Identification of Gross Output Production Functions with Nonseparable Productivity

Qingsong Pan*

School of Economics, Shandong University

Abstract

We study nonparametric identification of gross-output production functions in which productivity enters nonseparably, relaxing Hicks neutrality, and use the framework to measure the bias of technical change. Under perfect competition, we extend Gandhi et al. (2020) (GNR) to identify output elasticities, and then impose an empirically motivated homogeneity restriction to obtain full identification of the technology. Under imperfect competition with revenue data, markups and returns to scale (RTS) are difficult to separately identify. We therefore calibrate RTS for point identification and show that the implied *directions* and *relative magnitudes* of technological bias are *invariant* to this calibration. Applying the framework to Chinese manufacturing firms (1998–2007), we find that technical change is predominantly capital-biased and least favorable to labor. Yet in the realized data, the marginal product of labor (MPL) rises the *most* over the decade. A decomposition resolves this apparent paradox: MPL growth is driven primarily by capital and materials deepening through factor complementarities rather than by productivity growth, whereas for capital and materials the opposite pattern holds. These findings point to biased technical change as a distinct force behind the pronounced factor deepening observed over this period.

*Corresponding Author, Assistant Professor, School of Economics, Shandong University, Jinan, 250100, China. Fax: 86-531-88571371; Tel: 86-531-88364625; Email: qingsongpan@utexas.edu.

I extend my deepest gratitude to Dan Ackerberg, Brendan Kline, and Haiqing Xu for their invaluable guidance and support throughout this research. I am also deeply thankful to David Rivers, Andrey Ordin, Salvador Navarro, Mert Demirer, Amit Gandhi, and Xiaojun Song for their encouragement and insightful feedback. I thank Mert Demirer for providing access to the Chilean and Colombian datasets. Special thanks to Daniel Yi Xu for his excellent mentorship and constructive feedback through the ESWC 2025 Mentoring Initiative. Any errors or omissions in this paper remain my sole responsibility.

Key words: Production Functions, Productivity, Nonseparable Models, Technological Change

1 Introduction

Production functions are among the foundational tools of applied economics. Estimated using firm-level data, they deliver direct evidence on key supply-side objects, including total factor productivity (TFP), output elasticities, marginal products, and returns to scale. They also serve as inputs for the measurement of other economically important objects, such as markups, markdowns, and allocative efficiency. A pervasive restriction in the production function literature, however, has been the assumption of *Hicks-neutral* productivity, whereby latent productivity enters additively in logs and scales all marginal products proportionally. While analytically convenient, Hicks-neutral models fail to capture an important economic reality: technical change may be *biased*, raising the marginal product of some inputs more than others.

This is consequential for at least two reasons. First, Hicks-neutral models are inherently ill-suited to answer a central question: which inputs benefit most from technological progress? Measuring the direction and magnitude of technological bias is important in its own right, because it shapes relative input demand and the distribution of gains from technology across factors of production. It may also matter for aggregate distributional trends. In particular, a growing literature argues that biased technical change is a key force behind the secular decline in the labor share (e.g., Doraszelski and Jaumandreu, 2018; Zhang, 2019; Oberfield and Raval, 2021). Second, imposing Hicks neutrality when the true technology is non-Hicksian can induce production-function misspecification. This misspecification can distort estimates of fundamental supply-side objects and thereby contaminate downstream measures built from them. For instance, Demirer (2020) shows that ignoring non-Hicksian productivity leads to underestimation of variable-input elasticities and hence overstatement of markups, while Rubens et al. (2024) find that the Hicks-neutral restriction can substantially overstate markdowns.

Building on the conventional proxy-variable approach (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020), this paper develops a nonparametric framework for the identification and estimation of *gross-output* production functions in which productivity enters flexibly and nonseparably. By permitting unrestricted interactions between productivity and inputs, the framework relaxes the Hicks-neutrality restriction and allows for biased technical change. It therefore accommodates richer forms of technological heterogeneity and reduces the risk of misspecification relative to standard Hicks-neutral models.

Our way of relaxing Hicks neutrality differs from the factor-augmenting approach that has become prominent in the recent literature (e.g., Doraszelski and Jaumandreu, 2018; Zhang, 2019; Demirer, 2020; Raval, 2019; Oberfield and Raval, 2021). The two approaches entail different trade-offs. Factor-augmenting models can accommodate multiple, factor-specific productivity shifters, but they typically impose a CES form in inputs and restrict non-Hicksian productivity to operate through input augmentation rather than allowing it to alter, for instance, substitution patterns across inputs. Our framework instead allows a single scalar productivity term to enter the production function nonseparably, imposing weaker functional-form restrictions. The approaches also differ practically. Factor-augmenting models typically require input price data and at least two flexible inputs, including labor, making them less well suited to environments in which labor adjusts dynamically; they also generally assume perfectly competitive labor markets,[1] so markdowns may be observationally confounded with labor-augmenting productivity in the identifying first-order conditions. Our approach requires neither input price data nor a flexible labor input—only flexible intermediate inputs[2]—and is agnostic about labor-market structure.[3]

We begin with the benchmark case of perfect competition and then extend the analysis to imperfect competition, which underlies our main empirical application. Under perfect competition, we extend the first-order-condition approach of Gandhi et al. (2020) (GNR) and show that the output elasticities with respect to capital, labor, and materials are identified at each observation. We term this result *reduced-form identification*: the elasticities are identified along the equilibrium manifold generated by firms' input choices, even though the underlying structural production function remains unidentified absent further restrictions. We then establish that the structural technology is fully identified under an additional homogeneity restriction. A brief application to manufacturing data from Chile and Colombia (Appendix I) suggests that the proposed estimators perform well, that Hicks-neutral specifications systematically overstate labor elasticities, and that homogeneity is a reasonable empirical approximation.

Extending the framework to imperfect competition introduces a familiar data limitation:

[1] A notable exception is Rubens et al. (2024).

[2] In the main text, we treat labor as predetermined, but Appendix B explains how this timing assumption can be relaxed via an IV approach.

[3] Although our non-Hicksian framework can accommodate markdowns, they are not the focus of this paper, and we leave their explicit treatment to future work.

most production datasets report revenue but not physical quantities. To obtain consistent estimates in this environment, we follow Klette and Griliches (1996) and De Loecker (2011) in adopting a monopolistic-competition setting, while generalizing their CES demand structure to a nonparametric framework with heterogeneous markups. The central identification difficulty is that revenue combines prices and quantities, making it hard to separate production elasticities from demand elasticities. Equivalently, with revenue data alone, RTS and markups are difficult to disentangle. Rather than relying on variation in an aggregate demand shifter for identification, as in Klette and Griliches (1996) and De Loecker (2011), we proceed by calibrating RTS: that is, prespecifying a value of RTS based on estimates considered plausible in the prior literature. This strategy is attractive for two reasons: economists typically have a clear prior sense of the empirically plausible range of RTS, and, more importantly, our main empirical objects of interest—the directions and relative magnitudes of technological bias—are invariant to the particular RTS value chosen.

A key step in the identification strategy is to recover the intermediate-input demand function. We do so by leveraging the timing and information-set assumptions in Ackerberg et al. (2022), which pin down productivity up to a monotone transformation and thereby permit identification of a substantially more flexible class of production functions.

To implement the theory, we propose sieve-based estimators that closely track the identification arguments. We first estimate the intermediate-input demand function and then use the first-order conditions and cross-equation restrictions to recover output elasticities and the structural production function.

We apply the framework to Chinese manufacturing over 1998–2007, a period of extraordinary productivity growth (Brandt et al., 2012). A central question is whether this growth favored all factors equally or was biased in ways that help explain the concurrent capital and materials deepening observed in the data. We find that technical change was predominantly capital-biased and least favorable to labor: holding inputs fixed, productivity growth raised the marginal product of capital by 32%, that of materials by 22%, but that of labor by only 12%. Yet in the realized data, the marginal product of labor rose the *most* over the decade. A Divisia decomposition resolves this apparent paradox: MPL growth was driven overwhelmingly by capital and materials deepening through factor complementarities, whereas the growth in the marginal products of capital and materials was attributable mainly to productivity growth itself. These findings point to biased technical change as a distinct and economically meaningful force behind the pronounced capital and materials deepening observed over this period.

This paper contributes to three literatures. First, it contributes to the proxy-variable literature by extending the approach to nonseparable production-function models and thereby relaxing the Hicks-neutrality restriction. Second, it contributes to the literature on biased technical change by providing a flexible and practically implementable alternative to factor-augmenting approaches for measuring the direction and magnitude of technological bias, and by documenting capital-biased technical change in Chinese manufacturing. Third, it contributes to the literature on revenue-based production-function estimation under imperfect competition by showing that, although markups and returns to scale are challenging to be separately identified from revenue data alone, the direction of technological bias and the dispersion and trends of markups are invariant to the returns-to-scale calibration.

The remainder of the paper is organized as follows. Section 2 develops identification and estimation under perfect competition, first establishing reduced-form identification of output elasticities at each observation and then showing that an empirically motivated homogeneity restriction delivers point identification of the structural technology. Section 3 extends the framework to imperfect competition with revenue data, establishes identification up to returns to scale, and presents a calibrated returns-to-scale strategy for point identification; crucially, the directions and relative magnitudes of technological bias are invariant to the particular calibration chosen. Section 4 applies the framework to Chinese manufacturing over 1998–2007, studying the bias of technical change and decomposing the growth in marginal products into contributions from productivity and factor complementarities. Section 5 concludes. Appendix B develops an IV-based relaxation of the timing assumption for labor; the remaining appendices collect extensions, additional discussion, and technical proofs.

2 Perfect Competition

2.1 Setup

Data and Production Function. We observe a panel of firms $j = 1, \ldots, J$ over periods $t = 1, \ldots, T$. Let $(Y_{jt}, K_{jt}, L_{jt}, M_{jt})$ denote measured output, capital, labor, and intermediate inputs (hereafter, materials), with logarithms $(y_{jt}, k_{jt}, l_{jt}, m_{jt})$. Our asymptotic framework holds T fixed and lets $J \to \infty$, so the joint distribution of $\{(y_{jt}, k_{jt}, l_{jt}, m_{jt})\}_{t=1}^{T}$ is identified from the data.

For a generic (log) input $i_{jt} \in \{k_{jt}, l_{jt}, m_{jt}\}$, we call i_{jt} *dynamic* if its optimal choice in period t depends on its lagged value, and *flexible* otherwise.

Output is produced by technology F_t that maps observed inputs and an unobserved productivity component ω_{jt} into output, allowing ω_{jt} to enter in a fully flexible, nonseparable manner. Measured output may also include an ex-post shock or measurement error $e^{\epsilon_{jt}}$:

$$(1) \qquad Y_{jt} = F_t(K_{jt}, L_{jt}, M_{jt}, e^{\omega_{jt}})\, e^{\epsilon_{jt}} \quad \Leftrightarrow \quad y_{jt} = f_t(k_{jt}, l_{jt}, m_{jt}, \omega_{jt}) + \epsilon_{jt},$$

where F_t is strictly increasing in ω_{jt}, and ω_{jt} is complementary to inputs (k_{jt}, l_{jt}, m_{jt}) in the sense that an increase in ω_{jt} raises each input's marginal product.[4] All functions are continuously differentiable, and the support of observables provides the open-set variation required for the differentiation and integration steps used below.

Assumptions.

Assumption 2.1 (Markovian Productivity)**.** ω_{jt} follows a first-order Markov process,

$$\omega_{jt} = h_t(\omega_{jt-1}, \xi_{jt}),$$

where ξ_{jt} is an innovation independent of all variables determined prior to its realization.[5]

Assumption 2.2 (Timing and Information)**.** (i) (k_{jt}, l_{jt}) are chosen before observing ω_{jt}; (ii) m_{jt} is flexible and chosen after observing ω_{jt}; (iii) (k_{jt}, l_{jt}, m_{jt}) are chosen without knowledge of ϵ_{jt} and are independent of ϵ_{jt}.

Assumption 2.3 (Conditional Profit Maximization under Perfect Competition)**.** Firms are price takers in output and material markets. Conditional on $(K_{jt}, L_{jt}, \omega_{jt})$, the firm chooses M_{jt} to solve

$$(2) \qquad \max_{M_{jt}} \; P_t\, \mathbb{E}[F_t(K_{jt}, L_{jt}, M_{jt}, e^{\omega_{jt}}) e^{\epsilon_{jt}} \mid K_{jt}, L_{jt}, \omega_{jt}] - \rho_t M_{jt},$$

where P_t and ρ_t denote the output and material prices.

Assumption 2.4 (Strict Monotonicity)**.** Let $m_{jt} = \phi_t(k_{jt}, l_{jt}, \omega_{jt})$ denote the material policy implied by (2). Then ϕ_t is strictly increasing in ω_{jt}.

Assumption 2.1 is standard. The framework can be extended with only minor modifications to allow for higher-order Markov dynamics and fixed effects.[6] Assumption 2.2 requires

[4] This assumption is not required for identification given the assumptions in Section 2.1; we impose it in estimation to ensure that ω_{jt} has the standard property of a productivity index. Under the Hicks-neutral specification, it holds automatically.

[5] Without loss of generality, ξ_{jt} is scalar and h_t is strictly increasing in ξ_{jt}.

[6] See Lee et al. (2019) and Ackerberg (2021).

labor to be chosen prior to the innovation ξ_{jt}, which strengthens conventional proxy-variable timing. This restriction is motivated by identification concerns: even under functional-form restrictions, allowing l_{jt} to covary with ξ_{jt} may lead to under-identification (Ackerberg et al., 2020). Appendix B considers a relaxation that allows l_{jt} to be correlated with ξ_{jt}, building on the IV approach for nonseparable models developed by Chernozhukov and Hansen (2005) and extended to the timing-and-information framework by Ackerberg et al. (2022). Under Assumptions 2.1–2.2, inputs are correlated with ω_{jt} but orthogonal to ϵ_{jt}; moreover, only the flexible choice m_{jt} loads on ξ_{jt}.

Assumption 2.3 renders the material problem static and yields two implications used below: it delivers the policy ϕ_t (Assumption 2.4) and links the materials elasticity to the materials revenue share (Lemma 2.1). Perfect competition in materials markets is standard in the proxy-variable literature; we relax price-taking in output markets in Section 3.

Assumption 2.4 is the key proxy-variable condition. Strict monotonicity implies that, conditional on (k_{jt}, l_{jt}), m_{jt} is invertible in ω_{jt}, so observed materials serve as a proxy for productivity. A sufficient condition is strict concavity of the production function in M_{jt} together with complementarity between M_{jt} and ω_{jt} (i.e., $\partial^2 F_t/\partial M\,\partial\omega > 0$).

Assumption 2.4 also imposes a scalar latent state in the materials policy. It would be violated in the presence of unobserved heterogeneity in material prices or additional latent shocks (e.g., demand shocks or optimization errors) entering ϕ_t. In many applications, materials are close to commodities, so unobserved price dispersion may be limited. When material prices are heterogeneous but observed, they can be included as additional arguments in ϕ_t.

2.2 Identification

Under perfect competition, identification proceeds in five steps. First, the first-order condition for the static materials choice recovers the materials elasticity. Second, the derivatives of the materials demand function ϕ_t with respect to capital and labor are identified. Third, substituting $\omega_{jt} = \phi_t^{-1}(k_{jt}, l_{jt}, m_{jt})$ into the production function identifies the reduced-form production function $\bar{f}_t(k_{jt}, l_{jt}, m_{jt})$; the chain rule then delivers the output elasticities of capital and labor—and hence returns to scale—at each observation. We term this a *reduced-form* identification result, because it identifies output elasticities only along the equilibrium manifold without pinning down the structure of the production function. Fourth, under a support condition, the full demand function ϕ_t is identified, which in turn identifies the productivity index ω_{jt}. Finally, given ω_{jt}, an empirically motivated homogeneity restriction delivers full identification of the structural production function.

2.2.1 Identification of the Materials Elasticity

Lemma 2.1. *Under Assumption 2.3, the first-order condition for the firm's material choice is equivalent to*

$$s_{jt} = \ln C + \ln\left(\frac{\partial f_t(k_{jt}, l_{jt}, m_{jt}, \omega_{jt})}{\partial m_{jt}}\right) - \epsilon_{jt}, \tag{3}$$

where $s_{jt} \equiv \ln(\rho_t M_{jt}/(P_t Y_{jt}))$ is the log material revenue share and $C \equiv \mathbb{E}[e^{\epsilon_{jt}}]$.

Proof. The FOC of (2) implies

$$P_t \frac{\partial F_t(K_{jt}, L_{jt}, M_{jt}, e^{\omega_{jt}})}{\partial M_{jt}} C = \rho_t.$$

Multiplying by $M_{jt}/(P_t Y_{jt})$ and using $Y_{jt} = F_t(\cdot)\, e^{\epsilon_{jt}}$ gives

$$C \frac{M_{jt}}{F_t(\cdot)} \frac{\partial F_t(\cdot)}{\partial M_{jt}} e^{-\epsilon_{jt}} = \frac{\rho_t M_{jt}}{P_t Y_{jt}},$$

and taking logs yields (3). □

Equation (3) is the standard share-regression relationship (GNR). Since s_{jt} is observed, Lemma 2.1 implies that the materials elasticity $\partial f_t/\partial m_{jt}$ is identified at each observation once ϵ_{jt} is recovered (which is straightforward, as shown below).

A result similar to Lemma 3 is derived independently by Li and Sasaki (2024).

2.2.2 Identification of the Derivatives of the Materials Demand Function

Lemma 2.2. *Under Assumptions 2.1–2.4 and the normalization $\xi_{jt} \sim U(0,1)$, the derivatives $\partial\phi_t(k_{jt}, l_{jt}, \omega_{jt})/\partial k_{jt}$ and $\partial\phi_t(k_{jt}, l_{jt}, \omega_{jt})/\partial l_{jt}$ are identified at each observation.*

Proof. By Assumption 2.1, $\omega_{jt} = h_t(\omega_{jt-1}, \xi_{jt})$, and by Assumption 2.4, $m_{jt} = \phi_t(k_{jt}, l_{jt}, \omega_{jt})$ with $\omega_{jt-1} = \phi_{t-1}^{-1}(k_{jt-1}, l_{jt-1}, m_{jt-1})$. Substituting yields the reduced-form representation

$$m_{jt} = \bar{\phi}_t(k_{jt}, l_{jt}, k_{jt-1}, l_{jt-1}, m_{jt-1}, \xi_{jt}), \tag{4}$$

where $\bar{\phi}_t(\cdot) \equiv \phi_t\big(\cdot,\, h_t\big(\phi_{t-1}^{-1}(\cdot), \xi_{jt}\big)\big)$. Hence, for any (k, l, k', l', m', ξ), the partial derivatives of $\bar{\phi}_t$ with respect to (k, l) evaluated at (k, l, k', l', m', ξ) equal the corresponding partial derivatives of ϕ_t evaluated at (k, l, ω) with $\omega = h_t(\phi_{t-1}^{-1}(k', l', m'), \xi)$.

Under Assumption 2.2, $(k_{jt}, l_{jt}, k_{jt-1}, l_{jt-1}, m_{jt-1}) \perp \xi_{jt}$. With $\xi_{jt} \sim U(0,1)$, $\bar{\phi}_t$ is nonparametrically identified (e.g., Matzkin (2003)).[7] In particular, $\xi_{jt} = F(m_{jt} \mid k_{jt}, l_{jt}, k_{jt-1}, l_{jt-1}, m_{jt-1})$, so $\partial\bar{\phi}_t/\partial k_{jt}$ and $\partial\bar{\phi}_t/\partial l_{jt}$ are identified at each observation, which implies identification of $\partial\phi_t/\partial k_{jt}$ and $\partial\phi_t/\partial l_{jt}$ at each observation. □

The proof follows Ackerberg et al. (2022). They apply the same identification argument we use for ϕ_t to a value-added production function, with output y_{jt} (rather than m_{jt}) as the dependent variable.

Under the maintained assumptions, we identify these derivatives but not the full structural function ϕ_t, because h_t (and hence ω_{jt}) is not yet identified. This is sufficient for the reduced-form identification of output elasticities in Section 2.2.3.

2.2.3 Reduced-Form Identification of Output Elasticities

With $\partial\phi_t/\partial k_{jt}$ and $\partial\phi_t/\partial l_{jt}$ identified, we can recover output elasticities for each observed firm. We refer to this as *reduced-form* identification: elasticities are identified only on the equilibrium manifold

$$\{(k_{jt}, l_{jt}, m_{jt}, \omega_{jt}) : m_{jt} = \phi_t(k_{jt}, l_{jt}, \omega_{jt})\},$$

so the result does not support counterfactual analysis away from observed choices.[8] No assumptions beyond those in GNR are imposed, so the result nests GNR as the Hicks-neutral special case.

Theorem 2.1. *Under Assumptions 2.1–2.4, the output elasticities with respect to* (k_{jt}, l_{jt}, m_{jt}) *are identified for each observation in the data.*

Proof. Invert ϕ_t and substitute $\omega_{jt} = \phi_t^{-1}(k_{jt}, l_{jt}, m_{jt})$ into the production function:

$$y_{jt} = \bar{f}_t(k_{jt}, l_{jt}, m_{jt}) + \epsilon_{jt}, \qquad \bar{f}_t(k, l, m) \equiv f_t\big(k, l, m, \phi_t^{-1}(k, l, m)\big)\,. \tag{5}$$

By Assumption 2.2, $\epsilon_{jt} \perp (k_{jt}, l_{jt}, m_{jt})$, so $\bar{f}_t(k, l, m) = \mathbb{E}[y_{jt} \mid k_{jt} = k, l_{jt} = l, m_{jt} = m]$ is identified and ϵ_{jt} is the regression residual. With ϵ_{jt} identified, (3) identifies $\partial f_t/\partial m_{jt}$ pointwise.

[7] Any scalar ξ_{jt} with a strictly increasing CDF would suffice; in Section 2.3 we use $\xi_{jt} \sim N(0, 1)$ for estimation convenience.

[8] A related reduced-form result appears in Demirer (2020), who identifies output elasticities nonparametrically at observed data points for a factor-augmenting production function under an additional homothetic separability condition.

To recover the elasticities with respect to k_{jt} and l_{jt}, differentiate (5) (suppressing subscripts) to obtain

$$\bar{f}_k = f_k + f_\omega(\phi^{-1})_k, \quad \bar{f}_l = f_l + f_\omega(\phi^{-1})_l, \quad \bar{f}_m = f_m + f_\omega(\phi^{-1})_m.$$

Using the implicit-function identities $(\phi^{-1})_k/(\phi^{-1})_m = -\phi_k$ and $(\phi^{-1})_l/(\phi^{-1})_m = -\phi_l$, we get

$$f_k = \bar{f}_k + \phi_k(\bar{f}_m - f_m), \tag{6}$$

$$f_l = \bar{f}_l + \phi_l(\bar{f}_m - f_m), \tag{7}$$

which identifies f_k and f_l pointwise since $\bar{f}_k, \bar{f}_l, \bar{f}_m$ are identified from $\bar{f}_t$, f_m is identified from (3), and ϕ_k, ϕ_l are identified by Lemma 2.2. □

Theorem 2.1 does not pin down elasticities at counterfactual (k, l, m, ω) bundles off the equilibrium manifold, since off the manifold m need not satisfy $m = \phi_t(k, l, \omega)$. Indeed, the structural production function is not fully identified under the maintained assumptions.

Theorem 2.2. *Under Assumptions 2.1–2.4, the structural production function f_t is not identified.*

Proof. See Appendix F. □

Although counterfactual analysis is not available, Theorem 2.1 identifies economically relevant objects—marginal products, factor intensities, and returns to scale—at each observation, and it enables tests of additional restrictions such as Hicks neutrality or homogeneity. We next show that full identification follows once a homogeneity restriction is imposed.

2.2.4 Full Identification of the Materials Demand Function

To fully identify the structural production function, we first establish full identification of the materials demand function ϕ_t. Because ϕ_t contains a nonseparable unobservable ω_{jt}, it is generally identified only up to a monotone transformation; full identification therefore reduces to an *ordering* problem for the latent ω_{jt} (Matzkin, 2003; Ackerberg et al., 2022). A support condition is needed to pin down this ordering.

Let $x_{jt} = (k_{jt}, l_{jt})$ and $v_{jt-1} = (k_{jt-1}, l_{jt-1}, m_{jt-1})$. Write $\mathscr{S}_t^{xvm}$ for the support of $(x_{jt}, v_{jt-1}, m_{jt})$ (and analogously $\mathscr{S}_t^{xv}, \mathscr{S}_t^{x|v}$, etc.).

Assumption 2.5 (Support Condition)**.** (i) The support of ω_{jt} conditional on ω_{jt-1} does not depend on ω_{jt-1}. (ii) $\mathscr{S}_t^v$ is path-connected. (iii) For every $v_0 \in \mathscr{S}_t^v$, $\mathscr{S}_t^{x|v_0}$ has nonempty interior. (iv) For every $v_0 \in \mathscr{S}_t^v$, there exists $\varepsilon > 0$ such that $\bigcap_{\substack{v \in \mathscr{S}_t^v \\ \|v - v_0\| < \varepsilon}} \mathscr{S}_t^{x|v}$ has nonempty interior.

Part (i) is a regularity condition on the productivity process, of a type routinely imposed on unobservables in econometrics: it requires that h_t can steer ω_{jt} to any part of its support regardless of ω_{jt-1}, though the *distribution* of ω_{jt} given ω_{jt-1} remains unrestricted. Parts (ii) and (iii) are minimal connectivity and relevance requirements. Part (iv)—the key condition—says that nearby values of v share a common region of x-support; intuitively, $\mathscr{S}_t^{x|v}$ varies continuously with v, without discrete jumps. Parts (ii)–(iv) are strictly weaker than requiring a global common support point for all pairs $v^A, v^B \in \mathscr{S}_t^v$, and strictly weaker than convexity of $\mathscr{S}_t^{xv}$, while remaining primitive and stated entirely in terms of the observable support.[9] We view the support condition primarily as a regularity requirement: we do not impose it directly in estimation, nor do we attempt to enforce it in the data.

Under these conditions, $\phi_t(k_{jt}, l_{jt}, \omega_{jt})$ is fully identified, and hence the latent productivity ω_{jt} is recovered; as a byproduct, the Markov transition $h_t(\omega_{jt-1}, \xi_{jt})$ is identified as well. Because both ω_{jt} and ξ_{jt} are latent and enter nonseparably, two normalizations are required.

Lemma 2.3. *Under Assumptions 2.1–2.5 and the normalizations $\omega_{jt} \sim U(0,1)$ and $\xi_{jt} \sim U(0,1)$, the materials demand function ϕ_t is fully identified.*

Proof. See Appendix F. □

The proof follows Ackerberg et al. (2022) and proceeds in two steps. In the first step, identification of the reduced-form function $\bar{\phi}_t(x, v, \xi)$ is established via a control-function argument: once lagged inputs are conditioned on, current capital and labor are independent of current productivity. In the second step, the identified $\bar{\phi}_t$ is used to rank firms by their implied productivity levels under Assumption 2.5. The conditional-independence restriction is the main economic content; the support condition serves as a regularity requirement enabling the ranking.

Full identification of ϕ_t implies identification of productivity ω_{jt} even before imposing structure on the production function.[10] With a nonseparable technology, however, ω_{jt} is

[9] Even weaker—but less primitive—conditions that avoid part (i) entirely are available; see Conditions 4–6 of Ackerberg et al. (2022).

[10] Productivity can also be identified directly from the production function. Because

$$y_{jt} = f_t(k_{jt}, l_{jt}, m_{jt}, \omega_{jt}) + \epsilon_{jt} \equiv \bar{f}_t(k_{jt}, l_{jt}, m_{jt}) + \epsilon_{jt},$$

identified only up to a monotone transformation; its economic scale is determined jointly with the structural production function.

A distinctive feature of our approach, relative to the conventional proxy-variable literature, is that it identifies the intermediates demand function. Recent independent work by Kasahara and Sugita (2020) applies the transformation-model approach of Chiappori et al. (2015) to identify the input demand function, but still assumes Hicks neutrality. Zeng (2021) likewise uses results from Ackerberg et al. (2022) to identify an investment demand function, focusing on value-added production functions.

2.2.5 Full Identification of the Production Function under Homogeneity

With ϕ_t (and hence ω_{jt}) fully identified, an empirically motivated homogeneity restriction delivers full identification of the structural production function.[11]

Assumption 2.6 (Homogeneity)**.** For each fixed ω_{jt}, $F_t(K_{jt}, L_{jt}, M_{jt}, e^{\omega_{jt}})$ is homogeneous of arbitrary degree α in (K_{jt}, L_{jt}, M_{jt}).

Cobb–Douglas and CES technologies both satisfy Assumption 2.6. Moreover, the limited dispersion in returns-to-scale estimates obtained under the reduced-form identification (Figure 10) provides empirical support for this restriction. At the same time, homogeneity remains compatible with non-Hicks-neutral productivity and flexible substitution patterns across inputs. For example, the following nonseparable nested CES technology is homogeneous and nests the standard CES specification with labor-augmenting productivity:

$$Y_{jt} = \left(\beta_k[\kappa(\omega_{jt})K_{jt}]^{\rho} + (1-\beta_k)\Big((1-\beta_m)[\lambda(\omega_{jt})L_{jt}]^{\rho_1} + \beta_m[\mu(\omega_{jt})M_{jt}]^{\rho_1}\Big)^{\rho/\rho_1}\right)^{\alpha/\rho},$$

where $\kappa(\omega_{jt})$, $\lambda(\omega_{jt})$, and $\mu(\omega_{jt})$ are functions of the same productivity index ω_{jt}.

Under homogeneity, identification proceeds in two further steps. First, the share equation (3) identifies the *structural* materials elasticity as a function of inputs and productivity. Second, integrating this elasticity recovers the production function. The key is that homogeneity delivers a separable representation in which the materials elasticity depends on

identification of $\bar{f}_t$ implies identification of ϵ_{jt}. It follows that

$$\bar{y}_{jt} \equiv y_{jt} - \epsilon_{jt} = f_t\big(k_{jt}, l_{jt}, \phi_t(k_{jt}, l_{jt}, \omega_{jt}), \omega_{jt}\big) \equiv \tilde{f}_t(k_{jt}, l_{jt}, \omega_{jt}).$$

Under mild conditions, $\tilde{f}_t$ is strictly increasing in ω_{jt}, so ω_{jt} is identified in the same way as from the input-demand function ϕ_t.

[11] Alternatively, Navarro and Rivers (2018) provide conditions for full identification of a nonseparable production function under a separability restriction weaker than Hicks neutrality.

$(\tilde{l}_{jt}, \tilde{m}_{jt}, \omega_{jt})$ rather than on $(k_{jt}, l_{jt}, m_{jt}, \omega_{jt})$—breaking the functional dependence—so the share equation pins down the elasticity structurally rather than only along the equilibrium manifold. A more intuitive way to see this is that homogeneity imposes a functional-form restriction that enables extrapolation from the equilibrium manifold, identifying output elasticities at input–productivity combinations not observed in equilibrium and thereby pinning down the entire structural production function.

Under Assumption 2.6, the production function admits the separable representation

$$Y_{jt} = K_{jt}^{\alpha} F_t\left(1, \frac{L_{jt}}{K_{jt}}, \frac{M_{jt}}{K_{jt}}, e^{\omega_{jt}}\right) e^{\epsilon_{jt}} \quad \Leftrightarrow \quad y_{jt} = \alpha k_{jt} + g_t(\tilde{l}_{jt}, \tilde{m}_{jt}, \omega_{jt}) + \epsilon_{jt}, \tag{8}$$

where $\tilde{l}_{jt} \equiv l_{jt} - k_{jt}$ and $\tilde{m}_{jt} \equiv m_{jt} - k_{jt}$.

We impose a mild regularity condition to rule out perfect functional dependence of $\tilde{m}_{jt}$ on $(\tilde{l}_{jt}, \omega_{jt})$:

Assumption 2.7. With probability one,

$$\frac{\partial \phi_t(k_{jt}, l_{jt}, \omega_{jt})}{\partial k_{jt}} + \frac{\partial \phi_t(k_{jt}, l_{jt}, \omega_{jt})}{\partial l_{jt}} \neq 1.$$

Since ϕ_t is identified, Assumption 2.7 is directly testable.

The next result identifies the *structural* materials elasticity. Because this requires ω_{jt}, we invoke the support condition (Assumption 2.5) used to identify ϕ_t and ω_{jt}.

Theorem 2.3. *Under Assumptions 2.1–2.7, the share equation can be written as*

$$s_{jt} = \ln C + \ln\left(\frac{\partial g_t(\tilde{l}_{jt}, \tilde{m}_{jt}, \omega_{jt})}{\partial \tilde{m}_{jt}}\right) - \epsilon_{jt},$$

and the function $\partial g_t(\tilde{l}, \tilde{m}, \omega)/\partial \tilde{m}$ is identified.

Proof. Lemma 2.1 implies (3). Using (8), $\partial f_t/\partial m_{jt} = \partial g_t/\partial \tilde{m}_{jt}$. With ω_{jt} identified and Assumption 2.7 ruling out functional dependence, $\mathbb{E}[s_{jt} \mid \tilde{l}_{jt}, \tilde{m}_{jt}, \omega_{jt}]$ identifies $\ln C + \ln(\partial g_t/\partial \tilde{m}_{jt})$, and exponentiation yields $\partial g_t/\partial \tilde{m}_{jt}$. □

We now obtain full identification of the structural production function.

Theorem 2.4. *Under Assumptions 2.1–2.7, the structural production function f_t is identified.*

Proof. On the equilibrium manifold, $\tilde{m}_{jt} = \phi_t(\tilde{l}_{jt} + k_{jt},\, k_{jt},\, \omega_{jt}) - k_{jt}$. For fixed $(\tilde{l}_{jt}, \omega_{jt})$, $\partial \tilde{m}_{jt}/\partial k_{jt} = \partial\phi_t/\partial k + \partial\phi_t/\partial l - 1 \neq 0$ by Assumption 2.7, so varying k_{jt} sweeps $\tilde{m}_{jt}$ over an interval. Fix any $\tilde{m}_0(\tilde{l}_{jt}, \omega_{jt})$ in this interval. Integrating the identified $\partial g_t/\partial \tilde{m}$ (Theorem 2.3) gives

$$\int_{\tilde{m}_0(\tilde{l}_{jt},\omega_{jt})}^{\tilde{m}_{jt}} \frac{\partial g_t(\tilde{l}_{jt}, \tau, \omega_{jt})}{\partial \tau}\, d\tau = g_t(\tilde{l}_{jt}, \tilde{m}_{jt}, \omega_{jt}) - \delta_t(\tilde{l}_{jt}, \omega_{jt}), \tag{9}$$

where $\delta_t(\tilde{l}_{jt}, \omega_{jt}) \equiv g_t(\tilde{l}_{jt},\, \tilde{m}_0(\tilde{l}_{jt}, \omega_{jt}),\, \omega_{jt})$. Substituting (9) into (8) yields

$$y_{jt} - \int_{\tilde{m}_0(\tilde{l}_{jt},\omega_{jt})}^{\tilde{m}_{jt}} \frac{\partial g_t(\tilde{l}_{jt}, \tau, \omega_{jt})}{\partial \tau}\, d\tau = \alpha\, k_{jt} + \delta_t(\tilde{l}_{jt}, \omega_{jt}) + \epsilon_{jt},$$

a partially linear model with identified left-hand side. Since $\mathbb{E}[\epsilon_{jt} \mid k_{jt}, \tilde{l}_{jt}, \omega_{jt}] = 0$ and δ_t does not depend on k_{jt}, Robinson (1988) identifies α and δ_t pointwise. Substituting back into (9) recovers g_t, and (8) then identifies f_t. □

With f_t identified, counterfactual analysis becomes feasible.

2.3 Estimation

We now describe easy-to-implement sieve-based estimators.[12] The first step estimates the materials policy ϕ_t and the reduced-form production function $\bar{f}_t$, and recovers output elasticities using the reduced-form identification results. The second step specifies a homogeneous technology and estimates its parameters by matching the reduced-form elasticities to their model-implied counterparts. In principle, our identification arguments permit the production function to vary over time, but in the empirical application we hold the structure of the production function fixed over the sample period and let shifts in productivity ω_{jt} capture technical change. Accordingly, we suppress the time subscript t in what follows, though the estimation procedures developed below can be applied period by period if one wishes to allow the corresponding parameters to vary over time.

Estimation of Output Elasticities at Each Observation. We begin by estimating the materials policy function ϕ. The support condition is not imposed directly in estimation; instead, it serves as a regularity condition ensuring identification of ω_{jt}. The key restriction used for estimation is the conditional-independence assumption: conditional on lagged in-

[12] See Chen (2007) for a review of sieve estimators.

puts, ω_{jt} is independent of (k_{jt}, l_{jt}). Under this restriction, the conditional likelihood of m_{jt} can be expressed as the density of ξ_{jt} through a change of variables.

To avoid numerical inversion in ω, we specify the inverse policy ϕ^{-1} and estimate it directly. Specifically, we approximate

$$\begin{aligned}\omega_{jt} = \phi^{-1}(k_{jt}, l_{jt}, m_{jt}) &= p_k k_{jt} + p_l l_{jt} + m_{jt} + p_{kk} k_{jt}^2 + p_{kl} k_{jt} l_{jt} + p_{km} k_{jt} m_{jt} + p_{ll} l_{jt}^2 \\ &\quad + p_{lm} l_{jt} m_{jt} + p_{klm} k_{jt} l_{jt} m_{jt} + p_{kmm} k_{jt} m_{jt}^2 + p_{lmm} l_{jt} m_{jt}^2,\end{aligned} \tag{10}$$

imposing the normalization $\phi^{-1}(0, 0, m) = m$.[13] The sieve dimension can be adjusted to sample size.

For the Markov process, we likewise estimate the inverse transition:

$$\xi_{jt} = h^{-1}(\omega_{jt-1}, \omega_{jt}) = h_0 + h_c \omega_{jt} + h_l \omega_{jt-1} + h_{cc} \omega_{jt}^2 + h_{cl} \omega_{jt} \omega_{jt-1} + h_{ll} \omega_{jt-1}^2, \tag{11}$$

and normalize $\xi_{jt} \sim N(0, 1)$.

We estimate $(\mathbf{p}, \mathbf{h})$ by (partial) maximum likelihood:

$$\max_{\mathbf{p},\mathbf{h}} \sum_{j=1}^{J} \sum_{t=2}^{T} \log p\Big(m_{jt} \,\Big|\, {}^{k_{jt}, \ell_{jt}, k_{jt-1},}_{\ell_{jt-1}, m_{jt-1}}\Big) = \sum_{j=1}^{J} \sum_{t=2}^{T} \left[\log p(\xi_{jt}) + \log \frac{\partial h^{-1}(\omega_{jt})}{\partial \omega_{jt}} + \log \frac{\partial \phi^{-1}(x_{jt}, m_{jt})}{\partial m_{jt}} \right],$$

where the equality follows from a change of variables that introduces the Jacobian terms associated with h^{-1} and ϕ^{-1}.

The Jacobian terms also serve as implicit shape enforcement. Under the maintained monotonicity restrictions, $\partial h^{-1}(\omega)/\partial \omega > 0$ and $\partial \phi^{-1}(x, m)/\partial m > 0$, so the log-likelihood is well-defined only on the positive-derivative region; as either derivative approaches zero from above, its log term diverges to $-\infty$, strongly discouraging violations. For numerical stability, we implement these log-Jacobians using smoothed penalties based on a `log_softplus` transform.[14]

The second step estimates the reduced-form production function $\bar{f}$. Given $\omega_{jt} = \phi^{-1}(k_{jt}, l_{jt}, m_{jt})$, equation (5) is free of endogeneity, so $\bar{f}$ can be estimated by least squares. We approximate $\bar{f}$ with a complete third-order polynomial and estimate its coefficients $\mathbf{b}$

[13] With a nonseparable unobservable, such normalizations are without loss of generality; see Matzkin (2003).

[14] `log_softplus` is the smooth mapping $\log\big(\text{softplus}(x)\big)$ with $\text{softplus}(x) \equiv \log(1 + e^x)$. It behaves like $\log(x)$ when x is sufficiently positive (since $\text{softplus}(x) \approx x$), while remaining finite and smooth for $x \leq 0$ (since $\text{softplus}(x) \approx e^x$ when $x \ll 0$), which avoids undefined logs and improves numerical stability.

from

$$\text{(12)} \qquad \min_{\mathbf{b}} \sum_{j,t} \left(y_{jt} - \sum_{r_k+r_l+r_m \leq 3} b_{r_k,r_l,r_m}\, k_{jt}^{r_k} l_{jt}^{r_l} m_{jt}^{r_m} \right)^2, \qquad r_k, r_l, r_m \geq 0.$$

Let $\hat{\epsilon}_{jt}$ denote the fitted residual and set $\hat{C} \equiv (JT)^{-1} \sum_{j,t} \exp(\hat{\epsilon}_{jt})$.[15] The share equation (3) then yields the pointwise materials elasticity

$$\text{(13)} \qquad \frac{\partial \hat{f}(k_{jt}, l_{jt}, m_{jt}, \omega_{jt})}{\partial m_{jt}} = \exp\Big(s_{jt} + \hat{\epsilon}_{jt} - \ln \hat{C} \Big)\,.$$

The elasticities with respect to k and l follow from the identification results (6)–(7):

$$\text{(14)} \qquad \frac{\partial \hat{f}}{\partial k} = \frac{\partial \hat{\tilde{f}}}{\partial k} + \frac{\partial \hat{\phi}}{\partial k} \left(\frac{\partial \hat{\tilde{f}}}{\partial m} - \frac{\partial \hat{f}}{\partial m} \right),$$

$$\text{(15)} \qquad \frac{\partial \hat{f}}{\partial l} = \frac{\partial \hat{\tilde{f}}}{\partial l} + \frac{\partial \hat{\phi}}{\partial l} \left(\frac{\partial \hat{\tilde{f}}}{\partial m} - \frac{\partial \hat{f}}{\partial m} \right),$$

where derivatives of $\hat{\tilde{f}}$ and $\hat{\phi}$ are available in closed form under our sieve specifications.

Estimation of the Structural Production Function. We specify the log production function as a second-order polynomial in $(k_{jt}, l_{jt}, m_{jt}, \omega_{jt})$, allowing ω_{jt} to interact with inputs:

$$\begin{aligned} y_{jt} &= f(k_{jt}, l_{jt}, m_{jt}, \omega_{jt}) + \epsilon_{jt} \\ \text{(16)} \qquad &= \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \beta_\omega \omega_{jt} + \beta_{kk} k_{jt}^2 + \beta_{kl} k_{jt} l_{jt} + \beta_{km} k_{jt} m_{jt} + \beta_{k\omega} k_{jt} \omega_{jt} \\ &\quad + \beta_{ll} l_{jt}^2 + \beta_{lm} l_{jt} m_{jt} + \beta_{l\omega} l_{jt} \omega_{jt} + \beta_{mm} m_{jt}^2 + \beta_{m\omega} m_{jt} \omega_{jt} + \beta_{\omega\omega} \omega_{jt}^2 + \epsilon_{jt}. \end{aligned}$$

Note that ω_{jt} has already been estimated via the inversion of ϕ and can therefore be treated as an observed regressor. Nevertheless, (16) cannot simply be estimated by regressing y_{jt} on $(k_{jt}, l_{jt}, m_{jt}, \omega_{jt})$, because the regressors are functionally dependent: $m_{jt} = \phi_t(k_{jt}, l_{jt}, \omega_{jt})$. As our identification arguments make clear, recovering the structural production function requires combining the homogeneity condition with the materials first-order condition. We impose homogeneity on (16) via a set of linear equality restrictions on the parameters and solve out a subset of coefficients to reduce the dimension of the parameter vector; see Ap-

[15] Hats denote estimators throughout.

pendix G for details.

We then estimate $\boldsymbol{\beta}$ by minimum distance, matching the model-implied elasticities $(\check{f}_{k_{jt}}, \check{f}_{l_{jt}}, \check{f}_{m_{jt}}, \check{f}_{\omega_{jt}})$ from the homogeneous production function (16) to the reduced-form elasticities $(\hat{f}_{k_{jt}}, \hat{f}_{l_{jt}}, \hat{f}_{m_{jt}}, \hat{f}_{\omega_{jt}})$ obtained in Section 2.3:

$$\min_{\boldsymbol{\beta}} \sum_{j,t} \Big(w_k(\hat{f}_{k_{jt}} - \check{f}_{k_{jt}})^2 + w_l(\hat{f}_{l_{jt}} - \check{f}_{l_{jt}})^2 + w_m(\hat{f}_{m_{jt}} - \check{f}_{m_{jt}})^2 + w_\omega(\hat{f}_{\omega_{jt}} - \check{f}_{\omega_{jt}})^2 \Big), \tag{17}$$

with weights w_i set equal to the inverse of the estimated variance of $\hat{f}_{i_{jt}}$, $i \in \{k, l, m, \omega\}$. The criterion (17) identifies all elements of $\boldsymbol{\beta}$ except β_0, which is recovered from the moment restriction $E(\epsilon_{jt} \mid k_{jt}, l_{jt}, m_{jt}, \omega_{jt}) = 0$.

Note that the materials first-order condition has already been fully exploited in constructing the reduced-form elasticities. The minimum-distance matching therefore mirrors the logic of our identification argument: homogeneity enables us to extrapolate from the equilibrium manifold, thereby identifying output elasticities at input–productivity combinations not observed in equilibrium and pinning down the entire structural production function.

Empirically, minimizing (17) typically yields estimated structural production functions that largely respect standard theoretical regularities, including concavity and monotonicity/complementarity. To enforce these shape restrictions more tightly, we can optionally augment the objective with one-sided quadratic penalties for constraint violations; see Appendix G for details.

This two-step approach—estimate the reduced form first, then fit the structural model to the reduced-form objects subject to shape constraints—leaves the reduced-form estimator unchanged. The general strategy of estimating flexibly first and enforcing shape restrictions in a second step is well established (e.g., Mammen et al., 2001; Chernozhukov et al., 2009). With a penalty weight diverging at an appropriate rate, the penalized estimator is first-order asymptotically equivalent to the corresponding constrained extremum estimator (Gallant et al., 2022).

We apply these estimators to Chilean and Colombian manufacturing data in Appendix I; the results support the homogeneity restriction and document systematic overstatement of the labor elasticity under Hicks neutrality.

3 Imperfect Competition

We now extend the analysis to imperfect competition in the output market. The data and timing are as in Section 2, except that firms face downward-sloping demand. As in much of the production-function literature, we focus on the empirically common case in which only revenues—not physical quantities—are observed.[16]

With imperfect competition, revenues confound quantity and price variation, creating the familiar problem of disentangling markups from returns to scale (Klette and Griliches, 1996; De Loecker, 2011; Bond et al., 2020).[17] We address this by calibrating RTS for two reasons: (i) economically plausible values provide a tight prior on its magnitude; and (ii) our main empirical objects—the directions and relative magnitudes of technical-change bias—are *invariant* to the calibrated value. In addition, under homogeneity, markups are identified only up to a multiplicative constant, but this does not affect their dispersion or time trends.

3.1 Setup

Market Structure and Demand. With revenue data, consistent recovery of the production function and productivity requires modeling demand. Following Klette and Griliches (1996) and De Loecker (2011), we assume monopolistic competition and adopt a flexible nonparametric generalization of their CES demand system:

$$Q_{jt} = \tilde{D}_t\left(\frac{P_{jt}}{P_{Jt}}, e^{u_{jt}}\right), \tag{18}$$

where demand depends on the firm's relative price P_{jt}/P_{Jt} (nominal price deflated by the industry price index) and a demand shifter $e^{u_{jt}}$. Following Goldberg (1995) and De Loecker (2011), we decompose the demand shifter as

$$e^{u_{jt}} = G_t(Z_{jt}, e^{\epsilon_{jt}}), \tag{19}$$

where Z_{jt} collects observed demand shifters[18] and $e^{\epsilon_{jt}}$ captures an idiosyncratic disturbance. This setup extends Klette and Griliches (1996) and De Loecker (2011): their aggregate shifter (e.g., Q_{Jt}) can be included in Z_{jt}. In our baseline theoretical setup, we abstract from such aggregate shifters, with common aggregate variation represented parsimoniously

[16] Appendix A considers the case in which output prices and quantities are observed. In that setting, point identification can be achieved without calibrating returns to scale.

[17] For example, proportional increases in inputs and revenue may reflect constant returns under perfect competition, or increasing returns offset by lower prices under a downward-sloping demand curve.

[18] Examples of demand shifters in the literature include quota protection, export status, and tariff levels (De Loecker, 2011; De Loecker and Warzynski, 2012; Brandt et al., 2017). In our application to Chinese manufacturing, we use ownership type as a demand shifter to capture systematic differences in demand conditions between SOEs and private firms.

by the time index t.[19] Relative to CES, (18) allows demand elasticities to vary with both P_{jt}/P_{Jt} and $e^{u_{jt}}$, accommodating heterogeneous markups. Our modeling of demand is most closely related to Kasahara and Sugita (2020). We depart from their framework by allowing for non-neutral production functions on the supply side and by treating the demand shock differently.

Combining (18)–(19) yields

$$Q_{jt} = D_t\left(\frac{P_{jt}}{P_{Jt}}, Z_{jt}, e^{\epsilon_{jt}}\right).$$

Imposing the law of demand (strictly decreasing in P_{jt}/P_{Jt}), we invert for relative price:

$$\frac{P_{jt}}{P_{Jt}} = D_t^{-1}(Q_{jt}, Z_{jt}, e^{\epsilon_{jt}}), \tag{20}$$

so deflated revenue $R_{jt} \equiv (P_{jt}/P_{Jt})Q_{jt}$ satisfies

$$R_{jt} = D_t^{-1}(Q_{jt}, Z_{jt}, e^{\epsilon_{jt}})\ Q_{jt}. \tag{21}$$

Production Function. As in the perfect-competition case, we allow for a fully flexible, nonseparable technology:

$$Q_{jt} = F_t(K_{jt}, L_{jt}, M_{jt}, e^{\omega_{jt}}), \tag{22}$$

where Q_{jt} denotes unobserved physical output.[20]

Revenue Production Function. Taking logs and substituting the production function into (21) yields the revenue production function (in logs):

$$\begin{aligned} r_{jt} &= d_t^{-1}(q_{jt}, z_{jt}, \epsilon_{jt}) + q_{jt} \\ &= \gamma_t\big(f_t(k_{jt}, l_{jt}, m_{jt}, \omega_{jt}),\ z_{jt},\ \epsilon_{jt}\big) \\ &\equiv \psi_t(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \omega_{jt}, \epsilon_{jt}), \end{aligned} \tag{23}$$

[19] Pan (ongoing work) develops a complementary approach that exploits aggregate demand shifters to identify heterogeneous markups and returns to scale in a fully nonparametric environment.

[20] Unlike De Loecker (2011), we do not add measurement error to the production equation. Under CES demand and Hicks-neutral technology, measurement error and the idiosyncratic demand shock can be combined into a single reduced-form unobservable; in our nonseparable environment this aggregation typically fails without additional structure.

where $q_{jt} \equiv \log Q_{jt}$ and $r_{jt} \equiv \log R_{jt}$. Since ϵ_{jt} is an idiosyncratic demand disturbance, we normalize ψ_t to be strictly increasing in ϵ_{jt}. Equation (23) is the nonparametric analogue of the revenue-production-function representation in Klette and Griliches (1996) and De Loecker (2011). Our goal is to recover the underlying technology f_t from ψ_t and thereby study the factor bias of technological change.

Assumptions. We impose the following assumptions, which are the imperfect-competition counterparts of those maintained under perfect competition.

Assumption 3.1 (IPC-Markovian Productivity)**.** Productivity evolves according to a controlled first-order Markov process,

$$\omega_{jt} = h_t(\omega_{jt-1}, o_{jt-1}, \xi_{jt}),$$

where o_{jt-1} denotes predetermined control variables that may affect productivity dynamics, and ξ_{jt} is an innovation independent of all variables realized prior to it.

Assumption 3.2 (IPC-Timing and Information)**.** (i) (k_{jt}, l_{jt}, z_{jt}) are chosen before the firm observes ω_{jt}; (ii) m_{jt} is flexible and chosen after observing ω_{jt}; (iii) $(k_{jt}, l_{jt}, m_{jt}, z_{jt})$ are determined without knowledge of ϵ_{jt} and are independent of ϵ_{jt}.

Assumption 3.3 (IPC-Conditional Profit Maximization)**.** Conditional on $(K_{jt}, L_{jt}, Z_{jt}, \omega_{jt})$, the firm chooses M_{jt} each period to maximize expected profits,

$$\max_{M_{jt}} \mathbb{E}[\Gamma_t(F_t(K_{jt}, L_{jt}, M_{jt}, e^{\omega_{jt}}),\ Z_{jt},\ e^{\epsilon_{jt}}) \mid K_{jt}, L_{jt}, Z_{jt}, \omega_{jt}] - \rho_t M_{jt}, \tag{24}$$

where $\Gamma_t(\cdot)$ maps quantity and demand shifters into deflated revenue (as implied by (21)).

Assumption 3.4 (IPC-Strict Monotonicity)**.** Let $m_{jt} = \phi_t(k_{jt}, l_{jt}, z_{jt}, \omega_{jt})$ denote the materials policy implied by (24). The function ϕ_t is strictly increasing in ω_{jt}.

Assumption 3.5 (IPC-Support Condition)**.** Let x_{jt} additionally include z_{jt}, and let v_{jt-1} additionally include (o_{jt-1}, z_{jt-1}). Assumption 2.5 (i)–(iv) holds with these enlarged state vectors.

Assumption 3.1 generalizes the Markov specification to a controlled process, allowing the evolution of productivity to depend on additional state variables. In the literature, examples of the control variable o_{jt-1} include R&D, quota protection, and ownership status

(Doraszelski and Jaumandreu, 2013; De Loecker, 2011; Rubens et al., 2024). In our application to Chinese manufacturing, following Rubens et al. (2024), we specify o_{jt-1} as a vector of ownership indicators, allowing productivity dynamics to differ between SOEs and private firms.

Assumption 3.2 further imposes that z_{jt} be determined prior to the realization of ω_{jt}. This restriction can be relaxed using arguments analogous to those developed for labor; see Appendix B.

The main departure from perfect competition is Assumption 3.3: profits are now mediated by a nontrivial (and a priori unknown) revenue mapping Γ_t, which must be identified from the data, whereas under perfect competition Γ_t reduces to the identity.

Assumption 3.4 is unchanged except that ϕ_t now depends on z_{jt}. Under imperfect competition, sufficient conditions for strict monotonicity involve both the curvature of demand and the structure of the production function. A key insight of Biondi (2022) is that, under Hicks-neutral productivity, monotonicity is governed entirely by demand curvature, which alone can overturn it when markups are variable. With nonseparable productivity, the complementarity between intermediates and productivity introduces an additional force that interacts with demand curvature to determine whether monotonicity holds. We derive these conditions formally in Appendix C, and a post-estimation consistency check confirms that they are satisfied in our estimated model.

Assumption 3.5 is the same support requirement as before, with the state vectors enlarged to include z_{jt} in x_{jt} and (o_{jt-1}, z_{jt-1}) in v_{jt-1}.

3.2 Identification

Identification under imperfect competition proceeds in three steps. First, as in the perfect-competition case, we identify the materials policy ϕ_t. Second, we identify the reduced-form revenue production function $\bar{\psi}_t$ and show that output elasticities are identified up to a returns-to-scale (RTS) term. Third, calibrating RTS delivers point identification of the structural production function.

3.2.1 Identification of the Materials Demand Function

Lemma 3.1. *Under Assumptions 3.1–3.5, the materials demand function ϕ_t is identified.*[21]

The proof parallels that of Lemma 2.3 and is omitted.

[21] If the goal is only identification of derivatives at each observation (and hence elasticities and markups), the support condition in Assumption 3.5 is not required.

3.2.2 Identification up to Returns to Scale

With revenue data, markups and returns to scale (RTS) are not separately identified: different combinations can generate the same revenue production function. Nonetheless, under our maintained assumptions, output elasticities—and hence markups—are identified up to an RTS term, using arguments analogous to the reduced-form identification under perfect competition.

Invert the materials policy and substitute $\omega_{jt} = \phi_t^{-1}(k_{jt}, l_{jt}, z_{jt}, m_{jt})$ into (23) to obtain the reduced form

$$\begin{aligned} r_{jt} &= \psi_t(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \omega_{jt}, \epsilon_{jt}) \\ &= \psi_t\big(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \phi_t^{-1}(k_{jt}, l_{jt}, z_{jt}, m_{jt}), \epsilon_{jt}\big) \\ &\equiv \bar{\psi}_t(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon_{jt}), \end{aligned} \tag{25}$$

where $\bar{\psi}_t$ is identified—under the normalization $\epsilon_{jt} \sim N(0,1)$—by standard nonparametric arguments (e.g., Matzkin, 2003).

Differentiating $\bar{\psi}_t$, applying the chain rule, and invoking the materials first-order condition identifies the output elasticities up to RTS. Formally, let α_{jt} denote returns to scale and write the output elasticity as $f_{i_{jt}} = \alpha_{jt}\tilde{f}_{i_{jt}}$ for each $i \in \{k, l, m, \omega\}$.

Theorem 3.1. *Under Assumptions 3.1–3.5, $\tilde{f}_{i_{jt}}$ is identified for $i \in \{k, l, m, \omega\}$.*

Proof. See Appendix F. □

The proof parallels the reduced-form identification under perfect competition (Theorem 2.1), but the presence of the nonlinear unobservable ϵ_{jt} introduces substantial complications. We leverage the identified ϕ_t and $\bar{\psi}_t$ and exploit the materials first-order condition and cross-equation restrictions to solve for the output elasticities. However, the unknown revenue mapping Γ_t prevents point identification; elasticities are recovered only up to RTS.

Theorem 3.1 imposes no restriction on returns to scale: firms' RTS, α_{jt}, may be heterogeneous.

Identification up to RTS is economically meaningful. First, ratios of marginal products are identified without pinning down RTS. For instance,

$$\frac{MPK_{jt}}{MPL_{jt}} = \frac{f_{k_{jt}}}{f_{l_{jt}}}\frac{L_{jt}}{K_{jt}} = \frac{\tilde{f}_{k_{jt}}}{\tilde{f}_{l_{jt}}}\frac{L_{jt}}{K_{jt}},$$

since $f_{i_{jt}} = \alpha_{jt}\tilde{f}_{i_{jt}}$ and α_{jt} cancels. This implies that the directions and relative magnitudes of factor bias in technology are identified without specifying RTS. Second, by De Loecker and Warzynski (2012), markups satisfy $\mu_{jt} = f_{m_{jt}}/S_{jt}$; hence μ_{jt} is identified up to RTS as well. Below, we impose a homogeneity restriction and calibrate a common RTS α. Then μ_{jt} is identified up to the multiplicative constant α, so markup dispersion and time trends are identified and invariant to the calibrated value. This echoes, and extends beyond Cobb–Douglas, the result in De Ridder et al. (2024) that markup dispersion and trends are identifiable from revenue data. Appendix E provides a brief empirical analysis of markup dispersion and its evolution over time.

3.2.3 Point Identification under Calibrated Returns to Scale

Assumption 3.6 (IPC: Homogeneity with Known Degree)**.** The production function F_t is homogeneous in (K_{jt}, L_{jt}, M_{jt}) of known degree α.

Knowing α pins down returns to scale without firm-by-firm calibration. As in the perfect-competition case (Assumption 2.7), we impose a regularity condition ensuring that, under homogeneity, the ratio-form materials input $\tilde{m}_{jt} \equiv m_{jt} - k_{jt}$ is not a deterministic function of $(\tilde{l}_{jt}, \omega_{jt})$ along the equilibrium manifold.

Assumption 3.7 (IPC: No Functional Dependence)**.** With probability one,

$$\frac{\partial \phi_t(k_{jt}, l_{jt}, z_{jt}, \omega_{jt})}{\partial k_{jt}} + \frac{\partial \phi_t(k_{jt}, l_{jt}, z_{jt}, \omega_{jt})}{\partial l_{jt}} \neq 1.$$

Since ϕ_t is identified (Lemma 3.1), Assumption 3.7 is directly testable.

Theorem 3.2. *Under Assumptions 3.1–3.7, the structural production function f_t is identified up to an additive constant.*

Proof. See Appendix F. □

Under homogeneity, $f_t(k, l, m, \omega) = \alpha k + g_t(\tilde{l}, \tilde{m}, \omega)$. Since α is known, recovering f_t reduces to recovering the three-argument function g_t from its identified partial derivatives (f_l, f_m, f_ω). Assumption 3.7 guarantees that, for any fixed z_{jt}, the mapping $(k_{jt}, l_{jt}, \omega_{jt}) \mapsto (\tilde{l}_{jt}, \tilde{m}_{jt}, \omega_{jt})$ is a local diffeomorphism, so the support of $(\tilde{l}_{jt}, \tilde{m}_{jt}, \omega_{jt})$ contains a connected open subset of $\mathbb{R}^3$. The fundamental theorem of line integrals then recovers g_t up to an additive constant, paralleling the role of Assumption 2.7 in Section 2.2.5. When the demand shifter z_{jt} is continuous, homogeneity is not needed for the integration step: varying z_{jt} while holding $(k_{jt}, l_{jt}, \omega_{jt})$ fixed sweeps out an interval of m_{jt} values, so (k, l, m, ω) fills out a four-dimensional set and integration recovers f_t directly. When z_{jt} is discrete, as in our application, homogeneity of a known degree plays a dual role: it both pins down returns to scale and enables integration by reducing the problem to three dimensions.

3.3 Estimation

The estimation procedure under imperfect competition closely parallels that under perfect competition. We proceed in two steps. First, we estimate the materials policy ϕ_t and the reduced-form revenue production function $\bar{\psi}_t$, and recover output elasticities using the identification arguments above. Second, we specify a structural production function and estimate its parameters by matching the recovered elasticities to their model-implied counterparts. Throughout, we calibrate returns to scale (RTS) at a plausible value α. Our main empirical objects—the directions and relative magnitudes of technological bias—are invariant to the chosen RTS. In addition, the same invariance holds for markup trends and dispersion.

Estimation of Output Elasticities. The estimation of the materials policy ϕ follows Section 2.3, except that the policy includes the additional argument z_{jt} and the Markov process includes the additional control o_{jt-1}.

We then estimate the reduced-form revenue function $\bar{\psi}$. To avoid numerically inverting for ϵ_{jt}, we specify and estimate the inverse mapping $\bar{\psi}^{-1}$ directly. Specifically, we approximate $\bar{\psi}^{-1}$ with a second-order polynomial,

$$\begin{aligned}
\epsilon_{jt} &= \bar{\psi}^{-1}(k_{jt}, l_{jt}, m_{jt}, z_{jt}, r_{jt}) \\
&= \gamma_0 + \gamma_k k_{jt} + \gamma_l l_{jt} + \gamma_m m_{jt} + \gamma_z z_{jt} + \gamma_r r_{jt} + \gamma_{kk} k_{jt}^2 + \gamma_{ll} l_{jt}^2 + \gamma_{mm} m_{jt}^2 + \gamma_{rr} r_{jt}^2 \\
&\quad + \gamma_{kl} k_{jt} l_{jt} + \gamma_{km} k_{jt} m_{jt} + \gamma_{kr} k_{jt} r_{jt} + \gamma_{lm} l_{jt} m_{jt} + \gamma_{lr} l_{jt} r_{jt} + \gamma_{mr} m_{jt} r_{jt} \\
&\quad + \gamma_{kz} k_{jt} z_{jt} + \gamma_{lz} l_{jt} z_{jt} + \gamma_{mz} m_{jt} z_{jt} + \gamma_{rz} r_{jt} z_{jt},
\end{aligned} \tag{26}$$

imposing the normalization $\epsilon_{jt} \sim N(0,1)$. We estimate $\boldsymbol{\gamma}$ by maximum likelihood. Let $p(\cdot)$ denote the standard normal density. The log-likelihood contribution of observation (j,t) is

$$\log p(\epsilon_{jt}) + \log \frac{\partial \bar{\psi}^{-1}}{\partial r_{jt}},$$

so the estimator solves

$$\max_{\boldsymbol{\gamma}} \sum_{j,t} \left[\log p(\epsilon_{jt}) + \log \frac{\partial \bar{\psi}^{-1}}{\partial r_{jt}} \right].$$

Given $\hat{\boldsymbol{\gamma}}$, we obtain $\hat{\epsilon}_{jt} = \bar{\psi}^{-1}(k_{jt}, l_{jt}, m_{jt}, z_{jt}, r_{jt}; \hat{\boldsymbol{\gamma}})$.

To recover an estimator of the forward mapping $\bar{\psi}$, we refit r_{jt} on a second-order polynomial in $(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \hat{\epsilon}_{jt})$.[22] We summarize the fitted function as

$$\hat{\bar{\psi}}(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \hat{\epsilon}_{jt}) = a_{jt} + b_{jt}\hat{\epsilon}_{jt} + c\,\hat{\epsilon}_{jt}^2, \tag{27}$$

where a_{jt} and b_{jt} are (fitted) functions of $(k_{jt}, l_{jt}, m_{jt}, z_{jt})$, and c is a constant.

Mirroring the proof of Theorem 3.1, we derive closed-form expressions for $(d_{1,jt}, d_{2,jt})$ and hence for the output elasticities. Fix $(k_{jt}, l_{jt}, m_{jt}, z_{jt})$ and write

$$\hat{\bar{\psi}}(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon) = a_{jt} + b_{jt}\epsilon + c\,\epsilon^2, \qquad \epsilon \sim N(0,1).$$

Define

$$R_{jt}(\epsilon) \equiv \exp\big(\hat{\bar{\psi}}(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon)\big) = \exp(a_{jt} + b_{jt}\epsilon + c\,\epsilon^2), \qquad \Delta \equiv 1 - 2c,$$

where $\Delta > 0$ (equivalently, $c < 1/2$) is verified empirically. The following Gaussian moments are available in closed form:

$$\mathbb{E}\Big[e^{b\epsilon + c\epsilon^2}\Big] = \Delta^{-1/2} \exp\left(\frac{b^2}{2\Delta}\right), \tag{28}$$

$$\mathbb{E}\Big[\epsilon\, e^{b\epsilon + c\epsilon^2}\Big] = \frac{b}{\Delta}\,\mathbb{E}\Big[e^{b\epsilon + c\epsilon^2}\Big]. \tag{29}$$

Let $\bar{R}_{jt} \equiv \mathbb{E}_\epsilon[R_{jt}(\epsilon)]$. Using (28),

$$\bar{R}_{jt} = \exp(a_{jt})\,\Delta^{-1/2} \exp\left(\frac{b_{jt}^2}{2\Delta}\right).$$

For $i \in \{k, l, m\}$, the derivative of $\hat{\bar{\psi}}$ with respect to i is

$$\hat{\bar{\psi}}_i(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon) = a_{jt,i} + b_{jt,i}\,\epsilon,$$

[22] The second-order polynomial refit is a practical simplification that yields closed-form expressions for the estimated output elasticities. A higher-order polynomial could be used at the cost of requiring numerical integration. In practice, the refit approximates the data very well, with an R^2 close to one.

where $a_{jt,i} \equiv \partial a_{jt}/\partial i$ and $b_{jt,i} \equiv \partial b_{jt}/\partial i$ are implied by the refit in (27). Therefore,

$$\mathbb{E}_\epsilon\big[R_{jt}(\epsilon)\,\hat{\hat{\psi}}_i(\epsilon)\big] = \bar{R}_{jt}\left(a_{jt,i} + b_{jt,i}\frac{b_{jt}}{\Delta}\right). \tag{30}$$

Define, as in the proof,

$$d_{1,jt} \equiv \mathbb{E}_\epsilon\big[R_{jt}(\epsilon)\,\hat{\hat{\psi}}_m(\epsilon)\big], \qquad c_{1,jt}(\epsilon) \equiv \hat{\hat{\psi}}_k(\epsilon) + \hat{\hat{\psi}}_l(\epsilon) + \hat{\hat{\psi}}_m(\epsilon), \qquad d_{2,jt} \equiv \mathbb{E}_\epsilon\big[R_{jt}(\epsilon)\,c_{1,jt}(\epsilon)\big].$$

Thus $d_{1,jt}$ and $d_{2,jt}$ are computed in closed form from (30) (with $i = m$ and with $i \in \{k, l, m\}$ summed, respectively).

Next let

$$c_{2,jt} \equiv (\phi^{-1})_{k_{jt}} + (\phi^{-1})_{l_{jt}} + (\phi^{-1})_{m_{jt}},$$

and recall that ρ_t is the (deflated) materials price. The identified quantity $\tilde{f}_{\omega_{jt}} = f_{\omega_{jt}}/\alpha_{jt}$ is obtained from (51) in the proof of Theorem 3.1 as

$$\tilde{f}_{\omega_{jt}} = \frac{d_{1,jt} - \rho_t M_{jt}}{(\phi^{-1})_{m_{jt}}\, d_{2,jt} - c_{2,jt}\,(d_{1,jt} - \rho_t M_{jt})}.$$

For $i \in \{k, l, m\}$, we then recover $\tilde{f}_{i_{jt}} = f_{i_{jt}}/\alpha_{jt}$:

$$\tilde{f}_{i_{jt}} = \frac{\hat{\hat{\psi}}_i(\hat{\epsilon}_{jt})}{c_{1,jt}(\hat{\epsilon}_{jt})}\big(1 + c_{2,jt}\tilde{f}_{\omega_{jt}}\big) - \tilde{f}_{\omega_{jt}}\,(\phi^{-1})_{i_{jt}}.$$

Calibrating a common RTS α gives $f_{i_{jt}} = \alpha\,\tilde{f}_{i_{jt}}$ and $f_{\omega_{jt}} = \alpha\,\tilde{f}_{\omega_{jt}}$. Markups follow from De Loecker and Warzynski (2012): $\mu_{jt} = f_{m_{jt}}/S_{jt}$, where $S_{jt} \equiv \rho_t M_{jt}/R_{jt}$ is the materials expenditure share.

Estimation of the Structural Production Function. Given the estimated output elasticities $(\hat{f}_{k_{jt}}, \hat{f}_{l_{jt}}, \hat{f}_{m_{jt}}, \hat{f}_{\omega_{jt}})$, we specify the structural production function as a second-order polynomial and estimate its coefficients by matching these estimates to their model-implied counterparts, following the same minimum-distance procedure as in Section 2.3. In addition, we enforce complementarity between productivity and inputs via one-sided penalty terms, as described in Appendix G.[23] The implementation is otherwise identical to Section 2.3 and is omitted.

[23] In our application, this constraint binds only occasionally, and only for labor.

4 Empirical Results

4.1 Empirical Setting and Calibration

We apply our framework to study the directions and magnitudes of technical change in Chinese manufacturing over 1998–2007. Our empirical analysis is based on firm-level data from China's Annual Survey of Industrial Enterprises (ASIE), compiled by the National Bureau of Statistics (NBS). The ASIE covers all state-owned enterprises and all non-state firms above the administrative size threshold, defined as annual sales of at least 5 million RMB. It reports detailed production and accounting information, including total sales, intermediate input expenditures, capital stock measures, employment, and related firm characteristics. Notably, as is typical for large-scale manufacturing datasets, the survey does not contain physical output quantities or firm-level output prices. Details on the data and summary statistics are provided in Appendix H.

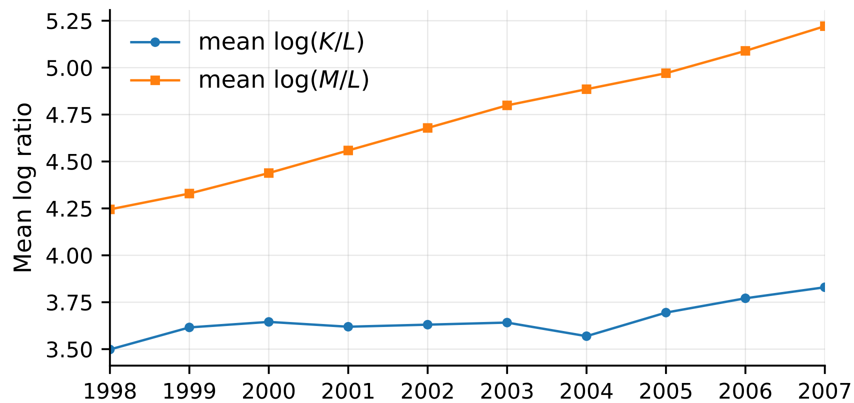
During the sample period, Chinese manufacturing experienced extraordinary productivity growth, as widely documented in the literature (e.g., Brandt et al., 2012; Zhu, 2012; Brandt et al., 2017). For example, Brandt et al. (2012) estimate that the weighted average annual productivity growth of incumbent firms was 2.85% under a gross output production function, one of the highest rates observed internationally. Consistent with these findings, Figure 2 also points to substantial productivity growth in Chinese manufacturing over our sample period under our nonseparable model.

Most existing empirical work on Chinese manufacturing imposes Hicks-neutral technology.[24] Under Hicks neutrality, productivity improvements scale all marginal products proportionally, leaving relative marginal products unchanged at any fixed input bundle. Hicks-neutral specifications therefore provide limited scope to assess whether productivity growth has been systematically more favorable to capital, labor, or intermediates. Moreover, under conventional homothetic production functions, Hicks-neutral productivity growth does not affect optimal input ratios when relative input prices are held fixed. In the data, however, the period of rapid productivity growth is accompanied by substantial capital and materials deepening: both the capital-to-labor ratio and the materials-to-labor ratio rise markedly over time (see Figure 1). Given the magnitude of productivity growth over this period, imposing Hicks neutrality would shut down an important technology-side channel through which the

[24] Notable exceptions include Zhang (2019) and Rubens et al. (2024), which consider factor-augmenting specifications in particular industries, namely steel and nonferrous metals. In contrast, we study the direction of technical change more broadly across Chinese manufacturing.

observed capital and materials deepening can be rationalized.

Figure 1: Trends in Capital and Materials Deepening



Note: This figure plots time trends in the mean log capital–labor and materials–labor ratios.

In light of these considerations, our primary objective is to identify whether technical change in Chinese manufacturing is biased toward particular factors of production. Our framework is designed to answer precisely this technology-side question. At the same time, because the paper is concerned with production-function estimation rather than a complete model of capital and labor demand, it remains agnostic about the forces governing input adjustment and the frictions operating in input markets. Accordingly, we do not interpret our estimates as delivering a precise decomposition of input deepening. Instead, we view the links between estimated technical bias and these broader outcomes as suggestive evidence on an important technology-side channel.

Throughout, we calibrate returns to scale (RTS) to 1.1. This choice is motivated by two considerations. First, revenue-based production-function estimation often yields modestly increasing returns (e.g., Klette and Griliches, 1996; De Loecker, 2011). Second, for Chinese manufacturing, existing evidence places RTS close to 1.1 (Hu et al., 2023). Importantly, our conclusions on the directions and relative magnitudes of factor bias and the evolution of marginal products are robust to plausible alternative RTS calibrations. In addition, while the level of estimated markups depends on the chosen RTS calibration, the time trend and dispersion of markups do not. Corresponding results on markups are reported in Appendix E.

Table 3 in the appendix reports the estimated mean output elasticities and implied mean markups. Overall, the resulting magnitudes are plausible and broadly consistent with existing evidence for Chinese manufacturing. Figure 7 further summarizes the distributions of the estimated second-order derivatives of the production function. The estimates conform well to standard theoretical regularities: inputs are predominantly complementary, marginal products are diminishing, and productivity is complementary to each input. These patterns provide useful discipline for the analysis that follows, which asks whether the substantial productivity growth documented above was neutral across inputs or systematically biased toward particular factors of production.

4.2 Bias in Technical Change

We formally define the bias of technical change. Let $\mathbf{I} = (K, L, M)$ denote the vector of inputs. Following Acemoglu (2002), we characterize factor bias in terms of how productivity shifts *relative* marginal products across inputs.

Definition 4.1 (Factor-biased technological change)**.** Technological change is biased toward I_1 relative to I_2 if

$$\frac{\partial}{\partial \omega}\left(\frac{F_{I_1}(\mathbf{I}, \omega)}{F_{I_2}(\mathbf{I}, \omega)}\right) > 0, \tag{31}$$

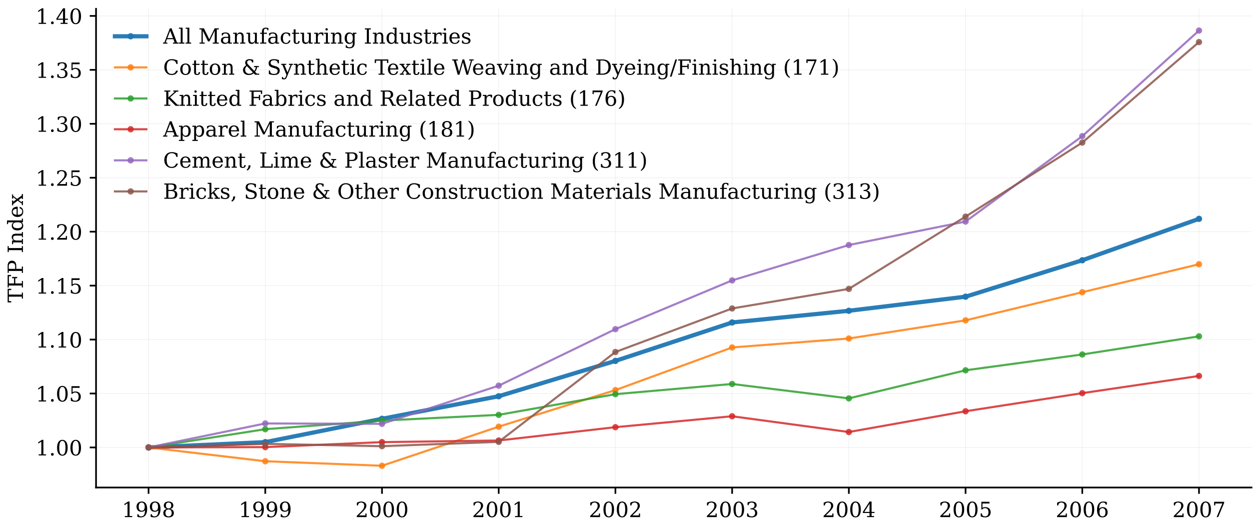
where $F_I(\mathbf{I}, \omega) \equiv \partial F(\mathbf{I}, \omega)/\partial I$. Condition (31) states that an increase in the productivity shifter ω raises the marginal product of I_1 proportionally more than that of I_2.

We focus on marginal products because they are the primitive objects governing firms' input choices. Under cost minimization or profit maximization, first-order conditions equate each input's shadow value to its marginal contribution to output, so relative marginal products pin down relative factor demands. When factor markets are competitive, equilibrium input prices equal marginal revenue products, and changes in F_I translate directly into changes in factor prices; under imperfect competition, the same holds up to a common markup (see De Loecker and Warzynski, 2012). Identifying which marginal products respond most to productivity improvements therefore provides a transparent summary of the direction of technical change and its distributional implications.

To convey the central message, we begin with a pooled specification that combines all manufacturing industries and estimate a single production function. We then examine heterogeneity by estimating separate production functions for each of the five largest three-digit industries.[25] The industry-specific results closely mirror those from the pooled specification, suggesting that the patterns emphasized below are not driven by any particular sector but

[25] We exclude the transport equipment industry from the analysis, as its production and market structures may not be well captured by our model.

Figure 2: Median Total Factor Productivity (TFP) Growth



Note: TFP growth is computed as output growth holding inputs fixed at their median levels, as productivity increases from its 1998 median to its 2007 median. Output is normalized to one in 1998.

instead arise broadly across manufacturing.

Figure 2 documents productivity growth itself, reporting the implied change in output as productivity rises from its 1998 median to its 2007 median, holding inputs fixed at their median levels.[26] Consistent with the existing literature, we find substantial productivity growth: under the pooled specification, TFP rises by nearly 20% over the decade, a magnitude that lies near the middle of the distribution across the five largest industries. Because we model demand explicitly, this estimate corresponds to quantity-based productivity growth (TFPQ), which is not confounded by demand-side variation, rather than the revenue-based productivity measure (TFPR) more commonly reported in the literature. For this reason, changes in this object are more appropriately interpreted as technical change.

The central empirical question, however, is not merely whether productivity rose, but whether the resulting technical change was neutral across inputs. Guided by Definition 4.1, Figure 3 traces the evolution of the ratios of marginal products,

$$\frac{\partial F(\mathbf{I}, \omega)/\partial I_1}{\partial F(\mathbf{I}, \omega)/\partial I_2},$$

as productivity increases from its 1998 median to its 2007 median, holding inputs fixed at their median levels. Two patterns stand out. First, across all industries, MPK and MPM rise

[26]The patterns highlighted in the paper are robust to using other representative values of the inputs.

more than MPL, indicating that technical change is least favorable to labor. Second, with the exception of Textiles (171), MPK also rises more than MPM, suggesting that technical change is predominantly capital-biased.

To quantify the magnitudes of these biases, Table 1 reports the ratios of marginal products evaluated at ω_{2007} relative to ω_{1998}, holding inputs fixed at their median levels. The differences are economically meaningful. In the pooled specification, productivity growth raises MPK by 32% and MPM by 22%, but MPL by only 12%. The industry-specific estimates reveal a similar pattern. Both the direction of factor bias and the relative magnitudes are robust to alternative RTS calibrations; Table 4 in the appendix confirms this by reporting the corresponding marginal-product ratios for RTS values of 1.00, 1.05, 1.10, 1.15, and 1.20.

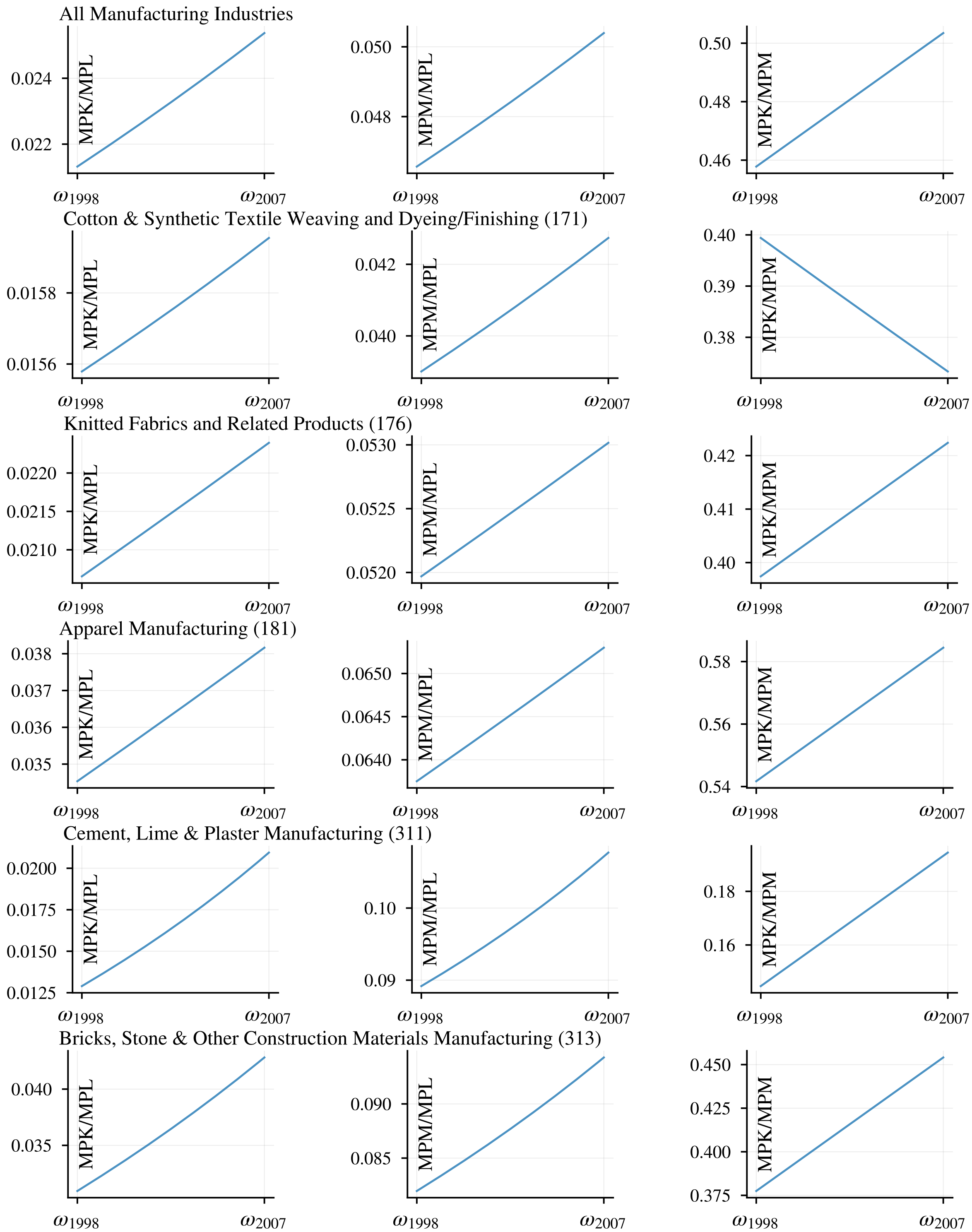
By contrast, under Hicks neutrality, all three marginal products would increase proportionally, implying that the entries within each row of Table 1 would be identical. These findings therefore provide direct evidence that productivity growth in Chinese manufacturing was systematically biased away from labor and, in most sectors, tilted most strongly toward capital.

Table 1: Bias in Technical Change: Productivity-Driven Growth in Marginal Products

	MPK(ω_{2007})/MPK(ω_{1998})	MPL(ω_{2007})/MPL(ω_{1998})	MPM(ω_{2007})/MPM(ω_{1998})
All	1.32	1.12	1.22
	(0.00)	(0.00)	(0.00)
Textiles (171)	1.12	1.08	1.19
	(0.02)	(0.01)	(0.01)
Fabrics (176)	1.16	1.08	1.10
	(0.02)	(0.01)	(0.01)
Apparel (181)	1.14	1.04	1.07
	(0.01)	(0.00)	(0.00)
Cement (311)	1.80	1.14	1.38
	(0.08)	(0.02)	(0.01)
Bricks (313)	1.64	1.24	1.37
	(0.06)	(0.02)	(0.02)

Note: The table reports ratios of marginal products evaluated at ω_{2007} and ω_{1998}, i.e., MPI(ω_{2007})/MPI(ω_{1998}) for $I \in \{K, L, M\}$, holding inputs fixed at their median levels. Bootstrap standard errors (200 replications) are reported in parentheses.

Figure 3: Bias in Technical Change: Marginal Rates of Technical Substitution (MRTS)



Note: The ratios MPK/MPL, MPM/MPL, and MPK/MPM are plotted against productivity as it rises from the 1998 median to the 2007 median, with inputs held fixed at the median levels.

4.3 Growth Decomposition of Marginal Products

The analysis in the previous subsection is a counterfactual exercise: we hold inputs fixed and trace how marginal products vary with productivity alone. The results indicate that productivity growth is least favorable to labor. Yet an observer examining the realized data over 1998–2007 would notice a seemingly contradictory pattern: the marginal product of labor rises the most among the three inputs over this period.

This apparent tension has substantive economic content. Under competitive factor markets, marginal products determine factor prices, so the sources of marginal-product growth speak directly to the determinants of wage growth relative to returns on capital and materials. The key question is whether MPL rose because productivity growth genuinely favored labor, or because workers came to operate with substantially more capital and materials in a technology featuring factor complementarities. The first channel would reflect a technology-driven improvement in labor's productive role; the second would indicate that the rise in MPL is a byproduct of the very capital and materials deepening that accompanies labor's *declining* relative importance.

To disentangle these channels, we decompose realized marginal-product growth into a productivity-growth component and an input-mix component using a Divisia/Törnqvist approximation.

For each $I \in \{K, L, M\}$, write $\mathrm{MP}_I(\mathbf{I}, \omega) \equiv F_I(\mathbf{I}, \omega) = \partial F(\mathbf{I}, \omega)/\partial I$. In continuous time, the Divisia decomposition implies

$$d \ln \mathrm{MP}_I = \eta_{IK}\, d \ln K + \eta_{IL}\, d \ln L + \eta_{IM}\, d \ln M + \eta_{I\omega}\, d\omega,$$

where the local elasticities are defined by

$$\eta_{IK}(\mathbf{I}, \omega) \equiv \frac{\partial \ln \mathrm{MP}_I(\mathbf{I}, \omega)}{\partial \ln K} = \frac{K}{\mathrm{MP}_I(\mathbf{I}, \omega)} \frac{\partial^2 F(\mathbf{I}, \omega)}{\partial I\, \partial K},$$

with analogous definitions for η_{IL}, η_{IM}, and $\eta_{I\omega}$. In practice, we use the discrete-time Divisia (Törnqvist/midpoint) approximation year by year:

$$\Delta \ln \mathrm{MP}_{I,jt} \approx \underbrace{\bar{\eta}_{IK,jt}\, \Delta \ln K_{jt} + \bar{\eta}_{IL,jt}\, \Delta \ln L_{jt} + \bar{\eta}_{IM,jt}\, \Delta \ln M_{jt}}_{\text{input mix and complementarity}} + \underbrace{\bar{\eta}_{I\omega,jt}\, \Delta\omega_{jt}}_{\text{productivity growth}} + \underbrace{\varepsilon_{I,jt}}_{\text{residual}},$$

where midpoint elasticities are computed by averaging endpoint values, for example

$$\bar{\eta}_{IK,jt} \equiv \tfrac{1}{2}\left[\eta_{IK}(\mathbf{I}_{jt}, \omega_{jt}) + \eta_{IK}(\mathbf{I}_{j,t-1}, \omega_{j,t-1})\right],$$

and similarly for the remaining terms. The residual $\varepsilon_{I,jt}$ captures higher-order discretization terms and measurement noise. Operationally, the "Total growth" reported below is the sum over years of the cross-sectional mean of $\Delta \ln \mathrm{MP}_{I,jt}$, and can therefore be interpreted as the cumulative change in $\ln \mathrm{MP}_I$ for the average plant over 1998–2007.

Table 2 reports the results. Three patterns emerge. First, among the three marginal products, MPL exhibits the largest increase over the decade: cumulative growth in ln MPL averages 0.76 in the pooled specification, compared with 0.34 for MPK and 0.20 for MPM. Second, the productivity-growth component is least favorable to labor: productivity growth contributes only 0.12 to MPL growth, compared with 0.32 for MPK and 0.22 for MPM. Third, MPL growth is driven predominantly by the input-mix component (0.64 of 0.76), whereas MPK and MPM growth is driven mainly by productivity growth, with Textiles (171) as the main exception for MPK. Put differently, the rise in MPL reflects primarily the fact that workers operate with more capital and materials in a technology featuring factor complementarities, while productivity growth itself accounts for only a modest share of the increase. The residual is negligible throughout, indicating that the midpoint approximation tracks the underlying decomposition closely. The industry-specific results reveal the same broad pattern, which is also robust to alternative RTS calibrations (see Table 5 in the appendix).

4.4 Biased Technical Change and Factor Deepening

A striking feature of Chinese manufacturing over 1998–2007 is the pronounced deepening of capital and materials relative to labor. As documented in Figure 1, both the capital–labor ratio and the materials–labor ratio rise markedly over this period. This deepening is also closely linked to the well-documented decline in the labor share over the same period (Berkowitz et al., 2015; Bai and Qian, 2010). Under a conventional homothetic production function with Hicks-neutral productivity, such deepening would be attributed entirely to changes in relative input prices. However, when productivity growth is biased, it shifts the relative marginal products of different inputs and can thereby contribute to factor deepening even holding input prices fixed. Our estimates point to a clear technology-side channel.

The mechanism is straightforward. Because productivity growth disproportionately raises the marginal products of capital and intermediates (Table 1), it increases the relative

Table 2: Growth Decomposition of Marginal Products, 1998–2007

	Total growth	Factor complementarity	Productivity growth	Residual
Panel A: MPK				
All	0.34	0.05	0.32	-0.02
Textiles (171)	0.48	0.32	0.16	-0.01
Fabrics (176)	0.00	-0.18	0.19	-0.01
Apparel (181)	-0.03	-0.19	0.18	-0.02
Cement (311)	0.78	0.20	0.60	-0.02
Bricks (313)	0.64	0.13	0.51	-0.00
Panel B: MPL				
All	0.76	0.64	0.12	0.00
Textiles (171)	0.94	0.82	0.12	0.00
Fabrics (176)	0.54	0.43	0.10	0.00
Apparel (181)	0.35	0.30	0.05	0.01
Cement (311)	0.91	0.79	0.12	0.00
Bricks (313)	0.89	0.73	0.15	0.01
Panel C: MPM				
All	0.20	-0.02	0.22	-0.00
Textiles (171)	0.21	-0.07	0.28	-0.00
Fabrics (176)	0.15	0.01	0.13	0.00
Apparel (181)	0.11	0.02	0.08	-0.00
Cement (311)	0.22	-0.10	0.32	0.00
Bricks (313)	0.30	-0.04	0.33	-0.00

Note: For each industry and each input $I \in \{K, L, M\}$, the table reports the cumulative average annual growth in $\log \mathrm{MP}_I$ over 1998–2007, and decomposes it—using the Divisia/Törnqvist procedure described above—into (i) a factor-complementarity component, (ii) a productivity-growth component, and (iii) a residual. When computing the averages, we exclude observations whose corresponding marginal products fall outside the 1st–99th percentile range.

profitability of reallocating toward those inputs. For cost-minimizing firms facing given input prices, and given the concavity of the estimated production function documented in Figure 7, this implies an outward shift in the relative demand for capital and materials vis-à-vis labor.

In particular, the biased technical change documented above helps rationalize an apparent tension in the data: the marginal product of labor rises the most among the three inputs over 1998–2007, yet labor's relative usage in the production process declines. The Divisia decomposition in Table 2 resolves this tension. Because productivity growth is least favorable to labor, it disproportionately raises the marginal products of capital and intermediates, encouraging capital and materials deepening. This deepening, in turn, raises MPL through factor complementarity—but it does so as a byproduct of the substitution away from labor, not as a reflection of productivity growth favoring labor. Indeed, the direct contribution of productivity growth to MPL growth is only 0.12, the smallest among the three inputs, whereas it accounts for nearly all of the growth in MPK (0.32 of 0.34) and MPM (0.22 of 0.20). The pronounced rise in MPL is therefore not evidence against the decline in labor's role in production; rather, it is a consequence of precisely the same factor deepening that underlies that decline.

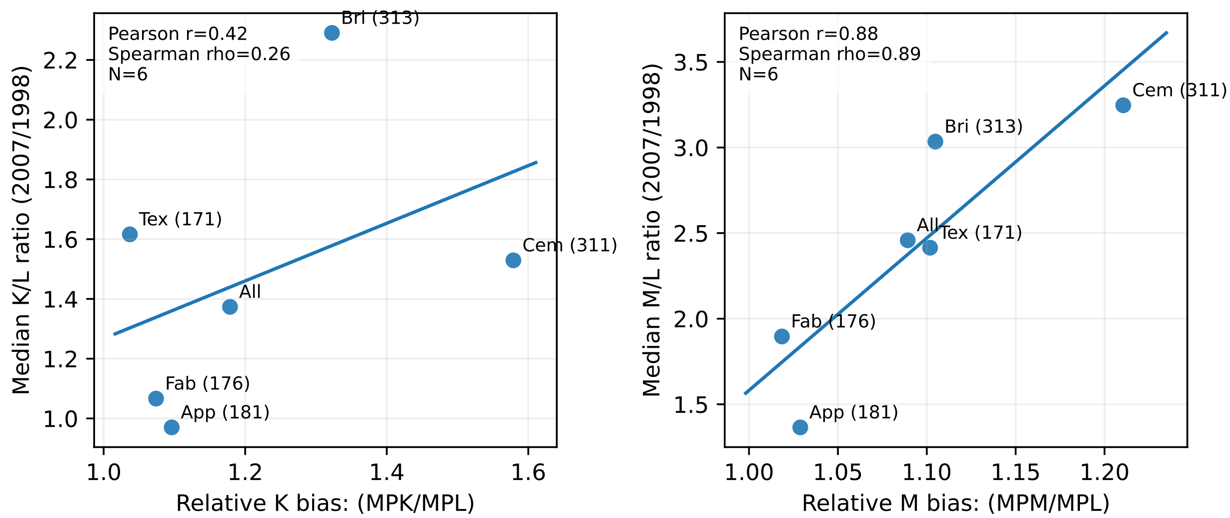
The cross-industry evidence in Figure 4 reinforces this interpretation. Industries experiencing stronger labor-saving productivity growth—as measured by the relative bias of technical change toward capital or intermediates vis-à-vis labor—also exhibit more pronounced capital and materials deepening. The association is particularly clear for materials: the Spearman rank correlation between relative M-bias and the change in the M/L ratio is 0.89 (right panel). The corresponding K-bias correlation is considerably weaker (left panel, Spearman $\rho = 0.26$), consistent with capital adjustment being subject to greater frictions than materials adjustment. These cross-industry correlations are descriptive and based on only six industry observations, so they should be interpreted as suggestive evidence rather than a formal test.

Of course, biased technical change is not the only force behind factor deepening. Available aggregate evidence suggests real manufacturing wages in China grew at around 10% per year over this period (Zhang and Liu, 2013; Banister and Cook, 2011), while the prices of capital and intermediate inputs exhibited no comparably strong secular trend (Kim and Kuijs, 2007). Rising relative labor costs would encourage capital and materials deepening regardless of the direction of technical change. Our framework does not separately identify the contributions of biased technical change and relative input price movements to the observed deepening, and we do not claim to do so. What our estimates do establish is that the technology side of the story is quantitatively important: productivity growth that differentially favors capital and intermediates constitutes a distinct and economically meaningful force consistent with the substitution away from labor observed in the data.

Finally, the documented factor deepening has a natural connection to the well-documented decline in the labor share in Chinese manufacturing over this period (Berkowitz et al., 2015; Bai and Qian, 2010),[27] which is itself part of a broader global trend (Karabarbounis and Neiman, 2014; Autor et al., 2020). Capital and materials deepening tends to put downward pressure on the labor share, other things equal. Our findings therefore suggest that biased

[27] For instance, Berkowitz et al. (2015) document a large decline in labor shares: the mean firm-level labor share falls from 53.3% in 1998 to 35.8% in 2007, while the aggregate labor share falls from 26.4% to 18.4%.

Figure 4: Relative Bias of Technical Change vs. Input Intensity Change, 1998–2007



Note: The figure compares cross-industry changes in input intensity with relative bias in technical change between 1998 and 2007. Input-intensity change is measured as the industry-level median ratio of inputs per labor, K/L and M/L, in 2007 relative to 1998. Relative bias is constructed as

$$\text{K-bias} \equiv \frac{\text{MPK}(\omega_{2007})/\text{MPK}(\omega_{1998})}{\text{MPL}(\omega_{2007})/\text{MPL}(\omega_{1998})}, \qquad \text{M-bias} \equiv \frac{\text{MPM}(\omega_{2007})/\text{MPM}(\omega_{1998})}{\text{MPL}(\omega_{2007})/\text{MPL}(\omega_{1998})}.$$

Values greater than one indicate stronger growth in the marginal product of capital (or intermediates) relative to labor. Correlations are descriptive and based on six industry observations.

technical change contributed to the decline in the labor share at least in part through its association with the factor-deepening patterns documented here. This complements work highlighting biased technical change as a force behind declining labor shares (Oberfield and Raval, 2021; Doraszelski and Jaumandreu, 2018; Zhang, 2019), while providing new evidence from a flexible, nonseparable framework that does not impose an *a priori* factor-augmenting structure.

5 Conclusion

This paper develops an identification and estimation framework for gross-output production functions in which productivity enters nonseparably, thereby relaxing the Hicks-neutrality restriction that underlies much of the production-function literature. The framework allows productivity to affect the marginal products of different inputs asymmetrically and thus provides a way to measure the directions and relative magnitudes of technological bias.

Under perfect competition, we show that conventional proxy-variable assumptions are sufficient to identify output elasticities at each observation, even when productivity interacts flexibly with inputs. This reduced-form identification delivers economically meaningful objects in its own right and provides a basis for testing additional restrictions. Under homogeneity, we further obtain point identification of the structural production function. Under imperfect competition with revenue data, we clarify the limits of nonparametric identification and show that a calibrated-returns-to-scale strategy yields a transparent practical route to point identification. Importantly, the directions and relative magnitudes of technological bias are invariant to the particular returns-to-scale calibration.

We apply the framework to Chinese manufacturing over 1998–2007 and find that the extraordinary productivity growth of this period was predominantly capital-biased and least favorable to labor. Although the marginal product of labor rose substantially in the realized data, this increase was driven mainly by capital and materials deepening through factor complementarities rather than by productivity growth itself. These findings help rationalize the pronounced shift in input mix observed over this period and illustrate the importance of moving beyond Hicks-neutral specifications when studying the nature and consequences of technical change.

Taken together, the results suggest that allowing for non-Hicksian productivity is not only conceptually appealing but also empirically important for credible measurement of production technologies and the bias of technical change.

References

Acemoglu, D. (2002). Directed technical change. *The review of economic studies*, 69(4):781–809. 29

Ackerberg, D., Hahn, J., and Pan, Q. (2022). Nonparametric identification using timing and information set assumptions with an application to non-hicks neutral productivity shocks. *Working Paper*. 4, 7, 9, 10, 11, 12, 45, 58

Ackerberg, D. A. (2021). Comment on "olley and pakes-style production function estimators with firm fixed effects". *Oxford Bulletin of Economics and Statistics*, 83(3):836–840. 6

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451. 2

Ackerberg, D. A., Frazer, G., Kim, K. I., Luo, Y., and Su, Y. (2020). *Under-Identification of Structural Models Based on Timing and Information Set Assumptions*. University of Toronto, Department of Economics. 7

Aguirregabiria, V. and Nevo, A. (2013). Recent developments in empirical io: Dynamic demand and dynamic games. In *Recent Developments in Empirical IO: Dynamic Demand*

and Dynamic Games (May 1, 2013). Advances in Economics and Econometrics. Tenth World Congress of the Econometric Society, volume 3. 44

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the parameter space. *Econometrica*, 69(3):683–734. 69

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly journal of economics*, 135(2):645–709. 36

Bai, C.-E. and Qian, Z. (2010). The factor income distribution in china: 1978–2007. *China Economic Review*, 21(4):650–670. 34, 36

Banister, J. and Cook, G. (2011). China's employment and compensation costs in manufacturing through 2008. *Monthly Labor Review*, 134(3):39–52. 36

Berkowitz, D., Ma, H., and Nishioka, S. (2015). Declining labour shares and heterogeneous firms. *Stanford Center for International Development Working Paper*. 34, 36

Biondi, F. (2022). Firm productivity and derived factor demand under variable markups. Technical report, Working paper, Katholieke Universiteit Leuven, Leuven. 21, 46, 47, 48

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2020). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. Technical report, National Bureau of Economic Research. 18

Brandt, L. and Rawski, T. G. (2008). *China's great economic transformation*. Cambridge university press. 66

Brandt, L., Van Biesebroeck, J., Wang, L., and Zhang, Y. (2017). Wto accession and performance of chinese manufacturing firms. *American Economic Review*, 107(9):2784–2820. 18, 27

Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2012). Creative accounting or creative destruction? firm-level productivity growth in chinese manufacturing. *Journal of development economics*, 97(2):339–351. 4, 27, 65, 66

Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2014). Challenges of working with the chinese nbs firm-level data. *China Economic Review*, 30:339–352. 66

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, 6:5549–5632. 14

Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575. 17, 45

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261. 7, 44, 45

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525. 45

Chiappori, P.-A., Komunjer, I., and Kristensen, D. (2015). Nonparametric identification and estimation of transformation models. *Journal of Econometrics*, 188(1):22–39. 12

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, 79(5):1407–1451. 4, 18, 19, 20, 21, 28

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, 84(2):445–510. 42, 43

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American economic review*, 102(6):2437–71. 18, 23, 26, 29, 51, 57

De Ridder, M., Grassi, B., and Morzenti, G. (2024). The hitchhiker's guide to markup estimation: assessing estimates from financial data. *Work. Pap., London Sch. Econ./Bocconi Univ./Analysis Group, UK/Milan, Italy/Paris, France*. 23

Demirer, M. (2020). Production function estimation with factor-augmenting technology: An application to markups. Technical report, MIT working paper. 2, 3, 9, 55

Doraszelski, U. and Jaumandreu, J. (2013). R&d and productivity: Estimating endogenous productivity. *Review of Economic Studies*, 80(4):1338–1383. 21

Doraszelski, U. and Jaumandreu, J. (2018). Measuring the bias of technological change. *Journal of Political Economy*, 126(3):1027–1084. 2, 3, 37, 42

Flynn, Z., Gandhi, A., and Traina, J. (2019). Measuring markups with production data. *Available at SSRN 3358472*. 64

Gallant, A. R., Hong, H., Leung, M. P., and Li, J. (2022). Constrained estimation using penalization and MCMC. *Journal of Econometrics*, 228(1):85–106. 17

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016. 1, 2, 3

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: The case of the us automobile industry. *Econometrica: Journal of the Econometric Society*, pages 891–951. 18

Hendel, I. and Nevo, A. (2006). Sales and consumer inventory. *The RAND Journal of Economics*, 37(3):543–561. 44

Hu, X., Jiang, Y., and Zhang, H. (2023). Returns to scale, productivity, and markup: Revisit the export premium. In *LACEA LAMES Annual Meeting 2023 (16/06/2023-18/06/2023, Bogotá)*. 28

Karabarbounis, L. and Neiman, B. (2014). The global decline of the labor share. *The Quarterly journal of economics*, 129(1):61–103. 36

Kasahara, H. and Sugita, Y. (2020). Nonparametric identification of production function, total factor productivity, and markup from revenue data. *Total Factor Productivity, and Markup from Revenue Data (October 28, 2020)*. 12, 19

Kim, S.-Y. and Kuijs, L. (2007). Raw material prices, wages, and profitability in china's industry—how was profitability maintained when input prices and wages increased so fast? Technical Report 8, World Bank Office, Beijing. 36

Klette, T. J. and Griliches, Z. (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of applied econometrics*, 11(4):343–361. 4, 18, 20, 28

Lee, Y., Stoyanov, A., and Zubanov, N. (2019). Olley and pakes-style production function estimators with firm fixed effects. *Oxford Bulletin of Economics and Statistics*, 81(1):79–97. 6

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2):317–341. 2, 46

Li, T. and Sasaki, Y. (2024). Identification of heterogeneous elasticities in gross-output production functions. *Journal of Econometrics*, 238(2):105637. 8

Lu, Y. and Yu, L. (2015). Trade liberalization and markup dispersion: evidence from china's wto accession. *American Economic Journal: Applied Economics*, 7(4):221–53. 55

Mammen, E., Marron, J. S., Turlach, B. A., and Wand, M. P. (2001). A general projection framework for constrained smoothing. *Statistical Science*, 16(3):232–248. 17

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, 71(5):1339–1375. 9, 10, 15, 22

Navarro, S. and Rivers, D. A. (2018). Nonparametric identification of productivity in non-separable production functions. *Working Paper*. 12

Oberfield, E. and Raval, D. (2021). Micro data and macro technology. *Econometrica*, 89(2):703–732. 2, 3, 37

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica: Journal of the Econometric Society*, pages 1263–1297. 2

Orr, S. (2022). Within-firm productivity dispersion: Estimates and implications. *Journal of Political Economy*, 130(11):2771–2828. 42

Raval, D. (2019). Testing the production approach to markup estimation. *Available at SSRN 3324849*. 3

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954. 14

Rubens, M., Wu, Y., and Xu, M. (2024). Exploiting or augmenting labor. *American Economic Review: Insights*. 2, 3, 21, 27

Valmari, N. (2023). Estimating production functions of multiproduct firms. *Review of Economic Studies*, 90(6):3315–3342. 42

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*. Chapman and Hall/CRC. 69

Zeng, J. (2021). Covid-19, productivity, and recovery: a nonseparable production function framework. *Available at SSRN 3707907*. 12

Zhang, H. (2019). Non-neutral technology, firm heterogeneity, and labor demand. *Journal of Development Economics*, 140:145–168. 2, 3, 27, 37

Zhang, J. and Liu, X. (2013). The evolving pattern of the wage–labor productivity nexus in china: Evidence from manufacturing firm-level data. *Economic Systems*, 37(3):354–368. 36

Zhu, X. (2012). Understanding china's growth: Past, present, and future. *Journal of Economic Perspectives*, 26(4):103–124. 27

Online Appendices

A Imperfect Competition with Output Prices Observed

In the main text, we work with firm-level financial records in which output prices are typically unobserved. This section revisits identification when firm-level output prices are available.[28] We restrict attention to single-product firms.[29]

[28]Output-price information is available in several widely used datasets; see, e.g., De Loecker et al. (2016); Doraszelski and Jaumandreu (2018).

[29]For treatments of multi-product firms, see De Loecker et al. (2016); Orr (2022); Valmari (2023).

A.1 Identification without Calibrating Returns to Scale

We maintain the primitives and monopolistic-competition structure of the main text. The key difference is that firm-level output quantities q_{jt} are now observed, since revenues r_{jt} can be deflated by firm-specific output prices p_{jt}. This allows us to disentangle markups from returns to scale (RTS) directly, without fixing RTS *ex ante*.

The first step is to identify the revenue function

$$r_{jt} = d_t^{-1}(q_{jt}, z_{jt}, \epsilon_{jt}) + q_{jt} \equiv \gamma_t(q_{jt}, z_{jt}, \epsilon_{jt}) \,,$$

where ϵ_{jt} is independent of (q_{jt}, z_{jt}). Under this assumption, $\gamma_t(\cdot)$ is identified, and so is the revenue elasticity with respect to quantity, $\gamma_{q_{jt}}$. Rearranging equation (52) in the proof of Theorem 3.1 then identifies RTS:

$$\alpha_{jt} = \frac{1}{\gamma_{q_{jt}}} \frac{c_{1,jt}(\epsilon_{jt})}{1 + c_{2,jt}\tilde{f}_{\omega_{jt}}},$$

where $(c_{1,jt}(\epsilon_{jt}), c_{2,jt}, \tilde{f}_{\omega_{jt}})$ are identified objects, as defined in the proof of Theorem 3.1. With α_{jt} identified, the output elasticities $\{f_k, f_l, f_m, f_\omega\}$ follow from the proof of Theorem 3.1, and the production function f_t is identified up to an additive constant.

A.2 Accommodating Unobserved Input-Price Heterogeneity

Our framework already accommodates potentially unobserved heterogeneity in capital and labor prices. In the main text, because neither output nor input prices are observed, we proxy intermediate inputs using material expenditures—implicitly abstracting from material-price heterogeneity. When output prices are available, we can instead follow De Loecker et al. (2016) and use them to control for the unobserved price component of materials.

The only modification relative to the baseline concerns identification of ϕ_t; all subsequent steps are unchanged.

Setup. Following De Loecker et al. (2016), model the (unobserved) firm-level material price as $\rho_{jt} = \rho_t(p_{jt}, z_{jt})$, with ρ_t strictly increasing in p_{jt}.[30] The intermediate-input demand function becomes $m_{jt} = \phi_t(k_{jt}, l_{jt}, \rho_{jt}, z_{jt}, \omega_{jt})$.

[30] See De Loecker et al. (2016) for a formal derivation.

Strategy. First, identify $\rho_t(\cdot)$ to recover ρ_{jt}. Second, construct material quantities as $m_{jt} = \tilde{m}_{jt}/\rho_{jt}$.[31] Third, use the constructed (m_{jt}, ρ_{jt}) to identify $\phi_t(k_{jt}, l_{jt}, \rho_{jt}, z_{jt}, \omega_{jt})$.

Identification of function ρ_t. Suppose ρ_{jt} follows a first-order Markov process.[32] If (p_{jt}, z_{jt}) is independent of the innovation in ρ_{jt}, then $\rho_t(\cdot)$ is identified by an argument analogous to that used for $\phi_t(\cdot)$ in the main text. If instead p_{jt} may be correlated with the innovation in ρ_{jt}, identification can be recovered using the IV strategy for nonseparable models developed by Chernozhukov and Hansen (2005), analogous to the approach in Appendix B.

Completing identification. With ρ_{jt} and m_{jt} in hand, ϕ_t is identified by the same logic as in the main text, provided $(k_{jt}, l_{jt}, \rho_{jt}, z_{jt})$ are independent of the innovation in ω_{jt}. The remainder of the identification argument proceeds unchanged.

B Flexible Labor

In the main text, we treat labor as predetermined. This section shows how to accommodate flexible labor—chosen within the period and potentially correlated with the productivity innovation ξ_{jt}.

The timing assumption on labor is needed only to identify the materials demand function $\phi_t(k_{jt}, \ell_{jt}, z_{jt}, \omega_{jt})$. Once ϕ_t is identified, all subsequent identification arguments—under both perfect and imperfect competition—go through unchanged. It therefore suffices to show how ϕ_t can be identified when labor is flexible.

Setup. Substituting $\omega_{jt} = h_t(\omega_{jt-1}, o_{jt-1}, \xi_{jt})$ and $\omega_{jt-1} = \phi_{t-1}^{-1}(k_{jt-1}, \ell_{jt-1}, z_{jt-1}, m_{jt-1})$ into ϕ_t yields the reduced form

$$m_{jt} = \bar{\phi}_t\big(k_{jt}, \ell_{jt}, z_{jt}, \underbrace{k_{jt-1}, \ell_{jt-1}, z_{jt-1}, m_{jt-1}, o_{jt-1}}_{\equiv \tilde{\mathbf{v}}_{jt-1}}, \xi_{jt}\big). \tag{32}$$

With labor flexible, ℓ_{jt} may be correlated with ξ_{jt}; all other arguments in (32) are predetermined. Our strategy identifies $\bar{\phi}_t$ via instrumental variables and then recovers ϕ_t from $\bar{\phi}_t$ using the support condition (Assumption 2.5 or 3.5), exactly as in the main text.

[31] An industry-level materials price index can be used to normalize the scale of ρ_{jt}.

[32] This assumption is most plausible when materials markets are competitive. Modeling prices as Markov is standard in empirical IO; see, e.g., Hendel and Nevo (2006); Aguirregabiria and Nevo (2013). The first-order restriction generalizes readily to higher-order dynamics.

Identification. Because $\bar{\phi}_t$ is strictly monotone in ξ_{jt}, we normalize $\xi_{jt} \sim U(0,1)$ and apply the IV framework for nonseparable models developed by Chernozhukov and Hansen (2005); see also Appendix C of Ackerberg et al. (2022).

Suppose an instrument w_{jt} for ℓ_{jt} is available such that $(k_{jt}, z_{jt}, \tilde{\mathbf{v}}_{jt-1}, w_{jt})$ is jointly independent of ξ_{jt}. Then, for each $\tau \in (0,1)$,

$$\Pr\big(m_{jt} \le \bar{\phi}_t(k_{jt}, \ell_{jt}, z_{jt}, \tilde{\mathbf{v}}_{jt-1}, \tau) \,\big|\, k_{jt}, z_{jt}, \tilde{\mathbf{v}}_{jt-1}, w_{jt}\big) = \tau. \tag{33}$$

Identification of $\bar{\phi}_t$ from (33) requires a completeness condition: variation in w_{jt} must generate sufficiently rich shifts in the conditional distribution of ℓ_{jt} given $(k_{jt}, z_{jt}, \tilde{\mathbf{v}}_{jt-1})$.[33]

Lemma B.1. *Under all maintained assumptions of the baseline model—except that ℓ_{jt} is now flexible—together with (i) the existence of an instrument w_{jt} satisfying independence from ξ_{jt} and (ii) the completeness condition above, the reduced-form function $\bar{\phi}_t$ is identified. Consequently, ϕ_t is identified by the same argument as in the main text.*

Proof. Under independence and completeness, the quantile restriction (33) identifies the structural quantile function $\bar{\phi}_t(k_{jt}, \ell_{jt}, z_{jt}, \tilde{\mathbf{v}}_{jt-1}, \tau)$ for each $\tau \in (0,1)$ by Theorem 1 of Chernozhukov and Hansen (2005). Given $\bar{\phi}_t$, the support condition recovers ϕ_t by the same ranking argument as in Lemma 2.3 (under PC) or Lemma 3.1 (under IPC). □

Estimation. We estimate $\bar{\phi}_t$ using the instrumental-variables quantile regression (IVQR) approach of Chernozhukov and Hansen (2006) on a fine grid of $\tau \in (0,1)$. We then recover the latent rank

$$\hat{\xi}_{jt} = \inf\Big\{\tau \in (0,1) : \, m_{jt} \le \widehat{\bar{\phi}}_t(k_{jt}, \ell_{jt}, z_{jt}, \tilde{\mathbf{v}}_{jt-1}, \tau)\Big\},$$

applying the monotone rearrangement of Chernozhukov et al. (2009) to enforce monotonicity across quantiles. With $\hat{\xi}_{jt}$ in hand, the structural primitives (ϕ, h) are estimated from

$$m_{jt} = \phi\Big(k_{jt}, \ell_{jt}, z_{jt},\, h\big(\phi^{-1}(k_{jt-1}, \ell_{jt-1}, z_{jt-1}, m_{jt-1}),\, o_{jt-1},\, \xi_{jt}\big)\Big), \tag{34}$$

which is free of endogeneity under the maintained assumptions. We approximate ϕ and h using sieve bases and estimate by nonlinear least squares.

[33] Formally, this is the completeness condition L1* or L2* of Chernozhukov and Hansen (2005). Intuitively, it is the nonparametric analogue of a standard rank condition for instrument relevance.

Instruments. Candidate instruments include: (i) wages, which shift labor demand holding other determinants fixed—if labor-market imperfections raise concerns about correlation with ξ_{jt}, lagged wages can be used instead; and (ii) lagged output y_{jt-1} (or lagged revenue), which is predetermined and relevant for current labor when employment adjusts to past shocks. Instrument relevance is assessed using standard first-stage diagnostics.

C Sufficient Conditions for Strict Monotonicity under Imperfect Competition

The strict monotonicity requirement that $m_{jt} = \phi(k_{jt}, \ell_{jt}, z_{jt}, \omega_{jt})$ is strictly increasing in productivity ω_{jt} can be imposed directly as a high-level condition. Under perfect competition, this requirement reduces to standard concavity and complementarity conditions on the production function and is well understood in the literature (see, e.g., Levinsohn and Petrin, 2003). Under imperfect competition, however, monotonicity is less automatic. Biondi (2022) shows that, when productivity is Hicks-neutral and markups are variable, factor demand can *decouple* from productivity—that is, monotonicity can fail—whenever demand is sufficiently inelastic relative to its curvature. Since our framework features nonseparable productivity, a richer set of forces governs monotonicity: both demand curvature and the interaction between productivity and the marginal product of the flexible input enter the condition. This section provides primitive sufficient conditions that decompose monotonicity into these two channels and verifies, as a post-estimation consistency check, that the sufficient conditions are satisfied by the estimated model.

Suppress time and firm subscripts and hold predetermined inputs and demand shifters fixed for exposition. Let output be $Q = F(M, e^{\omega})$, where M is the flexible input. Let the revenue function be $\Gamma(Q, Z)$, so that $\Gamma_Q(Q, Z)$ is marginal revenue (MR). Ignoring the ex post shock ϵ for the moment, the first-order condition for an interior optimum in M is

$$\Gamma_Q(F(M, e^{\omega}), Z)\, F_M(M, e^{\omega}) = \rho, \tag{35}$$

where ρ is the input price. Define

$$G(M, \omega) \equiv \Gamma_Q(F(M, e^{\omega}), Z)\, F_M(M, e^{\omega}) - \rho,$$

so that the optimal choice $M^*(\omega)$ solves $G(M, \omega) = 0$.

By the implicit function theorem,

$$\frac{\partial M^*}{\partial \omega} \;=\; -\frac{G_\omega}{G_M}.$$

A direct differentiation yields

$$G_M \;=\; \Gamma_{QQ} F_M^2 + \Gamma_Q F_{MM}, \qquad G_\omega \;=\; \Gamma_{QQ} F_\omega F_M + \Gamma_Q F_{M\omega},$$

where Γ_{QQ} is evaluated at (Q, Z) and the derivatives of F are evaluated at (M, e^ω). Note that G_M equals the second derivative of the firm's marginal-revenue-product schedule with respect to M, denoted Γ_{MM}.

A sufficient condition. It is natural to require the second-order condition

$$\Gamma_{MM} \equiv G_M < 0, \tag{36}$$

which reflects diminishing marginal productivity of M and downward-sloping demand. Given (36), the sign of $\partial M^*/\partial \omega$ is the same as the sign of G_ω. Hence a sufficient condition for strict monotonicity is $G_\omega > 0$, i.e.,

$$F_{M\omega} > -\frac{\Gamma_{QQ}}{\Gamma_Q} F_\omega F_M. \tag{37}$$

Under perfect competition, $\Gamma_{QQ} = 0$ and (37) reduces to $F_{M\omega} > 0$, the standard complementarity condition between productivity and the flexible input.

Hicks-neutral benchmark. Suppose productivity is Hicks-neutral, $Q = e^\omega \widetilde{F}(M)$. Then $F_\omega = Q$ and $F_{M\omega} = F_M$, so (37) becomes

$$1 \;>\; -\frac{\Gamma_{QQ}}{\Gamma_Q} Q \iff \frac{d \ln \Gamma_Q}{d \ln Q} \;>\; -1. \tag{38}$$

This is precisely the condition identified by Biondi (2022) (Proposition 1) for monotonicity under Hicks-neutral productivity: in his notation, monotonicity holds if and only if the elasticity of marginal revenue with respect to output exceeds -1, or equivalently $\varepsilon > 3 - \rho$, where ε is the price elasticity and ρ the curvature of demand. Condition (38) thus requires that MR may fall with Q under imperfect competition, but not faster than Q^{-1} in log terms. Under isoelastic (CES) demand, (38) is automatically satisfied whenever the price elasticity exceeds one, consistent with the finding of Biondi (2022) that CES demand precludes decoupling. A key insight of Biondi (2022) is that this robustness does not extend to richer demand systems: for specifications satisfying Marshall's Second Law (e.g., linear, CARA, logistic, or Kimball demand), condition (38) can be violated even in the elastic region of the demand curve, so that demand curvature alone can overturn monotonicity. With nonseparable productivity, condition (38) generalizes to (39) below, and the production side of the inequality introduces an additional force that interacts with demand curvature to determine whether monotonicity holds.

Nonseparable production: an elasticity form. For general $F(M, e^{\omega})$, it is convenient to rewrite (37) in terms of elasticities. Define

$$f_m \equiv \frac{\partial \ln Q}{\partial \ln M}, \qquad f_\omega \equiv \frac{\partial \ln Q}{\partial \omega}, \qquad \kappa_{M\omega} \equiv \frac{\partial^2 \ln Q}{\partial \ln M \, \partial \omega}, \qquad \theta \equiv -\frac{\partial \ln \Gamma_Q}{\partial \ln Q} = -\frac{Q\Gamma_{QQ}}{\Gamma_Q}.$$

Using the identity $\kappa_{M\omega} = (M/Q)F_{M\omega} - f_\omega f_m$ and $F_\omega = Q f_\omega$, one can verify that (37) is equivalent to

$$\frac{\kappa_{M\omega}}{f_\omega f_m} > \theta - 1. \tag{39}$$

Condition (39) decomposes monotonicity into a production-side channel (left) and a demand-side channel (right). A rise in productivity has two effects: it raises the marginal product of M, creating an incentive to expand intermediate use, but it also raises output, which under imperfect competition depresses marginal revenue and dampens that incentive. The left-hand side $\kappa_{M\omega}/(f_\omega f_m)$ measures the strength of the first effect; the right-hand side $\theta - 1$ measures the strength of the second. Monotonicity holds when complementarity dominates dampening.

When productivity is Hicks-neutral, $\kappa_{M\omega} = 0$ and condition (39) reduces to $\theta < 1$, so monotonicity depends entirely on demand curvature—the case analyzed by Biondi (2022). With nonseparable productivity, $\kappa_{M\omega}/(f_\omega f_m)$ enters the condition alongside θ, enriching the set of primitive forces at play. If the production side is sufficiently positive, monotonicity can be sustained even for demand systems where Biondi (2022) finds decoupling (i.e., $\theta \geq 1$). Conversely, a negative production side places a tighter requirement on demand curvature. Under CES demand, $\theta = 1/\sigma$ where $\sigma > 1$ is the price elasticity, so $\theta - 1 < 0$ and the condition is satisfied provided the production side is not too negative.

Allowing for an ex post demand shock. Now introduce an ex post shock ϵ and let the firm maximize expected profits given ω. The first-order condition becomes

$$\Gamma_Q^{exp}(Q,Z)\,F_M(M,e^{\omega})=\rho, \qquad \Gamma_Q^{exp}(Q,Z)\equiv\mathbb{E}_{\epsilon}\big[\Gamma_Q(Q,Z,e^{\epsilon})\big]. \tag{40}$$

Analogously, sufficient conditions for strict monotonicity are:

$$\Gamma_{MM}^{exp}<0, \qquad \Gamma_{MM}^{exp}\equiv\Gamma_{QQ}^{exp}F_M^2+\Gamma_Q^{exp}F_{MM}, \qquad \Gamma_{QQ}^{exp}\equiv\frac{\partial\Gamma_Q^{exp}}{\partial Q}, \tag{41}$$

and

$$\frac{\kappa_{M\omega}}{f_\omega f_m}>\theta^{exp}-1, \tag{42}$$

where

$$\theta^{exp}\equiv-\frac{\partial\ln\Gamma_Q^{exp}}{\partial\ln Q}.$$

These conditions are directly checkable once the revenue function and the production function have been estimated.

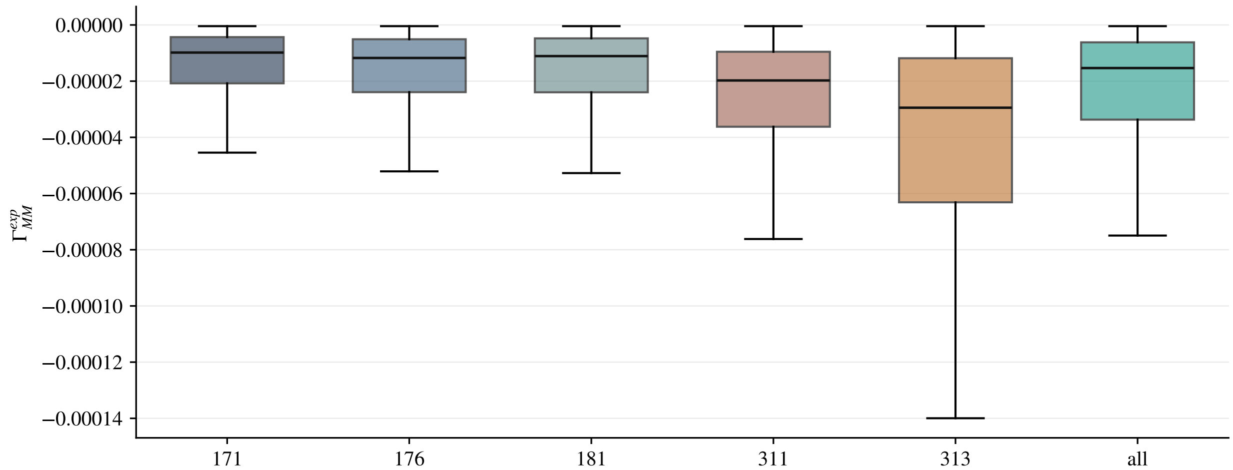
Post-estimation consistency check. Since the estimation procedure in the main text enforces monotonicity of ϕ in ω directly, the sufficient conditions derived above cannot be tested independently of the maintained assumption. Nevertheless, it is informative to check whether the estimated production function and revenue function satisfy these more primitive conditions. If they did not—for example, if the estimated demand curvature were large enough to violate (42) despite monotonicity being imposed—it would indicate that the constraint is doing substantial work and that the plausibility of the monotonicity assumption rests on features of the model not captured by these primitive conditions.

Since $f_\omega f_m>0$, condition (42) is equivalently written as

$$sign\equiv\kappa_{M\omega}-\big(\theta^{exp}-1\big)f_\omega f_m \;>\; 0.$$

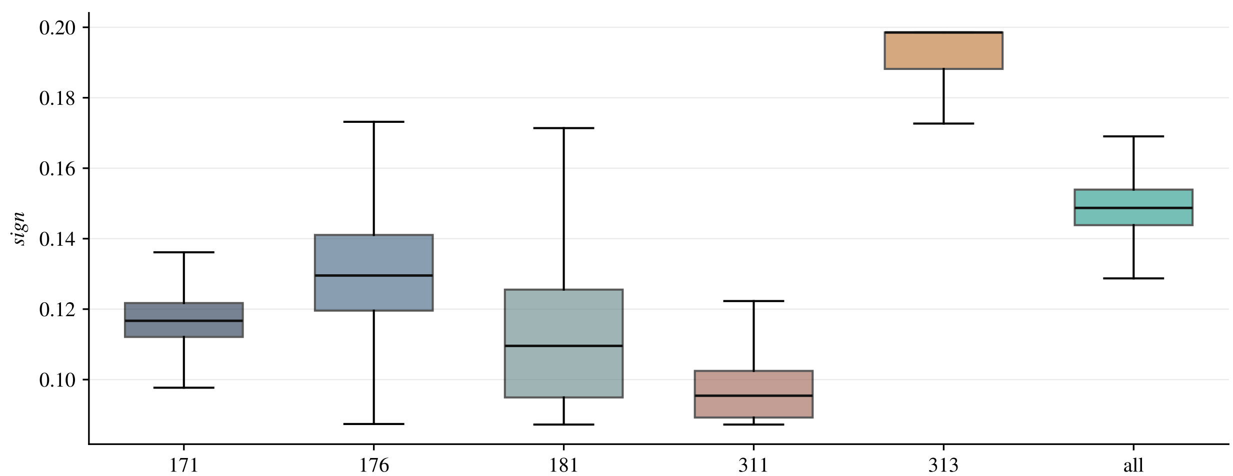
Figures 5 and 6 plot the empirical distributions of estimated Γ_{MM}^{exp} and *sign* by industry. Across all three-digit industries, the estimated second-order condition $\Gamma_{MM}^{exp}<0$ holds for virtually all observations, confirming that firms operate at interior optima. More importantly, the estimated *sign* is strictly positive for the vast majority of observations in every industry. While this does not constitute an independent test of monotonicity, it provides reassurance that the maintained assumption is consistent with economically plausible primitive conditions in the estimated model.

Figure 5: Industry Distributions of Γ^{exp}_{MM}



Note: The figure summarizes, for each three-digit industry, the distribution of the estimated Γ^{exp}_{MM}, as defined above. The distributions are shown using boxplots (center line: median; box: interquartile range; whiskers: $1.5 \times$ IQR). For readability, the underlying values are winsorized at the pooled 1st–99th percentiles.

Figure 6: Industry Distributions of $\text{sign}(\partial m_{jt}/\partial \omega_{jt})$



Note: The figure summarizes, for each three-digit industry, the distribution of the estimated $sign \equiv \kappa_{M\omega} - \left(\theta^{exp} - 1\right) f_\omega f_m$, as defined above. The distributions are shown using boxplots (center line: median; box: interquartile range; whiskers: $1.5 \times$ IQR). For readability, the underlying values are winsorized at the pooled 1st–99th percentiles.

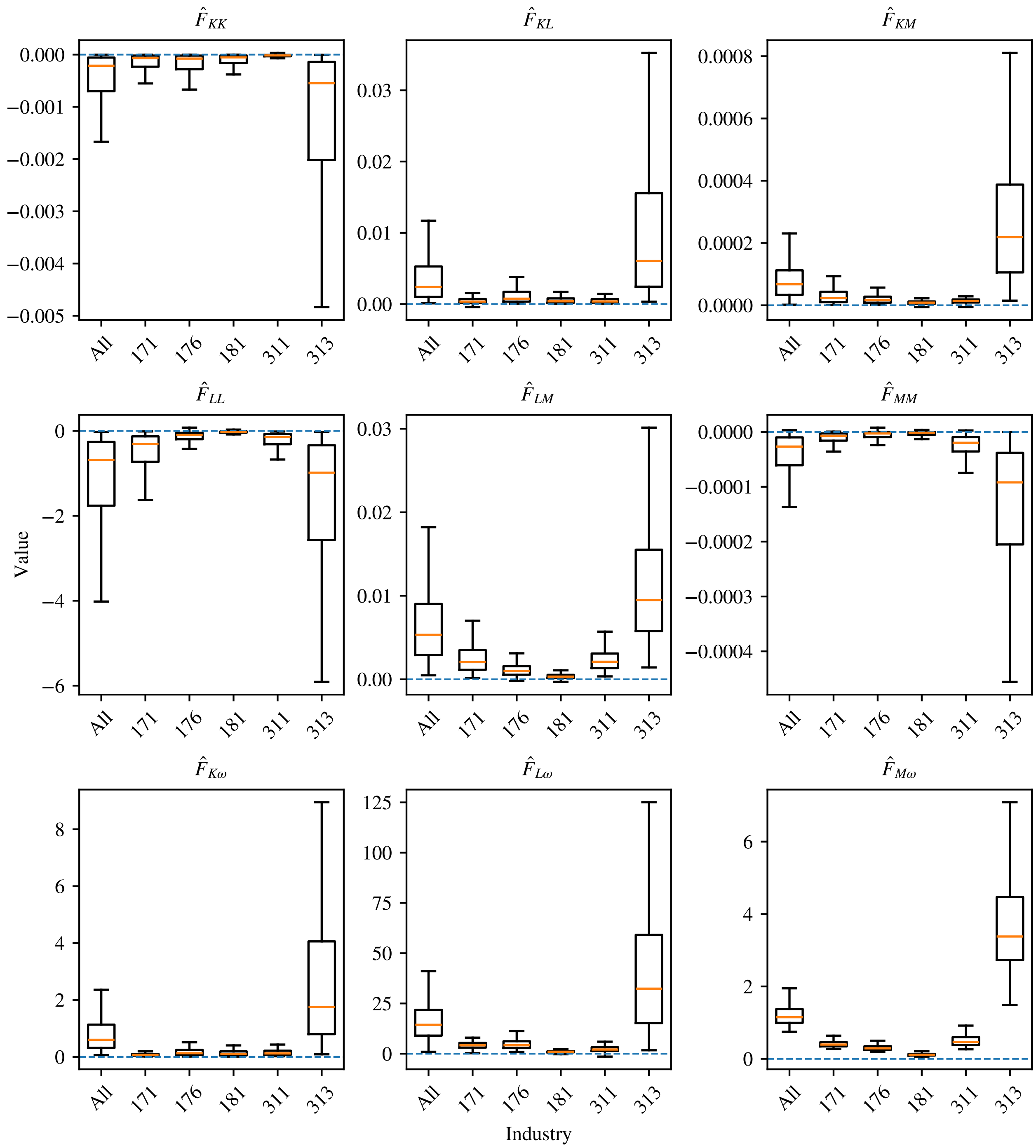
D Additional Tables and Figures

Table 3: Mean Output Elasticities and Markups in Chinese Manufacturing

	Capital	Labor	Materials	Markups
All	0.10	0.14	0.86	1.17
	(0.00)	(0.00)	(0.00)	(0.00)
Textiles (171)	0.07	0.13	0.89	1.18
	(0.00)	(0.00)	(0.00)	(0.00)
Fabrics (176)	0.06	0.15	0.88	1.18
	(0.00)	(0.00)	(0.00)	(0.00)
Apparel (181)	0.08	0.18	0.85	1.17
	(0.00)	(0.00)	(0.00)	(0.00)
Cement (311)	0.12	0.14	0.85	1.16
	(0.01)	(0.01)	(0.01)	(0.01)
Bricks (313)	0.12	0.13	0.84	1.21
	(0.00)	(0.00)	(0.01)	(0.01)

Note: The table reports sample means of the estimated output elasticities with respect to capital, labor, and materials, together with the implied markups. Estimates are obtained using the procedure in Section 3.3, with returns to scale calibrated to 1.1. Markups are computed following De Loecker and Warzynski (2012). Bootstrap standard errors based on 200 replications are reported in parentheses.

Figure 7: Distribution of Second-Order Derivatives



Note: This figure summarizes, by industry, the distributions of the estimated second-order derivatives implied by the same production-function estimates reported in Table 3. The distributions are shown using boxplots (center line: median; box: interquartile range; whiskers: 1.5 × IQR). For readability, the underlying values are winsorized at the pooled 1st–99th percentiles.

Table 4: Bias in Technical Change: Robustness to RTS Calibration

Industry	RTS	MPK(ω_{2007})/MPK(ω_{1998})	MPL(ω_{2007})/MPL(ω_{1998})	MPM(ω_{2007})/MPM(ω_{1998})
All	1.00	1.30	1.11	1.19
	1.05	1.31	1.12	1.20
	1.10	1.32	1.12	1.22
	1.15	1.33	1.13	1.23
	1.20	1.34	1.14	1.24
Textiles (171)	1.00	1.10	1.07	1.17
	1.05	1.11	1.08	1.18
	1.10	1.12	1.08	1.19
	1.15	1.12	1.09	1.20
	1.20	1.13	1.10	1.21
Fabrics (176)	1.00	1.15	1.07	1.09
	1.05	1.15	1.08	1.10
	1.10	1.16	1.08	1.10
	1.15	1.16	1.09	1.11
	1.20	1.17	1.09	1.11
Apparel (181)	1.00	1.13	1.04	1.06
	1.05	1.14	1.04	1.06
	1.10	1.14	1.04	1.07
	1.15	1.14	1.04	1.07
	1.20	1.14	1.04	1.07
Cement (311)	1.00	1.75	1.12	1.34
	1.05	1.77	1.13	1.36
	1.10	1.80	1.14	1.38
	1.15	1.83	1.15	1.40
	1.20	1.86	1.17	1.43
Bricks (313)	1.00	1.60	1.20	1.33
	1.05	1.62	1.22	1.35
	1.10	1.64	1.24	1.37
	1.15	1.67	1.25	1.39
	1.20	1.69	1.27	1.41

Note: The table reports ratios of marginal products evaluated at ω_{2007} and ω_{1998}, i.e., MPI(ω_{2007})/MPI(ω_{1998}) for $I \in \{K, L, M\}$, holding inputs fixed at their median levels. RTS denotes the calibrated value of returns to scale. The benchmark specification in Table 1 corresponds to RTS = 1.10. Standard errors are comparable to those reported in Table 1 and are therefore omitted.

Table 5: Growth Decomposition of Marginal Products: Robustness to RTS Calibration

	Entries shown as RTS = 1.05 / 1.10 / 1.15			
Industry	Total	Comp.	Prod.	Res.
Panel A: MPK				
All	0.29/0.34/0.39	0.01/0.05/0.09	0.31/0.32/0.33	-0.02/-0.02/-0.03
Textiles (171)	0.42/0.48/0.53	0.28/0.33/0.37	0.15/0.16/0.17	-0.01/-0.01/-0.01
Fabrics (176)	-0.04/0.00/0.04	-0.21/-0.18/-0.15	0.19/0.19/0.20	-0.01/-0.01/-0.01
Apparel (181)	-0.06/-0.03/0.00	-0.22/-0.19/-0.17	0.18/0.18/0.19	-0.02/-0.02/-0.02
Cement (311)	0.74/0.78/0.82	0.18/0.20/0.23	0.58/0.60/0.61	-0.02/-0.02/-0.02
Bricks (313)	0.58/0.64/0.69	0.09/0.13/0.16	0.49/0.51/0.53	0.00/0.00/0.00
Panel B: MPL				
All	0.71/0.76/0.81	0.60/0.64/0.68	0.11/0.12/0.13	0.00/0.00/0.00
Textiles (171)	0.90/0.94/1.00	0.79/0.82/0.87	0.11/0.12/0.13	0.00/0.00/0.00
Fabrics (176)	0.49/0.54/0.57	0.40/0.43/0.47	0.09/0.10/0.11	0.00/0.00/0.00
Apparel (181)	0.32/0.35/0.39	0.27/0.30/0.33	0.04/0.05/0.05	0.01/0.01/0.01
Cement (311)	0.87/0.91/0.95	0.76/0.79/0.82	0.11/0.12/0.13	0.00/0.00/0.00
Bricks (313)	0.87/0.92/0.98	0.68/0.72/0.76	0.18/0.20/0.21	0.00/0.00/0.00
Panel C: MPM				
All	0.15/0.20/0.25	-0.06/-0.02/0.02	0.21/0.22/0.23	0.00/0.00/0.00
Textiles (171)	0.15/0.21/0.26	-0.12/-0.07/-0.03	0.27/0.28/0.29	0.00/0.00/0.00
Fabrics (176)	0.10/0.15/0.19	-0.02/0.01/0.05	0.12/0.13/0.14	0.00/0.00/0.00
Apparel (181)	0.08/0.11/0.14	-0.01/0.02/0.05	0.08/0.09/0.09	0.00/0.00/0.00
Cement (311)	0.18/0.22/0.26	-0.12/-0.10/-0.08	0.30/0.32/0.33	0.00/0.00/0.00
Bricks (313)	0.24/0.29/0.35	-0.08/-0.04/0.00	0.31/0.33/0.34	0.00/0.00/0.00

Note: Each cell reports the decomposition estimates for RTS = 1.05 / 1.10 / 1.15, in that order. The middle entry therefore corresponds to the benchmark specification reported in Table 2. *Comp.* abbreviates factor complementarity, *Prod.* productivity growth, and *Res.* the residual.

E Markups

Under our model and the available data, markup levels are identified only up to the assumed returns to scale (RTS). If, in addition, we impose the homogeneity restriction, firm-level markups are pinned down up to a common multiplicative constant that depends on RTS. As a result, both markup dispersion and the time trend in markups are identified without committing to a particular RTS value.

We use the estimators developed in the main text to address two descriptive questions regarding markups in Chinese manufacturing. First, how did markup dispersion evolve over 1998–2007? Because dispersion in markups is often interpreted as reflecting dispersion in distortions and hence potential allocative inefficiency (see, e.g., Lu and Yu, 2015), this exercise is informative about whether resource allocation improved over the decade. Second, how did revenue-weighted markups—an aggregate measure of market power—evolve over the same period?

Figure 8 plots the time series of markup dispersion, measured by the Theil index. For comparison, we also report the corresponding series implied by a Hicks-neutral specification estimated within the same framework. Both specifications indicate a broad decline in markup dispersion over the decade, consistent with an improvement in allocative efficiency. Quantitatively, however, the Hicks-neutral specification generally yields higher Theil indices than our nonseparable model.

Figure 9 reports the evolution of revenue-weighted markups. The Hicks-neutral specification implies a clear upward trend: in the pooled sample, revenue-weighted markups rise from slightly above 1.15 to slightly above 1.25, corresponding to an increase of roughly 10 percent. By contrast, our nonseparable model delivers a much more muted pattern; in the pooled specification, revenue-weighted markups remain nearly flat over the decade. The industry-specific panels reveal somewhat more variation—for instance, markups decline modestly in Cement (311) under the nonseparable model—but the overall contrast between the two specifications is consistent across industries. This pattern suggests that imposing Hicks neutrality may overstate the upward trend in markups, consistent with the mechanism emphasized by Demirer (2020).

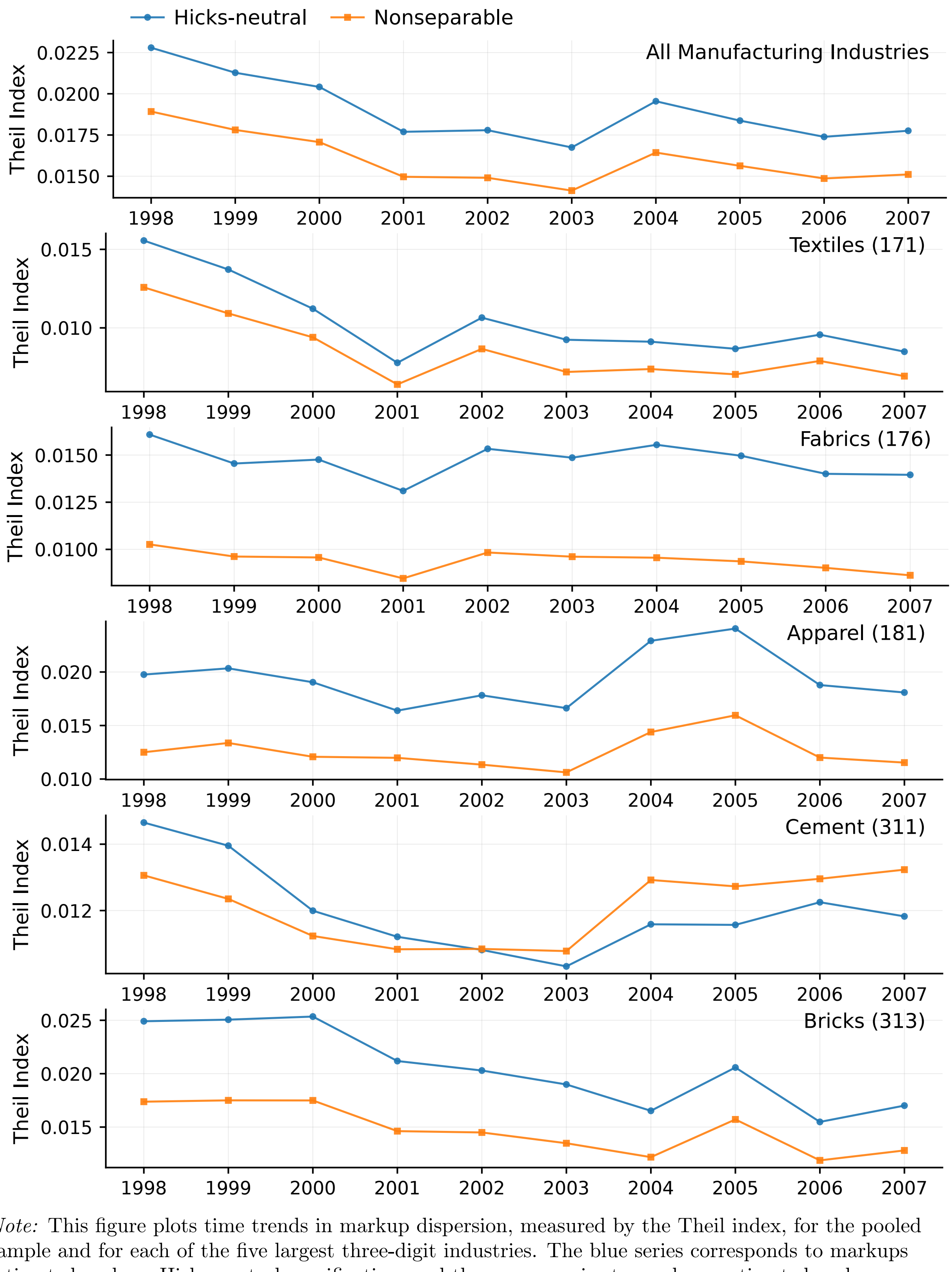
F Additional Proofs

F.1 Proof of Theorem 2.2

Proof. We construct two distinct structural production functions that generate the same identified objects ϕ_t, $\bar{f}_t$, and $\bar{g}_t$ (where $\bar{g}_t(k,l,m) \equiv \frac{\partial f_t}{\partial m}\big|_{\omega=\phi_t^{-1}(k,l,m)}$), and that both satisfy Assumptions 2.1–2.4.

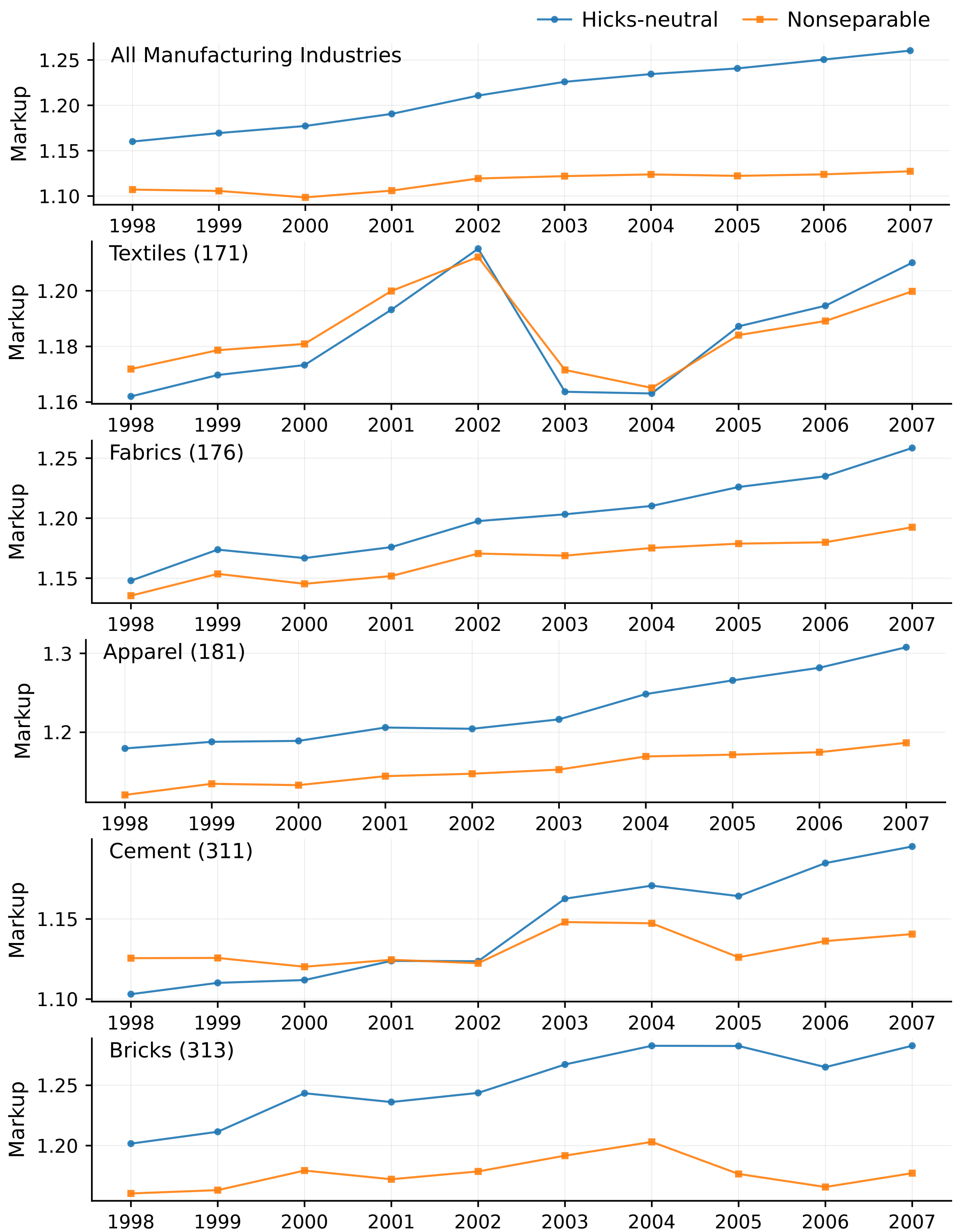
For transparency, we present the construction for a single pair (k,l); the argument extends immediately to all (k,l). Fix (k,l) and suppress them from the notation. Sup-

Figure 8: Trends in Markup Dispersion



Note: This figure plots time trends in markup dispersion, measured by the Theil index, for the pooled sample and for each of the five largest three-digit industries. The blue series corresponds to markups estimated under a Hicks-neutral specification, and the orange series to markups estimated under our nonseparable model; both are estimated within the same framework described in the main text.

Figure 9: Trends in Revenue-Weighted Markups



Note: This figure plots time trends in revenue-weighted markups, where weights are firm-level revenue shares within each industry-year cell. Markups are computed from the estimated production functions using the De Loecker and Warzynski (2012) approach.

pose $\omega = 2m$ is the identified inverse input demand, and consider the family of structural production functions

$$f(m, \omega) = b_1 m^2 + b_2\, m\omega + b_3\, \omega^2,$$

with materials elasticity $f_m(m, \omega) = 2b_1 m + b_2\, \omega$ and productivity elasticity $f_\omega(m, \omega) = b_2\, m + 2b_3\, \omega$. Substituting $\omega = 2m$ yields the reduced-form production function and reduced-form materials derivative:

$$\bar{f}(m) = (b_1 + 2b_2 + 4b_3)\, m^2, \qquad \bar{g}(m) = 2(b_1 + b_2)\, m.$$

These two identified objects impose only two restrictions on the three structural parameters (b_1, b_2, b_3):

$$b_1 + 2b_2 + 4b_3 = C_1, \qquad b_1 + b_2 = C_2,$$

leaving a one-parameter family of structural functions that are observationally equivalent on the equilibrium manifold $\{\omega = 2m\}$. Any two members of this family agree on $\bar{f}$, $\bar{g}$, and the associated ϕ, yet differ in their structural elasticities — and hence in the factor bias of technological change — at input bundles off the manifold.

The source of nonidentification is a dimension mismatch: the structural function $f_t(k, l, m, \omega)$ has four arguments, but data constrain it only on the three-dimensional equilibrium manifold $\{m = \phi_t(k, l, \omega)\}$, leaving its behavior in the transverse direction undetermined. □

F.2 Proof of Lemma 2.3

Proof. By Lemma 1 of Ackerberg et al. (2022),[34] it suffices to show that, for any two points (x^A, m^A) and (x^B, m^B) in $\mathscr{S}_t^{xm}$, the associated productivity indices can be ranked; that is,

$$\phi_t^{-1}(x^A, m^A) \lesseqgtr \phi_t^{-1}(x^B, m^B).$$

The proof of Lemma 2.2 has already established identification of the reduced-form function $\bar{\phi}_t(x, v, \xi)$. We use this to construct a chain of iso-productivity transfers under Assumption 2.5.

Step 0: A consequence of Assumption 2.5(i). Under Assumption 2.5(i) and the normalization $\omega_{jt} \sim U(0, 1)$, the support of ω_{jt} is $(0, 1)$ regardless of ω_{jt-1}. Since $m_{jt} = \phi_t(x_{jt}, \omega_{jt})$ and ϕ_t is continuous and strictly increasing in ω_{jt}, the conditional support $\mathscr{S}_t^{m|x,v} = \{\phi_t(x, \omega) :$

[34] Their outcome variable y corresponds to our m.

$\omega \in (0,1)\}$ depends on x but not on v. That is,

$$\mathscr{S}_t^{m|x,v} = \mathscr{S}_t^{m|x} \quad \text{for all } v \in \mathscr{S}_t^{v|x}. \tag{43}$$

Step 1: Choose v^A and v^B. Since $(x^A, m^A) \in \mathscr{S}_t^{xm}$, there exists v^A such that $(x^A, m^A, v^A) \in \mathscr{S}_t^{xmv}$. Similarly, choose v^B with $(x^B, m^B, v^B) \in \mathscr{S}_t^{xmv}$.

Step 2: Construct a chain from v^A to v^B. By Assumption 2.5(ii), $\mathscr{S}_t^v$ is path-connected, so there exists a continuous path $\gamma : [0,1] \to \mathscr{S}_t^v$ with $\gamma(0) = v^A$ and $\gamma(1) = v^B$. By Assumption 2.5(iv), for each $s \in [0,1]$ there exists $\varepsilon(\gamma(s)) > 0$ such that

$$\bigcap_{\substack{v \in \mathscr{S}_t^v \\ \|v-\gamma(s)\| < \varepsilon(\gamma(s))}} \mathscr{S}_t^{x|v}$$

has nonempty interior. The open balls $\{B(\gamma(s), \varepsilon(\gamma(s)))\}_{s\in[0,1]}$ cover the compact set $\gamma([0,1])$. Extract a finite subcover indexed by $0 = s_0 < s_1 < \cdots < s_K = 1$, with consecutive balls overlapping. Write $v_k \equiv \gamma(s_k)$, so $v_0 = v^A$ and $v_K = v^B$.

For each k, Assumption 2.5(iv) yields a point $\hat{x}_k$ in the interior of $\bigcap_{\|v-v_k\|<\varepsilon(v_k),\, v\in\mathscr{S}_t^v} \mathscr{S}_t^{x|v}$. Because consecutive balls overlap, for each $k = 0, \ldots, K-1$ there exists $\tilde{v}_k \in B(v_k, \varepsilon(v_k)) \cap B(v_{k+1}, \varepsilon(v_{k+1})) \cap \mathscr{S}_t^v$, and by construction both $\hat{x}_k$ and $\hat{x}_{k+1}$ lie in $\mathscr{S}_t^{x|\tilde{v}_k}$. Together with $x^A \in \mathscr{S}_t^{x|v^A}$ and $x^B \in \mathscr{S}_t^{x|v^B}$, this gives the chain

$$x^A \overset{v_0=v^A}{\longleftrightarrow} \hat{x}_0 \overset{\tilde{v}_0}{\longleftrightarrow} \hat{x}_1 \overset{\tilde{v}_1}{\longleftrightarrow} \cdots \overset{\tilde{v}_{K-1}}{\longleftrightarrow} \hat{x}_K \overset{v_K=v^B}{\longleftrightarrow} x^B,$$

where $x \overset{v}{\leftrightarrow} x'$ means both (x, v) and (x', v) lie in $\mathscr{S}_t^{xv}$.

Step 3: Iso-productivity transfer along the chain. At each link, the transfer mechanism is the following. Given (x_j, m_j) at a control value v with $(x_j, v), (x_{j+1}, v) \in \mathscr{S}_t^{xv}$, compute $\xi = \bar{\phi}_t^{-1}(x_j, v, m_j)$ and construct $m_{j+1} = \bar{\phi}_t(x_{j+1}, v, \xi)$, which is feasible since $\bar{\phi}_t(x, v, \xi)$ is identified and strictly monotone in ξ. The transfer preserves ξ, and since v is the same at both endpoints, $\omega = h_t(\phi_{t-1}^{-1}(v), \xi)$ is also preserved. Hence $\phi_t^{-1}(x_{j+1}, m_{j+1}) = \phi_t^{-1}(x_j, m_j)$.

To move from one link to the next requires re-anchoring at a different v': the constructed (x_{j+1}, m_{j+1}) must satisfy $m_{j+1} \in \mathscr{S}_t^{m|x_{j+1},v'}$. This is guaranteed by (43), since $m_{j+1} = \phi_t(x_{j+1}, \omega)$ for some $\omega \in (0,1)$ and hence lies in $\mathscr{S}_t^{m|x_{j+1}}$ regardless of v'.

Applying this iteratively along the chain from Step 2, starting at (x^A, m^A) with v^A and terminating at x^B with v^B, we obtain a point $(x^B, m^{A\to B})$ satisfying $\phi_t^{-1}(x^B, m^{A\to B}) =$

$\phi_t^{-1}(x^A, m^A) \equiv \omega^A$.

Step 4: Ordering. We now have $(x^B, m^{A\to B})$ with $\phi_t^{-1}(x^B, m^{A\to B}) = \omega^A$ and the original (x^B, m^B) with $\phi_t^{-1}(x^B, m^B) = \omega^B$. Both points share the same $x = x^B$, and ϕ_t is strictly increasing in ω, so

$$\omega^A \lesseqgtr \omega^B \quad \Longleftrightarrow \quad m^{A\to B} \lesseqgtr m^B.$$

Since $m^{A\to B}$ is identified using the identified $\bar{\phi}_t$ function, the ordering is identified. □

F.3 Proof of Theorem 3.1

Proof. Fix an arbitrary value of $(k_{jt}, l_{jt}, m_{jt}, z_{jt})$ in its support. By the monotonicity (invertibility) of the materials policy, this pins down—and hence identifies—the corresponding ω_{jt}, which in turn pins down q_{jt} through the production function. We keep (j,t) subscripts when they help avoid ambiguity.

Let

$$R_{jt}(\epsilon) \equiv \Gamma_t\big(Q_{jt}, Z_{jt}, e^{\epsilon}\big) = \exp\big(\bar{\psi}(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon)\big), \qquad q_{jt} \equiv \log Q_{jt},$$

and define the (log) revenue elasticity with respect to quantity,

$$\gamma_q(q_{jt}, z_{jt}, \epsilon) \equiv \frac{\partial \log R_{jt}(\epsilon)}{\partial q_{jt}}.$$

Since $\partial R_{jt}/\partial q_{jt} = R_{jt}\gamma_q$ and $\partial q_{jt}/\partial Q_{jt} = 1/Q_{jt}$,

$$\Gamma_Q\big(Q_{jt}, Z_{jt}, e^{\epsilon}\big)\, Q_{jt} = \gamma_q(q_{jt}, z_{jt}, \epsilon)\, R_{jt}(\epsilon). \tag{44}$$

Step 1: a key moment from the materials FOC. The FOC of (24) implies

$$\mathbb{E}_{\epsilon}\big[\Gamma_Q\big(Q_{jt}, Z_{jt}, e^{\epsilon}\big) \mid K_{jt}, L_{jt}, Z_{jt}, \omega_{jt}\big]\ F_{M_{jt}} = \rho_t.$$

Let $f_{m_{jt}} \equiv F_{M_{jt}} M_{jt}/Q_{jt}$ denote the materials elasticity in the quantity technology. Multiplying the FOC by M_{jt} and using $F_{M_{jt}} M_{jt} = f_{m_{jt}} Q_{jt}$ gives

$$f_{m_{jt}}\, \mathbb{E}_{\epsilon}\big[\Gamma_Q\big(Q_{jt}, Z_{jt}, e^{\epsilon}\big)\, Q_{jt}\big] = \rho_t M_{jt}. \tag{45}$$

Using (44), define

$$X_{jt} \equiv \mathbb{E}_{\epsilon}[\gamma_q(q_{jt}, z_{jt}, \epsilon)\, R_{jt}(\epsilon)]\,,$$

so that (45) becomes

$$f_{m_{jt}} X_{jt} = \rho_t M_{jt}. \tag{46}$$

Step 2: derivative moments from $\bar{\psi}$. Because $\bar{\psi}(\cdot, \epsilon)$ is identified (under the normalization $\epsilon \sim N(0,1)$), so are its derivatives with respect to (k, l, m). Differentiating with respect to m_{jt} and applying the chain rule yields

$$\bar{\psi}_m(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon) = \gamma_q(q_{jt}, z_{jt}, \epsilon) \left(f_{m_{jt}} + f_{\omega_{jt}} (\phi^{-1})_{m_{jt}} \right), \tag{47}$$

where $(\phi^{-1})_{m_{jt}}$ is evaluated at $(k_{jt}, l_{jt}, m_{jt}, z_{jt})$. Multiplying by $R_{jt}(\epsilon)$ and taking expectations gives

$$d_{1,jt} \equiv \mathbb{E}_\epsilon \Big[R_{jt}(\epsilon)\, \bar{\psi}_m(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon) \Big] = \left(f_{m_{jt}} + f_{\omega_{jt}} (\phi^{-1})_{m_{jt}} \right) X_{jt}. \tag{48}$$

Next define

$$c_{1,jt}(\epsilon) \equiv \bar{\psi}_k(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon) + \bar{\psi}_l(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon) + \bar{\psi}_m(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon),$$

$$\alpha_{jt} \equiv f_{k_{jt}} + f_{l_{jt}} + f_{m_{jt}}, \qquad c_{2,jt} \equiv (\phi^{-1})_{k_{jt}} + (\phi^{-1})_{l_{jt}} + (\phi^{-1})_{m_{jt}}.$$

Summing $\bar{\psi}_i = \gamma_q \big(f_i + f_{\omega_{jt}} (\phi^{-1})_i \big)$ over $i \in \{k, l, m\}$ yields

$$c_{1,jt}(\epsilon) = \gamma_q(q_{jt}, z_{jt}, \epsilon) \left(\alpha_{jt} + c_{2,jt} f_{\omega_{jt}} \right), \tag{49}$$

and multiplying by $R_{jt}(\epsilon)$ and taking expectations gives

$$d_{2,jt} \equiv \mathbb{E}_\epsilon \Big[c_{1,jt}(\epsilon)\, R_{jt}(\epsilon) \Big] = \left(\alpha_{jt} + c_{2,jt} f_{\omega_{jt}} \right) X_{jt}. \tag{50}$$

Step 3: identify $f_{\omega_{jt}}/\alpha_{jt}$ and γ_q up to scale. Equations (46), (48), and (50) imply

$$f_{m_{jt}} X_{jt} = \rho_t M_{jt}, \qquad d_{1,jt} = \big(f_{m_{jt}} + f_{\omega_{jt}} (\phi^{-1})_{m_{jt}} \big) X_{jt}, \qquad d_{2,jt} = \big(\alpha_{jt} + c_{2,jt} f_{\omega_{jt}} \big) X_{jt}.$$

Eliminating $f_{m_{jt}} X_{jt}$ using (46) gives $d_{1,jt} - \rho_t M_{jt} = f_{\omega_{jt}} (\phi^{-1})_{m_{jt}} X_{jt}$. Substitute $X_{jt} =$

$d_{2,jt}/(\alpha_{jt} + c_{2,jt} f_{\omega_{jt}})$ to obtain

$$f_{\omega_{jt}} = \frac{\alpha_{jt}\,(d_{1,jt} - \rho_t M_{jt})}{(\phi^{-1})_{m_{jt}}\, d_{2,jt} - c_{2,jt}\,(d_{1,jt} - \rho_t M_{jt})} \;\equiv\; \alpha_{jt}\,\tilde{f}_{\omega_{jt}}. \tag{51}$$

Hence $\tilde{f}_{\omega_{jt}}$ is identified. Let ϵ_{jt} denote the implied (and hence identified) idiosyncratic demand disturbance obtained by inverting the reduced-form revenue function,

$$\epsilon_{jt} = \bar{\psi}^{-1}(k_{jt}, l_{jt}, m_{jt}, z_{jt}, r_{jt}).$$

Using (49) and $\alpha_{jt} + c_{2,jt} f_{\omega_{jt}} = \alpha_{jt}(1 + c_{2,jt}\tilde{f}_{\omega_{jt}})$, we obtain

$$\gamma_q(q_{jt}, z_{jt}, \epsilon_{jt}) = \frac{1}{\alpha_{jt}}\,\frac{c_{1,jt}(\epsilon_{jt})}{1 + c_{2,jt}\tilde{f}_{\omega_{jt}}}, \tag{52}$$

so γ_q is identified up to the scale factor $1/\alpha_{jt}$.

Step 4: recover elasticities up to RTS. For $i \in \{k, l, m\}$, the identity $\bar{\psi}_i = \gamma_q\big(f_{i_{jt}} + f_{\omega_{jt}}(\phi^{-1})_{i_{jt}}\big)$ implies

$$\begin{aligned} f_{i_{jt}} &= \frac{\bar{\psi}_i(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon_{jt})}{\gamma_q(q_{jt}, z_{jt}, \epsilon_{jt})} - f_{\omega_{jt}}(\phi^{-1})_{i_{jt}} \\ &= \alpha_{jt}\left(\frac{\bar{\psi}_i(k_{jt}, l_{jt}, m_{jt}, z_{jt}, \epsilon_{jt})}{c_{1,jt}(\epsilon_{jt})}\,(1 + c_{2,jt}\tilde{f}_{\omega_{jt}}) - \tilde{f}_{\omega_{jt}}(\phi^{-1})_{i_{jt}}\right) \;\equiv\; \alpha_{jt}\,\tilde{f}_{i_{jt}}. \end{aligned} \tag{53}$$

Hence $\tilde{f}_{i_{jt}}$ is identified for $i \in \{k, l, m\}$; together with (51), this yields identification of $\tilde{f}_{i_{jt}}$ for $i \in \{k, l, m, \omega\}$. □

F.4 Proof of Theorem 3.2

Proof. Under Assumption 3.6, the production function admits the representation

$$f_t(k, l, m, \omega) = \alpha k + g_t(\tilde{l}, \tilde{m}, \omega),$$

where $\tilde{l} \equiv l - k$, $\tilde{m} \equiv m - k$, and α is the known degree of homogeneity. The partial derivatives of g_t satisfy

$$\frac{\partial g_t}{\partial \tilde{l}} = f_l, \qquad \frac{\partial g_t}{\partial \tilde{m}} = f_m, \qquad \frac{\partial g_t}{\partial \omega} = f_\omega.$$

By Theorem 3.1, (f_l, f_m, f_ω) are identified up to RTS; with α known, they are point-identified at each observation. Hence the gradient $\nabla g_t \equiv (f_l, f_m, f_\omega)$ is identified at each point on the equilibrium manifold.

It remains to show that the identified gradient determines g_t up to an additive constant. On the equilibrium manifold, $\tilde{m}_{jt} = \phi_t(k_{jt}, \tilde{l}_{jt} + k_{jt}, z_{jt}, \omega_{jt}) - k_{jt}$. For any fixed (z_{jt}, ω_{jt}), the Jacobian of the mapping $(k_{jt}, l_{jt}) \mapsto (\tilde{l}_{jt}, \tilde{m}_{jt})$ is

$$\frac{\partial(\tilde{l}, \tilde{m})}{\partial(k, l)} = \begin{pmatrix} -1 & 1 \\ \dfrac{\partial \phi_t}{\partial k} - 1 & \dfrac{\partial \phi_t}{\partial l} \end{pmatrix},$$

whose determinant equals $1 - \partial\phi_t/\partial k - \partial\phi_t/\partial l \neq 0$ by Assumption 3.7. Hence $(k_{jt}, l_{jt}) \mapsto (\tilde{l}_{jt}, \tilde{m}_{jt})$ is a local diffeomorphism for each fixed (z_{jt}, ω_{jt}), so $(\tilde{l}_{jt}, \tilde{m}_{jt})$ fills out an open subset of $\mathbb{R}^2$. Because ω_{jt} varies over an interval, the support of $(\tilde{l}_{jt}, \tilde{m}_{jt}, \omega_{jt})$ contains a connected open set $\mathscr{S} \subseteq \mathbb{R}^3$.

Since g_t is C^1 and its gradient is identified on $\mathscr{S}$, the fundamental theorem of line integrals gives, for any reference point $(\tilde{l}_0, \tilde{m}_0, \omega_0) \in \mathscr{S}$ and any path $\gamma \subset \mathscr{S}$ from $(\tilde{l}_0, \tilde{m}_0, \omega_0)$ to $(\tilde{l}, \tilde{m}, \omega)$,

$$g_t(\tilde{l}, \tilde{m}, \omega) - g_t(\tilde{l}_0, \tilde{m}_0, \omega_0) = \int_\gamma \nabla g_t \cdot \mathrm{d}s,$$

where the right-hand side is identified because ∇g_t is identified on $\mathscr{S}$ and the integral is path-independent (since ∇g_t is by construction a gradient field). Therefore g_t is identified on $\mathscr{S}$ up to the additive constant $g_t(\tilde{l}_0, \tilde{m}_0, \omega_0)$, and $f_t = \alpha k + g_t$ is identified up to an additive constant. □

G Estimation Details

G.1 Imposing Homogeneity

Under specification (16), homogeneity of degree r in (K, L, M) requires that the second-order interaction terms respect the constraint that scaling all inputs by a common factor λ scales output by λ^r. At the coefficient level, this is equivalent to the following linear equality constraints:[35]

$$
\begin{aligned}
2\beta_{kk} + \beta_{kl} + \beta_{km} &= 0, \\
2\beta_{ll} + \beta_{kl} + \beta_{lm} &= 0, \\
2\beta_{mm} + \beta_{km} + \beta_{lm} &= 0, \\
\beta_{k\omega} + \beta_{l\omega} + \beta_{m\omega} &= 0.
\end{aligned} \tag{54}
$$

In implementation, we impose (54) by reparameterization: we solve these linear restrictions for a subset of coefficients and substitute them back into the objective, thereby reducing the dimension of the parameter vector to be estimated.

Specifically, we solve the first three constraints for β_{kk}, β_{ll}, β_{mm} and the fourth for $\beta_{k\omega}$:

$$
\beta_{kk} = -\tfrac{1}{2}(\beta_{kl}+\beta_{km}), \qquad \beta_{ll} = -\tfrac{1}{2}(\beta_{kl}+\beta_{lm}), \qquad \beta_{mm} = -\tfrac{1}{2}(\beta_{km}+\beta_{lm}), \qquad \beta_{k\omega} = -(\beta_{l\omega}+\beta_{m\omega}).
$$

The free parameter vector is therefore $(\beta_0, \beta_k, \beta_l, \beta_m, \beta_\omega, \beta_{kl}, \beta_{km}, \beta_{lm}, \beta_{l\omega}, \beta_{m\omega}, \beta_{\omega\omega})$, which has eleven elements instead of the original fifteen.

Substituting into (16) and collecting terms by free parameter, the reparameterized production function is

$$
\begin{aligned}
f(k_{jt}, l_{jt}, m_{jt}, \omega_{jt}) = {} & \beta_0 + \beta_k\, k_{jt} + \beta_l\, l_{jt} + \beta_m\, m_{jt} + \beta_\omega\, \omega_{jt} \\
& - \tfrac{\beta_{kl}}{2}(k_{jt} - l_{jt})^2 - \tfrac{\beta_{km}}{2}(k_{jt} - m_{jt})^2 - \tfrac{\beta_{lm}}{2}(l_{jt} - m_{jt})^2 \\
& + \beta_{l\omega}(l_{jt} - k_{jt})\omega_{jt} + \beta_{m\omega}(m_{jt} - k_{jt})\omega_{jt} + \beta_{\omega\omega}\, \omega_{jt}^2.
\end{aligned} \tag{55}
$$

Note that all second-order input interactions now enter through pairwise log-ratios $(k_{jt} - l_{jt})$, $(k_{jt} - m_{jt})$, and $(l_{jt} - m_{jt})$, while the productivity interactions enter through $(l_{jt} - k_{jt})$ and $(m_{jt} - k_{jt})$.

Differentiating (55) with respect to each argument yields the model-implied elasticities

[35] Flynn et al. (2019) impose analogous linear restrictions when imposing constant returns to scale under imperfect competition.

used in the minimum-distance criterion (17):

$$\check{f}_{k_{jt}} = \beta_k + \beta_{kl}(l_{jt} - k_{jt}) + \beta_{km}(m_{jt} - k_{jt}) - (\beta_{l\omega} + \beta_{m\omega})\,\omega_{jt}, \tag{56}$$

$$\check{f}_{l_{jt}} = \beta_l + \beta_{kl}(k_{jt} - l_{jt}) + \beta_{lm}(m_{jt} - l_{jt}) + \beta_{l\omega}\,\omega_{jt}, \tag{57}$$

$$\check{f}_{m_{jt}} = \beta_m + \beta_{km}(k_{jt} - m_{jt}) + \beta_{lm}(l_{jt} - m_{jt}) + \beta_{m\omega}\,\omega_{jt}, \tag{58}$$

$$\check{f}_{\omega_{jt}} = \beta_\omega + \beta_{l\omega}(l_{jt} - k_{jt}) + \beta_{m\omega}(m_{jt} - k_{jt}) + 2\beta_{\omega\omega}\,\omega_{jt}. \tag{59}$$

As a consistency check, observe that $\check{f}_{k_{jt}} + \check{f}_{l_{jt}} + \check{f}_{m_{jt}} = \beta_k + \beta_l + \beta_m$ for every $(k_{jt}, l_{jt}, m_{jt}, \omega_{jt})$, confirming that returns to scale are constant across observations, as required by homogeneity.

G.2 Imposing Optional Shape Restrictions

To ensure that the estimated production function conforms to standard theoretical regularities, we optionally augment the minimum-distance criterion (17) with *one-sided quadratic penalties* that are incurred only when a restriction is violated.

Let $[a]_+ \equiv \max\{a, 0\}$. Denote by $F(\cdot; \boldsymbol{\beta})$ the structural production function in (16), with first and second partial derivatives F_a and F_{ab} evaluated at $(K_{jt}, L_{jt}, M_{jt}, \omega_{jt})$. For each input $I \in \{K, L, M\}$, we consider the following restrictions: (i) concavity in input I, $F_{II} \leq 0$; (ii) monotonicity in productivity, $F_\omega \geq 0$; and (iii) complementarity between input I and productivity, $F_{I\omega} \geq 0$. The penalized objective is

(60)

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & Q(\boldsymbol{\beta}) \;+\; \lambda_{II} \sum_{j,t} \big[\, F_{II}(K_{jt}, L_{jt}, M_{jt}, \omega_{jt}; \boldsymbol{\beta}) \,\big]_+^2 \\ & + \; \lambda_\omega \sum_{j,t} \big[-F_\omega(K_{jt}, L_{jt}, M_{jt}, \omega_{jt}; \boldsymbol{\beta}) \,\big]_+^2 \;+\; \lambda_{I\omega} \sum_{j,t} \big[-F_{I\omega}(K_{jt}, L_{jt}, M_{jt}, \omega_{jt}; \boldsymbol{\beta}) \,\big]_+^2, \end{aligned}$$

where $Q(\boldsymbol{\beta})$ denotes the baseline minimum-distance objective in (17), and $\lambda_{II}, \lambda_\omega, \lambda_{I\omega} > 0$ are user-chosen penalty weights (tuning parameters, not estimated).

The penalties are one-sided: when a restriction is satisfied, the corresponding term equals zero and imposes no cost; when it is violated, the objective increases quadratically with the magnitude of the violation.

H Data Details and Summary Statistics

We construct the variables and estimation sample following Brandt et al. (2012). Nominal revenue is deflated using industry-specific output deflators, and nominal expenditures on intermediate inputs are deflated using the corresponding input deflators, both provided in Brandt et al. (2012). Real capital is constructed using the perpetual inventory method described therein: the initial nominal capital stock is estimated from a firm's reported fixed assets at original purchase prices and the year the firm was established, and subsequent real capital stocks are obtained by deflating annual investment using the investment price deflator constructed by Brandt and Rawski (2008) and applying a depreciation rate of 9%.[36] Labor is measured as the number of employees.

The ASIE covers all state-owned enterprises and all non-state firms with annual revenue from principal business exceeding 5 million RMB. As documented in Brandt et al. (2014), the data from the 2004 Economic Census year differ from the regular annual surveys in coverage and variable availability. Most notably, the 2004 wave includes a broader set of firms captured by the census enumeration, which leads to a noticeable increase in the number of observed firms relative to adjacent years (Table 7). Because the firms captured in 2004 satisfy the same above-scale threshold and report the same core production variables used in our analysis, we retain the 2004 observations.

The panel is unbalanced: firms enter and exit the sample as they cross the reporting threshold or begin and cease operations. We do not restrict attention to a balanced panel, so our estimates reflect the full cross-section of above-scale firms in each year. Industry classification follows the concordance procedure in Brandt et al. (2012), which harmonizes the change from the GB/T 1994 to the GB/T 2002 national industry classification system that took effect in 2003.

We classify firms into three ownership categories. A firm is defined as an SOE if its registered ownership type is recorded as state-owned, as foreign if it is recorded as foreign-invested or Hong Kong/Macau/Taiwan-invested, and as domestic private otherwise. In our estimation, the resulting vector of ownership indicators enters both as a demand shifter and as a control variable in the Markov process for productivity.

We implement several standard data-cleaning steps. We first drop observations with nonpositive values for the key variables in the analysis, including revenue, capital, labor, intermediate inputs, wages, and value added. We then exclude firms with fewer than seven employees. Finally, for each industry-specific sample as well as for the pooled sample, we regress revenue on capital, labor, and materials using OLS, and winsorize observations based on the 1st and 99th percentiles of the resulting residuals.

[36] See Brandt et al. (2014) for a detailed discussion of capital stock measurement and other variable construction issues in the ASIE data.

The following two tables present summary statistics and time trends for the key variables in our sample.

Table 6: Summary Statistics for the Chinese Manufacturing Panel

Variable	*N*	Mean	SD	P25	Median	P75
Panel A: Continuous Variables						
Revenue	2,080,589	76,284.51	727,010.75	7,534.00	16,146.00	41,111.00
Real Capital	2,080,589	38,383.52	550,893.80	1,597.32	4,359.67	13,414.85
Employment	2,080,589	278.24	1,313.72	50.00	105.00	230.00
Material Expenditure	2,080,589	58,235.87	557,971.25	5,700.00	12,368.00	31,540.00
Panel B: Ownership Types						
Ownership Type	# Firms	# Observations				
SOE	73,901	298,141				
Foreign	95,622	412,903				
Domestic Private	390,033	1,369,545				
Total firms	559,556					
Total observations		2,080,589				

Note: All monetary values are expressed in thousands of Chinese yuan.

Table 7: Time Trends in Firm Medians and Ownership Composition for the Chinese Manufacturing Panel

Year	Revenue	Real Capital	Employment	Material Exp.	SOE	Foreign	D. Private	Total Firms
1998	10,180.00	4,569.63	132	8,190.00	53,981	25,616	74,770	154,367
1999	10,800.00	4,990.90	128	8,519.00	48,798	25,923	75,447	150,168
2000	12,000.00	5,022.88	125	9,352.00	41,048	27,449	81,213	149,710
2001	12,634.00	4,670.47	120	9,855.00	33,560	30,251	92,303	156,114
2002	13,800.00	4,539.42	117	10,696.00	28,814	33,152	103,580	165,546
2003	15,900.00	4,470.36	113	12,194.00	23,125	37,106	119,575	179,806
2004	14,375.00	3,468.02	92	11,055.00	25,974	55,237	181,016	262,227
2005	19,085.00	4,186.13	100	14,438.00	16,579	54,211	183,161	253,951
2006	21,287.00	4,305.32	95	15,960.00	15,191	58,721	212,554	286,466
2007	25,170.00	4,403.06	90	18,718.00	11,071	65,237	245,926	322,234

Note: Total Firms is the number of distinct firms observed in that year.

I Empirical Appendix for Perfect Competition: Chile and Colombia

Our main empirical analysis focuses on imperfect competition. In this appendix, we briefly report the corresponding results under perfect competition. This exercise serves three purposes: to assess the performance of our estimators, to document the misspecification bias induced by Hicks-neutral models, and to provide empirical support for the homogeneity condition.

We apply our estimators to the Chilean and Colombian manufacturing panels used by GNR, which have also been widely used in the IO and international trade literatures. Descriptive statistics and estimation results are reported in Table 8.

Table 9 reports output elasticities estimated using our nonseparable reduced-form estimator (NS; Section 2.3), our homogeneous structural estimator (HM; Section 2.3), and GNR's Hicks-neutral procedure (GNR). When implementing the HM estimator, we further enforce concavity ($F_{MM} < 0$) and complementarity ($F_{M\omega} > 0$) via the one-sided penalty terms described in Appendix G.

Both NS and HM produce plausible output elasticities. Because the NS specification nests Hicks neutrality as a special case, the comparison between NS and GNR isolates the consequences of Hicks-neutral misspecification. Across industries in both countries, the

Table 8: Descriptive Statistics for Chilean and Colombian Manufacturing Estimation Samples

Variable	N	Mean	SD	P25	Median	P75
Panel A: Chile (1979–1996; firms with >10 employees)						
Sales	70,446	134,365.38	978,174.30	3,047.13	11,525.05	46,549.62
Capital	70,446	20,377.95	148,443.33	420.97	1,416.29	6,048.61
Labor	70,446	85.54	190.05	19.82	33.26	71.75
Material Expenditure	70,446	34,917.39	239,225.70	1,854.43	4,344.52	14,572.38
Number of firms	9,827					
Number of observations	70,446					
Panel B: Colombia (1981–1991; plants with >10 employees)						
Sales	41,493	24,316.79	102,345.12	983.00	3,013.27	12,536.32
Capital	41,493	8,178.67	140,260.28	222.40	786.48	3,210.73
Labor	41,493	119.37	268.46	18.10	40.50	106.23
Material Expenditure	41,493	15,088.74	74,643.60	511.58	1,691.70	7,647.72
Number of firms	6,633					
Number of observations	41,493					

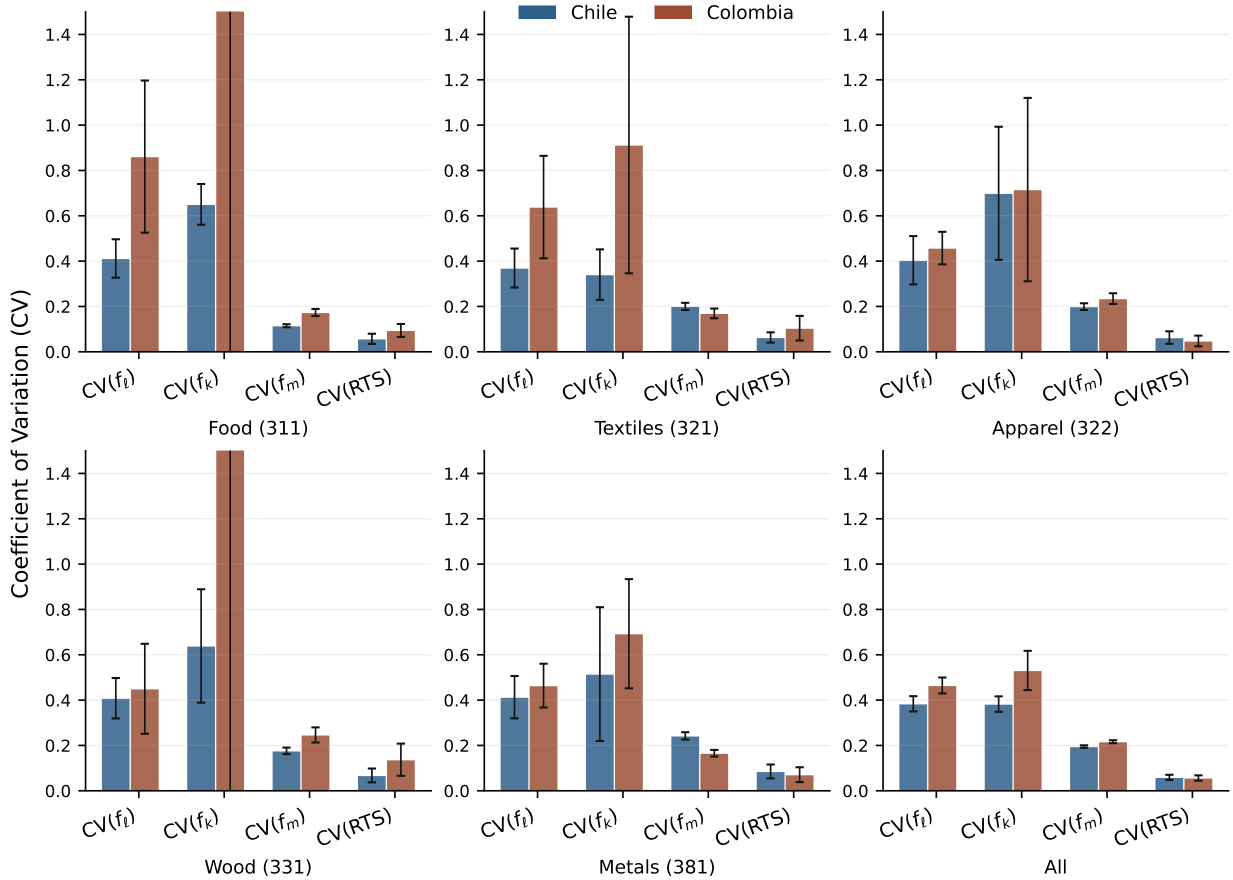
Note: We construct all variables following the procedures in GNR. The Chilean sample is based on the manufacturing census compiled by Chile's Instituto Nacional de Estadística and covers firms with more than 10 employees during 1979–1996. The Colombian sample comes from the manufacturing census and covers plants with more than 10 employees during 1981–1991. Monetary values are measured in local pesos. Labor is measured as a wage-weighted sum of blue- and white-collar workers, where blue-collar employment is weighted by the ratio of the average blue-collar wage to the average white-collar wage.

Hicks-neutral specification systematically overstates the labor elasticity, by about 22% on average, and therefore also overstates returns to scale. In some cases, this changes the implied classification of the technology as exhibiting increasing, constant, or decreasing returns to scale.

Figure 10 shows that the NS-based estimates of returns to scale display little dispersion across firms. This pattern is consistent with approximate homogeneity and therefore provides empirical support for the HM specification.[37]

[37] A formal test of homogeneity based on whether the standard deviation of returns to scale is zero is nonstandard, because zero lies on the boundary of the parameter space (Andrews, 2001). One practical way to address this boundary issue is to test whether the standard deviation exceeds a small positive threshold δ (Wellek, 2010).

Figure 10: Coefficients of Variation of Output Elasticities and Returns to Scale: Chile and Colombia



Note: This figure reports the coefficients of variation (CVs) of the estimated output elasticities and the implied returns to scale (RTS) across manufacturing industries in the Chilean and Colombian samples, obtained using the reduced-form nonseparable estimator described in Section 2.3. Error bars indicate 95% confidence intervals. Importantly, no restrictions on returns to scale are imposed in this estimation.

Table 9: Average Output Elasticities in Chilean and Colombian Manufacturing

	Industry (ISIC Code)																	
	Food (311)			Textiles (321)			Apparel (322)			Wood Products (331)			Fabricated Metals (381)			All		
	NS	HM	GNR	NS	HM	GNR	NS	HM	GNR	NS	HM	GNR	NS	HM	GNR	NS	HM	GNR
Chile																		
Labor	0.23	0.23	0.28	0.36	0.36	0.45	0.36	0.35	0.45	0.33	0.33	0.40	0.42	0.42	0.52	0.29	0.29	0.38
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)
Capital	0.10	0.10	0.11	0.12	0.13	0.11	0.09	0.09	0.06	0.08	0.09	0.07	0.13	0.14	0.13	0.16	0.16	0.16
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Materials	0.67	0.68	0.67	0.55	0.55	0.54	0.56	0.57	0.56	0.60	0.60	0.59	0.50	0.50	0.50	0.56	0.56	0.55
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
RTS	1.00	1.00	1.05	1.03	1.03	1.10	1.01	1.01	1.08	1.02	1.02	1.06	1.06	1.06	1.15	1.01	1.01	1.09
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.01)
Cap. int.	0.41	0.41	0.39	0.35	0.35	0.24	0.24	0.26	0.14	0.26	0.26	0.18	0.32	0.35	0.25	0.54	0.55	0.43
	(0.03)	(0.03)	(0.01)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)
Colombia																		
Labor	0.18	0.20	0.22	0.26	0.26	0.32	0.36	0.35	0.42	0.36	0.35	0.44	0.32	0.32	0.43	0.29	0.29	0.35
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)	(0.05)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Capital	0.04	0.05	0.12	0.08	0.09	0.16	0.06	0.06	0.05	0.04	0.07	0.04	0.10	0.11	0.10	0.09	0.09	0.14
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.00)	(0.00)	(0.01)
Materials	0.71	0.70	0.67	0.56	0.56	0.54	0.53	0.54	0.52	0.53	0.54	0.51	0.55	0.55	0.53	0.56	0.56	0.54
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
RTS	0.93	0.94	1.01	0.90	0.91	1.01	0.94	0.94	0.99	0.94	0.96	0.99	0.97	0.98	1.06	0.94	0.94	1.04
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)
Cap. int.	0.22	0.24	0.55	0.33	0.37	0.49	0.16	0.18	0.12	0.12	0.21	0.08	0.33	0.36	0.23	0.32	0.32	0.40
	(0.07)	(0.05)	(0.08)	(0.08)	(0.07)	(0.09)	(0.03)	(0.03)	(0.02)	(0.07)	(0.06)	(0.05)	(0.05)	(0.06)	(0.04)	(0.02)	(0.02)	(0.03)

Note: The table reports mean output elasticities of labor, capital, and materials under three estimators (NS, HM, and GNR) for the five largest manufacturing industries in Chile and Colombia, as well as a pooled specification (*All*). Returns to scale (RTS) are defined as the sum of the three elasticities. Capital intensity (Cap. int.) is measured as the ratio of the mean capital elasticity to the mean labor elasticity. Parentheses report bootstrap standard errors based on 200 replications. NS denotes our nonseparable estimator under perfect competition (see Section 2.3 for details). HM imposes an additional homogeneity restriction relative to NS (Section 2.3). GNR denotes the Hicks-neutral estimator proposed by GNR.