psi(\cdot) = f(\cdot) + \varphi(\cdot)$, we also obtain consistent estimates of the composite error term u_{it} which, in addition to the *i.i.d.* noise, includes the endogeneity-inducing unobservable productivity innovation ζ_{it}. With this information, we can now separably recover production function $f(\cdot)$ from (3.3) via least squares. More concretely, recognizing that the inconsistency of least-squares estimation of our main equation of interest (3.3) [or, equivalently, equation (3.8)] is in essence due to the presence of an "omitted variable" ζ_{it}, we can tackle this problem by proxying for the latter using the consistently estimated residuals $\widehat{u}_{it}$ from Step 1, which effectively play a role of the control function $u(\zeta_{it})$. Thus, we estimate $f(\cdot)$ via an additive nonparametric least-squares regression of y_{it} on $(K_{it}, L_{it}, M_{it})'$, $(K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})'$ and $\widehat{u}_{it}$. Such a regression is no longer subject to endogeneity because the formerly omitted variation in ζ_{it} is now being explicitly controlled for:

$$y_{it} = f(K_{it}, L_{it}, M_{it}) + \varphi(K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1}) + \widehat{u}(\zeta_{it}) + e_{it}, \tag{4.12}$$

where e_{it} is mean-independent of the regressors by construction.[14] Also note that, as an alternative, we can separably recover $f(\cdot)$ and $\varphi(\cdot)$ by regressing $\widehat{\psi}_{it}$ itself, which is now rid of the endogeneity-inducing correlation with ζ_{it}, on $(K_{it}, L_{it}, M_{it})'$ and $(K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})'$.

We estimate (4.12) using sieve methods because of the relative ease with which the separability between $f(\cdot)$ and $\varphi(\cdot)$ can be imposed. Specifically, we use the single-layer feed-forward artificial neural network (ANN) sieves to approximate $f(\cdot)$ and $\varphi(\cdot)$:

$$f(\cdot) \approx \alpha_{t,0} + \sum_{r \geq 1} \alpha_{t,r}\mathcal{A}\left(\gamma_{t,r}^0 + \gamma_{t,r}^k k_{it} + \gamma_{t,r}^l l_{it} + \gamma_{t,r}^m m_{it}\right) \tag{4.13a}$$

[14] Obviously, since $u_{it} = \zeta_{it} + \eta_{it}$, the proxy $\widehat{u}_{it}$ also includes the estimate of random noise η_{it}. However, owing to the *i.i.d.* nature of η_{it}, this should not affect the validity of our procedure because the inclusion of irrelevant regressors is not to jeopardize consistency of the estimator, although one would expect a larger variance.

$$\varphi(\cdot) \approx \beta_{t,0} + \sum_{d\geq 1} \beta_{t,d}\mathcal{A}\left(\delta^0_{t,d} + \delta^k_{t,d}k_{it-1} + \delta^l_{t,d}l_{it-1} + \delta^X_{t,d}X_{it-1} + \delta^m_{t,d}m_{it}\right), \tag{4.13b}$$

where $\mathcal{A}(\cdot)$ is the so-called "activation function", which can be any known function (except polynomial) of fixed finite degree (Hornik et al., 1989); $r = 1, \ldots, R_n$ and $d = 1, \ldots, D_n$, and both R_n and D_n increase slowly with the sample size n.

To better grasp the ANN sieve estimation procedure, consider the approximation of the production function $f(\cdot)$ in (4.13a). Here, we essentially relate "input units" (k_{it}, l_{it}, m_{it}), the so-called "neurons", to the "output unit" f by building a network in which input neurons send "signals" to R_n "hidden intermediate units", each of which produces an "activation" $\mathcal{A}\left(\gamma^0_{t,r} + \gamma^k_{t,r}k_{it} + \gamma^l_{t,r}l_{it} + \gamma^m_{t,r}m_{it}\right)$ that is then sent to the output unit f. In turn, the output unit treats these hidden intermediate units as input units, the single index of whose signals are then transformed into the output via the activation mapping $\mathcal{C}(\cdot)$, i.e., $f(\cdot) = \mathcal{C}\left(\alpha_{t,0} + \sum_{r\geq 1}\alpha_{t,r}\mathcal{A}\left(\cdot\right)\right)$. In this paper, we resort to the popular logistic specification of $\mathcal{A}(\cdot)$, i.e., $\mathcal{A}(x) = (1 + \exp\{-x\})^{-1}$, and set the activation mapping $\mathcal{C}(\cdot)$ to the identity link function. We specify a network with only a single layer of hidden intermediate units. Further, since the information travels strictly in one direction from input units to the output unit with no feedback loops, the network we build is said to be the "feed-forward network". For more details on the ANN sieves and their use in econometrics, we refer the reader to an excellent discussion in White (1989), Kuan & White (1994a) and Kuan (2008).

The single-layer feed-forward ANN structure with a sigmoid activation function $\mathcal{A}(\cdot)$ like the logistic and the identity map $\mathcal{C}(\cdot)$ that we use here can approximate any function to any desired accuracy with the use of a sufficiently large number of hidden intermediate units regulating the degree of approximation complexity (Hornik et al., 1989; White, 1989). We however note that other sieves such as Hermite polynomials or B-splines are also valid alternatives for the nonparametric estimation of $f(\cdot)$. Here, we choose the ANN sieves primarily due to the following three reasons. First, ANN approximators possess a superior ability to accommodate nonlinearities and non-separabilities between the elements of an approximated function (Hornik et al., 1989; Chen & Ludvigson, 2009). Second, other approximators such as linear tensor product Fourier sieves or spline sieves usually exhibit slower convergence rates requiring a higher degree of approximation complexity to attain comparable accuracy (Kuan, 2008; Chen & Ludvigson, 2009). Lastly, ANN sieves do not suffer from the "curse of dimensionality" in the same fashion as do nonparametric series approximators like splines or Fourier sieves. Specifically, the degree of ANN approximation depends only on the complexity of the network and not on the dimension of the covariate vector, whereas splines or Fourier sieves are subject to the "curse of dimensionality" on both of these fronts (Kuan & White, 1994b).

The approximated $f(\cdot) + \varphi(\cdot)$ are estimated via nonlinear least-squares.[15] A commonly used algorithm for "training" ANN networks, i.e., for the estimation of "weights" (parameters) in (4.13), is the recursive gradient-based error back-propagation, the mechanism of which in its spirit resem-

[15]Note that we can only identify the *sum* of $\alpha_{t,0}$ and $\beta_{t,0}$.

bles the learning process (White, 1989). However, gradient-based algorithms are meant for local searches and are likely to fail to find global extrema. Therefore, in the estimation, we make use of a more robust global-search simulated annealing algorithm, which has been found to outperform traditional back-propagation in training feed-forward neural networks (Goffe et al., 1994; Sexton et al., 1999). Just like earlier, we select smoothing parameters D_n and R_n via generalized cross-validation. Since the estimating equation is nonlinear (in parameters) in this case, we use the following generalized cross-validation function (also see Gao, 2007):

$$(\bar{D}_n, \bar{R}_n)' = \underset{(D_n,R_n)'}{\operatorname{argmin}} \frac{\frac{1}{n}||\widehat{\mathbf{e}}(D_n, R_n)||^2}{\left[1 - \frac{1}{n}\kappa(D_n, R_n)\right]^2}, \tag{4.14}$$

where $\widehat{\mathbf{e}}(\cdot)$ is the vector of residuals from (4.12), and $\kappa(\cdot)$ is the total number of parameters in the ANN approximators effectively substituting for the trace of a projection matrix. Thus, we have the estimator of the production function: $\widehat{f}(\cdot) = \sum_{r\geq 1} \widehat{\alpha}_{t,r}\mathcal{A}\left(\widehat{\gamma}^0_{t,r} + \widehat{\gamma}^k_{t,r}k_{it} + \widehat{\gamma}^l_{t,r}l_{it} + \widehat{\gamma}^m_{t,r}m_{it}\right)$.[16]

Step 3. Using the estimates of $f(\cdot)$ obtained in the second step, we next compute $\widehat{y}^*_{it} = y_{it} - \widehat{f}(\cdot)$ and proceed with the second-stage estimation of ω_{it} from (3.13), where $g(\cdot)$ is also approximated via the single-layer ANN sieve with the logistic activation function $\mathcal{A}(\cdot)$:

$$g(\cdot) \approx \theta_{t,0} + \sum_b \theta_{t,b}\mathcal{A}\left(\nu^0_{t,b} + \nu^k_{t,b}k_{it} + \nu^l_{t,b}l_{it} + \nu^X_{t,b}X_{it} + \nu^m_{t,b}m_{it}\right), \tag{4.15}$$

where $b = 1, \ldots, B_n$ and the number of hidden units $B_n \to \infty$ slowly as $n \to \infty$ (cross-validated as well). Here, we use the least-squares-like orthogonality conditions in (3.14), where the error term $\eta_{it} = \widehat{y}^*_{it} - g(\cdot)$ is constructed using the ANN approximation above. This produces estimates of productivity $\widehat{\omega}_{it} = \widehat{g}(\cdot)$.

Consistency and limit normality of the estimators in Steps 2 and 3 are expected to follow from the results in Hahn et al. (2018) and Chen et al. (2015).

Remark 4 Before applying our proposed estimator to the real data, we first examine our methodology in a small set of Monte Carlo experiments where we compare its performance to that of the traditional estimator. The results are encouraging, and simulation experiments show that our approach recovers the true parameters well, thereby lending strong support to the validity of our identification strategy. As expected of a consistent estimator, the estimation becomes more stable as the sample size grows. In contrast, consistent with Gandhi et al.'s (2017) critique, the under-identified traditional proxy estimator exhibits non-vanishing biases in the estimates of production function. To conserve space, we relegate the discussion of simulation experiments to Appendix D.

[16] The estimated intercept $\widehat{\alpha}_{t,0} + \widehat{\beta}_{t,0}$ is attributed in its entirety to the productivity proxy.

4.1 Bias-Corrected Bootstrap Inference

For inference, we rely on Efron's (1987) accelerated bias-corrected bootstrap percentile confidence intervals, which are generally 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 estimators via wild residual block bootstrap that takes into account a panel structure of the data, with all the steps bootstrapped jointly owing to a sequential nature of our estimation procedure. The bootstrap algorithm is described in Appendix C. We repeat it $B = 400$ times. We then use the empirical distribution of B bootstrap estimates of f_{it} and ω_{it} as well as the functionals thereof to construct accelerated bias-corrected percentile confidence intervals as described below.

Let the estimand of interest be denoted by $\widehat{Q}$, e.g., the mean returns to scale defined as the industry average of the firm-specific sums of partials of $\widehat{f}_{it}$. We use the empirical distribution of B bootstrap estimates $\{\widehat{Q}^1, \ldots, \widehat{Q}^B\}$ to estimate $(1-a)\times 100\%$ confidence bounds for $\widehat{Q}$ as intervals between the $a_1 \times 100$th and $a_2 \times 100$th percentiles of its bootstrap distribution with

$$a_1 = \Phi\left(\widehat{\phi}_0 + \big(\widehat{\phi}_0 + \phi_{a/2}\big)/\big[1 - \widehat{c}\big(\widehat{\phi}_0 + \phi_{a/2}\big)\big]\right) \tag{4.16}$$

$$a_2 = \Phi\left(\widehat{\phi}_0 + \big(\widehat{\phi}_0 + \phi_{(1-a/2)}\big)/\big[1 - \widehat{c}\big(\widehat{\phi}_0 + \phi_{(1-a/2)}\big)\big]\right), \tag{4.17}$$

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

$$\widehat{\phi}_0 = \Phi^{-1}\left(\#\{\widehat{Q}^b < \widehat{Q}\}/B\right) \tag{4.18}$$

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

$$\widehat{c} = \sum_{j=1}^{J}\left(\sum_{s=1}^{J}\widehat{Q}^s - \widehat{Q}^j\right)^3 \Bigg/ \left(6\left[\sum_{j=1}^{J}\left(\sum_{s=1}^{J}\widehat{Q}^s - \widehat{Q}^j\right)^2\right]^{3/2}\right), \tag{4.19}$$

where $\widehat{Q}^j$ is the $j(=1,\ldots,J)$th jackknife estimate of Q.[17] Note that both the acceleration and bias-correction factors are different for each estimator, denoted here generically by $\widehat{Q}$. That is, the bias-correction procedure is estimand-specific. Also, the estimated confidence intervals may not contain the original estimates if the finite-sample bias is large.

Remark 5 The justification of the described bootstrap procedure for our estimator is complicated given the complexity of our multi-step regularized estimation procedure, and establishing the va-

[17] 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.

lidity of such a bootstrap is not trivial and beyond the scope of our paper. We therefore call for caution in practical implementation. Having said that, we investigate performance of the outlined bootstrap procedure in Monte Carlo experiments in Appendix D. The simulations show a satisfactory performance of our bootstrap confidence intervals in finite samples. The results indicate that there may be size distortions for small n, which is common for nonparametric tests. However, for a sample size modestly large enough, the estimated coverage is close to the correct coverage. The intervals exhibit good power, which improves as $n \to \infty$ as anticipated of a consistent test.

5 Data

The data come from the Chinese Industrial Enterprises Database survey conducted by China's National Bureau of Statistics (NBS). This database covers all state-owned firms and all non-state-owned firms with sales above 5 million yuan (about 0.6 million in U.S. dollar). The firms in the database account for more than 90% of the gross industrial output value of the entire country. The covered industries include mining, manufacturing and public utilities. In this paper, we limit our analysis to manufacturing firms only. Specifically, we consider 28 two-digit China Standard Industrial Classification (CSIC) manufacturing industries, the list of which is provided in Table E.1 in Appendix E. The same appendix also contains additional details about our data.

To mitigate potential distortionary impacts of the 1997–1998 Asian financial crisis and the 2007–2008 global economic and financial turmoil on our results, we choose the sample period from 1999 to 2006. The raw data start with about 160,000 firms in 1999 and grow to about 300,000 firms in 2006: a total of 615,652 unique firms.

The firm's capital stock is defined as the self-reported net fixed asset deflated by the price index of investment in fixed assets. Labor is measured as the total wage bill plus benefits deflated by the GDP deflator. Materials are defined as the total intermediate inputs, including raw materials and other production-related inputs, deflated by the purchasing price index for industrial inputs. The output is defined as the gross industrial output value which, in line with (2.5), we deflate using an export-intensity-weighted average of domestic consumer and export price indices available from China Statistical Yearbooks. The price indices are obtained from the NBS and the World Bank. Thus, the four variables are measured in thousands of real RMB. We exclude observations with missing values for these key variables entering the firm's production function. Following Guariglia et al. (2011), we also truncate the top and the bottom 0.5th percentiles of these four variables to rule out outliers, including the observations with negative values. The operable sample includes 328,130 unique firms with a total of 1,286,530 observations.

We measure the firm's degree of export orientation using export intensity defined as the ratio of the firm's reported total export value to its total output value. The variable is bounded and lies between zero and one by construction. Figure 1(a) plots a histogram of export intensity for the entire sample which, as expected, is distinctly bimodal indicating that the industry is dominated by the firms with predominantly domestic or foreign orientation. Figure 1(b) on the right plots the

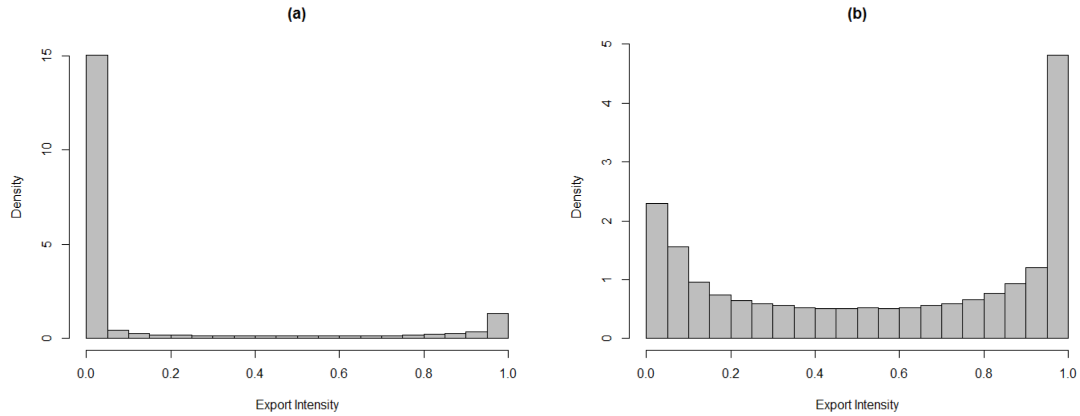

Figure 1. Empirical Distribution of Export Intensity, 1999–2006:
(a) All Firms; (b) Exporters Only

distribution exclusively for exporting firms.

We classify each firm as state-invested and/or foreign-invested if its equity includes state and/or foreign capital. We also have information if the firm is reported to have received state subsidy. Lastly, we define exporter firms as those with the reported export intensity greater than zero. See Table E.2 in Appendix E for summary statistics of our data and a brief discussion thereof.

6 Results

This section reports the results on firm productivity for 28 two-digit CSIC manufacturing industries in China. The nonparametric productivity estimates are obtained via the three-step estimation procedure outlined in Section 4. Also note that, since our identification strategy requires the use of first-order lags, in what follows we report the results for the 2000–2006 period, where the data for 1999 are used up to proxy for latent productivity in the year 2000.

We first take a look at the returns to scale and scale efficiency in Chinese Manufacturing. The discussion of these results is relegated to Appendix F. In what follows, we proceed to the analysis of firm-level productivity, which is the primary focus of our paper.

6.1 Aggregate Productivity Growth

We construct the estimates of unobserved firm-level productivity (in logs) using the definition of the gross production function (2.1), i.e., $\mathrm{p}_{it} = \widehat{\omega}_{it} + \widehat{\eta}_{it}$, where $\widehat{\omega}_{it}$ and $\widehat{\eta}_{it}$ are the third-step estimates from (3.13) and (4.15). Thus, in what follows, we analyze the composite firm-level productivity defined as the sum of both the persistent and random productivity. The left column of Table 1 reports estimates of the annual growth in the weighted-average aggregate productivity p_t for the entire manufacturing sector (all firms pooled across industries in our sample). Here, the aggregate productivity is computed as the firm-output-weighted average of firm-level productivity measures

Table 1. (Weighted) Aggregate Productivity Growth Decomposition for all Industries

Year	$\Delta \mathrm{p}_t$	$\Delta \overline{\mathrm{p}}_t$	$\Delta \mathrm{cov}_t$
2001	0.004	0.006	–0.002
2002	0.008	0.014	–0.006
2003	0.044	0.041	0.002
2004	0.064	0.061	0.003
2005	0.060	0.062	–0.002
2006	0.051	0.047	0.005
Cum.	0.231	0.230	0.001

Notes: Reported are the annual growth rates of the (weighted) aggregate productivity p_t and of its two components: the (unweighted) average productivity $\overline{\mathrm{p}}_t$ and the covariance between the firm-level output and productivity cov_t.

for each year, i.e.,

$$\mathrm{p}_t = \sum_i \varpi_{it} \mathrm{p}_{it}, \quad \text{where} \quad \varpi_{it} \equiv Y_{it} / \sum_j Y_{jt} \quad \forall\ t, \tag{6.1}$$

with the corresponding annual aggregate productivity growth rates computed as the log difference, i.e., $\Delta \mathrm{p}_t = \mathrm{p}_t - \mathrm{p}_{t-1}$.

The average annual productivity growth rate in China's manufacturing sector from 2000 to 2006 was around 3.85%. This included a period of unusually slow growth (2001–2002) following which the aggregate productivity growth picked up in 2003 and was steadily in the 4.4–6.4% range. The average post-2002 trend in the aggregate productivity *level* in our sample was fairly stable. We observe a similar pattern across most individual two-digit CSIC industries, as can be seen in Figure 2 which plots the estimated aggregate productivity indices for all 28 individual industry groups[18] along with that for all industries combined. From Figure 2, it is also apparent that productivity growth has not been uniform across industries. Depending on the industry, the cumulative growth (over the 6-year period) in aggregate productivity ranges from 3.1% (Communication and Computer Equipment) to 41.1% (Nonferrous Metals) with the cumulative increase in the all-industry weighted average amounting to 23.1%.

The documented growth in the (weighted) aggregate productivity can be attributed to two primary sources: *(i)* a secular increase in the average productivity across firms in the industry and *(ii)* the reallocation of fixed factors towards more productive firms which would enable the latter to produce more output. To differentiate between these two sources, we decompose the growth in the aggregate productivity (i.e., $\Delta \mathrm{p}_t$) into two components à la Olley & Pakes (1996). To begin,

[18] Computed in a fashion analogous to that defined in (6.1) using data on firms in a given industry only.

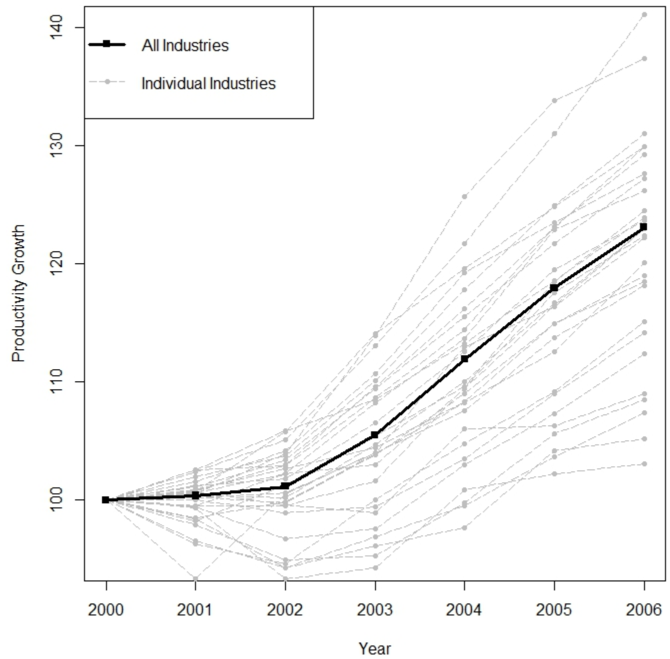

Figure 2. (Weighted) Aggregate Productivity Indices across Industries

we first decompose the aggregate productivity p_t itself:

$$\begin{aligned}\mathrm{p}_t &= \sum_i \varpi_{it}\mathrm{p}_{it} = \sum_i \left[\overline{\varpi}_t + (\varpi_{it} - \overline{\varpi}_t)\right]\left[\overline{\mathrm{p}}_t + (\mathrm{p}_{it} - \overline{\mathrm{p}}_t)\right] \\ &= \overline{\mathrm{p}}_t + \sum_i (\varpi_{it} - \overline{\varpi}_t)(\mathrm{p}_{it} - \overline{\mathrm{p}}_t) \equiv \overline{\mathrm{p}}_t + \mathrm{cov}_t \quad \forall\ t,\end{aligned} \tag{6.2}$$

where $\overline{\mathrm{p}}_t = 1/n_t \sum_i \mathrm{p}_{it}$ and $\overline{\varpi}_t = 1/n_t$ are the unweighted average productivity and unweighted output share (a uniform weight), respectively. According to the above decomposition, the aggregate productivity p_t is a sum of the (unweighted) average of firm-level productivity $\overline{\mathrm{p}}_t$ and a sample covariance between the firm-level output and productivity $\mathrm{cov}_t \equiv \sum_i (\varpi_{it} - \overline{\varpi}_t)(\mathrm{p}_{it} - \overline{\mathrm{p}}_t)$. From (6.2), it immediately follows that $\Delta\mathrm{p}_t = \Delta\overline{\mathrm{p}}_t + \Delta\mathrm{cov}_t\ \forall\ t$, where the average productivity growth $\Delta\overline{\mathrm{p}}_t$ represents a secular change in productivity capturing temporal shifts in the productivity distribution via the change in its first moment, and $\Delta\mathrm{cov}_t$ measures the reallocation of market share from less productive to more productive firms thus capturing "reshuffling" within the joint distribution of the productivity and market share. The larger the covariance term, the larger the output share of more productive firms in the industry.

The decomposition results for the entire manufacturing sector are presented in Table 1 (columns 2 and 3). The empirical evidence indicates that the measured increase in the weighted aggregate productivity in 2000–2006 can be attributed entirely to an increase in the average productivity $\overline{\mathrm{p}}_t$. The reallocation of resources towards more productive firms has been largely anemic and, in fact, in

Table 2. Median (Log) Productivity Differentials by Category, 2000–2006

CSIC	Export	State	Foreign	Subsidy	CSIC	Export	State	Foreign	Subsidy
13	–0.020	–0.254	–0.039	–0.056	27	0.037	–0.150	0.086	0.056
14	0.012	–0.285	0.021	0.040	28	–0.068	–0.192	–0.050	0.052
15	0.058	–0.198	0.110	0.056	29	–0.045	–0.149	*0.004*	–0.039
16	0.249	0.129	*0.103*	0.279	30	–0.080	–0.177	–0.013	*0.009*
17	–0.073	–0.120	–0.042	*–0.002*	31	0.043	–0.210	0.054	–0.014
18	–0.129	–0.076	–0.012	–0.011	32	*–0.006*	–0.192	0.064	*0.006*
19	–0.158	–0.088	–0.013	*–0.008*	33	–0.037	–0.195	*–0.016*	0.063
20	–0.090	–0.241	–0.081	–0.048	34	–0.089	–0.203	*–0.002*	–0.017
21	–0.079	–0.282	–0.030	–0.048	35	–0.041	–0.244	0.048	*0.000*
22	*–0.001*	–0.174	0.022	–0.012	36	*0.004*	–0.295	0.082	*–0.005*
23	0.076	–0.278	0.094	0.025	37	0.032	–0.197	0.094	0.026
24	–0.101	–0.170	0.022	*0.001*	39	–0.060	–0.165	0.018	0.031
25	*0.022*	–0.147	0.109	*0.030*	40	–0.037	–0.103	0.059	0.039
26	–0.009	–0.210	0.033	*0.000*	41	–0.060	–0.213	0.080	*0.008*
All	–0.063	–0.180	*0.001*	0.004					

Notes: Reported are the differences in the unconditional median productivity estimates of exporters vs. non-exporters, state-invested vs. wholly privately owned, foreign-invested vs. wholly domestically owned, subsidized vs. non-subsidized, respectively. All point estimates, except those in *italic*, are statistically significant at the 5% level.

some years has contributed negatively, albeit negligibly, towards the productivity growth in China's manufacturing. Cumulatively over our sample period, we find that the aggregate productivity growth is entirely (23.0 out of 23.1% points) due to a rightward shift in the productivity distribution reflective of an increase in the average firm productivity. This relative insignificance of the cross-firm reallocation effect is especially interesting given multiple findings to the contrary for other countries (e.g., Bernard et al., 2003; Bernard & Jensen, 2004). The unimportance of reallocation for the productivity growth during our sample period may however be unique to the Chinese manufacturing sector featuring a sizable presence of state-owned/invested firms that might be able to shelter their market shares despite being significantly less productive (as we do document below) by, say, relying on subsidies. For instance, similar findings pertaining to the Chinese manufacturing have been documented by Hashiguchi (2015) who also reports negligible and negative reallocation (or misallocation, in this case) effects. Relatedly, Brandt et al. (2013) also document a misallocation of factors of production in China.

Since pooling all industries together may mask heterogeneous experiences of individual industry groups, we also conduct a similar productivity decomposition for each individual two-digit CSIC industry. These additional results are reported in Appendix F.

6.2 (Unconditional) Productivity Differentials

Table 2 reports the median estimates of raw (unconditional) productivity differentials across some firm types. Specifically, we compare the median productivity estimates for firms along the following indicators: whether a firm is an exporter, state-invested, foreign-invested or subsidized. We opt to look at the *median* differentials in order to minimize distortionary effects of outliers.[19] Consistent with one's expectations, we find that the state-invested firms tend to be less productive than wholly privately owned firms with no state capital across all industries except one, with the pooled all-industries median differential of –18%. When examining firms based on whether they are foreign-invested or subsidized, the evidence on productivity differentials is more mixed, highlighting heterogeneity across individual industries. We find that, in 16 out of 28 manufacturing industries including such major recipients of inbound foreign direct investment like Computer Equipment or Stationery, foreign-invested firms are statistically significantly more productive, at the median, than their wholly domestically owned counterparts. In 8 industries that includes the also heavily foreign-invested Apparel Production, the foreign-investment productivity premium is actually statistically negative, with the rest of industries exhibiting no significant productivity differentials across foreign and domestic firms. The corresponding pooled all-industries-combined median premium is estimated at statistically insignificant 0.1%. Similarly, we document mixed findings about the median productivity differential by the subsidy tabulation: out of 28 industries, 8/10/10 exhibit statistically negative/positive/zero productivity premium.

Of more interest are the results about productivity differentials by the exporter status of the firm. While there is some heterogeneity in exporter/non-exporter productivity differential estimates across industries, the overall finding is that exporting firms tend to exhibit statistically *lower* productivity than non-exporting firms, at least at the unconditional median. We observe this in 17 industries, with the unconditional exporter productivity differential ranging from –0.9 to –15.8%. Overall, the pooled median exporter premium estimate across all industries is –6.3%. The negative exporter productivity differential may seem puzzling given the widely held view among economists, whereby exporting firms are usually *more* productive than their domestically oriented counterparts because they are exposed to tougher competition, which induces them to raise their productivity, as well as they are more likely to enjoy productivity gains due to learning, quality and variety effects, and the absorption of new technologies from abroad (Delgado et al., 2002; Pavcnik, 2002; Melitz, 2003; Amiti & Konings, 2007; De Loecker, 2007, 2013). This seems to not apply to China's manufacturing sector. Similar findings of *lower* productivity levels exhibited by exporters in China have also been reported by Lu et al. (2010), Lu (2010) and Dai et al. (2016), who provide various plausible explanations for the phenomenon although their approaches to measuring productivity differ from ours. In what follows, we therefore take a closer look at the exporter productivity differential in China.

6.3 Conditional Exporter Productivity Differentials

Until now, the presented findings about the largely negative exporter productivity premium in China's manufacturing sector have relied on the comparison of *unconditional* medians of firm-level productivity. Clearly, to allow for a more meaningful analysis of exporter productivity differentials,

[19] (Unconditional) mean differentials suggest qualitatively similar findings.

Table 3. Median (Log) Exporter Productivity Premia
by Ownership Type, 2000–2006

CSIC	*Privately Owned*	*State-Invested*	*Foreign-Invested*	CSIC	*Privately Owned*	*State-Invested*	*Foreign-Invested*
13	–0.114	–0.099	–0.082	27	–0.085	–0.117	–0.085
14	–0.112	–0.074	–0.064	28	–0.091	*–0.022*	–0.048
15	–0.066	*–0.014*	–0.042	29	–0.145	–0.086	–0.101
16	*–0.050*	*–0.064*	*0.126*	30	–0.158	–0.048	–0.136
17	–0.119	–0.098	–0.112	31	–0.085	–0.103	–0.077
18	–0.156	–0.126	–0.135	32	–0.118	–0.140	–0.085
19	–0.166	–0.113	–0.145	33	–0.159	–0.138	–0.128
20	–0.120	*–0.031*	–0.081	34	–0.155	–0.104	–0.108
21	–0.170	–0.137	–0.160	35	–0.132	–0.086	–0.123
22	–0.089	–0.047	–0.084	36	–0.125	–0.097	–0.110
23	–0.129	*–0.025*	–0.104	37	–0.109	–0.079	–0.098
24	–0.165	–0.134	–0.149	39	–0.163	–0.112	–0.138
25	–0.123	–0.261	*–0.023*	40	–0.173	–0.072	–0.167
26	–0.118	–0.088	–0.084	41	–0.160	–0.079	–0.162
All	–0.148	–0.115	–0.126				

Notes: Reported are the conditional median estimates of the exporter productivity premium tabulated by three equity type categories: wholly privately owned, state-invested and foreign-invested firms. All point estimates, except those in *italic*, are statistically significant at the 5% level.

it is imperative to also account for cross-firm differences in characteristics other than the exporter status alone, which would enable us to study the exporter productivity premium *conditional* on relevant firm controls.

Ownership Type. We first examine the exporter productivity differentials conditional on the firm's equity type. Specifically, for each two-digit CSIC industry group in our sample, we estimate the following median regression:

$$\begin{aligned}\mathbb{Q}_{0.5}\left[\mathrm{p}_{it}|\cdot\right] = {} & \alpha_0 + \alpha_1\mathrm{EXP}_{it} + \alpha_2\mathrm{STATE}_{it} + \alpha_3\mathrm{FOREIGN}_{it} + \\ & \alpha_{12}\mathrm{EXP}_{it}\times\mathrm{STATE}_{it} + \alpha_{13}\mathrm{EXP}_{it}\times\mathrm{FOREIGN}_{it} + \\ & \alpha_4\mathrm{SUBSIDY}_{it} + \alpha_5\log(Y_{it}) + \alpha_6\log(Y_{it})^2 + \lambda_t,\end{aligned}$$

where EXP_{it}, STATE_{it}, $\mathrm{FOREIGN}_{it}$ and $\mathrm{SUBSIDY}_{it}$ are indicator variables for the "exporter", "state-invested", "foreign-invested" and "subsidized" types of firms, respectively. A wholly privately-owned domestic firm is our reference group. In addition to time fixed effects λ_t, we also include the log (and its square) of output to proxy for the firm size. Instead of the traditional mean regression, we estimate conditional median regressions to minimize distortionary impacts of outliers to which quantiles are fairly robust. Based on the results from these regressions, we compute the (conditional) median exporter productivity premiums for each firm type as defined by the tabulation of indicators for the exporter status and the state and foreign participation in firm equity.

Table 3 reports these premium estimates, with most being statistically significant at the 5%

level. The major takeaway here is that, once we control for firm characteristics, exporters now almost universally exhibit a statistically negative productivity premium across all manufacturing industries no matter their ownership type. Among the 24 industries, Tobacco is the only one for which we find no statistically significant evidence of the negative exporter productivity differential. Pooling all industries together (but controlling for industry dummies), the conditional median exporter productivity differential is estimated at –14.8, –11.5 and –12.6% for wholly privately-owned, state-invested and foreign-invested firms, respectively.

Export Intensity. We next explore potential heterogeneity in the exporter productivity premium along the intensive margin of exporting, whereby we differentiate firms based on their varying degree of export intensity. To rationalize China's exporter productivity puzzle, Dai et al. (2016) argue in favor of differentiating between the "ordinary" and "processing" exports, where the latter is defined as the assembly of tariff-exempted imported inputs into final goods for the purpose of subsequent export to foreign markets. Given the exemption from input tariffs and the high labor intensity of the assembly process, such processing exporters may be expected to be less productive than their domestically oriented counterparts. Once these processing exporter firms (which also tend to be foreign-invested) are accounted for, the productivity puzzle should be resolved: ordinary exporters are then found to be more productive (Dai et al., 2016). While, due to data limitations, we cannot differentiate between ordinary and processing exporter firms, we are however able to indirectly capture this difference by differentiating between the exporters of lower and higher intensity. The argument here is as follows. Processing exporter firms' export intensity is reasonably expected to be significantly higher than that of ordinary exporters who tend to sell in both the foreign and domestic markets as opposed to being fully foreign-market-oriented. If the negative exporter productivity differential is primarily driven by processing exporters, one would then expect to see it largely vanish as the firm's export intensity decreases. To investigate this, we estimate the following median regression for each industry:

$$\begin{aligned}\mathbb{Q}_{0.5}\left[\mathrm{p}_{it}|\cdot\right] = \gamma_0 + \gamma_1\mathrm{STATE}_{it} + \gamma_2\mathrm{FOREIGN}_{it} + \gamma_3\mathrm{SUBSIDY}_{it} + \\ \sum_q \gamma_{4q}\mathrm{D}^X_{q,it} + \gamma_5\log(Y_{it}) + \gamma_6\log(Y_{it})^2 + \lambda_t,\end{aligned}$$

where $\mathrm{D}^X_{q,it} \equiv \mathbb{1}\left\{\mathbb{Q}_{0.1(q-1)}[X|X>0] < X_{it} \le \mathbb{Q}_{0.1q}[X|X>0] \;\middle|\; X>0\right\}$ $\forall\ q = 1,\ldots,10$ is the indicator variable that takes a unit value if the *exporter* firm's export intensity falls between the $(q-1)$th and qth deciles of the empirical distribution of $X|X>0$. Note that, since these indicator variables take zero values for non-exporting firms by construction, all 10 of them are included in the regression. Thus, non-exporters serve as our reference group.

The corresponding exporter productivity differential estimates are reported in Table F.3 of Appendix F. Most estimates are significantly different from zero at the 5% level. With the sole exception of Tobacco industry just like earlier, we fairly consistently find an increasingly negative productivity differential between non-exporters and exporter with the higher degree of export ori-

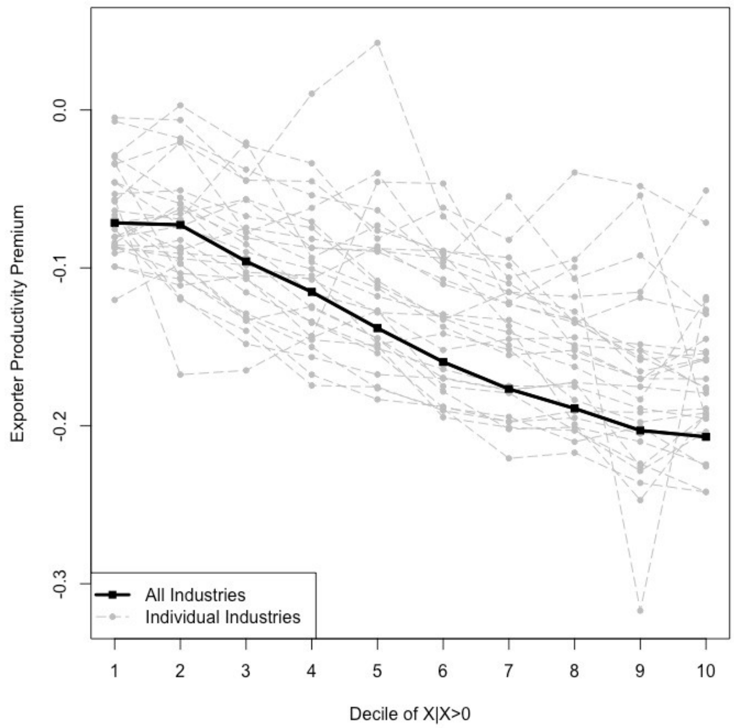

Figure 3. Median (Log) Exporter Productivity Premia by Export Intensity Deciles

entation. This can be seen most prominently by looking at the results for the whole manufacturing sector (the "All" row at the bottom of Table F.3). The (statistically significant) pooled exporter productivity premium estimates decline from –7.1% for the first decile to –20.7% for the tenth decile of the $X|X > 0$ distribution. Notably, the exporter productivity premium is negative for exporters along the entire distribution of export intensity, indicating that even the likely-to-be ordinary exporters from the bottom deciles are less productive than the wholly domestically-oriented non-exporters. At the disaggregated industry level, the evidence however is a bit less clear-cut, with few premium estimates being statistically insignificant for exporters from the left tail of the intensity distribution. But more broadly, the disaggregate data continue to lend support to the premise of high-intensity exporters, which are more likely to be engaged in processing exports, being less productive than their low-intensity counterparts. Although, we largely fail to find empirical evidence of low-export-intensity firms (that are more likely to partake in ordinary exports) exhibiting a positive productivity differential over non-exporters. At most, some exporters of this kind may be only as productive as their domestically-oriented counterparts. Our findings are succinctly summarized in Figure 3 that plots the estimated exporter premiums for all individual industry groups.

7 Conclusion

Motivated by the longstanding interest of economists in understanding the nexus between firm productivity and export behavior, we develop a novel structural framework for control-function-based nonparametric identification of the gross production function and latent firm productivity in the presence of endogenous export opportunities that is robust to recent unidentification critiques of proxy estimators. We generalize the standard behavioral assumptions about the firm-level production customarily assumed in the literature to explicitly formalize firms' endogenous export decisions along with their potential learning-by-exporting effects on productivity. We model exports in an agnostic "input-allocation" paradigm in which the firm chooses the degree of its export orientation along with inputs essentially letting us conceptualize the change in the former as another "investment". We show that our structural framework provides a workable identification strategy, where the firm's degree of export orientation is shown to provide the needed (excluded) relevant independent exogenous variation in endogenous freely varying inputs, thereby allowing us to successfully identify the model.

In order to avoid model misspecification, our methodology employs the nonparametric formulation not only for the control function but also for the production process itself. We estimate our fully nonparametric IV model using the Landweber-Fridman regularization (to tackle the ill-posedness) with the unknown functions approximated via artificial neural network sieves with a sigmoid activation function which are known for their superior performance relative to other popular sieve approximators, including the polynomial series favored in the literature. We first study our methodology in a small set of Monte Carlo experiments. The results are encouraging, and simulation experiments show that our approach recovers the true parameters well, thereby lending strong support to the validity of our identification strategy. As expected of a consistent estimator, the estimation also becomes more stable as the sample size grows. We then apply our methodology to the data on manufacturing firms from 28 industries in China during the 1999–2006 period to take a fresh look at China's exporter productivity puzzle.

Appendix

A Under-identification Issues

Gandhi et al. (2017) show that the traditional control-function-based production function estimators, which utilize freely varying inputs (such as materials M_{it}) to proxy for unobserved latent productivity, suffer from an inherent under-identification. As shown in Section 3, our estimator is immune to this problem owing to our ability to employ the firm's export intensity X_{it} as an additional source of *relevant independent* exogenous variation from outside the production function to help us identify the latter. In what follows, we show why our model would have been under-identified *if* we had adopted the standard conceptual framework by ignoring the firm's export decisions altogether. In that case, *(i)* the persistent productivity would evolve according to the exogenous Markov process with no learning by exporting effects whereby $\omega_{it} = \mathbb{E}[\omega_{it}|\ \omega_{it-1}] + \zeta_{it}$, and *(ii)* the conditional material demand function would be given by $M_{it} = \mathbb{M}_t(K_{it}, L_{it}, \omega_{it})$. The analogue of the production function in (3.3) is then:

$$y_{it} = f(K_{it}, L_{it}, M_{it}) + \varphi(K_{it-1}, L_{it-1}, M_{it-1}) + \zeta_{it} + \eta_{it}. \tag{A.1}$$

Next, note that the endogenous M_{it} entering the production function $f(\cdot)$:

$$M_{it} = \mathbb{M}_t(K_{it}, L_{it}, \omega_{it}) = \mathbb{M}_t(K_{it}, L_{it}, \varphi(K_{it-1}, L_{it-1}, M_{it-1}) + \zeta_{it}), \tag{A.2}$$

is a function of the following observables $(K_{it}, L_{it}, K_{it-1}, L_{it-1}, M_{it-1})$ and the unobservable ζ_{it}. Comparing these arguments of M_{it} with the variables entering $f(\cdot)$ directly as well as appearing inside the proxy for productivity $\varphi(\cdot)$, it is evident that the only extra source of variation for M_{it}, which has not already been included on the right-hand side of (A.1), is the unobservable ζ_{it}. In other words, conditional on the already included self-instrumenting variables $(K_{it}, L_{it}, K_{it-1}, L_{it-1}, M_{it-1})$, there is *no* other relevant exogenous variable that may be used to instrument for the endogenous M_{it} because, for it to be relevant in predicting M_{it}, it would have to correlate with ζ_{it}, which is the only source of "free" variation left in M_{it}. The correlation with ζ_{it} would however violate the exogeneity requirement thereby invaliding the instrument. Notably, this conundrum also applies to (weakly exogenous) excluded lags (of the second and higher order) of inputs which, following the bulk of the literature, one may be tempted to utilize in hopes of identifying the model. As evident from (A.2), any such additional (excluded) lag is *ir*relevant for predicting M_{it} once conditioned on the already included $(K_{it}, L_{it}, K_{it-1}, L_{it-1}, M_{it-1})$. Therefore, M_{it} still lacks an excluded relevant instrument from outside of the equation, and the production function $f(\cdot)$ in (A.1) is unidentified under the *standard* structural assumptions.

B Separable Identifiability Issues

Ackerberg et al.'s (2015) critique of control-function-based production function estimators effectively boils down to one's potential inability to *separably* identify the production and productivity proxy functions due to the presence of endogenous freely varying inputs inside the production function. This problem occurs primarily when *(i)* labor is assumed to be a freely varying input (like M_{it}) thereby becoming endogenous and *(ii)* one attempts to "control" for contemporaneous ω_{it} as opposed to lagged ω_{it-1}, as originally done by Olley & Pakes (1996) and Levinsohn & Petrin (2003). If we were to go this route, *(i)* the conditional material demand function would no longer be a function of L_{it}, i.e., $M_{it} = \mathbb{M}_t(K_{it}, X_{it}, \omega_{it})$, and *(ii)* unobserved ω_{it} would be proxied by the contemporaneous inverted demand for materials without making use of its Markovian nature. In this case, the analogue of the production function in (3.3) is given by

$$\begin{aligned} y_{it} &= f(K_{it}, L_{it}, M_{it}) + \omega_{it} + \eta_{it} \\ &= f(K_{it}, L_{it}, M_{it}) + \mathbb{M}_t^{-1}(K_{it}, X_{it}, M_{it}) + \eta_{it} \\ &= f(K_{it}, \mathbb{L}_t(K_{it}, X_{it}, \omega_{it}), M_{it}) + \mathbb{M}_t^{-1}(K_{it}, X_{it}, M_{it}) + \eta_{it} \\ &= f\left(K_{it}, \mathbb{L}_t\left(K_{it}, X_{it}, \mathbb{M}_t^{-1}(K_{it}, X_{it}, M_{it})\right), M_{it}\right) + \mathbb{M}_t^{-1}(K_{it}, X_{it}, M_{it}) + \eta_{it} \\ &= g_t(K_{it}, X_{it}, M_{it}) + \mathbb{M}_t^{-1}(K_{it}, X_{it}, M_{it}) + \eta_{it}, \end{aligned} \tag{B.1}$$

where we have made use of the firm's conditional demand for now freely varying labor $L_{it} = \mathbb{L}_t(K_{it}, X_{it}, \omega_{it})$ in the third line. From (B.1), it is obvious that the two unknown functions $g_t(\cdot)$ and $\mathbb{M}_t^{-1}(\cdot)$ are *not* separably identified since they depend on the same covariates.

However, this separable unidentifiability problem no longer applies if, instead of ω_{it}, one were to proxy for ω_{it-1} (like we do in our paper). Specifically, from

$$\begin{aligned} y_{it} &= f(K_{it}, L_{it}, M_{it}) + h_t[\omega_{it-1}, X_{it-1}] + \zeta_{it} + \eta_{it} \\ &= f(K_{it}, L_{it}, M_{it}) + \varphi(K_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it} + \eta_{it} \\ &= f(K_{it}, \mathbb{L}_t(K_{it}, X_{it}, \omega_{it}), \mathbb{M}_t(K_{it}, X_{it}, \omega_{it})) + \varphi(K_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it} + \eta_{it} \\ &= f\left(K_{it}, \mathbb{L}_t\left(K_{it}, X_{it}, \varphi(K_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it}\right), \mathbb{M}_t\left(K_{it}, X_{it}, \varphi(K_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it}\right)\right) + \\ &\quad \varphi(K_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it} + \eta_{it} \end{aligned} \tag{B.2}$$

it is evident that, conditional on exogenous covariates $(K_{it}, X_{it}, K_{it-1}, X_{it-1}, M_{it-1})$, the production function $f(\cdot)$ is separably identified from the proxy function $\varphi(\cdot)$ because both endogenous inputs L_{it} and M_{it} still vary independently from all arguments of $\varphi(\cdot)$ owing to the presence of ζ_{it}.

Based on the above results, it might appear that, in our model, we too could have allowed labor to be freely varying. Unfortunately, while our model would still have been immune to Ackerberg et al.'s (2015) critique in this case, we would however have faced the under-identification problem à la Gandhi et al. (2017) not only with respect to M_{it} but also L_{it}. Since we only have *one* valid excluded instrument (X_{it}), we generally would have been unable to identify our model in the presence of *two* endogenous freely varying inputs due to the failure to meet the order condition. We are therefore bound to the assumption of quasi-fixity of labor.

C Bootstrap Algorithm

We approximate sampling distributions of the estimators via wild residual block bootstrap that takes into account a panel structure of the data, with all the steps bootstrapped jointly owing to a sequential nature of our estimation procedure. To make matters more concrete, we use the following wild residual panel bootstrap algorithm:

1. Compute the three steps of our estimation procedure using the original data. Denote the obtained estimates as $\widehat{\psi}_{it} = \widehat{\psi}(K_{it}, L_{it}, M_{it}, K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$ in the first step, $\widehat{f}_{it} = \widehat{f}(K_{it}, L_{it}, M_{it})$ and $\widehat{\varphi}_{it} = \widehat{\varphi}(K_{it-1}, L_{it-1}, X_{it-1}, M_{it-1})$ in the second step, and $\widehat{\omega}_{it} = \widehat{g}(K_{it}, L_{it}, X_{it}, M_{it})$ in the third step, for all $i = 1, \ldots, n$ and $t = 1, \ldots, T$. Let the residuals from each of these three sequential steps be respectively denoted by $\{\widehat{u}_{it} = y_{it} - \widehat{\psi}_{it}\}$, $\{\widehat{e}_{it} = y_{it} - \widehat{f}_{it} - \widehat{\varphi}_{it} - \widehat{u}_{it}\}$ and $\{\widehat{\eta}_{it} = y_{it} - \widehat{f}_{it} - \widehat{\omega}_{it}\}$. Recenter the residuals.

2. Generate bootstrap weights ξ_i^b for all cross-sectional units $i = 1, \ldots, n$ from the Mammen (1993) two-point mass distribution:

$$\xi_i^b = \begin{cases} \frac{1+\sqrt{5}}{2} & \text{with prob. } \frac{\sqrt{5}-1}{2\sqrt{5}} \\ \frac{1-\sqrt{5}}{2} & \text{with prob. } \frac{\sqrt{5}+1}{2\sqrt{5}}. \end{cases} \tag{C.1}$$

   Next, for each observation (i, t) with $i = 1, \ldots, n$ and $t = 1, \ldots, T$, generate a new bootstrap first-step disturbance: $u_{it}^b = \xi_i^b \widehat{u}_{it}$.

3. Generate a new bootstrap outcome variable. From the first step, we have $y_{it}^b = \widehat{\psi}_{it} + u_{it}^b$ for all $i = 1, \ldots, n$ and $t = 1, \ldots, T$. Recompute the first-step estimator using $\{y_{it}^b\}$ in place of $\{y_{it}\}$ and denote the obtained estimates as $\{\widehat{\psi}_{it}^b\}$ with the corresponding residuals $\{\widehat{u}_{it}^b = y_{it} - \widehat{\psi}_{it}^b\}$. Use these residuals in the next step of the algorithm.

4. Recompute the second-step estimator using $\{y_{it}^b\}$ in place of $\{y_{it}\}$ and $\{\widehat{u}_{it}^b\}$ in place of $\{\widehat{u}_{it}\}$. Denote the obtained estimates as $\{\widehat{f}_{it}^b\}$ and $\{\widehat{\varphi}_{it}^b\}$. Use these estimates in the next step of the algorithm.

5. Recompute the third-step estimator using $\{y_{it}^b\}$ in place of $\{y_{it}\}$ and $\{\widehat{f}_{it}^b\}$ in place of $\{\widehat{f}_{it}\}$ and denote the obtained estimates as $\{\widehat{\omega}_{it}^b\}$.

6. Repeat steps 2 through 5 of the algorithm B times. Use the empirical distribution of B bootstrap estimates of f_{it} and ω_{it} as well as the functionals thereof to construct accelerated bias-corrected percentile confidence intervals as described in Section 4.1.

D Monte-Carlo Experiments

In this section, we describe Monte Carlo experiments that evaluate the performance of our proposed estimator and demonstrate its ability, owing to its use of the excluded exogenous variation in the firm's degree of export orientation, to successfully identify the production function. We then compare its performance to that of the traditional proxy estimator à la Levinsohn & Petrin (2003) based on the lagged instrumentation of endogenous materials. In addition, we also investigate the finite-sample performance of our proposed procedure for bootstrap inference.

Our data generating process (DGP) is essentially a fusion of those used by Grieco et al. (2016) and Gandhi et al. (2017). More specifically, we consider a balanced panel of $n = \{50, 100, 200, 400, 800\}$ firms operating during 10 time periods.[20] Each panel is simulated 1,000 times. To simplify matters, we dispense with labor and consider the production process with two inputs only: a quasi-fixed capital and freely varying materials. We also let that the true technology take a simple constant-input-elasticity Cobb-Douglas form:

$$Y_{it} = K_{it}^{\alpha_1} M_{it}^{\alpha_2} \exp\{\omega_{it} + \eta_{it}\}, \tag{D.1}$$

where we assume decreasing returns to scale and set $\alpha_1 = 0.25$ and $\alpha_2 = 0.65$.

The productivity components are generated as follows. We model the persistent productivity as a controlled AR(1) process:

$$\omega_{it} = \rho_0 + \rho_1 \omega_{it-1} + \rho_2 X_{it-1} + \zeta_{it}, \tag{D.2}$$

where we set $\rho_0 = 0.2$, $\rho_1 = \rho_2 = 0.8$, and $\zeta_{it} \sim$ i.i.d. $\mathbb{N}(0, \sigma_\zeta^2)$ with $\sigma_\zeta = 0.04$. The initial level of productivity ω_{i1} is drawn from $\mathbb{U}(1, 3)$ identically and independently distributed over i. The random transitory shocks η_{it} are drawn from $\eta_{it} \sim$ i.i.d. $\mathbb{N}(0, \sigma_\eta^2)$ with $\sigma_\eta = 0.07$.

The exogenous export-associated cost shifter $\mathcal{V}_{it}$ is assumed to exhibit first-order Markov persistence, i.e.,

$$\mathcal{V}_{it} = \varrho_0 + \varrho_1 \mathcal{V}_{it-1} + \nu_{it}, \tag{D.3}$$

where $\varrho_0 = 0.2$, $\varrho_1 = 0.5$, and $\nu_{it} \sim$ i.i.d. $\mathbb{N}(0, \sigma_\nu^2)$ with $\sigma_\nu = 0.5$. Note that the time-persistence in $\mathcal{V}_{it}$ is unimportant, and making it be a white noise would have sufficed.

We assume the following about evolution of the firm's state variables. Capital and the export intensity, both quasi-fixed, are set to respectively evolve according to

$$K_{it} = I_{it-1} + (1 - \delta_i) K_{it-1} \tag{D.4}$$

$$X_{it} = \mathcal{X}_{it-1} + X_{it-1}, \tag{D.5}$$

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

with the corresponding decision rules taking the following forms:[21]

$$I_{it-1} = K_{it-1}^{\beta_1} X_{it-1}^{\beta_2} [\exp\{-\mathcal{V}_{it-1}\}]^{\beta_3} [\exp\{\omega_{it-1}\}]^{\beta_4} \tag{D.6}$$

$$\mathcal{X}_{it-1} = \Phi\left(\gamma_0 + \gamma_1 \log K_{it-1} + \gamma_2 X_{it-1} - \gamma_3 \mathcal{V}_{it-1} + \gamma_4 \omega_{it-1}\right) - X_{it-1}, \tag{D.7}$$

where $\beta_1 = 0.8$, $\beta_2 = \beta_3 = \beta_4 = 0.1$, $\gamma_0 = -1$, $\gamma_1 = 0.1$, $\gamma_2 = 0.5$, $\gamma_3 = \gamma_4 = 0.1$ and $\Phi(\cdot)$ is the standard normal cdf. The firm-specific depreciation rates $\delta_i \in \{0.05, 0.075, 0.10, 0.125, 0.15\}$ are distributed uniformly across i. The initial levels of capital K_{i1} and export intensity X_{i1} are respectively drawn from $\mathbb{U}(10, 200)$ and $\mathbb{U}(0, 1)$ identically and independently distributed over firms. A particular form of the export adjustment rule in (D.7) along with the chosen $\{\gamma_j; j = 0, \ldots, 4\}$ ensures that the firm's export intensity stays bounded within the unit interval during the considered time periods, given our data generating process. Also, note that (D.6)–(D.7) imply an intuitively negative effect of the export-associated cost shifter $\mathcal{V}_{it}$ on both the investment and export intensity.

The optimal materials series are generated solving the firm's static restricted profit maximization problem along the lines of (2.5) after having already generated the series of $(K_{it}, X_{it}, \omega_{it})$ for each firm and time period. The conditional demand for M_{it} is given by

$$\begin{aligned} M_{it} &= \arg\max_{\mathcal{M}_{it}} \left\{ \left[P_t^X X_{it} + P_t^D (1 - X_{it})\right] K_{it}^{\alpha_1} \mathcal{M}_{it}^{\alpha_2} \exp\{\omega_{it}\}\mathcal{E} - P_t^M \mathcal{M}_{it} \right\} \\ &= \left(\frac{1}{P_t^M} \left[P_t^X X_{it} + P_t^D (1 - X_{it})\right] \alpha_2 K_{it}^{\alpha_1} \exp\{\omega_{it}\}\mathcal{E} \right)^{1/(1-\alpha_2)}, \end{aligned} \tag{D.8}$$

where we normalize $P_t^M = \mathcal{E} \ \forall \ t$ and intentionally assume away any temporal variation in the output prices: $P_t^X = 2$ and $P_t^D = 1$ for all t. Our decision to impose zero time variation in prices is dictated by Gandhi et al.'s (2017) recent findings whereby, at least in theory, the time-series variation in prices may help identify the production function even in the absence of an excluded relevant exogenous variable (in our case, X_{it} subject to the exogenous variation from $\mathcal{V}_{it}$).[22] By suppressing any such variation in prices, we are therefore able to restrict the source of identification of the production technology exclusively to the new element in the model, namely X_{it}.

Estimator. In Step 1 of our estimation procedure, we use the log-polynomial series of degree 2 to estimate r, $\mathbb{T}$ and $\mathbb{T}^*$. We do not cross-validate to minimize computational burden of our simulations. For the same reason, in Step 2, we approximate $f(\cdot)$ and $\varphi(\cdot)$ also using second-degree log-polynomial sieves thereby ensuring the second-step estimator has a closed-form solution, which obviates the need to numerically solve the least-squares optimization problem (as in the case of nonlinear ANN sieves used in our empirical application). For each simulation repetition, we compute the median, the root mean squared error (RMSE) and the mean absolute deviation

[21] For the sake of tractability, we follow Grieco et al. (2016) in postulating such policy rules instead of numerically deriving them from *each* firm's Bellman equation.

[22] Although their simulations suggest that the identification strategy based on temporal price variability performs poorly in practice.

Table D.1. Second-Step Estimates of Input Elasticities

	IV in Step 1: X_{it}		IV in Step 1: M_{it-2}	
	Capital	*Material*	*Capital*	*Material*
$n = 50$				
Median	0.2944	0.6031	0.7525	0.1107
RMSE	0.2614	0.1901	0.5283	0.5412
MAD	0.2142	0.1576	0.5016	0.5376
$n = 100$				
Median	0.2673	0.6367	0.8141	0.0546
RMSE	0.1807	0.1261	0.5744	0.5962
MAD	0.1481	0.1061	0.5638	0.5948
$n = 200$				
Median	0.2545	0.6486	0.8463	0.0288
RMSE	0.1226	0.0848	0.6017	0.6210
MAD	0.1004	0.0713	0.5956	0.6206
$n = 400$				
Median	0.2463	0.6522	0.8610	0.0150
RMSE	0.0822	0.0587	0.6131	0.6348
MAD	0.0671	0.0484	0.6111	0.6347
$n = 800$				
Median	0.2473	0.6528	0.8693	0.0082
RMSE	0.0546	0.0390	0.6201	0.6417
MAD	0.0455	0.0326	0.6188	0.6415

Notes: The true values of input elasticities are 0.25 for capital and 0.65 for materials. $T = 10$ throughout.

(MAD) of observation-specific[23] input elasticities (across firm-years) and report the medians of these metrics over 1,000 simulations.

We first estimate the production function via our proposed estimator using X_{it} to instrument for the endogenous M_{it}. The two-input analogues of (3.3) and (3.7) are

$$y_{it} = f(K_{it}, M_{it}) + \varphi(K_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it} + \eta_{it},\ \mathbb{E}[\zeta_{it} + \eta_{it}|\ K_{it}, X_{it}, K_{it-1}, X_{it-1}, M_{it-1}] = 0,$$

with the corresponding second-step simulation results reported in the left panel of Table D.1. The results are very encouraging and show that our approach recovers the true parameters well, thereby lending strong support to the validity of our identification strategy. As expected of a consistent estimator, the estimation becomes more stable with both the RMSE and MAD declining as the sample size grows.

Expectedly, the proxy approach however fails to identify the input elasticities if, in line with the convention in the literature, we employ exogenous but irrelevant higher-order lags such as M_{it-2} to instrument for M_{it} instead of using X_{it}. The right panel of Table D.1 reports the results for this

[23]The estimates are observation-specific despite that our DGP assumes a constant-elasticity Cobb-Douglas technology because our methodology estimates the unknown production function nonparametrically thereby delivering firm-year-specific estimates of the gradients of $f(\cdot)$.

Table D.2. Second-Step Estimates of Input Elasticities for the DGP with no Exports

	Capital	*Material*
$n = 50$		
Median	0.6524	0.0760
RMSE	0.4255	0.5759
MAD	0.4036	0.5718
$n = 100$		
Median	0.6781	0.0391
RMSE	0.4408	0.6099
MAD	0.4264	0.6092
$n = 200$		
Median	0.6987	0.0199
RMSE	0.4514	0.6295
MAD	0.4477	0.6292
$n = 400$		
Median	0.7055	0.0105
RMSE	0.4572	0.6392
MAD	0.4553	0.6392
$n = 800$		
Median	0.7099	0.0046
RMSE	0.4613	0.6452
MAD	0.4599	0.6452

Notes: The true values of input elasticities are 0.25 for capital and 0.65 for materials. Both the DGP and model feature no export-related aspects; M_{it} is instrumented via M_{it-2} in Step 1.

model, i.e.,

$$y_{it} = f(K_{it}, M_{it}) + \varphi(K_{it-1}, X_{it-1}, M_{it-1}) + \zeta_{it} + \eta_{it},\ \mathbb{E}[\zeta_{it} + \eta_{it}|\ K_{it}, K_{it-1}, X_{it-1}, M_{it-1}, M_{it-2}] = 0.$$

This model grossly overestimates the capital elasticity and produces near-zero estimates of the material elasticity, showing no improvement in the estimation as n grows.

Lastly, to establish baseline for the under-identification problem in the model of firm production under the standard structural assumptions, we also examine the consequences of completely ignoring the exporting aspect of the firm's production behavior on the estimation of production function as oftentimes done in the literature. That is, we estimate the model assuming the more standard conceptual framework which ignores the firm's export decisions (including the learning by exporting effects) altogether. Appendix A discusses such a model in detail. The corresponding data generating process features no export-related aspects: we drop X_{it} from all equations and set $\rho_2 = \beta_2 = \beta_3 = 0$ and $P_t^X = P_t^X = 1$. The estimating equation where, following the tradition, M_{it-2} is employed to

Table D.3. Coverage Probability and Width of the 95% Bootstrap Confidence Intervals for Average Input Elasticities

	Capital	*Material*
$n = 200$		
Coverage Prob.	0.8767	0.8567
Interval Width	0.2726	0.2307
$n = 400$		
Coverage Prob.	0.9103	0.8970
Interval Width	0.1888	0.1597
$n = 600$		
Coverage Prob.	0.9467	0.9233
Interval Width	0.1535	0.1283

Notes: $T = 10$ throughout.

instrument for M_{it} is given by

$$y_{it} = f(K_{it}, M_{it}) + \varphi(K_{it-1}, M_{it-1}) + \zeta_{it} + \eta_{it},\ \mathbb{E}[\zeta_{it} + \eta_{it}|\ K_{it}, K_{it-1}, M_{it-1}, M_{it-2}] = 0,$$

with the corresponding results presented in Table D.2. In line with Gandhi et al.'s (2017) Monte-Carlo evidence, this under-identified—due to the conditional irrelevance of M_{it-2}—proxy estimator exhibits systematic biases in the estimates of production function, with the input elasticity estimates being nowhere near the true values and no improvement therein as n increases.

Bootstrap Inference. Using the same DGP, we also investigate a finite-sample performance of the accelerated bias-corrected bootstrap procedure that we employ for our proposed estimator. This is of interest because of the complexity of our multi-step regularized estimation procedure, which makes establishing the validity of such a bootstrap nontrivial. Just like earlier in this section, here our focus is on unknown input elasticities, which are estimated in the second step of our estimation procedure. Hence, we bootstrap only the first two stages, skipping step 5 of the bootstrap algorithm described in Section 4.1 of the paper. The estimation details are unchanged and the same as those used to populate the left panel of Table D.1.

We consider the 95% confidence intervals (i.e., $a = 0.05$) for the average elasticities of capital and labor. We focus on the *average* elasticities because our methodology estimates the unknown production function nonparametrically thereby delivering observation-specific estimates of the gradients of $f(\cdot)$ despite that our DGP assumes a constant-elasticity Cobb-Douglas technology. We average these estimates across firm-years to obtain scalar measures of elasticities denoted by $\bar{f}_K$ and $\bar{f}_M$ that we then study against the true $\beta_K = 0.25$ and $\beta_M = 0.65$.

Of interest are both the size and power. Table D.3 reports coverage probabilities and the median interval widths for $n = \{200, 400, 600\}$ and fixed $T = 10$. To conserve computational time, the number of simulations is set at $S = 300$ with $B = 200$ bootstrap replications per each simulation. The coverage probability is estimated as the relative frequency (over S simulations)

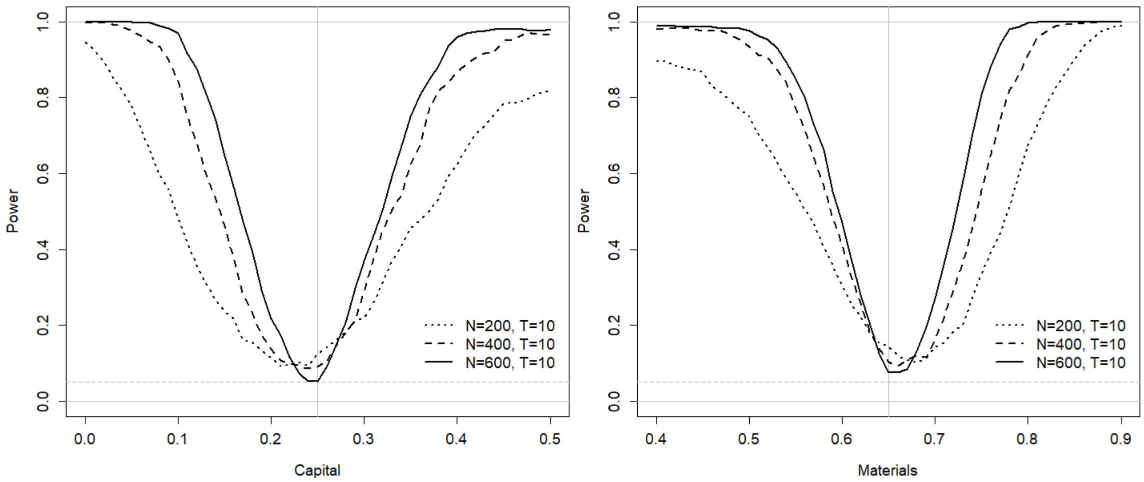

Figure D.1. Power of the Two-Sided Test for Average Input Elasticities using the 95% Bootstrap Confidence Intervals

of the estimated 95% accelerated bias-corrected confidence interval containing the true elasticity parameter. The reported interval widths are the medians (across S simulations) of the distances between the upper and lower bounds of the estimated confidence intervals.

The results indicate that there may be size distortions for small n, which is quite common for nonparametric tests. However, with the sample size growing, the estimated coverage of our intervals approaches the correct coverage and, for n modestly large enough, the estimated rejection probabilities seem close to the nominal size. The employed bootstrap intervals exhibit good power that increases with the distance between the null and the true parameter value as can be seen in Figure D.1 which plots power curves for the estimated confidence intervals. Here, power is computed as the relative rejection frequency. Importantly, the power improves as the sample size increases which is anticipated of a consistent test. We see this both in the shrinking intervals widths (Table D.3) as well as in the increasing speed with which power curves attain unity (Figure D.1).

E Data

In this section, we provide additional details about construction of the data sample. The 28 two-digit China Standard Industrial Classification (CSIC) manufacturing industries that we include on our analysis are listed in Table E.1.

In 2002, the CSIC code system underwent slight changes at the two-digit level. The industry codes for the manufacturing of electrical machinery and equipment, manufacturing of communication equipment, computers and other electronic equipment, and manufacturing of measuring instruments and machinery for cultural activity and office work changed from 40, 41, 42 to 39, 40, 41, respectively. In addition, the original CSIC 43 industry ("other manufacturing") was split into two new, different industry groups. We match firm-year observations in the pre-2002 years with

Table E.1. Manufacturing Industries

CSIC	Description
13	Food Processing (from Agricultural Products)
14	Food
15	Beverages
16	Tobacco
17	Textile
18	Apparel and Other Fabric Products
19	Leather, Fur, Feather and Related Products
20	Timber
21	Furniture
22	Paper and Other Paper Products
23	Printing, Reproduction of Recording Media
24	(Office) Stationery, Educational and Sports Goods
25	Petroleum Refining, Coking and Processing of Nuclear Fuel
26	Raw Chemical Materials and Chemical Products
27	Medical and Pharmaceutical Products
28	Chemical Fiber
29	Rubber Products
30	Plastic Products
31	Nonmetal Mineral Products
32	Smelting and Pressing of Ferrous Metals
33	Smelting and Pressing of Nonferrous Metals
34	Metal Products
35	General Purpose Machinery
36	Special Purpose Machinery
37	Transport Equipment
39	Electric Machinery and Equipment
40	Communication Equipment, Computers and Other Electronics
41	Precision Machinery, Measuring Instruments and Meters

those classified in accordance with the new CSIC code (GB\T 4754 – 2002). Since the CSIC 42 and CSIC 43 industries did not exist in the old CSIC code (GB\T 4754 – 1994), we exclude them from our analysis.

Table E.2 reports basic summary statistics for our data, including the sample mean and standard deviation estimates for key variables tabulated by the CSIC industry groups. The table clearly points to significant heterogeneity across industries. For instance, the average firm size, as measured by the output value, ranges from 27.8 million RMB in the Printing industry to 97.4 million RMB in Petroleum Refining. Industries with the largest fraction of firms participating in exports include the manufacturing of Leather Products (64.6%), Apparel (66.1%) and Computers (53.3%). Interestingly, the heaviest presence of the state capital is seen in the production of tobacco goods (68.6%) with the second most state-invested industry being Printing (26.9%). Foreigners tend to invest more in the production of Computers, Office Goods and Apparel: more than 20% of firms in these industries reports having foreign capital in the equity.

Table E.2. Data Summary Statistics, 1999–2006

CSIC	Y	K	L	M	$X\|X>0$	Exporter	State	Foreign	Subsidy
13	55,683	11,726	1,581	38,895	0.601	0.216	0.195	0.098	0.128
	(91,049)	(24,692)	(2,870)	(64,771)	(0.412)				
14	48,688	15,812	2,458	33,089	0.548	0.248	0.171	0.143	0.127
	(85,595)	(32,403)	(4,330)	(58,834)	(0.439)				
15	59,485	27,064	2,799	38,696	0.420	0.135	0.250	0.104	0.118
	(99,480)	(48,397)	(4,812)	(66,165)	(0.434)				
16	121,697	64,351	7,597	74,720	0.179	0.080	0.686	0.018	0.152
	(158,099)	(67,079)	(7,413)	(102,147)	(0.278)				
17	49,571	15,143	2,944	35,107	0.646	0.410	0.069	0.098	0.139
	(76,274)	(31,366)	(4,534)	(55,229)	(0.379)				
18	36,638	6,334	3,455	25,158	0.859	0.661	0.035	0.209	0.122
	(59,243)	(13,612)	(4,478)	(41,807)	(0.299)				
19	48,011	7,035	3,751	33,401	0.856	0.646	0.024	0.190	0.118
	(76,481)	(14,259)	(5,557)	(54,737)	(0.316)				
20	34,127	10,448	1,728	23,464	0.719	0.293	0.082	0.096	0.155
	(56,708)	(28,551)	(2,762)	(40,317)	(0.375)				
21	41,420	9,511	2,851	28,548	0.787	0.463	0.048	0.157	0.113
	(65,747)	(21,140)	(4,726)	(45,953)	(0.367)				
22	39,756	13,388	1,984	27,864	0.482	0.132	0.083	0.065	0.121
	(63,213)	(30,754)	(3,209)	(44,484)	(0.407)				
23	27,795	13,307	2,151	18,268	0.498	0.107	0.269	0.060	0.110
	(53,211)	(25,723)	(3,697)	(35,124)	(0.417)				
24	39,341	7,955	3,726	27,340	0.852	0.749	0.034	0.208	0.149
	(61,771)	(16,777)	(5,423)	(43,882)	(0.304)				
25	97,379	28,138	2,724	68,461	0.305	0.059	0.121	0.060	0.108
	(145,996)	(53,831)	(4,681)	(106,756)	(0.375)				
26	52,981	15,410	2,293	36,941	0.415	0.225	0.134	0.088	0.163
	(87,251)	(35,219)	(4,027)	(62,066)	(0.359)				
27	58,718	22,182	3,351	37,543	0.395	0.222	0.207	0.105	0.181
	(87,346)	(38,701)	(5,346)	(58,173)	(0.365)				
28	77,280	32,057	2,687	56,694	0.444	0.197	0.086	0.105	0.176
	(113,607)	(62,971)	(4,322)	(83,797)	(0.378)				
29	44,127	12,376	2,946	30,326	0.601	0.341	0.081	0.116	0.129
	(78,259)	(28,431)	(4,616)	(55,164)	(0.409)				
30	38,863	11,468	2,154	27,304	0.669	0.330	0.051	0.125	0.126
	(63,048)	(25,916)	(3,413)	(45,200)	(0.402)				
31	36,878	15,987	2,389	24,809	0.553	0.159	0.140	0.058	0.160
	(57,787)	(33,493)	(3,591)	(39,350)	(0.420)				
32	83,951	16,410	2,478	59,758	0.403	0.097	0.083	0.040	0.111
	(130,908)	(33,855)	(4,249)	(95,700)	(0.365)				
33	83,154	17,405	2,362	59,761	0.389	0.178	0.088	0.062	0.171
	(126,831)	(41,760)	(4,221)	(93,077)	(0.368)				
34	40,393	9,029	2,279	28,335	0.689	0.331	0.061	0.106	0.126
	(71,140)	(20,513)	(3,650)	(51,348)	(0.386)				
35	39,959	10,252	2,618	27,354	0.485	0.268	0.108	0.094	0.144
	(69,351)	(23,179)	(4,227)	(48,675)	(0.397)				
36	43,249	11,787	3,033	29,317	0.357	0.239	0.167	0.099	0.148
	(74,656)	(24,839)	(5,080)	(52,616)	(0.362)				
37	54,702	15,739	3,479	38,035	0.457	0.222	0.155	0.101	0.160
	(93,397)	(32,781)	(5,546)	(66,975)	(0.397)				
39	59,328	11,622	3,087	41,718	0.621	0.326	0.079	0.121	0.166
	(99,034)	(24,713)	(4,929)	(70,752)	(0.410)				
40	76,803	20,297	5,249	53,144	0.678	0.533	0.104	0.247	0.178
	(121,183)	(40,980)	(7,087)	(88,139)	(0.394)				
41	45,410	11,724	3,885	30,396	0.654	0.444	0.136	0.188	0.194
	(79,157)	(25,411)	(5,739)	(55,665)	(0.412)				
All	48,418	13,357	2,756	33,531	0.646	0.314	0.110	0.112	0.144
	(82,518)	(29,920)	(4,487)	(58,731)	(0.403)				

Notes: Reported are the sample means with sample standard deviations in parentheses. Y – the gross output; K – capital stock; L – labor; M – materials; X – the export intensity. Y, K, L and M are measured in thousands of RMB. X is unit-free proportions. “Exporter”, “State”, “Foreign” and “Subsidy” are binary indicators equal to one if the firm is an exporter ($X_{it} > 0$), state-invested, foreign-invested and subsidized, respectively.

F Additional Results

F.1 Returns to Scale

We first take a look at the scale efficiency of manufacturing firms in China, where it is defined as the industry's attainment of the efficient firm scale size associated with (unitary) constant returns to scale. Table F.1 reports the average estimates of the returns to scale exhibited by firms across industries and years. We compute the firm-year-specific returns to scale as the sum of log-derivatives of the gross production function $f(\cdot)$ estimated in the second step: $\text{RTS}_{it} = \sum_{a\in\{k,l,m\}} \partial\widehat{f}(K_{it}, L_{it}, M_{it})/\partial a_{it}$, where $\widehat{f}(\cdot)$ is the estimated sieve approximation given in (4.13). The instance when the RTS estimate is significantly less than/equal to/greater than one corresponds to decreasing/constant/increasing returns to scale.

From Table F.1, we observe that all confidence intervals for the industry-level mean RTS estimates include unity suggesting that an *average* firm in all manufacturing industries operated at the constant returns to scale (at the 5% significance level) in 2000–2006. Thus, the manufacturing sector in China is generally scale efficient. There is also some evidence that, in most industries, the returns to scale have been increasing (albeit statically insignificantly) over the course of our sample period. We now proceed to the analysis of firm-level productivity, which constitutes the primary focus of our paper.

F.2 Industry-Level Productivity Growth

To explore potentially heterogeneous experiences of individual industry groups, we perform a productivity growth decomposition for each individual two-digit CSIC industry.

Table F.2 reports the results, and they generally bolster the narrative whereby China's manufacturing industries have experienced no tangible reallocation of factors from less to more productive firms during the 2000–2006 period. The only notable exceptions are the Apparel and Metals (both ferrous and nonferrous) industries which exhibited positive reallocation effects that contributed at least 3.5% points to the cumulative aggregate productivity growth in these industries over 6 years. The Nonferrous Metals industry (CSIC 33) is the leader in this respect having experienced a cumulative reallocation effect of 7.7% which amounted to slightly less than a fifth of the total productivity improvement observed in the industry. In contrast, the majority of industries (17 out of 28) experienced a worsening in allocation of productive factors with a cumulative *decline* in the sample covariance of firm-level output and productivity, which attests to the reallocation of output from *more to less* productive firms in these industries. For instance, in industries such as Printing, Timber, Special Purpose Machinery, Communication Equipment and Measuring Instruments, this reallocation effect shaved 8% points or more off the cumulative cumulative productivity growth. In case of the Computer industry (CSIC 40), the total contribution of negative reallocation effects is estimated at a non-negligible 11.4% which has brought the cumulative growth in the industry productivity from 14.5 down to 3.1% only.

Table F.1. Mean Returns to Scale Estimates across Industries, 2000–2006

CSIC	2000	2001	2002	2003	2004	2005	2006
13	0.885	0.887	0.889	0.890	0.889	0.888	0.888
	(0.438, 1.561)	(0.433, 1.545)	(0.434, 1.547)	(0.432, 1.532)	(0.430, 1.508)	(0.429, 1.532)	(0.427, 1.520)
14	0.893	0.894	0.895	0.896	0.896	0.897	0.897
	(0.668, 1.331)	(0.658, 1.337)	(0.637, 1.342)	(0.626, 1.346)	(0.611, 1.352)	(0.617, 1.358)	(0.610, 1.361)
15	0.881	0.881	0.884	0.885	0.891	0.890	0.891
	(0.522, 1.313)	(0.527, 1.306)	(0.529, 1.309)	(0.530, 1.312)	(0.534, 1.311)	(0.534, 1.320)	(0.535, 1.318)
16	0.900	0.906	0.912	0.911	0.906	0.901	0.899
	(0.748, 2.689)	(0.751, 2.725)	(0.758, 2.618)	(0.765, 2.722)	(0.753, 2.809)	(0.746, 2.812)	(0.747, 2.886)
17	0.891	0.892	0.894	0.894	0.896	0.894	0.895
	(0.282, 1.457)	(0.283, 1.457)	(0.285, 1.466)	(0.287, 1.469)	(0.270, 1.473)	(0.279, 1.497)	(0.281, 1.487)
18	0.906	0.906	0.906	0.905	0.903	0.901	0.902
	(0.669, 1.286)	(0.671, 1.283)	(0.674, 1.281)	(0.673, 1.281)	(0.673, 1.284)	(0.676, 1.270)	(0.671, 1.277)
19	0.883	0.886	0.886	0.889	0.889	0.889	0.889
	(0.678, 1.219)	(0.682, 1.218)	(0.683, 1.219)	(0.682, 1.217)	(0.680, 1.225)	(0.681, 1.228)	(0.683, 1.224)
20	0.881	0.884	0.889	0.892	0.893	0.895	0.896
	(0.385, 1.346)	(0.382, 1.337)	(0.392, 1.347)	(0.393, 1.339)	(0.394, 1.344)	(0.394, 1.358)	(0.399, 1.346)
21	0.902	0.903	0.904	0.904	0.904	0.905	0.907
	(0.634, 2.304)	(0.635, 2.322)	(0.637, 2.268)	(0.640, 2.272)	(0.640, 2.327)	(0.641, 2.397)	(0.651, 2.363)
22	0.912	0.913	0.914	0.914	0.914	0.914	0.914
	(0.770, 1.225)	(0.770, 1.229)	(0.771, 1.230)	(0.772, 1.232)	(0.767, 1.235)	(0.775, 1.232)	(0.774, 1.230)
23	0.909	0.910	0.911	0.911	0.913	0.911	0.912
	(0.852, 1.008)	(0.844, 1.013)	(0.840, 1.017)	(0.827, 1.028)	(0.824, 1.043)	(0.823, 1.037)	(0.821, 1.048)
24	0.899	0.901	0.903	0.905	0.906	0.905	0.906
	(0.719, 1.269)	(0.720, 1.271)	(0.733, 1.280)	(0.734, 1.282)	(0.734, 1.286)	(0.731, 1.280)	(0.734, 1.283)
25	0.896	0.900	0.899	0.902	0.902	0.903	0.904
	(0.268, 1.487)	(0.271, 1.487)	(0.272, 1.479)	(0.276, 1.464)	(0.281, 1.457)	(0.292, 1.459)	(0.289, 1.447)
26	0.893	0.894	0.897	0.899	0.901	0.899	0.901
	(0.409, 1.723)	(0.402, 1.730)	(0.405, 1.743)	(0.406, 1.740)	(0.406, 1.743)	(0.412, 1.738)	(0.414, 1.732)
27	0.879	0.881	0.883	0.883	0.884	0.884	0.884
	(0.689, 1.161)	(0.686, 1.168)	(0.688, 1.168)	(0.680, 1.177)	(0.669, 1.191)	(0.676, 1.185)	(0.676, 1.187)
28	0.918	0.917	0.915	0.920	0.920	0.919	0.920
	(0.681, 1.338)	(0.681, 1.336)	(0.681, 1.342)	(0.685, 1.356)	(0.685, 1.351)	(0.689, 1.352)	(0.687, 1.364)
29	0.905	0.904	0.907	0.909	0.909	0.909	0.910
	(0.639, 1.419)	(0.637, 1.425)	(0.639, 1.421)	(0.639, 1.423)	(0.638, 1.433)	(0.644, 1.432)	(0.642, 1.434)
30	0.894	0.894	0.894	0.896	0.898	0.894	0.896
	(0.367, 1.553)	(0.366, 1.550)	(0.366, 1.554)	(0.367, 1.555)	(0.368, 1.554)	(0.369, 1.562)	(0.371, 1.558)
31	0.901	0.902	0.903	0.903	0.904	0.904	0.905
	(0.460, 1.986)	(0.460, 1.971)	(0.453, 1.922)	(0.452, 1.913)	(0.446, 1.847)	(0.449, 1.828)	(0.448, 1.840)
32	0.902	0.904	0.905	0.907	0.909	0.907	0.908
	(0.390, 1.712)	(0.389, 1.666)	(0.410, 1.746)	(0.423, 1.720)	(0.426, 1.714)	(0.424, 1.749)	(0.425, 1.717)
33	0.903	0.905	0.906	0.906	0.908	0.904	0.907
	(0.378, 1.508)	(0.380, 1.505)	(0.371, 1.499)	(0.376, 1.505)	(0.373, 1.505)	(0.370, 1.512)	(0.375, 1.504)
34	0.901	0.903	0.905	0.906	0.907	0.906	0.908
	(0.680, 1.203)	(0.682, 1.204)	(0.683, 1.205)	(0.683, 1.204)	(0.682, 1.207)	(0.684, 1.208)	(0.683, 1.206)
35	0.902	0.905	0.908	0.910	0.912	0.908	0.910
	(0.101, 1.751)	(0.102, 1.739)	(0.088, 1.732)	(0.083, 1.720)	(0.055, 1.723)	(0.073, 1.757)	(0.055, 1.737)
36	0.899	0.900	0.903	0.906	0.906	0.906	0.907
	(0.572, 1.310)	(0.572, 1.309)	(0.574, 1.311)	(0.571, 1.312)	(0.566, 1.314)	(0.573, 1.321)	(0.573, 1.320)
37	0.906	0.907	0.908	0.909	0.909	0.908	0.908
	(0.608, 1.439)	(0.597, 1.443)	(0.601, 1.441)	(0.600, 1.444)	(0.589, 1.445)	(0.590, 1.448)	(0.589, 1.445)
39	0.904	0.905	0.906	0.908	0.910	0.908	0.909
	(0.755, 1.287)	(0.755, 1.288)	(0.758, 1.290)	(0.756, 1.294)	(0.757, 1.297)	(0.768, 1.291)	(0.759, 1.294)
40	0.884	0.887	0.890	0.891	0.892	0.889	0.891
	(0.401, 1.376)	(0.402, 1.379)	(0.415, 1.376)	(0.417, 1.371)	(0.421, 1.359)	(0.406, 1.404)	(0.417, 1.391)
41	0.901	0.901	0.904	0.905	0.904	0.904	0.904
	(0.599, 1.259)	(0.600, 1.261)	(0.600, 1.262)	(0.601, 1.259)	(0.592, 1.266)	(0.605, 1.265)	(0.601, 1.262)
All	0.898	0.899	0.901	0.902	0.903	0.901	0.902
	(0.843, 1.411)	(0.841, 1.423)	(0.845, 1.425)	(0.846, 1.432)	(0.848, 1.460)	(0.846, 1.509)	(0.844, 1.504)

Notes: Reported are the average nonparametric estimates of the returns to scale with the 95% accelerated bias-corrected percentile bootstrap confidence intervals in parentheses.

Table F.2. (Weighted) Aggregate Productivity Growth Decomposition across Industries

Year	$\Delta \mathrm{p}_t$	$\Delta \overline{\mathrm{p}}_t$	$\Delta \mathrm{cov}_t$	$\Delta \mathrm{p}_t$	$\Delta \overline{\mathrm{p}}_t$	$\Delta \mathrm{cov}_t$	$\Delta \mathrm{p}_t$	$\Delta \overline{\mathrm{p}}_t$	$\Delta \mathrm{cov}_t$	$\Delta \mathrm{p}_t$	$\Delta \overline{\mathrm{p}}_t$	$\Delta \mathrm{cov}_t$
	CSIC 13			**CSIC 14**			**CSIC 15**			**CSIC 16**		
2001	0.015	0.029	–0.013	–0.017	0.009	–0.027	0.026	0.013	0.013	–0.067	–0.007	–0.060
2002	0.026	0.039	–0.013	0.030	0.044	–0.014	0.003	0.033	–0.030	0.072	0.058	0.014
2003	0.066	0.069	–0.003	0.042	0.046	–0.004	0.058	0.051	0.007	0.035	–0.030	0.064
2004	0.085	0.065	0.020	0.041	0.058	–0.017	0.057	0.080	–0.023	0.090	0.093	–0.003
2005	0.043	0.069	–0.027	0.089	0.084	0.006	0.088	0.087	0.000	0.034	0.037	–0.003
2006	0.041	0.045	–0.004	0.060	0.059	0.000	0.060	0.058	0.002	0.059	0.051	0.007
Cum.	0.277	0.316	–0.040	0.245	0.300	–0.055	0.292	0.323	–0.030	0.223	0.203	0.020
	CSIC 17			**CSIC 18**			**CSIC 19**			**CSIC 20**		
2001	0.002	–0.009	0.011	–0.021	–0.030	0.010	–0.016	–0.007	–0.009	–0.006	0.003	–0.009
2002	0.004	0.005	–0.001	–0.030	–0.031	0.001	–0.041	–0.028	–0.013	–0.027	0.017	–0.045
2003	0.034	0.022	0.012	0.003	–0.001	0.004	0.027	0.000	0.027	0.009	0.031	–0.022
2004	0.036	0.041	–0.004	0.045	0.039	0.006	0.026	0.027	–0.001	0.054	0.045	0.009
2005	0.062	0.067	–0.006	0.058	0.054	0.004	0.041	0.041	0.000	0.044	0.070	–0.027
2006	0.044	0.043	0.001	0.029	0.019	0.009	0.038	0.030	0.008	0.051	0.053	–0.002
Cum.	0.181	0.169	0.013	0.085	0.050	0.035	0.074	0.063	0.011	0.124	0.220	–0.096
	CSIC 21			**CSIC 22**			**CSIC 23**			**CSIC 24**		
2001	0.007	0.009	–0.002	0.012	0.009	0.003	0.009	0.009	0.000	–0.035	–0.024	–0.011
2002	–0.011	–0.007	–0.004	0.006	0.019	–0.013	0.018	0.034	–0.016	–0.023	–0.019	–0.003
2003	–0.007	0.007	–0.014	0.048	0.036	0.011	0.018	0.044	–0.026	0.019	0.007	0.012
2004	0.071	0.049	0.022	0.061	0.056	0.005	0.039	0.062	–0.023	0.015	0.039	–0.023
2005	0.003	0.044	–0.040	0.069	0.057	0.013	0.066	0.086	–0.019	0.065	0.046	0.018
2006	0.027	0.033	–0.006	0.042	0.044	–0.002	0.036	0.047	–0.011	0.011	0.015	–0.004
Cum.	0.090	0.134	–0.044	0.237	0.220	0.017	0.185	0.280	–0.094	0.052	0.064	–0.011
	CSIC 25			**CSIC 26**			**CSIC 27**			**CSIC 28**		
2001	0.007	0.020	–0.012	0.020	0.021	–0.001	–0.005	0.020	–0.024	0.007	–0.019	0.026
2002	0.050	0.015	0.035	0.018	0.023	–0.005	0.000	0.003	–0.003	0.016	0.028	–0.012
2003	0.084	0.096	–0.013	0.064	0.047	0.016	0.021	0.037	–0.016	0.060	0.083	–0.022
2004	0.054	0.093	–0.038	0.077	0.069	0.008	0.080	0.061	0.018	0.054	0.058	–0.004
2005	0.053	0.038	0.015	0.071	0.077	–0.006	0.080	0.066	0.014	0.092	0.073	0.019
2006	0.051	0.060	–0.009	0.062	0.055	0.007	0.049	0.037	0.012	0.033	0.045	–0.011
Cum.	0.299	0.322	–0.023	0.311	0.291	0.020	0.224	0.224	0.001	0.262	0.268	–0.005
	CSIC 29			**CSIC 30**			**CSIC 31**			**CSIC 32**		
2001	–0.015	–0.026	0.011	0.001	0.000	0.000	0.008	0.003	0.005	0.005	0.003	0.002
2002	0.013	0.028	–0.015	–0.011	–0.011	0.000	0.022	0.026	–0.004	0.034	0.038	–0.004
2003	0.041	0.041	–0.001	0.005	0.032	–0.027	0.066	0.048	0.018	0.100	0.077	0.023
2004	0.043	0.032	0.012	0.040	0.054	–0.014	0.066	0.066	0.000	0.118	0.104	0.014
2005	0.043	0.064	–0.021	0.055	0.057	–0.003	0.069	0.076	–0.007	0.081	0.052	0.029
2006	0.076	0.065	0.011	0.052	0.037	0.015	0.069	0.059	0.010	0.036	0.054	–0.018
Cum.	0.202	0.205	–0.003	0.142	0.170	–0.028	0.300	0.277	0.023	0.374	0.328	0.046
	CSIC 33			**CSIC 34**			**CSIC 35**			**CSIC 36**		
2001	0.024	0.022	0.002	0.003	0.002	0.001	0.011	0.009	0.002	–0.004	0.010	–0.015
2002	0.027	0.019	0.008	–0.005	0.007	–0.012	0.023	0.023	–0.001	0.027	0.037	–0.010
2003	0.080	0.038	0.042	0.041	0.038	0.003	0.060	0.057	0.003	0.008	0.069	–0.061
2004	0.087	0.099	–0.012	0.051	0.051	0.000	0.061	0.074	–0.013	0.064	0.072	–0.008
2005	0.093	0.057	0.037	0.059	0.054	0.005	0.062	0.061	0.002	0.070	0.074	–0.003
2006	0.101	0.099	0.001	0.041	0.041	0.001	0.055	0.052	0.003	0.065	0.055	0.010
Cum.	0.411	0.334	0.077	0.190	0.193	–0.003	0.272	0.275	–0.003	0.230	0.316	–0.086
	CSIC 37			**CSIC 39**			**CSIC 40**			**CSIC 41**		
2001	0.025	0.020	0.005	0.005	0.003	0.002	–0.007	0.013	–0.019	–0.037	0.002	–0.039
2002	0.034	0.027	0.007	–0.002	0.004	–0.007	–0.060	–0.019	–0.041	–0.017	0.004	–0.021
2003	0.026	0.052	–0.027	0.045	0.044	0.001	0.009	0.038	–0.029	0.054	0.044	0.010
2004	0.048	0.065	–0.017	0.053	0.051	0.002	0.066	0.070	–0.003	0.047	0.079	–0.032
2005	0.053	0.064	–0.011	0.067	0.054	0.012	0.014	0.017	–0.003	0.044	0.065	–0.021
2006	0.054	0.043	0.010	0.063	0.049	0.014	0.008	0.026	–0.018	0.059	0.037	0.022
Cum.	0.240	0.272	–0.032	0.230	0.205	0.025	0.031	0.145	–0.114	0.151	0.231	–0.080

Notes: Reported are the annual growth rates of the (weighted) aggregate productivity p_t and of its two components: the (unweighted) average productivity $\overline{\mathrm{p}}_t$ and the covariance between the firm-level output and productivity cov_t.

F.3 (Conditional) Exporter Productivity Differentials by Export Intensity

Table F.3. Median (Log) Exporter Productivity Premia by Export Intensity, 2000–2006

CSIC	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
13	–0.053	–0.051	–0.067	–0.075	–0.111	–0.132	–0.149	–0.163	–0.183	–0.120
14	–0.070	–0.089	–0.075	–0.082	–0.088	–0.089	–0.093	–0.135	–0.119	–0.129
15	*–0.029*	*0.003*	*–0.022*	*–0.034*	–0.081	–0.062	–0.082	–0.040	–0.048	–0.071
16	*–0.005*	*–0.006*	*–0.045*	*0.010*	*0.042*	*–0.067*	*–0.106*	*–0.095*	*–0.054*	*–0.195*
17	–0.046	–0.073	–0.090	–0.101	–0.118	–0.129	–0.133	–0.149	–0.152	–0.159
18	–0.067	–0.119	–0.148	–0.157	–0.167	–0.170	–0.178	–0.173	–0.175	–0.179
19	–0.034	–0.106	–0.133	–0.168	–0.183	–0.188	–0.197	–0.195	–0.205	–0.204
20	–0.046	–0.059	–0.085	–0.116	–0.128	–0.152	–0.145	–0.144	–0.148	–0.153
21	–0.099	–0.107	–0.129	–0.146	–0.149	–0.195	–0.201	–0.203	–0.229	–0.204
22	–0.034	*–0.020*	–0.045	–0.045	–0.076	–0.089	–0.098	–0.151	–0.170	–0.170
23	–0.064	–0.069	*–0.021*	–0.094	–0.073	–0.091	–0.137	–0.199	–0.247	–0.192
24	–0.075	–0.120	–0.140	–0.174	–0.175	–0.190	–0.197	–0.191	–0.191	–0.189
25	*–0.057*	–0.168	–0.165	–0.143	*–0.046*	*–0.047*	–0.122	–0.100	–0.317	–0.119
26	–0.080	–0.063	–0.081	–0.087	–0.090	–0.107	–0.123	–0.135	–0.156	–0.176
27	–0.071	–0.066	–0.056	–0.071	–0.108	–0.133	–0.115	–0.118	–0.115	–0.051
28	–0.058	*–0.021*	–0.078	–0.062	*–0.040*	–0.096	–0.055	–0.107	–0.092	–0.126
29	–0.084	–0.060	–0.103	–0.105	–0.148	–0.142	–0.152	–0.184	–0.198	–0.192
30	–0.088	–0.091	–0.094	–0.125	–0.145	–0.178	–0.202	–0.201	–0.210	–0.224
31	*–0.007*	–0.018	–0.038	–0.054	–0.064	–0.099	–0.115	–0.128	–0.166	–0.158
32	*–0.030*	–0.056	–0.077	–0.096	–0.113	–0.137	–0.155	–0.152	–0.170	–0.145
33	–0.120	–0.097	–0.133	–0.124	–0.145	–0.170	–0.177	–0.172	–0.226	–0.193
34	–0.086	–0.087	–0.116	–0.134	–0.151	–0.164	–0.175	–0.175	–0.189	–0.195
35	–0.087	–0.094	–0.098	–0.107	–0.128	–0.130	–0.141	–0.156	–0.170	–0.158
36	–0.099	–0.111	–0.105	–0.107	–0.086	–0.110	–0.123	–0.133	–0.152	–0.177
37	–0.081	–0.074	–0.057	–0.087	–0.089	–0.094	–0.110	–0.134	–0.158	–0.154
39	–0.087	–0.103	–0.107	–0.135	–0.154	–0.188	–0.194	–0.210	–0.201	–0.226
40	–0.069	–0.097	–0.131	–0.150	–0.176	–0.189	–0.221	–0.217	–0.236	–0.242
41	–0.091	–0.082	–0.106	–0.144	–0.127	–0.175	–0.179	–0.203	–0.224	–0.242
All	–0.071	–0.073	–0.096	–0.115	–0.138	–0.160	–0.177	–0.189	–0.203	–0.207

Notes: Reported are the conditional median estimates of the exporter productivity premium tabulated by deciles of the empirical distribution of non-zero export intensity. All point estimates, except those in *italic*, are statistically significant at the 5% level.

References

Ackerberg, D. A., Benkard, C. L., Berry, S., & Pakes, A. (2007). Econometric tools for analyzing market outcomes. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics*, volume 6A. North Holland.

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

Ackerberg, D. A. & Hahn, J. (2015). Some nonparametric identification results using timing and information set assumptions. Mimeo, University of California Los Angeles.

Amiti, M. & Konings, J. (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from indonesia. *American Economic Review*, *97*, 1611–1638.

Andrews, D. W. K. (2017). Examples of L^2-complete and boundedly-complete distributions. *Journal of Econometrics*, *199*, 213–220.

Aw, B.-Y. & Hwang, A. R. (1995). Productivity and the export market: A firm-level analysis. *Journal of Development Economics*, *47*, 313–332.

Baldwin, J. R. & Gu, W. (2003). Export-market participation and productivity performance in Canadian manufacturing. *Canadian Journal of Economics*, *36*, 634–657.

Bernard, A. B., Eaton, J., Jensen, J. B., & Korum, S. (2003). Plants and productivity in international trade. *American Economic Review*, *93*, 1268–1290.

Bernard, A. B. & Jensen, J. B. (1995). Exporters, jobs, and wages in U.S. manufacturing: 1976–1987. *Brookings Papers on Economic Activity: Microeconomics*, 67–119.

Bernard, A. B. & Jensen, J. B. (2004). Exporting and productivity in the USA. *Oxford Review of Economic Policy*, *20*, 343–357.

Blundell, R., Chen, X., & Kristensen, D. (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica*, *75*, 1613–1669.

Brandt, L., Tombe, T., & Zhu, Z. (2013). Factor market distortions across time, space and sectors in China. *Review of Economic Dynamics*, *16*, 39–58.

Caballero, R. J., Engel, E. M. R. A., & Haltiwanger, J. (1997). Aggregate employment dynamics: Building from microeconomic evidence. *American Economic Review*, *87*, 115–137.

Carrasco, M., Florens, J. P., & Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics*, volume 6B. North Holland.

Centorrino, S. (2016). Data-driven selection of the regularized parameter in additive nonparametric instrumental regressions. Working Paper, Stony Brook University.

Centorrino, S., Fève, F., & Florens, J. P. (2017). Additive nonparametric instrumental regressions: A guide to implementation. *Journal of Econometric Methods*, *6*.

Centorrino, S., Fève, F., & Florens, J. P. (2019). Nonparametric instrumental regressions with (potentially discrete) instruments independent of the error term. Working Paper, Stony Brook University.

Chen, X., Hahn, J., Liao, Z., & Ridder, G. (2015). Asymptotic properties of nonparametric two-step sieve estimates. Working Paper, University of California Los Angeles.

Chen, X. & Ludvigson, S. C. (2009). Land of addicts? An empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics*, *24*, 1057–1093.

Clerides, S. K., Lach, S., & Tybout, J. R. (1998). Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco. *Quarterly Journal of Economics*, *113*, 903–947.

Cooper, R. & Willis, J. (2003). The cost of labor adjustment: Inference from the gap. NBER Working Paper 10006.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, *13*, 377–403.

Dai, M., Maitra, M., & Yu, M. (2016). Unexceptional exporter performance in China? The role of processing trade. *Journal of Development Economics*, *121*, 177–189.

Darolles, S., Fan, Y., Florens, J. P., & Renualt, E. (2011). Nonparametric instrumental regression. *Econometrica*, *79*, 1541–1565.

De Loecker, J. (2007). Do exports generate higher productivity? Evidence from Slovenia. *Journal of International Economics*, *73*, 69–98.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, *79*, 1407–1451.

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

Delgado, M. A., Farinas, J. C., & Ruano, S. (2002). Firm productivity and export markets: A non-parametric approach. *Journal of International Economics*, *57*, 397–422.

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 technical change. *Journal of Political Economy*, *126*, 1027–1084.

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

Ericson, R. & Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, *62*, 53–82.

Feenstra, R. C., Li, Z., & Yu, M. (2014). Exports and credit constraints under incomplete information: Theory and evidence from China. *Review of Economics and Statistics*, *96*(4), 729–744.

Florens, J. P., Racine, J. S., & Centorrino, S. (2018). Nonparametric instrumental variable derivative estimation. *Journal of Nonparametric Statistics*, *30*, 368–391.

Freyberger, J. (2017). On completeness and consistency in nonparametric instrumental variable models. *Econometrica*, *85*, 1629–1644.

Gandhi, A., Navarro, S., & Rivers, D. (2013). On the identification of gross output production functions: How heterogeneous is productivity? Working Paper.

Gandhi, A., Navarro, S., & Rivers, D. (2017). On the identification of gross output production functions. *Journal of Political Economy*. forthcoming.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Taylor & Francis Group, LLC.

Goffe, W. L., Ferrier, G. D., & Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, *60*, 65–99.

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

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.

Guariglia, A., Liu, X., & Song, L. (2011). Internal finance and growth: Microeconometric evidence on Chinese firms. *Journal of Development Economics*, *96*, 79–94.

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

Hall, P. & Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics*, *33*, 2904–2929.

Hashiguchi, Y. (2015). Allocation efficiency in China: An extension of the dynamic Olley-Pakes productivity decomposition. Discussion Paper No. 544, Institute of Developing Economies.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multi-layer feedforward networks are universal approximators. *Neural Networks*, *2*, 359–366.

Horowitz, J. L. (2011). Applied nonparametric instrumental variable estimation. *Econometrica*, *79*, 347–394.

Horowitz, J. L. (2014). Ill-posed inverse problems in economics. *Annual Review of Economics*, *6*, 21–51.

Johannes, J., Van Bellegem, S., & Vanhems, A. (2013). Iterative regularisation in nonparametric instrumental regression. *Journal of Statistical Planning and Inference*, *143*, 24–39.

Kuan, C.-M. (2008). Artificial neural networks. In S. N. Durlauf & L. E. Blume (Eds.), *New Palgrave Dictionary of Economics*. Palgrave Macmillan.

Kuan, C.-M. & White, H. (1994a). Artificial neural networks: An econometric perspective. *Econometric Reviews*, *13*, 1–91.

Kuan, C.-M. & White, H. (1994b). Reply to comments on "Artificial neural networks: An econometric perspective". *Econometric Reviews*, *13*, 139–143.

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

Lu, D. (2010). Exceptional exporter performance? Evidence from Chinese manufacturing firms. Working Paper, University of Chicago.

Lu, J., Lu, Y., & Tao, Z. (2010). Exporting behavior of foreign affiliates: Theory and evidence. *Journal of International Economics*, *81*, 197–205.

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

Marschak, J. & Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, *12*, 143–205.

Matzkin, R. L. (2004). Unobservable instruments. Mimeo, University of California Los Angeles.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, *71*, 1695–1725.

Newey, W. K. & Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, *71*(5), 1565–1578.

Newey, W. K., Powell, J. L., & Vella, F. (1999). Nonparametric estimation of triangular equations models. *Econometrica*, *67*(3), 565–603.

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

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. *Review of Economic Studies*, *69*, 245–276.

Sexton, R. S., Dorsey, R. E., & Johnson, J. D. (1999). Optimization of neural networks: A comparative analysis of the genetic algorithm and simulated annealing. *European Journal of Operational Research*, *114*, 589–601.

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

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

Wagner, J. (2007). Exports and productivity. *World Economy*, *30*, 60–82.

Wang, Z. & Wei, S.-J. (2010). What accounts for the rising sophistication of China's exports? In *China's Growing Role in World Trade* (pp. 63–104). University of Chicago Press.

White, H. (1989). Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical Association*, *84*, 1003–1013.

Yu, M. (2014). Processing trade, tariff reductions and firm productivity: Evidence from Chinese firms. *Economic Journal*. forthcoming.