Beyond Labor Augmentation: Identifying the Bias of Technical Change

Daniel Ackerberg* Qingsong Pan†

First draft: July 2026 This draft: September 2026

Abstract

We study the nonparametric identification of production functions with both Hicks-neutral and factor-biased productivity. Our framework relaxes the conventional restriction that the factor-biased component be labor augmenting, as in Doraszelski and Jaumandreu (2018) and Demirer (2020), by allowing a scalar non-neutral productivity shock to interact flexibly with capital, labor, and materials. We develop a two-step estimator. In the first step, we recover an index of the non-neutral productivity component by estimating the structural function governing the ratio of the materials and labor output elasticities. In the second step, we substitute this index into the production function and use the ratio relation to reduce the dimensionality of the estimation problem, thereby lowering the number of instruments required. When output quantities are observed, our estimator is robust to arbitrary imperfect-output-market conditions. When only revenue is observed, it can remain consistent under additional structural assumptions, as in Ackerberg and De Loecker (2026). In our application to China's textile industry, our estimated model implies that moving the factor-biased productivity shock from its 1999 median to its 2007 median decreases output elasticities of capital and labor, and increases the output elasticity of materials. Fitting a general factor-augmenting model suggests that labor experienced the largest efficiency gains, capital efficiency also improved substantially, and materials efficiency was comparatively flat.

*Department of Economics, The University of Texas at Austin. Email: daniel.ackerberg@utexas.edu.
†School of Economics, Zhejiang University. Email: qingsongpan@utexas.edu.

Key words: Production functions; capital augmenting technical change; labor augmenting technical change; factor-biased productivity

1 Introduction

Technical change need not affect all inputs proportionally. A Hicks-neutral innovation shifts the production frontier without changing relative marginal products, whereas factor-biased technical change can alter the relative importance of capital, labor, and materials in production. This distinction is central to the Hicks–Acemoglu analysis of technical change (Hicks, 1932; Acemoglu, 2002, 2003), but it is difficult to recover empirically when productivity shocks are unobserved and firms choose inputs after observing them. In particular, an observed change in an input mix does not by itself reveal whether the underlying non-neutral technology shock is labor augmenting, capital augmenting, or more generally biased across factors.

Recent production-function methods make this problem tractable by adding a non-Hicks-neutral component to the standard Hicks-neutral productivity shock under homothetic separability, with CES as a special case.[1] Doraszelski and Jaumandreu (2018) and Demirer (2020) provide important examples in which the non-neutral component enters as a labor-augmenting productivity shock. The labor-augmentation restriction is economically transparent, but it is also substantive. A rise in the materials–labor ratio, for example, is consistent with labor augmentation in CES models, yet it need not imply that labor augmentation is the primitive source of the change. As a counterexample, under nested-CES technology, a purely capital-augmenting technical change can also raise the cost-minimizing materials–labor ratio when the capital–materials elasticity of substitution is lower than the elasticity of substitution between the capital–materials composite and labor. In addition, the homothetic-separability restrictions that allow flexible-input ratios alone to control for labor-augmenting productivity can rule out differential substitution patterns between capital and the two flexible inputs. An important empirical question is therefore whether one can identify the direction of technical change without prespecifying its factor-augmenting form or restricting its implied substitution patterns.

We develop such an approach. Our production technology is

$$q_{jt} = f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega^B_{jt}) + \omega^H_{jt},$$

[1] Zhao et al. (2024) relax the assumption of homothetic separability in a translog production-function model with labor-augmenting productivity.

where ω_{jt}^H is a Hicks-neutral productivity shock and ω_{jt}^B is a scalar non-Hicksian productivity shock. By definition, the Hicks-neutral productivity shock shifts all *marginal products* proportionally and therefore leaves their ratios unchanged. On the other hand, we treat the function f_t nonparametrically: ω_{jt}^B may interact with capital, labor, and materials in arbitrary ways subject to one key shape restriction. This shape restriction has a direct economic interpretation: conditional on a fixed input vector, the scalar shock ω_{jt}^B monotonically affects the materials–labor relative *marginal-product ratio*. In particular, an increase in ω_{jt}^B either strictly raises that ratio, indicating a materials-biased effect relative to labor, or strictly lowers it, indicating a labor-biased effect relative to materials. It therefore neither assigns ω_{jt}^B a primitive labor-, capital-, or materials-augmenting interpretation nor restricts the otherwise flexible way in which it interacts with inputs. For example, it does not place monotonicity restrictions on marginal-product ratios regarding capital (or other fixed inputs).

The restriction has an important implication. When labor and materials are flexible inputs chosen by cost-minimizing firms, their first-order conditions imply that the relative expenditure ratio $w_{jt}^M M_{jt}/(w_{jt}^L L_{jt})$ equals the ratio of the corresponding log-output elasticities. The common cost multiplier and the Hicks-neutral component, as well as demand factors, cancel from this ratio. Thus, under the maintained monotonicity condition, relative materials–labor expenditure provides an index for the biased productivity shock, conditional on inputs.

We combine this inversion with assumed first-order Markov processes for ω_{jt}^B and ω_{jt}^H in a two-step identification argument. The first step combines timing assumptions, exclusion and completeness conditions, and an IV-quantile restriction to identify the structural elasticity-ratio function and recover an index of biased productivity; the second applies related additive NPIV conditions to identify the production function and the Hicks-neutral productivity component. The argument builds on the timing- and information-set approach of Ackerberg et al. (2026), but allows two distinct productivity shocks instead of one.

A key feature of our two-step approach is that it uses the identified first-stage ratio function to reduce the dimensionality of the second-stage problem and, hence, the number of required instruments. This is particularly useful when an excluded instrument for materials is difficult to find, as emphasized by Gandhi et al. (2020). Gandhi et al. (2020) likewise use a first-order condition to reduce the dimensionality of the second stage. They use the profit-maximization first-order condition in levels for one flexible input under perfect output-market competition, whereas we use the ratio of cost-minimization first-order conditions for two flexible inputs under imperfect competition. They integrate the material first-order

condition to eliminate materials from the second-stage problem; we instead solve an ordinary differential equation (ODE) implied by the FOC ratio to compress the two flexible inputs (labor and materials) into a scalar index.

Our framework nests the labor-augmenting models of Doraszelski and Jaumandreu (2018) and Demirer (2020), since their maintained assumptions imply our relative-marginal-product monotonicity condition. This generality comes at a cost: without their purely labor-augmenting functional form and homothetic separability, the relative-expenditure inversion must condition on the full vector of inputs rather than on a lower-dimensional input ratio.

Other approaches to relaxing a purely labor-augmenting specification typically adopt a parametric CES form and impose additional structural restrictions. Oberfield and Raval (2021) allow for plant-level capital- and labor-augmenting productivity differences within a CES framework. Their benchmark treats capital as a statically adjustable input, and they separately examine the implications of adjustment costs. Zhang (2019) estimates a CES gross-output production function with capital-, labor-, and materials-augmenting efficiency terms. To accommodate the additional unobserved factor-efficiency term, he also relies on an inversion of the investment policy function in the spirit of Olley and Pakes (1996).

Our approach is complementary. On one hand it has two technology shocks (the scalar non-neutral productivity component and the Hicks-neutral productivity component), less than the three structural shocks in Zhang (2019). On the other hand, our biased productivity shock can interact almost arbitrarily with inputs in our flexible production function, while Zhang's biased productivity shocks are strictly factor augmenting in a parametric CES model. Our biased productivity shock may not only augment inputs—that is, scale their effective units up or down—but also alter the shape and curvature of the production function, for example through changes in the elasticities of substitution. Unlike the standard proxy-variable approaches of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015), our baseline model does not rely on inverting an input- or investment-demand policy to control for unobserved Hicks-neutral productivity. The cost is that this baseline assumes away a separate ex-post output shock, which proxy-variable approaches commonly accommodate. As a result, the baseline is more robust to unobserved demand shocks and demand-side market conditions. Alternatively, it could be used as a post-inversion procedure. In other words, when an application warrants a separate ex-post output shock, an optional OP/LP/ACF-style proxy-first-stage extension could purge that shock before applying our

estimator by nonparametrically regressing output on inputs and demand shifters.[2]

We apply the method to China's textile industry. Our principal empirical finding concerns the implications of the biased productivity shock. Moving only this component from its 1999 median to its 2007 median, while holding observed inputs fixed, lowers the labor output elasticity and raises the materials output elasticity in our flexible baseline specification and its two extensions. The capital output elasticity also declines in all three specifications. These changes are therefore directionally robust, although their magnitudes remain somewhat sensitive to specification, particularly for capital.

Under cost minimization, the first-order conditions for flexible inputs equate the materials–labor expenditure ratio with the corresponding output-elasticity ratio. Under these maintained conditions, the decline in the labor elasticity and the increase in the materials elasticity therefore correspond to a shift in the model-implied shares of variable costs away from labor and toward materials.[3]

Our finding of material-biased technical change in China's textile industry is consistent with that of Doraszelski and Jaumandreu (2018) for Spanish manufacturing data. However, we obtain this result under a more general specification of factor-biased technical change, i.e. without imposing ex ante that the source of the factor bias arises solely through labor augmentation. To shed light on the broader sources of this bias, we further decompose the non-neutral productivity component into factor-augmenting terms, which also provides a test of capital versus labor augmentation.

The coefficients of our sieve-based flexible production functions are hard to directly interpret. For an interpretable low-dimensional summary, we fit factor-augmenting CES surfaces to the three flexible production functions and compare those refits with a directly estimated CES specification with linear factor-augmenting loadings. After combining each factor-specific efficiency index with the Hicks-neutral index,[4] the four multi-factor-augmenting CES specifications imply 2007 composite indexes—normalized to 100 in 1999—of 322.22–469.11 for labor, 187.12–228.96 for capital, and 89.14–107.34 for materials; the Hicks-neutral index itself ranges from 105.96 to 116.01. Thus, labor experienced the largest efficiency gains, capital efficiency also improved substantially, and materials efficiency was comparatively flat. The deliberately restrictive pure labor-augmenting CES is reported below only as a diagnostic

[2]The base model of Zhang (2019) has this additional first stage, i.e., allowing an ex-post output shock.

[3]When capital is adjustable in the long run, the same reasoning may extend to capital's cost share, to the extent that the corresponding first-order condition holds.

[4]By factor efficiency, we mean the multiplier that converts observed units of a factor into effective units. In a purely labor-augmenting model, for example, labor efficiency is simply labor-augmenting productivity.

of the potential restriction-induced misspecification bias from imposing labor-only augmentation; it is neither a preferred specification nor a robustness check. The three CES refits are downstream projections of the flexible production-function estimates, whereas the direct CES specification imposes linear factor-augmenting loadings during estimation. Despite their different roles, the four multi-factor-augmenting CES specifications agree qualitatively on the factor-specific efficiency pattern.

The remainder of the paper proceeds as follows. Section 2 presents the firm problem, the productivity decomposition, and the relative-marginal-product restriction. Section 3 establishes identification, and Section 4 develops the sequential estimator and thefactor-augmenting CES projection. Section 5 describes the data, Section 6 reports the empirical evidence, and the conclusion summarizes the implications for measuring factor-biased technical change.

2 Model

We consider the model

$$q_{jt} = f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega^B_{jt}) + \omega^H_{jt}, \tag{1}$$

where q_{jt} is log output, k_{jt} is log capital input, ℓ_{jt} is log labor input, m_{jt} is log materials input, ω^B_{jt} is the unobserved non-Hicksian (biased) productivity shock, and ω^H_{jt} is a Hicks-neutral productivity shock. Let L_{jt} and M_{jt} denote labor and materials in levels, and let w^L_{jt} and w^M_{jt} denote their input prices. The baseline production function contains no separate ex-post output shock.[5] Thus the productivity components in (1) are the structural productivity terms observed by the firm when period-t flexible inputs are chosen.

Let $\mathcal{I}_{j,t-1}$ denote the firm's information set after period $t-1$ decisions and before the period-t idiosyncratic productivity shocks. Capital is predetermined, so $k_{jt} \in \mathcal{I}_{j,t-1}$. After observing current productivity, input prices, and capital, the firm chooses labor and materials as flexible inputs. Conditional on capital, productivity, and the desired log output level $\bar{q}_{jt}$, the static cost-minimization problem is

$$\min_{L,M} \; w^L_{jt} L + w^M_{jt} M \quad \text{subject to} \quad \bar{q}_{jt} = f_t(k_{jt}, \log L, \log M, \omega^B_{jt}) + \omega^H_{jt}. \tag{2}$$

[5] One can generalize the framework to accommodate ex-post output shocks at the cost of additional inversion-related assumptions, similar to those in Zhang (2019).

The Lagrangian is

$$\mathcal{L} = w_{jt}^{L} L + w_{jt}^{M} M + \lambda_{jt} \left[\bar{q}_{jt} - f_t(k_{jt}, \log L, \log M, \omega_{jt}^{B}) - \omega_{jt}^{H} \right]. \tag{3}$$

At an interior optimum, the first-order conditions with respect to L and M are

$$w_{jt}^{L} = \lambda_{jt} \frac{\partial f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B})}{\partial \ell} \frac{1}{L_{jt}}, \tag{4}$$

$$w_{jt}^{M} = \lambda_{jt} \frac{\partial f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B})}{\partial m} \frac{1}{M_{jt}}. \tag{5}$$

The labor and materials output elasticities are, respectively,

$$\frac{\partial f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B})}{\partial \ell}, \qquad \frac{\partial f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B})}{\partial m}. \tag{6}$$

Multiplying (4) by L_{jt} and (5) by M_{jt} gives

$$\begin{aligned} w_{jt}^{L} L_{jt} &= \lambda_{jt} \frac{\partial f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B})}{\partial \ell}, \\ w_{jt}^{M} M_{jt} &= \lambda_{jt} \frac{\partial f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B})}{\partial m}. \end{aligned} \tag{7}$$

Taking the ratio eliminates the multiplier and yields

$$\frac{w_{jt}^{M} M_{jt}}{w_{jt}^{L} L_{jt}} = \frac{\partial f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B})/\partial m}{\partial f_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B})/\partial \ell}. \tag{8}$$

The Hicks-neutral shock ω_{jt}^{H} does not enter the derivatives with respect to the flexible inputs. Any effect it has on the shadow value λ_{jt} is common to both first-order conditions and therefore cancels from (8). Conditional on $(k_{jt}, \ell_{jt}, m_{jt})$, this is the useful feature of the flexible-input ratio: it isolates variation in relative marginal products generated by the non-Hicksian shock.

Define the observable log relative expenditure of materials to labor as

$$\chi_{jt}^{ML} \equiv \log \left(\frac{w_{jt}^{M} M_{jt}}{w_{jt}^{L} L_{jt}} \right). \tag{9}$$

The term materials–labor denotes the ordered $M{:}L$ margin: materials relative to labor.

Using (8),

$$\chi_{jt}^{ML} = \rho_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^B), \tag{10}$$

where, writing f_ℓ and f_m for the partial derivatives of f_t with respect to ℓ and m,

$$\rho_t(k, \ell, m, \omega^B) \equiv \log f_m(k, \ell, m, \omega^B) - \log f_\ell(k, \ell, m, \omega^B). \tag{11}$$

Assumption 2.1 (Relative flexible-input bias)**.** Let MPL_{jt} and MPM_{jt} denote the marginal products of labor and materials implied by (1). For each t and each (k, ℓ, m) in the support of inputs, hold inputs fixed and vary only ω^B. The mapping

$$\omega^B \mapsto \frac{\text{MPM}_{jt}}{\text{MPL}_{jt}} \tag{12}$$

is strictly monotone on the conditional support of ω^B; that is, it is either strictly increasing or strictly decreasing conditional on (k, ℓ, m).[6]

Recall that Hicks-neutral productivity shifts all marginal products proportionally and therefore drops out of every relative marginal product. Assumption 2.1 can thus be understood as a restriction on the two-component productivity decomposition. The remaining scalar non-Hicksian productivity component, ω^B, is the dimension of technology that moves the prespecified materials–labor margin. Conditional on each fixed input vector, the assumption requires ω^B to shift MPM/MPL strictly monotonically. Thus, it is not a restriction that an arbitrary residual productivity term be intrinsically material- or labor-augmenting. Rather, after separating the Hicks-neutral component, the non-neutral productivity component indexes a one-dimensional monotone tilt in relative marginal products along the materials–labor margin. At a fixed input vector, an increase in ω^B either raises the materials-to-labor marginal-product ratio (material bias relative to labor) or lowers it (labor bias relative to materials). The direction may vary across input vectors and across periods. Assumption 2.1′ instead fixes an orientation within each period and bounds the slope away from zero. For interpretation and estimation it is useful to consider a stronger version of the same restriction.

Assumption 2.1′ (Uniform relative flexible-input bias)**.** The function $\rho_t(k, \ell, m, \omega^B)$ is

[6] The Hicks-neutral productivity component drops out of this ratio because it scales both marginal products by the same factor.

continuously differentiable in ω^B on the relevant support. For every t, there exist a sign $s_t \in \{-1, 1\}$ and a constant $c_t > 0$ such that, for every (k, ℓ, m, ω^B) in the period-t support,

$$s_t \frac{\partial \rho_t(k, \ell, m, \omega^B)}{\partial \omega^B} \geq c_t. \tag{13}$$

Assumption 2.1′ implies Assumption 2.1, but the converse need not hold. Assumption 2.1 is the pointwise injectivity condition used in Lemma 2.1. The strengthened version fixes a direction of bias within each period and rules out regions in which the relative marginal-product schedule is nearly flat in ω^B. The lower bound keeps the inverse well conditioned when the relative-expenditure index is estimated with sampling error.

Assumption 2.1 states the monotonicity that the inversion uses directly: conditional on observed inputs, the scalar non-Hicksian shock must move a relative marginal product one-to-one. The same monotonicity object also arises inside the labor-augmenting frameworks of Demirer (2020) and Doraszelski and Jaumandreu (2018). To see the connection with Demirer (2020), let $A = \exp(\omega^B)$ and suppose

$$Q = F_t(K, h_t(K, AL, M)) \exp(\omega^H), \tag{14}$$

where $h_t(K, \cdot, \cdot)$ is homogeneous and the elasticity of substitution $\sigma_t(K, AL, M)$ between effective labor and materials is everywhere on one side of one. Then

$$\frac{\text{MPM}}{\text{MPL}} = \frac{h_M(K, AL, M)}{A h_E(K, AL, M)}. \tag{15}$$

The outer aggregator and Hicks-neutral component cancel. By homogeneity, h_M/h_E depends on (AL, M) only through $M/(AL)$, so holding (K, L, M) fixed,

$$\frac{\partial}{\partial \omega^B} \log\left(\frac{\text{MPM}}{\text{MPL}}\right) = \frac{1 - \sigma_t(K, AL, M)}{\sigma_t(K, AL, M)}. \tag{16}$$

Thus if $\sigma_t < 1$ everywhere, the materials-to-labor marginal-product ratio is strictly increasing in ω^B; if $\sigma_t > 1$ everywhere, it is strictly decreasing. If σ_t is also bounded away from one, as in Demirer's corresponding regularity condition, the derivative is bounded away from zero and Assumption 2.1′ holds with the appropriate sign. The familiar non-Cobb–Douglas CES labor-augmenting specification in Doraszelski and Jaumandreu (2018) implies the same kind of strict monotonicity: holding inputs fixed, MPM/MPL is strictly monotone in the

labor-augmenting term away from the Cobb–Douglas case.

The point of Assumption 2.1 is to state this operative monotonicity directly rather than impose labor augmentation. This can be seen within a familiar homothetic nested-CES technology. Let $A = \exp(\omega^B)$ augment capital and consider the nested-CES technology

$$Q = [\beta X^\rho + (1-\beta)L^\rho]^{\gamma/\rho}, \tag{17}$$

$$X = [\alpha(AK)^\nu + (1-\alpha)M^\nu]^{1/\nu}, \tag{18}$$

with $\alpha, \beta \in (0,1)$, $\gamma > 0$, $K, L, M > 0$, and CES parameters $\rho, \nu < 1$ with $\rho, \nu \neq 0$. The parameter γ is the degree of returns to scale; $\gamma = 1$ is the constant-returns special case. The inner elasticity between capital and materials is $\sigma_{KM} = 1/(1-\nu)$, and the outer elasticity between the capital-materials composite and labor is $\sigma_{XL} = 1/(1-\rho)$. Direct differentiation gives

$$\frac{\text{MPM}}{\text{MPL}} = \frac{\beta(1-\alpha)}{1-\beta} L^{1-\rho} M^{\nu-1} X^{\rho-\nu}, \tag{19}$$

and hence, holding (K, L, M) fixed,

$$\frac{\partial}{\partial \omega^B} \log\left(\frac{\text{MPM}}{\text{MPL}}\right) = (\rho - \nu)\frac{\alpha(AK)^\nu}{\alpha(AK)^\nu + (1-\alpha)M^\nu}. \tag{20}$$

The second factor is positive, so the sign is the sign of $\rho - \nu = 1/\sigma_{KM} - 1/\sigma_{XL}$. Hence the mapping from ω^B to MPM/MPL is strictly monotone whenever $\sigma_{KM} \neq \sigma_{XL}$. The equality case is the knife-edge case in which the two nests have the same substitution elasticity and the capital-augmenting shock leaves the materials–labor marginal-product ratio unchanged. The direction of monotonicity depends on which elasticity is larger: if $\sigma_{KM} < \sigma_{XL}$, the shock strictly increases MPM/MPL and is material-biased relative to labor; if $\sigma_{KM} > \sigma_{XL}$, it strictly decreases this ratio and is labor-biased relative to materials. Thus Assumption 2.1 requires unequal nest elasticities, not a particular ranking between them. The strengthened Assumption 2.1′ requires the absolute derivative in the last display to be uniformly bounded away from zero on the estimation support, so the unequal-elasticity condition must hold away from the knife edge relevant for the observed support. The ω^B coordinate in these examples is the natural log-augmentation scale. An increasing relabeling into the reference scale used below preserves the monotonicity direction, although it changes cardinal derivative magnitudes.

The example also clarifies how to interpret movements in observed input ratios. Assumption 2.1 is a fixed-input restriction on relative marginal products; it is not, by itself, a statement about the re-optimized materials–labor ratio. For a cost-minimizing firm that reoptimizes labor and materials at fixed output, fixed capital, and fixed input prices, Appendix B.1 shows that

$$\left.\frac{d(m-\ell)}{d\omega^B}\right|_{\bar{q},w,k} = \sigma_{LM,t}\left[\rho_{t,\omega^B} - f_{t,\omega^B}\frac{\rho_{t,\ell}+\rho_{t,m}}{f_{t,\ell}+f_{t,m}}\right], \tag{21}$$

where subscripts denote partial derivatives and $\sigma_{LM,t} \equiv -\partial(m-\ell)/\partial\log(w^M/w^L) > 0$ is the local Hicksian elasticity of substitution between labor and materials. Local homotheticity of the flexible-input tradeoff sets $\rho_{t,\ell} + \rho_{t,m} = 0$, so the fixed-input bias and the Hicksian input-ratio response have the same sign. Beyond homotheticity, the direct-bias term must be large enough to offset the scale-recomposition term.

For the nested-CES technology above, Appendix B.1 shows that the response in (21) has the same sign as $\rho - \nu$. Consequently, when $\sigma_{KM} < \sigma_{XL}$, a purely capital-augmenting shock locally raises the cost-minimizing materials–labor ratio under the maintained fixed-output, fixed-capital, fixed-price comparison. This distinction helps interpret patterns such as those emphasized by Doraszelski and Jaumandreu (2018): a rising materials–labor ratio is consistent with labor augmentation, but is not by itself diagnostic of that primitive source of non-neutrality.

The material-biased direction is also empirically plausible in manufacturing data. In Chinese manufacturing, Pan (2026) documents that both the capital-to-labor ratio and the materials-to-labor ratio rise markedly over 1998–2007. He also finds that, holding inputs fixed, productivity growth raises the marginal product of materials more than the marginal product of labor, and that industries with stronger materials bias exhibit larger increases in the materials–labor ratio. For Spanish manufacturing, Doraszelski and Jaumandreu (2018) document a shift from labor to materials and emphasize that changes in the input mix contain information about biased technical change once relative prices, outsourcing, and labor-market composition are accounted for. These patterns are not by themselves a test of Assumption 2.1, since relative input prices and other firm choices also affect input ratios, but they make the material-biased direction a natural one to allow in manufacturing data.

In the Hicks–Acemoglu sense of biased technical change (Hicks, 1932; Acemoglu, 2002), Assumption 2.1′ gives a period-specific global interpretation: if $s_t = 1$, higher ω^B is material-biased relative to labor throughout the relevant period-t support; if $s_t = -1$, higher ω^B is

labor-biased relative to materials throughout that support. The weaker Assumption 2.1 is enough for conditional inversion but should be read pointwise, since its monotone direction can in principle vary with the input vector. The cost of stating the operative monotonicity directly is dimensionality: absent restrictions that collapse the flexible-input tradeoff, the inversion generally conditions on the full input vector (k, ℓ, m).

Lemma 2.1 (ω^B inversion)**.** *Suppose f_t is differentiable in (ℓ, m), $L_{jt}, M_{jt}, w^L_{jt}, w^M_{jt} > 0$, and labor and materials are interior flexible inputs satisfying the cost-minimization first-order conditions in (4)–(5), with $f_\ell(k_{jt}, \ell_{jt}, m_{jt}, \omega^B_{jt}) > 0$ and $f_m(k_{jt}, \ell_{jt}, m_{jt}, \omega^B_{jt}) > 0$ on the relevant support. Under Assumption 2.1, for each fixed (k, ℓ, m) the map*

$$\omega^B \mapsto \rho_t(k, \ell, m, \omega^B)$$

has an inverse on its image. Hence there is a function ψ_t on the support of $(k_{jt}, \ell_{jt}, m_{jt}, \chi^{ML}_{jt})$ such that

$$\omega^B_{jt} = \psi_t(k_{jt}, \ell_{jt}, m_{jt}, \chi^{ML}_{jt}). \tag{22}$$

Proof. The first-order conditions imply (8); taking logs of that equality gives (10). It remains to verify that Assumption 2.1 implies that $\rho_t(k, \ell, m, \omega^B)$ is strictly monotone in ω^B, with either monotone direction allowed, conditional on (k, ℓ, m).

Let $Q = \exp(q)$ denote output in levels. Since $L = \exp(\ell)$ and $M = \exp(m)$, the level marginal products implied by (1) are

$$\text{MPL} = \frac{\partial Q}{\partial L} = Q\frac{f_\ell(k, \ell, m, \omega^B)}{L}, \qquad \text{MPM} = \frac{\partial Q}{\partial M} = Q\frac{f_m(k, \ell, m, \omega^B)}{M}. \tag{23}$$

The common factor Q contains the Hicks-neutral component and cancels from this ratio. Thus

$$\log\left(\frac{\text{MPM}}{\text{MPL}}\right) = \log\left(\frac{f_m(k, \ell, m, \omega^B)}{f_\ell(k, \ell, m, \omega^B)}\right) + \ell - m = \rho_t(k, \ell, m, \omega^B) + \ell - m. \tag{24}$$

Holding (k, ℓ, m) fixed also holds $\ell - m$ fixed. Therefore, if the materials-to-labor marginal-product ratio is strictly increasing in ω^B, then $\rho_t(k, \ell, m, \omega^B)$ is strictly increasing in ω^B; if the ratio is strictly decreasing, then $\rho_t(k, \ell, m, \omega^B)$ is strictly decreasing. Thus the map $\omega^B \mapsto \rho_t(k, \ell, m, \omega^B)$ is one-to-one on the conditional support of ω^B. The inverse map exists on the image of ρ_t, and applying it to the observed value χ^{ML}_{jt} gives (22). □

This is the same relative-marginal-product object used in Lemma 2.1. In the homothetic-separable labor-augmenting model of Demirer (2020), written in his notation as ω^L, cost minimization reduces the relevant flexible-input tradeoff to $M/L = r_t(K, \omega^L)$, strictly monotone in ω^L; with heterogeneous input prices, the corresponding mapping is conditional on the materials–labor price ratio. Here, because f_t is not restricted to deliver that lower-dimensional reduction and input prices may vary across firms, the inverse is allowed to depend on the full input vector (k, ℓ, m) and on the observed relative expenditure ratio χ^{ML}.

Assumption 2.2 (Productivity dynamics)**.** Let $o^B_{j,t-1}, o^H_{j,t-1} \in \mathcal{I}_{j,t-1}$ be observed lagged controls for the biased and Hicks-neutral productivity, respectively. For each $s \in \{B, H\}$, the productivity component ω^s follows a separate controlled first-order, time-inhomogeneous Markov process. For every x,

$$\Pr\left(\omega^s_{jt} \leq x \mid \mathcal{I}_{j,t-1}\right) = P^s_t\left(x \mid \omega^s_{j,t-1}, o^s_{j,t-1}\right), \qquad s \in \{B, H\}. \tag{25}$$

For the biased component, use the nonseparable rank representation

$$\omega^B_{jt} = H^B_t(\omega^B_{j,t-1}, o^B_{j,t-1}, u^B_{jt}), \tag{26}$$

where $H^B_t(\omega, o^B, u)$ is continuous and strictly increasing in u, and $u^B_{jt} \mid \mathcal{I}_{j,t-1} \sim U(0,1)$. For the Hicks-neutral component, assume the conditional mean exists and use the additive representation

$$\omega^H_{jt} = g^H_t(\omega^H_{j,t-1}, o^H_{j,t-1}) + \xi^H_{jt}, \qquad E[\xi^H_{jt} \mid \mathcal{I}_{j,t-1}] = 0. \tag{27}$$

Assumption 2.2 follows the separate productivity laws of motion in Doraszelski and Jaumandreu (2018), while allowing each transition to condition on its own observed lagged control, $o^B_{j,t-1}$ or $o^H_{j,t-1}$. The distinction between the two representations is deliberate. The nonseparable rank representation for ω^B supports the conditional-quantile argument in the first identification step. Additive separability is imposed only on the Hicks-neutral transition and delivers the additive NPIV equation used in the second step. The time subscripts on P^s_t, g^H_t, and H^B_t allow those objects to vary across calendar time. The first-stage argument below retains this period-specific formulation: for each transition period, it identifies ρ_t and the observable-control map induced by H^B_t from that period's conditional-quantile restriction. It does not require first-stage structural or reduced-form maps to coincide across calendar time. Time-invariant maps enter only in the estimation section, where they are explicit

implementation restrictions. This formulation differs from the joint Markov law in Demirer (2020) by making the two controlled Markov processes and their distinct identifying roles explicit. The separate Markov laws do not by themselves restrict the joint distribution of the two current innovations. For the baseline identification result, Assumption 3.5 adds conditional independence of those innovations given the common second-step conditioning set. Appendix F considers the weaker alternative that leaves their conditional-mean dependence unrestricted.

Remark 2.1 (Hicks-neutral productivity and the proxy-first-stage extension)**.** The environment is otherwise close to that of Demirer (2020): capital is predetermined, labor and materials are static flexible inputs, and productivity is two-dimensional. The key difference is that the baseline does not impose the materials-demand monotonicity condition in Demirer (2020, Assumption 2.5), under which materials demand is strictly monotone in Hicks-neutral productivity. Accordingly, the baseline does not use materials demand as a proxy, as in Olley and Pakes (1996) or Levinsohn and Petrin (2003), to recover ω^H. Instead, the relative flexible-input condition identifies the non-Hicksian component ω^B. Conditional on f_t and ω^B, the production equation then recovers ω^H, whose dynamics are governed by Assumption 2.2.

This choice comes at a cost: the baseline excludes a separate ex-post output shock. It can therefore be interpreted as a post-inversion model, in which the conventional OP/LP/ACF-style proxy-variable first stage nonparametrically regresses output on inputs to purge such a shock before our estimator is applied. As a robustness check in the empirical application, we implement this first stage before applying our estimator to the resulting data. The main empirical findings are robust to this extension.

3 Identification

This section gives the identification argument. The model contains two scalar productivity components, but they enter two different equations. The relative flexible-input expenditure equation identifies the non-Hicksian component ω^B and the ratio of the labor and materials elasticities. The identified ratio field then supplies a scalar index that collapses the two flexible inputs to one endogenous direction in the production equation. Under baseline conditional independence of the two productivity innovations, the recovered ω^B can serve as its own instrument; the reduced production equation identifies the production function and the Hicks-neutral component ω^H. Because labor and materials are chosen after current productivity is observed, both steps use instruments, but the two equations have different

structures. The first step uses the nonseparable IV quantile argument in Ackerberg et al. (2026, Section 5.2 and Appendix C), which builds on Chernozhukov and Hansen (2005). In the second step, the Hicks-neutral Markov shock enters additively, so the relevant framework is the additive nonparametric IV model of Newey and Powell (2003).

We state the argument for the first-order Markov case in Assumption 2.2; with higher-order Markov processes, the lagged control vectors are enlarged accordingly. As in Ackerberg et al. (2026), identification is for periods $t \geq 2$, with the initial period used to form lagged controls. All latent productivity objects below are understood subject to the normalizations and support qualifications stated below.

Assumption 3.1 (Rank, biased-productivity-scale, and location normalizations)**.** Maintain the conditional rank normalization for u_{jt}^B in (26). After orienting each period's non-neutral productivity component so that ρ_t is increasing in its ω^B argument, fix a reference input vector x° in the domain on which the period-specific structural functions and the estimation sieve are defined, and recenter inputs so that $x^\circ = 0$. We report biased productivity in the reference scale

$$\rho_t(0, \omega) = \omega, \qquad \omega \in \Omega_t^B, \quad \text{for every observed } t, \tag{28}$$

where Ω_t^B is the range of normalized ω^B in period t. This equality is a scale convention, not a distributional restriction on ω^B. If the reference slice x° is supported, or lies in the continuity closure of the relevant support over the ω^B range of interest, the reparameterization relation used below is identified at x°. Otherwise, we maintain the corresponding continuation restriction: whenever two admissible structural maps are related by a reparameterization of ω^B on the covered support, the same relation holds at x°. Thus, an off-support reference point fixes the ω^B reference scale within the maintained function class; it does not supply additional identifying variation. For the Hicks-neutral component, impose the location normalization $E[\omega_{jt}^H] = 0$.

3.1 The Biased Component

For each period, the Ackerberg–Hahn–Pan quantile argument uses an increasing orientation of the scalar unobservable. Therefore, in this section we use the strengthened Assumption 2.1′. If $s_t = -1$ in period t, reverse that period's innovation rank by $u^B \mapsto 1 - u^B$ and apply an order-reversing relabeling of the non-neutral productivity component before imposing (28). We henceforth use these oriented coordinates, so for every t, both ρ_t in its ω^B argument and the induced map h_t in its rank argument are increasing.

Let

$$V_{j,t-1}^{B} \equiv (k_{j,t-1}, \ell_{j,t-1}, m_{j,t-1}, \chi_{j,t-1}^{ML}, o_{j,t-1}^{B}) \tag{29}$$

denote the lagged observables that control for lagged biased productivity in the relative-expenditure equation. Let Z_{jt}^{B} denote excluded shifters for current flexible-input choices. In applications, Z_{jt}^{B} may include predetermined variables such as lagged wage levels and lagged output. Lagged output is useful in this first step because it is not needed to control for $\omega_{j,t-1}^{B}$.

Write $X_{jt}^{B} \equiv (k_{jt}, \ell_{jt}, m_{jt})$. For every observed period s, including a period used only to construct lagged controls, let $\psi_s(x, \chi)$ denote the inverse of $\omega \mapsto \rho_s(x, \omega)$ on its image under the increasing orientation. The structural model then implies that, for each transition period $t \geq 2$ and $v = (x_-, \chi_-, o^B)$,

$$h_t(v, u) \equiv H_t^B\big(\psi_{t-1}(x_-, \chi_-), o^B, u\big)\,. \tag{30}$$

It follows that

$$\bar{\rho}_t(x, v, u) \equiv \rho_t(x, h_t(v, u))\,. \tag{31}$$

These are model-implied representations, rather than additional restrictions. They are useful for stating the support condition below.

For each transition period t, let $\mathscr{S}_t^{\bar{\rho}}$ denote the observed support on which the IV-quantile restriction in Assumption 3.3 identifies $\bar{\rho}_t$, and let

$$\mathscr{D}_t \equiv \mathrm{proj}_{(x,\chi)}\, \mathscr{S}_t^{\bar{\rho}}.$$

For $(x, \chi) \in \mathscr{D}_t$, define

$$\mathscr{V}_t(x, \chi) \equiv \left\{v : (x, v, \chi) \in \mathscr{S}_t^{\bar{\rho}}\right\}, \qquad v_t(x, v, \chi) \equiv [\bar{\rho}_t(x, v, \cdot)]^{-1}(\chi),$$

where the inverse is taken in the rank argument. Also define

$$\mathscr{X}_t(v, u) \equiv \left\{x : (x, v, \bar{\rho}_t(x, v, u)) \in \mathscr{S}_t^{\bar{\rho}}\right\}.$$

Thus, $\mathscr{X}_t(v, u)$ is the implied conditional support at rank u, constructed from observed

support and the identified reduced form. For $A \subseteq \mathscr{D}_t$, let

$$\mathscr{W}_t(A) \equiv \big\{(x', \bar{\rho}_t(x', v, u)) : (x, \chi) \in A,\ v \in \mathscr{V}_t(x, \chi),\\ u = \upsilon_t(x, v, \chi),\ x' \in \mathscr{X}_t(v, u)\big\},$$

and let $\mathscr{W}_t^r$ denote its r-fold iteration, with $\mathscr{W}_t^0(A) = A$.

Assumption 3.2 (First-stage support condition)**.** For each transition period $t \geq 2$, let $\mathscr{C}_t \subseteq \mathscr{D}_t$ be a target support set on which structural identification is sought. For every two points $a, b \in \mathscr{C}_t$, there are finite nonnegative integers r_a and r_b such that

$$\operatorname{proj}_x \mathscr{W}_t^{r_a}(\{a\}) \cap \operatorname{proj}_x \mathscr{W}_t^{r_b}(\{b\}) \neq \varnothing. \tag{32}$$

This is an Ackerberg–Hahn–Pan-style sufficient finite-expansion condition. Each expansion step uses only observed support and the identified period-t reduced-form map; it holds the lagged controls and rank fixed while varying current inputs. The condition is imposed separately by period and neither requires support overlap nor equates any structural or reduced-form maps across calendar periods. The maps above merely define the expansion; their forms are not restricted by this assumption.

Assumption 3.3 (Excluded variation for relative expenditure)**.** Capital is predetermined, while (ℓ_{jt}, m_{jt}) may be chosen with knowledge of current productivity and current input-market conditions. For each transition period $t \geq 2$, the researcher observes a vector Z_{jt}^B such that

$$(Z_{jt}^B, k_{jt}, V_{j,t-1}^B) \perp\!\!\!\perp u_{jt}^B. \tag{33}$$

For every $\tau \in (0, 1)$ and each transition period $t \geq 2$, the conditional law

$$(\ell_{jt}, m_{jt}) \mid k_{jt}, V_{j,t-1}^B, Z_{jt}^B$$

satisfies the required support, rank, and completeness conditions for the period-t conditional-quantile restriction

$$\Pr\big(\chi_{jt}^{ML} \leq \bar{\rho}_t(X_{jt}^B, V_{j,t-1}^B, \tau) \mid k_{jt}, V_{j,t-1}^B, Z_{jt}^B\big) = \tau \tag{34}$$

to have a unique solution $\bar{\rho}_t(\cdot, \cdot, \tau)$ in the maintained period-t function class.

Proposition 3.1 (Identification of biased-productivity maps)**.** *Suppose the strengthened Assumption 2.1′ and Assumptions 2.2, 3.1, 3.2, and 3.3 hold. Fix a transition period $t \geq 2$. Then $\bar{\rho}_t(x, v, \tau)$ is identified on the relevant period-t support. On each target support set $\mathscr{C}_t$ satisfying Assumption 3.2, (ρ_t, h_t) are identified on the covered support.*

Proof. Fix a transition period $t \geq 2$. By the model-implied identities immediately preceding Assumption 3.2, substituting the biased-productivity transition and the lagged inverse into the relative-expenditure equation gives

$$\begin{aligned} \chi_{jt}^{ML} &= \rho_t\big(X_{jt}^B, h_t\big(V_{j,t-1}^B, u_{jt}^B\big)\big) \\ &\equiv \bar{\rho}_t(X_{jt}^B, V_{j,t-1}^B, u_{jt}^B). \end{aligned} \tag{35}$$

Both ρ_t in its ω^B argument and h_t in its rank argument are strictly increasing. Hence $\bar{\rho}_t$ is strictly increasing in u_{jt}^B.

Current flexible inputs may depend on u_{jt}^B, so they need not be valid instruments. Assumption 3.3 supplies the excluded variation. Together with the conditional uniform-rank normalization, it yields (34). The period-t IV-quantile uniqueness condition therefore identifies $\bar{\rho}_t$.

Consider any two admissible period-t decompositions

$$\bar{\rho}_t(x, v, u) = \rho_t(x, h_t(v, u)) = \widetilde{\rho}_t\Big(x, \widetilde{h}_t(v, u)\Big)\,.$$

The finite-expansion condition in Assumption 3.2 permits the associated non-neutral productivity components to be ordered throughout $\mathscr{C}_t$: each expansion holds (v, u) fixed, and overlap at a shared current-input value converts the order of relative expenditure into the order of the non-neutral productivity component. The Ackerberg–Hahn–Pan argument therefore implies that the two decompositions can be related by a strictly increasing bijection ϕ_t on the relevant ω^B range, where

$$\widetilde{h}_t(v, u) = \phi_t(h_t(v, u))\,, \qquad \widetilde{\rho}_t(x, a) = \rho_t\big(x, \phi_t^{-1}(a)\big)\,. \tag{36}$$

Because both maps satisfy (28) and the reparameterization relation extends to x° by Assumption 3.1, for every ω^B value a in the relevant range,

$$\phi_t(a) = \widetilde{\rho}_t(x^\circ, \phi_t(a)) = \rho_t(x^\circ, a) = a.$$

Hence ϕ_t is the identity on that range, so $\widetilde{h}_t = h_t$ and $\widetilde{\rho}_t = \rho_t$ on the covered support. Strict monotonicity of $\rho_t(x, \cdot)$ then identifies its inverse $\psi_t(x, \cdot)$ on its image, and

$$\omega_{jt}^B = \psi_t(X_{jt}^B, \chi_{jt}^{ML})$$

recovers the non-neutral productivity component for every observation in the covered period-t support. □

3.2 Using the Ratio Restriction in the Second Step

Proposition 3.1 identifies not only the non-neutral productivity component ω^B, but also the period-t flexible-input elasticity-ratio field. The central role of this restriction in the Hicks-neutral stage is to reduce dimension. Conditional on $z = (k, \omega^B)$, it allows us to replace the two-dimensional current flexible-input block (ℓ, m) with a scalar index that labels the isoquants it traces. The second-stage nonparametric IV problem therefore needs to handle one endogenous flexible-input direction rather than two. This reduction in dimension can improve precision and computational efficiency and lower the number of instrument variables required. Write, for each period t,

$$z \equiv (k, \omega^B), \qquad R_t(z, \ell, m) \equiv \exp\{\rho_t(z, \ell, m)\} = \frac{f_{t,m}(z, \ell, m)}{f_{t,\ell}(z, \ell, m)}. \tag{37}$$

The ratio field and its support may vary across periods. Holding z fixed, an isoquant in the (ℓ, m) plane has slope

$$\left.\frac{d\ell}{dm}\right|_{f_t \text{ constant}} = -\frac{f_{t,m}}{f_{t,\ell}} = -R_t. \tag{38}$$

Thus the identified ratio field determines the local geometry of conditional isoquants, although it does not determine the output level attached to each isoquant. To turn that geometry into a scalar index, we trace the characteristic through a point (z, ℓ, m), defined by $d\ell/dm = -R_t$, back to a reference section $m = m_{0t}(z)$ and label the characteristic by its labor coordinate a at that intersection. Local ODE regularity alone is not enough: the path must exist over the required range, stay in the relevant domain, and meet the reference section once and only once. The following assumption separates the conditions that provide a safe reference section, local ODE well-posedness, and a globally unique characteristic label.

Assumption 3.4 (Ratio-index characteristic-flow domain)**.** For each period t, maintain the

following conditions.

(i) Interior domain and reference section. Fix a compact continuation domain $\mathscr{D}_t^c$ and a compact target characteristic strip $\mathscr{U}_t^c \subset \mathrm{int}(\mathscr{D}_t^c)$ containing the relevant observed support. Fix a compact relevant set

$$\mathrm{proj}_z(\mathscr{U}_t^c) \subseteq \mathscr{X}_t^c \subset \mathrm{int}(\mathrm{proj}_z(\mathscr{D}_t^c)) .$$

Fix a continuously differentiable reference rule $m_{0t}(z)$ and a compact reference-label set

$$\Gamma_{0t} = \{(z, a, m_{0t}(z)) : z \in \mathscr{X}_t^c,\ a \in \mathscr{L}_{0t}(z)\}$$

where $\mathscr{L}_{0t}(z)$ is the set of admissible labor labels at the reference section. Assume that Γ_{0t} has an open neighborhood whose closure lies in $\mathrm{int}(\mathscr{D}_t^c)$. Thus, the target strip, the relevant values of z, and the reference-section landing points are kept away from the boundary. These conditions provide a safe interior anchor from which to trace the characteristics.

(ii) Local field and joint continuation. Fix a field R_t^c on an open neighborhood of $\mathscr{D}_t^c$. This field is finite, strictly positive, and continuously differentiable jointly in (z, ℓ, m), and it agrees with the identified period-t ratio field wherever the latter is supported. The smoothness of R_t^c gives local existence, uniqueness, and smooth dependence on initial conditions for the characteristic ODE below: in particular, it makes the ODE's right-hand side locally Lipschitz in ℓ. In addition, fix a particular continuously differentiable structural continuation f_t^c that agrees with f_t on its supported domain, has $f_{t,\ell}^c > 0$, and satisfies

$$\frac{f_{t,m}^c(z, \ell, m)}{f_{t,\ell}^c(z, \ell, m)} = R_t^c(z, \ell, m) \qquad \text{on } \mathscr{D}_t^c.$$

Once the supported functions are given, the off-support values of (R_t^c, f_t^c) are fixed by the continuation specification; they are not free nuisance functions to be identified from off-support data.

(iii) Global reachability and unique labeling. For every $(z, \ell, m) \in \mathscr{U}_t^c$, there is a unique $a \in \mathscr{L}_{0t}(z)$ for which the differential equation

$$\frac{\partial \Lambda_t(s; z, a)}{\partial s} = -R_t^c(z, \Lambda_t(s; z, a), s) , \qquad \Lambda_t(m_{0t}(z); z, a) = a, \tag{39}$$

has a unique solution on the material-coordinate interval between $m_{0t}(z)$ and m and satisfies $\Lambda_t(m; z, a) = \ell$. Each required characteristic segment has uniformly bounded length, remains

in the interior of $\mathcal{D}_t^c$ with a uniform positive buffer, and intersects the reference section $m = m_{0t}(z)$ exactly once. The resulting flow is continuously differentiable in (z, a) and, on the relevant strip, satisfies

$$\partial_a \Lambda_t(s; z, a) = \exp\left\{-\int_{m_{0t}(z)}^{s} \partial_\ell R_t^c(z, \Lambda_t(u; z, a), u)\, du\right\} > 0.$$

Together, the existence, bounded-length, and interior-buffer requirements ensure that every required path can be followed to the reference section without exploding or exiting the continuation domain. The unique-intersection clause makes the label single-valued, while strict positivity of $\partial_a \Lambda_t$ makes the inversion smooth.

On a characteristic component in $\mathcal{U}_t^c$ whose complete path, including its reference-section landing point, is empirically supported, write $R_t^c = R_t$ and $f_t^c = f_t$. If a point in $\mathcal{U}_t^c$ or its required path traverses an unsupported portion of $\mathcal{D}_t^c$, the fixed values of (R_t^c, f_t^c) constitute a specified joint continuation rather than nonparametrically identified counterfactuals. Continuing R_t alone would determine off-support isoquant slopes but would not attach mutually consistent output levels to isoquants joined through unsupported regions. Thus level statements for the composed production function and Hicks-neutral productivity component on those portions are conditional on that particular continuation f_t^c (or, equivalently, on the particular G_t^c in the representation below). The baseline estimator treats the continuation domain and joint continuation as maintained and does not select observations using an estimated characteristic-path support set.

The global-flow condition turns the locally well-posed ODE into a globally single-valued index on $\mathcal{U}_t^c$. It guarantees that the characteristic through every target point can be traced to the reference section without leaving the continuation domain and has one, rather than several, landing labels. For a point $(z, \ell, m) \in \mathcal{U}_t^c$, let a be the labor coordinate at which its path reaches $m = m_{0t}(z)$ and define

$$h_{R,t}(z, \ell, m) = a \quad \Longleftrightarrow \quad \Lambda_t(m; z, a) = \ell. \tag{40}$$

The normalization is therefore

$$h_{R,t}(z, \ell, m_{0t}(z)) = \ell. \tag{41}$$

In words, all points on the same conditional isoquant receive the labor coordinate at which

that isoquant crosses the reference section.

Proposition E.1, proved in Appendix E, establishes that this construction is identified on supported characteristic components of $\mathcal{U}_t^c$ and, elsewhere on that strip, defined conditional on the maintained continuation. It satisfies

$$\frac{\partial_m h_{R,t}(z,\ell,m)}{\partial_\ell h_{R,t}(z,\ell,m)} = R_t^c(z,\ell,m). \tag{42}$$

It also gives the exact representation

$$f_t^c(k,\ell,m,\omega^B) = G_t^c\big(k,\omega^B,h_{R,t}(k,\omega^B,\ell,m)\big) \tag{43}$$

for an otherwise unrestricted continuously differentiable function G_t^c with $G_{t,h}^c > 0$. On the supported domain we omit the superscript c. Conversely, any function with the representation in (43) satisfies the continued ratio restriction, and hence (37) on the supported domain. Here a supported characteristic component is one whose complete path to the reference section, including its landing point, lies in the support on which R_t is identified. Conditional on the specified continuation, the representation imposes no additional restriction beyond the continued ratio relation and the normalization used to label its isoquants.

Define the normalized scalar index and its associated current argument vector by

$$H_{jt}^R \equiv h_{R,t}(k_{jt},\omega_{jt}^B,\ell_{jt},m_{jt}), \qquad X_{jt}^R \equiv (k_{jt},\omega_{jt}^B,H_{jt}^R). \tag{44}$$

The production equation can then be written as

$$q_{jt} = G_t(X_{jt}^R) + \omega_{jt}^H. \tag{45}$$

The ratio restriction therefore replaces the two-dimensional flexible-input block (ℓ,m) by the single endogenous index H^R. It does not by itself identify G_t: the production equation must still determine the output level attached to each isoquant, and valid excluded variation is still required for H^R. Its identifying contribution is dimensional. Conditional on (k,ω^B), the second step needs completeness for one flexible-input direction rather than unrestricted two-dimensional variation in (ℓ,m). This reduction can make the second-step nonparametric IV problem better conditioned and can improve precision by reducing sieve and moment dimensionality, although it is not an automatic efficiency gain because H^R is itself estimated in the first step.

3.3 The Hicks-Neutral Component

After Proposition 3.1, use the identified reference scale and define

$$X_{jt}^{H} \equiv (k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B}). \tag{46}$$

The current and lagged production equations are

$$q_{jt} = f_t(X_{jt}^{H}) + \omega_{jt}^{H}, \qquad q_{j,t-1} = f_{t-1}(X_{j,t-1}^{H}) + \omega_{j,t-1}^{H}. \tag{47}$$

The lagged equation therefore gives

$$\omega_{j,t-1}^{H} = q_{j,t-1} - f_{t-1}(X_{j,t-1}^{H}). \tag{48}$$

Collect the observables entering this expression and the Hicks-neutral transition in

$$V_{j,t-1}^{H} \equiv (q_{j,t-1}, k_{j,t-1}, \ell_{j,t-1}, m_{j,t-1}, \chi_{j,t-1}^{ML}, o_{j,t-1}^{B}, o_{j,t-1}^{H}). \tag{49}$$

This vector is observable even in the initial period used to form lags. It includes $o_{j,t-1}^{H}$ for the Hicks-neutral transition and $o_{j,t-1}^{B}$ to hold fixed the biased transition when conditioning on current biased productivity. By (22), the lagged non-neutral productivity component is a function of the components $(k_{j,t-1}, \ell_{j,t-1}, m_{j,t-1}, \chi_{j,t-1}^{ML})$, so the unknown composition involving $\omega_{j,t-1}^{B}$ can be absorbed into a function of $V_{j,t-1}^{H}$. Substituting (27) into the current production equation and then using (48) yields

$$\begin{aligned} q_{jt} &= f_t(X_{jt}^{H}) + g_t^{H}(\omega_{j,t-1}^{H}, o_{j,t-1}^{H}) + \xi_{jt}^{H} \\ &= f_t(X_{jt}^{H}) + g_t^{H}\big(q_{j,t-1} - f_{t-1}(X_{j,t-1}^{H}), o_{j,t-1}^{H}\big) + \xi_{jt}^{H}. \end{aligned} \tag{50}$$

Define the lag-control component

$$a_t^{H}(V_{j,t-1}^{H}) \equiv g_t^{H}\big(q_{j,t-1} - f_{t-1}\big(k_{j,t-1}, \ell_{j,t-1}, m_{j,t-1}, \psi_{t-1}\big(k_{j,t-1}, \ell_{j,t-1}, m_{j,t-1}, \chi_{j,t-1}^{ML}\big)\big), o_{j,t-1}^{H}\big) \tag{51}$$

For $t = 2$, neither ψ_1 nor the inner composition in (51) needs to be separately recovered: a_2^{H} is an unrestricted nuisance function of the observable vector V_{j1}^{H}. This avoids requiring identification of the initial-period non-neutral productivity component before the second step

begins. Define the composite function

$$\bar{f}_t(X_{jt}^H, V_{j,t-1}^H) \equiv f_t(X_{jt}^H) + a_t^H(V_{j,t-1}^H). \tag{52}$$

Equation (50) can thus be written in the additive form

$$q_{jt} = f_t(X_{jt}^H) + a_t^H(V_{j,t-1}^H) + \xi_{jt}^H = \bar{f}_t(X_{jt}^H, V_{j,t-1}^H) + \xi_{jt}^H. \tag{53}$$

Using the identified representation in (43), the same equation has the lower-dimensional form

$$q_{jt} = G_t(X_{jt}^R) + a_t^H(V_{j,t-1}^H) + \xi_{jt}^H. \tag{54}$$

This is the baseline second-step equation. The unrestricted form in (53) remains useful for comparison, but it requires instruments to identify separate movements in both flexible inputs. Equation (54) uses the already identified ratio restriction and requires excluded variation only for the scalar H_{jt}^R.

$$Z_{jt}^H \equiv (k_{jt}, \omega_{jt}^B, Z_{jt}^{LM}), \qquad C_{jt} \equiv (k_{jt}, Z_{jt}^{LM}, V_{j,t-1}^H), \tag{55}$$

where Z_{jt}^{LM} denotes excluded shifters for the current flexible-input index H_{jt}^R.

Assumption 3.5 (Baseline cross-innovation independence)**.** The innovation ξ_{jt}^H is integrable and

$$E[\xi_{jt}^H \mid C_{jt}] = 0. \tag{56}$$

Conditional on the same information set C_{jt}, the current innovations in the two productivity processes are independent:

$$u_{jt}^B \perp\!\!\!\perp \xi_{jt}^H \mid C_{jt}. \tag{57}$$

The inherited non-neutral productivity component $\omega_{j,t-1}^B$ and its control $o_{j,t-1}^B$ are measurable with respect to C_{jt}.

The last condition is a restriction on the two current innovations, not on the two productivity levels. The levels may remain correlated through their histories and through observed

controls.

Lemma 3.1 (Validity of the recovered non-neutral productivity component)**.** *Under Assumptions 2.2 and 3.5,*

$$E[\xi_{jt}^H \mid k_{jt}, \omega_{jt}^B, Z_{jt}^{LM}, V_{j,t-1}^H] = 0. \tag{58}$$

Equivalently, using the vector in (55),

$$E[\xi_{jt}^H \mid Z_{jt}^H, V_{j,t-1}^H] = 0. \tag{59}$$

Hence the identified non-neutral productivity component ω_{jt}^B can serve as its own instrument in the baseline second step.

Proof. Conditional on C_{jt}, the transition

$$\omega_{jt}^B = H_t^B(\omega_{j,t-1}^B, o_{j,t-1}^B, u_{jt}^B)$$

is strictly monotone in u_{jt}^B and therefore has a measurable inverse on its conditional image. It follows that

$$\sigma(\omega_{jt}^B, C_{jt}) = \sigma(u_{jt}^B, C_{jt}).$$

Conditional independence and (56) then give

$$E[\xi_{jt}^H \mid \omega_{jt}^B, C_{jt}] = E[\xi_{jt}^H \mid u_{jt}^B, C_{jt}] = E[\xi_{jt}^H \mid C_{jt}] = 0,$$

which is (58). □

Assumption 3.6 (Reduced additive NPIV conditions)**.** Let $\mathscr{D}_t^R$ be the admissible class of square-integrable additive differences

$$\mathscr{D}_t^R = \left\{\delta_t^R(x, v) = d_{G,t}(x) + d_{a,t}(v)\right\}, \qquad x = (k, \omega^B, h). \tag{60}$$

The conditional-expectation operator is injective on $\mathscr{D}_t^R$: for every $\delta_t^R \in \mathscr{D}_t^R$,

$$\begin{aligned} &E\left[\delta_t^R(X_{jt}^R, V_{j,t-1}^H) \mid Z_{jt}^H, V_{j,t-1}^H\right] = 0 \\ &\implies \qquad \delta_t^R(X_{jt}^R, V_{j,t-1}^H) = 0 \quad \text{a.s.} \end{aligned} \tag{61}$$

Finally, the support $\mathscr{S}_t^{R,H}$ of $(X_{jt}^R, V_{j,t-1}^H)$ is additively connected: for any x, x' in the marginal support of X_{jt}^R, there exist $x = x_0, \ldots, x_J = x'$ and $v_1, \ldots, v_J$ such that $(x_{r-1}, v_r), (x_r, v_r) \in \mathscr{S}_t^{R,H}$ for every r. A Cartesian product support is sufficient.

Proposition 3.2 (Identification of f_t and ω^H)**.** *Suppose Proposition 3.1 holds. If Assumptions 3.4, 2.2, 3.1, 3.5, and 3.6 hold, then, for $t \geq 2$, the additive composite*

$$G_t(X_{jt}^R) + a_t^H(V_{j,t-1}^H)$$

is identified on the relevant support. Its components are identified under the location normalization in Assumption 3.1. Consequently, f_t is identified through (43) at points whose complete characteristics and reference intersections are supported. At points whose construction traverses an unsupported region, the data alone do not identify the extension; f_t^c and the corresponding residual are defined only after imposing the particular joint continuation (R_t^c, f_t^c) fixed in Assumption 3.4. On supported components,

$$\omega_{jt}^H = q_{jt} - f_t(X_{jt}^H) \tag{62}$$

is identified for every firm-period on the relevant support; elsewhere the same formula is a continuation-dependent reporting object.

Proof. Equations (54) and (59) imply

$$E\left[q_{jt} - G_t(X_{jt}^R) - a_t^H(V_{j,t-1}^H) \mid Z_{jt}^H, V_{j,t-1}^H\right] = 0. \tag{63}$$

Injectivity in (61) identifies this additive composite by the nonparametric IV argument of Newey and Powell (2003).

Suppose (G_t, a_t^H) and $(\widetilde{G}_t, \widetilde{a}_t^H)$ generate the same identified composite. Their differences satisfy

$$\Delta G_t(x) + \Delta a_t^H(v) = 0 \qquad \text{on } \mathscr{S}_t^{R,H}.$$

At a common v, this equality gives $\Delta G_t(x) = \Delta G_t(x')$. Additive connectedness extends the equality across the support, so $\Delta G_t(x) = c_t$ and $\Delta a_t^H(v) = -c_t$.

By (27) and (51),

$$\omega_{jt}^H = a_t^H(V_{j,t-1}^H) + \xi_{jt}^H.$$

The conditional-mean restriction on ξ_{jt}^H and $E[\omega_{jt}^H] = 0$ imply

$$E[a_t^H(V_{j,t-1}^H)] = 0.$$

This condition sets $c_t = 0$. Proposition E.1 then maps the identified G_t and $h_{R,t}$ into f_t; when the complete characteristic is not supported, this mapping is conditional on the specified joint continuation. The production residual then gives (62) subject to the same qualification. □

At interior points of supported characteristic components, the identified representation also gives the three input elasticities:

$$f_{t,\ell} = G_{t,h}\,\partial_\ell h_{R,t}, \qquad f_{t,m} = G_{t,h}\,\partial_m h_{R,t}, \tag{64}$$

$$f_{t,k} = G_{t,k} + G_{t,h}\,\partial_k h_{R,t}. \tag{65}$$

The second term in (65) is required because the ratio index generally depends on capital. With twice continuous differentiability and common interior support as ω^B varies, the derivatives of these elasticities with respect to ω^B, and hence pairwise Hicks–Acemoglu bias measures, are identified as well.

Remark 3.1 (Instrument sets)**.** Variables already included in $V_{j,t-1}^H$, including $q_{j,t-1}$, $o_{j,t-1}^B$, and $o_{j,t-1}^H$, are controls rather than excluded instruments. Lagged output can nevertheless enter Z_{jt}^B in the first step because it is not included in $V_{j,t-1}^B$.

Lemma 3.1 explains why no separate excluded instrument is needed for the recovered ω_{jt}^B in the baseline. Mere zero correlation between u_{jt}^B and ξ_{jt}^H would not suffice when H_t^B is nonlinear. The excluded variables Z_{jt}^{LM} must instead provide relevant variation for the single endogenous flexible-input index H_{jt}^R, conditional on $(k_{jt}, \omega_{jt}^B, V_{j,t-1}^H)$. The ratio restriction reduces the dimension of this relevance problem; it does not establish exclusion or completeness.

When labor and material prices or external supply-side price shifters are observed, they are natural candidates for Z_{jt}^{LM}. Appendix C distinguishes input-market price taking from instrument exogeneity and compares a direct price-IV route with the lag-controlled hybrid used here. Appendix D develops a more restrictive share-based second step under perfect competition in the output market. Appendix F shows what the period-specific second-step conditional-moment system identifies when the baseline cross-innovation independence restriction is relaxed, and explains how additional cross-period restrictions could further discipline the remaining location ambiguity.

Candidate excluded variables include predetermined liquidity, trade-credit, inventory, tax, debt, subsidy, and input-cost shifters. Their relevance can be examined empirically, but the conditional exclusion restriction and the Newey–Powell completeness condition remain identifying assumptions.

4 Estimation

This section translates the identification argument into a sieve estimator. There are two structural stages. The first recovers biased productivity from the relative-expenditure equation. The second converts the identified materials–labor elasticity ratio into a scalar characteristic index and estimates the production surface and Hicks-neutral productivity. We then project the estimated surface onto a factor-augmenting CES specification for interpretation. The discussion here is deliberately nonparametric; the finite-dimensional specifications, support rules, regularization constants, and numerical implementation used for the empirical results are reported in Appendix A.

The identification argument permits the first-stage maps $(\rho_t, H_t^B, h_t, \bar{\rho}_t)$, the reduced production surface G_t, and the Hicks-neutral transition component a_t^H to vary by period. For notational economy, the generic sieve formulas below suppress those time subscripts. The reported estimator combines the retained transitions and, for convenience and precision, imposes time-invariant sieve representations for the first-stage ratio and transition maps, the reduced production surface, and the Hicks-neutral transition map. These are explicit cross-period implementation restrictions rather than identification assumptions; Appendix A states their finite-dimensional form.

4.1 Stage One: Biased Productivity

Index the retained transition observations by $i = (j,t)$, write $x_i = (k_i, \ell_i, m_i)$, and let χ_i^{ML} denote the log materials–labor relative-expenditure ratio. Approximate the time-invariant structural ratio map and the monotone controlled transition imposed in the empirical implementation by sieve functions

$$\rho_{K_\rho}(x, b; \theta_\rho) \quad \text{and} \quad h^B_{K_h}(b_-, o_-, u; \theta_h),$$

where b is biased productivity on the reference scale and $h^B_{K_h}$ is strictly increasing in its innovation coordinate u. The sieve spaces impose

$$\rho_{K_\rho}(0, b; \theta_\rho) = b, \qquad \partial_b \rho_{K_\rho}(x, b; \theta_\rho) \geq \underline{c}_{\rho,N} > 0 \tag{66}$$

on a prespecified domain. The first equality implements the biased-productivity reference normalization in Assumption 3.1; the second makes the inversion well defined and fixes its orientation.

For every trial value θ_ρ, recover current and lagged non-neutral productivity components inside the objective:

$$b_i(\theta_\rho) = \rho_{K_\rho}^{-1}(x_i, \chi_i^{ML}; \theta_\rho), \qquad b_{i-}(\theta_\rho) = \rho_{K_\rho}^{-1}(x_{i-}, \chi_{i-}^{ML}; \theta_\rho). \tag{67}$$

The implied biased-productivity innovation coordinate is obtained by inverting the transition in its last argument:

$$u_i^B(\theta) = \left[h_{K_h}^B(b_{i-}(\theta_\rho), o_{i-}, \cdot\,; \theta_h)\right]^{-1}[b_i(\theta_\rho)], \qquad \theta = (\theta_\rho', \theta_h')'. \tag{68}$$

Thus the non-neutral productivity component is re-inverted at every candidate parameter vector; it is not a generated dependent variable held fixed in a separate transition regression. The additive autoregressive transition used empirically is a low-dimensional special case of this representation.

The estimator uses distributional independence, not only a conditional-mean restriction. Let p_i^B be a sieve of instruments and controls, let $\widetilde{p}_i^B = p_i^B - \overline{p}^B$, and let

$$S_v(u; s, a) = \mathscr{K}\left(\frac{c_v - u/s}{a}\right)$$

be a bounded smooth-CDF score, where $\mathscr{K}$ is an integrated kernel, s is a robust innovation scale, a is a bandwidth, and c_v is a reference cutoff. With $\widetilde{S}_{iv}(\theta) = S_v(u_i^B(\theta); s, a) - N^{-1}\sum_r S_v(u_r^B(\theta); s, a)$, form

$$\widehat{g}_v^B(\theta) = \frac{1}{N}\sum_i \widetilde{p}_i^B \widetilde{S}_{iv}(\theta). \tag{69}$$

Centering both factors makes these covariance restrictions invariant to the unknown marginal distribution and location of u^B. With a sufficiently rich instrument and score sieve, they approximate the independence restriction that underlies the first identification step.

The generic penalized sieve minimum-distance estimator is

$$\widehat{\theta} = \arg\min_{\theta\in\Theta_N}\left\{\sum_{v=1}^{V_N}\pi_v \widehat{g}_v^B(\theta)'\widehat{W}^B\widehat{g}_v^B(\theta) + \lambda_{\rho,N}\mathscr{J}_\rho(\theta_\rho) + \lambda_{h,N}\mathscr{J}_h(\theta_h)\right\}, \tag{70}$$

subject to (66). The penalties control sieve complexity; they need not be positive for a low-dimensional sieve. The estimated non-neutral productivity component is

$$\widehat{\omega}_i^B = \rho_{K_\rho}^{-1}(x_i, \chi_i^{ML}; \widehat{\theta}_\rho).$$

4.2 Stage Two: Characteristic Index and Production Surface

Analysis sample and ratio index. Let $O_i = 1$ indicate that all variables required by the second stage are observed. The second-stage sample is determined by a prespecified biased-productivity and input-support rule:

$$\mathcal{A}_N^B = \left\{ i : O_i = 1,\ (x_i, \widehat{\omega}_i^B, x_{i-}, \widehat{\omega}_{i-}^B) \in \widehat{\mathcal{S}}_N^B \right\}, \qquad n_B = |\mathcal{A}_N^B|. \tag{71}$$

The rule is fixed before estimating the production surface. In particular, it does not select observations using the fitted output residual or a second-stage elasticity.

On the maintained continuation domain, define

$$\widehat{R}(k, \ell, m, b) = \exp\{\rho_{K_\rho}(k, \ell, m, b; \widehat{\theta}_\rho)\}.$$

Choose a prespecified reference section $m_0(k, b)$. For a point (k, ℓ, m, b), solve the reverse characteristic

$$\frac{dL(\mu)}{d\mu} = -\widehat{R}(k, L(\mu), \mu, b), \qquad L(m) = \ell, \qquad \widehat{h}_R(k, \ell, m, b) = L\{m_0(k, b)\}. \tag{72}$$

This labels an isoquant by its landing coordinate on the reference section. Any smooth section that intersects each relevant characteristic once gives an equivalent index after relabeling. The empirical implementation uses a diagonal section because it is more stable numerically; Appendix A.5 gives the exact ODE.

Reduced NPIV problem. Let

$$\widehat{X}_i^R = (k_i, \widehat{\omega}_i^B, \widehat{H}_i^R), \qquad \widehat{H}_i^R = \widehat{h}_R(k_i, \ell_i, m_i, \widehat{\omega}_i^B).$$

Approximate the reduced production surface and the Hicks-neutral transition component by

$$G(x) \approx b_{K_G}^G(x)'\beta, \qquad a^H(v) \approx b_{K_a}^a(v)'\gamma. \tag{73}$$

For instruments p_i^H, the conditional-mean moments are

$$\widehat{g}^H(\beta,\gamma) = \frac{1}{n_B}\sum_{i\in\mathcal{A}_N^B} p_i^H\left[q_i - b_{K_G}^G(\widehat{X}_i^R)'\beta - b_{K_a}^a(V_{i-}^H)'\gamma\right]. \tag{74}$$

When the stronger full-independence restriction is imposed, we add centered smooth-CDF covariance moments of the same Hicks-neutral innovation with the nonconstant instruments. Denote their stacked vector by $\widehat{g}^F(\beta,\gamma)$.

The general second-stage estimator solves

$$\begin{aligned}(\widehat{\beta},\widehat{\gamma}) = \arg\min_{\beta,\gamma}\Big\{&\widehat{g}^H(\beta,\gamma)'\widehat{W}^H\widehat{g}^H(\beta,\gamma) + \eta_F\widehat{g}^F(\beta,\gamma)'\widehat{W}^F\widehat{g}^F(\beta,\gamma)\\ &+ \lambda_{G,N}\mathscr{J}_G(\beta) + \lambda_{a,N}\mathscr{J}_a(\gamma) + \lambda_{S,N}\mathscr{P}_S(\beta)\Big\},\end{aligned} \tag{75}$$

subject to a location normalization and

$$\partial_h G(k,b,h) \geq \underline{c}_{G,N} > 0. \tag{76}$$

Here $\mathscr{J}_G$ and $\mathscr{J}_a$ are roughness penalties, while $\mathscr{P}_S$ is an optional soft shape penalty. The ratio restriction is not penalized: it is imposed by the characteristic construction itself.

The fitted structural objects are

$$\begin{aligned}\widehat{f}(k,\ell,m,b) &= \widehat{G}\Big(k,b,\widehat{h}_R(k,\ell,m,b)\Big),\\ \widehat{\omega}_i^H &= q_i - \widehat{G}(\widehat{X}_i^R).\end{aligned} \tag{77}$$

This composition is important: G is not itself the structural production function in the original inputs. Differentiating the complete composition gives

$$\begin{aligned}\widehat{f}_\ell &= \widehat{G}_h\,\widehat{h}_{R,\ell}, & \widehat{f}_m &= \widehat{G}_h\,\widehat{h}_{R,m},\\ \widehat{f}_k &= \widehat{G}_k + \widehat{G}_h\,\widehat{h}_{R,k}.\end{aligned} \tag{78}$$

The last term in $\widehat{f}_k$ is generally nonzero because the ratio index depends on capital. The variational equations associated with (72) deliver the three derivatives of $\widehat{h}_R$. They also imply $\widehat{f}_m/\widehat{f}_\ell = \widehat{R}$ up to numerical integration tolerance.

4.3 Factor-Augmenting CES Projection

For interpretation, we project the composed structural surface onto a three-input factor-augmenting CES. Let $x_K = k$, $x_L = \ell$, $x_M = m$ and let $g_r(b)$ be a factor-specific log-efficiency loading. The projection is

(79)

$$f^{CES}(k, \ell, m, b) = \delta + \frac{\Gamma}{\varrho} \log \left[\sum_{r \in \{K,L,M\}} s_r \exp\{\varrho[x_r + g_r(b)]\} \right], \qquad s_r > 0, \quad \sum_r s_r = 1,$$

with the continuous Cobb–Douglas limit at $\varrho = 0$. Γ is returns to scale and the common elasticity of substitution is $\sigma = 1/(1 - \varrho)$.

The CES is fitted to two targets from the flexible estimator: the composed structural level (equal to $\widehat{G}(\widehat{X}_i^R)$ at an observed point) and the materials–labor log elasticity ratio. The latter has CES representation

$$r_{ML,i}^{CES} = \log(s_M/s_L) + \varrho\{m_i - \ell_i + g_M(\widehat{\omega}_i^B) - g_L(\widehat{\omega}_i^B)\}.$$

Using prespecified weights and robust scales, the projection minimizes

$$Q^{CES}(\vartheta; \mu_R) = Q_G(\vartheta) + \mu_R Q_R(\vartheta). \tag{80}$$

It does not use observed output as a new structural equation, and it does not re-estimate either productivity component. Consequently, the CES factor labels are a low-dimensional interpretation of the estimated flexible surface, not an additional source of identification.

4.4 Inference

We use a whole-firm-history bootstrap. Each replication samples firms with replacement, keeps each sampled firm's preformed transitions intact, and re-estimates the declared first-stage, support, characteristic-flow, second-stage, and CES-projection blocks. Tuning rules are fixed before resampling: absolute penalty weights selected for a reported specification are held fixed rather than retuned in each draw. Bootstrap standard errors are sample standard deviations across 200 replications. The external city-wage instrument and already generated proxy-interaction columns are held fixed, so the reported uncertainty is conditional on those upstream objects and is not weak-identification-robust.

5 Data

We use firm-level data from China's Annual Survey of Industrial Enterprises (ASIE), whose construction and limitations are discussed by Brandt et al. (2014), and restrict attention to CIC 171 (textiles). The source covers 1998–2007, and our one-calendar-year transitions run from 1999 through 2007. Accounting and complete-case filters, a prespecified pooled-residual first-to-ninety-ninth-percentile screen, and the requirement of an exact one-calendar-year predecessor yield 42,619 observations from 13,793 firms. Seven observations lack the lagged city identifier needed for the city-wage measure, leaving a common analytical sample of 42,612 observations. The baseline bilinear-index specification retains 42,609 observations after its second-stage support check.

The production variables are quantities in logs: q_{jt} is log real gross output, k_{jt} is log real capital, ℓ_{jt} is log employment, and m_{jt} is log real intermediate inputs. These differ from the nominal materials–labor expenditure ratio,[7]

$$\chi_{jt}^{ML} = \log\left(\frac{E_{jt}^{M}}{E_{jt}^{L}}\right), \qquad E_{jt}^{M} = \mathrm{NI}_{jt},$$

We also retain ownership, location, and lagged city-wage variables for estimation.

6 Empirical Results

The empirical analysis proceeds in three steps. First, we compare local production-function objects across the flexible estimates and the CES comparisons. Second, we report fixed-input comparative statics from the three flexible production-function estimates (a baseline and its two extensions), which provide the paper's primary structural evidence. Third, we use the CES specifications to provide a parsimonious factor-efficiency interpretation. The CES refits are downstream projections of the flexible production-function estimates, whereas the directly estimated CES with linear factor-augmenting loadings is a more restrictive benchmark, but is free of the projection step. The labor-augmenting CES is included only as a deliberately restrictive diagnostic. The fixed-input exercises in Tables 2 and 3 therefore do not impose a factor-augmenting CES representation. Unless otherwise noted, uncertainty is measured by 200 whole-firm bootstrap replications that re-estimate the relevant specification.

[7] Following Hsieh and Klenow (2009), we multiply reported nominal labor compensation by $c_{\mathrm{HK}} = 2.3134422909506$ when constructing labor expenditure: $E_{jt}^{L} = c_{\mathrm{HK}}(\mathrm{wage}_{jt} + \mathrm{wfr}_{jt})$. Thus, relative to the raw nominal expenditure ratio, χ_{jt}^{ML} includes $-\log c_{\mathrm{HK}}$.

6.1 Specifications and Local Production-Function Objects

Table 1 compares eight estimators, one per row. Panel A reports the baseline flexible production-function estimator and its two extensions. The three flexible estimators share a common second-stage specification: a second-order Bernstein polynomial in capital, the non-neutral productivity component, and the scalar flexible-input index; they differ only in the first stage. For estimation of the elasticity-ratio function in the first stage, the baseline uses a normalized bilinear index with linear input terms and interactions between each input and ω^B; the quadratic-input extension adds the full quadratic in inputs while retaining these interactions; and the proxy-first-stage extension augments the baseline with the conventional OP/LP/ACF-style proxy-variable first stage (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015). It nonparametrically regresses output on inputs and demand shifters,[8] thereby purging the ex-post output shock before the baseline estimation procedure is applied. Panel B reports the three corresponding Bernstein-loading CES refits, together with a directly estimated CES model with linear factor-augmenting loadings and a deliberately restrictive labor-augmenting CES diagnostic. Appendix A provides the exact finite-dimensional specifications and estimation details.

The flexible estimates consistently assign the largest mean output elasticity to materials (0.771–0.807), followed by labor (0.142–0.160) and capital (0.026–0.053); mean returns to scale range from 0.966 to 0.993. Their CES refits preserve this ranking, with materials elasticities of 0.761–0.825, labor elasticities of 0.139–0.159, capital elasticities of 0.019–0.070, and returns to scale of 0.970–1.003. The three refits also imply similar substitution elasticities, 0.680–0.698, although the quadratic-input refit is considerably less precise. The direct CES with linear factor-augmenting loadings preserves the input ranking but implies a lower materials elasticity (0.661), lower returns to scale (0.888), and a much smaller substitution elasticity (0.102). The labor-augmenting CES is included solely as a deliberately restrictive diagnostic. Thus, the input ranking is stable across the flexible estimates, their CES refits, and the direct linear-loading CES, whereas the magnitude of substitution is sensitive to whether CES structure is imposed directly or used as a projection of a flexible surface. This difference plausibly reflects the linear factor-augmenting loadings imposed on a common non-neutral productivity component in the direct CES specifications, which tightly restrict the admissible joint variation in the capital-, labor-, and materials-augmenting components.

[8] In our application, input and output tariffs serve as demand shifters, providing the variation needed to break the nonidentification problem highlighted by Gandhi et al. (2020).

Table 1: Industry 171: Mean Production-Function Objects Across Estimators

	Mean output elasticity			Mean	
Specification	Capital	Labor	Materials	RTS	σ
Panel A: Flexible production-function estimates					
Baseline	0.033	0.154	0.793	0.981	–
	(0.004)	(0.002)	(0.005)	(0.004)	
Quadratic-input extension	0.053	0.142	0.771	0.966	–
	(0.010)	(0.005)	(0.024)	(0.020)	
Proxy-first-stage extension	0.026	0.160	0.807	0.993	–
	(0.006)	(0.003)	(0.010)	(0.008)	
Panel B: CES estimates					
Baseline CES refit	0.028	0.155	0.808	0.992	0.682
	(0.007)	(0.002)	(0.009)	(0.006)	(0.014)
Quadratic-input CES refit	0.070	0.139	0.761	0.970	0.698
	(0.015)	(0.005)	(0.024)	(0.017)	(0.150)
Proxy-first-stage CES refit	0.019	0.159	0.825	1.003	0.680
	(0.006)	(0.002)	(0.010)	(0.008)	(0.014)
Direct CES (linear loadings)	0.093	0.134	0.661	0.888	0.102
	(0.028)	(0.008)	(0.043)	(0.023)	(0.027)
Labor-augmenting CES diagnostic	0.069	0.145	0.702	0.917	0.190
	(0.009)	(0.001)	(0.007)	(0.010)	(0.010)

NOTE.—The table reports firm-year means of the capital, labor, and materials output elasticities and returns to scale for Industry 171. In parentheses are bootstrap standard errors from 200 whole-firm bootstrap replications that resample complete firm histories and re-estimate the declared estimator. Panel A reports the three flexible estimators. They share a common second-stage specification: a second-order Bernstein polynomial in capital, the non-neutral productivity component, and the scalar flexible-input index; they differ only in the elasticity-ratio first stage. The baseline uses a normalized bilinear index with linear input terms and input–ω^B interactions; the quadratic-input extension adds the full input quadratic; and the proxy-first-stage extension first nonparametrically regresses output on inputs and demand shifters, in the conventional OP/LP/ACF style, to purge the ex-post output shock before applying the baseline procedure. The first three rows of Panel B are the corresponding downstream Bernstein-loading CES refits. Direct CES (linear loadings) estimates the CES structure directly under mean-independence moments, whereas the labor-augmenting CES diagnostic restricts factor augmentation to labor. For all CES estimates, $\sigma = 1/(1-\rho)$. Dashes indicate that a flexible specification does not impose a single scalar elasticity of substitution.

These local-object comparisons summarize average production-function properties across specifications and provide reassurance that the estimated production functions have plausible local properties. However, the primary structural evidence comes from the fixed-input comparative statics of the flexible estimates. We turn to those results next.

6.2 Material-Biased Technical Change

Table 2 reports the fixed-input elasticity responses to biased productivity in the three flexible estimates. Because Hicks-neutral productivity does not affect output elasticities, this exercise varies only ω^B, from each lineage's 1999 median to its 2007 median, at a common fixed input vector. Each reported 2007/1999 index is normalized to 100 in 1999, and the ratio index and all associated model-implied objects are recomputed after moving the non-neutral productivity component.

Table 2: Industry 171: Fixed-Input Elasticity Responses to Biased Productivity

Specification	Capital	Labor	Materials	RTS
Baseline	48.78	47.34	120.52	100.52
	(11.29)	(3.36)	(2.55)	(0.61)
Quadratic-input extension	70.57	51.98	113.06	99.07
	(5.06)	(5.48)	(2.16)	(0.62)
Proxy-first-stage extension	57.51	46.69	118.87	99.99
	(14.28)	(3.36)	(2.39)	(0.62)

NOTE.—The table reports 2007 fixed-input comparative-static indexes for the three flexible production-function estimates, with each specification's 1999 value normalized to 100. Raw log inputs are fixed at the common vector $(k, l, m) = (8.930, 5.056, 10.056)$. Because Hicks-neutral productivity does not affect output elasticities, the exercise moves only ω^B, from each lineage's 1999 median to its 2007 median. In parentheses are bootstrap standard errors from 200 whole-firm bootstrap replications. The ω^B medians are draw-specific, and each index is normalized within its draw before the standard error is computed. Each draw retains its estimated reference section and replication-specific Bernstein domain. These are within-model fixed-input comparative statics, not observed changes.

All three flexible specifications yield the same directional pattern. The capital-elasticity index falls to 48.78–70.57 and the labor-elasticity index to 46.69–51.98, whereas the materials-elasticity index rises to 113.06–120.52. Returns to scale remain close to their 1999 levels, with indexes between 99.07 and 100.52. Bootstrap standard errors range from 5.06 to 14.28 for capital, 3.36 to 5.48 for labor, 2.16 to 2.55 for materials, and 0.61 to 0.62 for returns to scale. The cross-row range describes sensitivity across the maintained flexible lineages, not the response to a common realized shock.

Under the maintained flexible-input first-order conditions,

$$\frac{w^M M}{w^L L} = \frac{\partial_m f_t}{\partial_\ell f_t}.$$

The elasticity movements therefore raise the model-implied materials–labor expenditure ratio

and shift the shares of combined labor-and-materials expenditure away from labor and toward materials. This is a fixed-input, within-model comparison, not a statement about total cost shares or realized, reoptimized input bundles.

6.3 Output Effects of the Productivity Components

Table 3 applies the same fixed-input design to output for the three flexible estimates. Raw log inputs are held at the common comparison vector, and each lineage's 1999 medians of ω^B and ω^H define its baseline. The three columns move only ω^B, only ω^H, or both productivity components to their corresponding 2007 medians; every reported 2007/1999 output index is normalized to 100 in 1999.

Table 3: Industry 171: Fixed-Input Output Responses to Productivity Components

Specification	Only ω^B	Only ω^H	Both components
Baseline	108.54	115.94	125.84
	(0.97)	(0.72)	(1.02)
Quadratic-input extension	105.16	116.01	122.00
	(0.97)	(1.38)	(1.12)
Proxy-first-stage extension	107.66	113.96	122.69
	(1.16)	(0.73)	(0.94)

NOTE.—The table reports 2007 fixed-input output indexes for the three flexible production-function estimates, with each specification's own 1999 output normalized to 100. Raw log inputs are fixed at the common vector $(k, l, m) = (8.930, 5.056, 10.056)$. Within each flexible lineage, the first two columns move only the named productivity component from its lineage-specific 1999 median to its 2007 median; the third moves both productivity components. In parentheses are bootstrap standard errors from 200 whole-firm bootstrap replications. The medians of the productivity components are draw-specific, and every draw is normalized by its own 1999 output before the standard error is computed. Each draw retains its estimated reference section and replication-specific Bernstein domain. The joint-component index equals the product of the two single-component indexes divided by 100, draw by draw. These are within-model comparative statics, not observed changes.

Moving only ω^B raises the fixed-input output index by 5.2–8.5 percent across the three flexible specifications. Moving only ω^H produces a larger increase of 14.0–16.0 percent. When both productivity components move, the output index rises by 22.0–25.8 percent. Thus, within these fixed-input comparisons, the Hicks-neutral component produces the larger direct output shift, while the biased component changes both output and the relative marginal importance of labor and materials.

The joint-component column provides an internal consistency check: within each bootstrap draw, its index equals the product of the two single-component indexes divided by 100.

The cross-row ranges reflect sensitivity across the maintained flexible lineages, not responses to a common realized shock. Taken together, the three flexible estimates deliver the paper's primary structural result: moving ω^B from each lineage's 1999 median to its 2007 median at fixed inputs shifts the materials–labor margin toward materials, whereas ω^H produces the larger direct fixed-input output effect. These are model-implied, lineage-specific comparative statics; they do not describe reoptimized firm choices, decompose observed output growth, or establish causal calendar-time effects.

These results do not decompose the scalar non-neutral productivity component into factor-specific efficiency terms. We therefore turn to the CES specifications only for a parsimonious, conditional factor-efficiency interpretation.

6.4 Factor-Efficiency Interpretation from the CES Specifications

The CES specifications provide a complementary, parsimonious factor-efficiency interpretation rather than a second measure of the flexible production-function estimates. Table 4 reports 2007 factor-efficiency and Hicks-neutral indexes for the four multi-factor-augmenting CES specifications, with each index normalized to 100 in 1999. It separately reports the deliberately restrictive labor-augmenting CES only as a diagnostic of the labor-only restriction. Let E^j_{2007} denote the factor-specific efficiency index for factor j and let H_{2007} denote the Hicks-neutral index. For CES returns to scale Γ, the reported Hicks-adjusted factor index is $\widetilde{E}^j_{2007} = E^j_{2007}(H_{2007}/100)^{1/\Gamma}$. The exponent $1/\Gamma$ converts the Hicks-neutral output shift into an equivalent common shift in input efficiency. The index preserves the 1999 normalization; the final column reports H_{2007} separately. Panel A contains the three CES refits of the flexible production-function estimates, Panel B contains the directly estimated CES with linear factor-augmenting loadings, and Panel C contains the labor-augmentation diagnostic. The flexible Bernstein production functions do not impose separable factor-specific efficiency loadings, so they do not themselves produce these indexes.

Table 4: Industry 171: Hicks-Adjusted Factor-Efficiency and Hicks-Neutral Indexes

	Hicks-adjusted factor-efficiency index			Hicks-neutral index
Specification	Capital	Labor	Materials	H
Panel A: CES refits				
Baseline CES refit	228.96	469.11	90.48	115.94
	(29.58)	(31.31)	(1.43)	(0.72)
Quadratic-input CES refit	193.47	447.72	93.63	116.01
	(34.92)	(79.29)	(10.16)	(1.38)
Proxy-first-stage CES refit	208.61	457.85	89.14	113.96
	(25.51)	(30.80)	(1.31)	(0.73)
Panel B: Direct CES with linear factor-augmenting loadings				
Direct CES (linear loadings)	187.12	322.22	107.34	105.96
	(4.09)	(9.36)	(1.39)	(1.30)
Panel C: Labor-augmentation diagnostic				
Labor-augmenting CES diagnostic	106.91	332.01	106.91	106.32
	(0.55)	(10.18)	(0.55)	(0.46)

NOTE.—The table reports 2007 indexes normalized to 100 in 1999. For factor $j \in \{k, l, m\}$, let E^j_{2007} denote the factor-specific efficiency index and let H_{2007} denote the Hicks-neutral index, each separately normalized to 100 in 1999. The first three columns report the composite index $\widetilde{E}^j_{2007} = E^j_{2007}(H_{2007}/100)^{1/\Gamma}$, where Γ is the estimated CES returns to scale; the final column reports H_{2007} separately. The underlying indexes are $E^j_{2007} = 100 \operatorname{med}_{i:t=2007} \exp\{g_j(\omega^B_{it})\} / \operatorname{med}_{i:t=1999} \exp\{g_j(\omega^B_{it})\}$ and $H_{2007} = 100 \exp\{\operatorname{med}_{i:t=2007}(\omega^H_{it}) - \operatorname{med}_{i:t=1999}(\omega^H_{it})\}$. In parentheses are bootstrap standard errors from 200 whole-firm bootstrap replications that resample complete firm histories and re-estimate the declared estimator. For each adjusted factor index, the formula is evaluated within every bootstrap replication using that replication's estimated Γ, E^j_{2007}, and H_{2007}; its standard error is the sample standard deviation of the resulting 200 indexes. Panel A reports the three Bernstein-loading CES refits, Panel B reports the direct CES with linear factor-augmenting loadings, and Panel C reports the deliberately restrictive labor-augmentation diagnostic. In Panel C, the capital- and materials-specific efficiency indexes are fixed at 100, so their adjusted indexes both equal $100(H_{2007}/100)^{1/\Gamma}$ by construction.

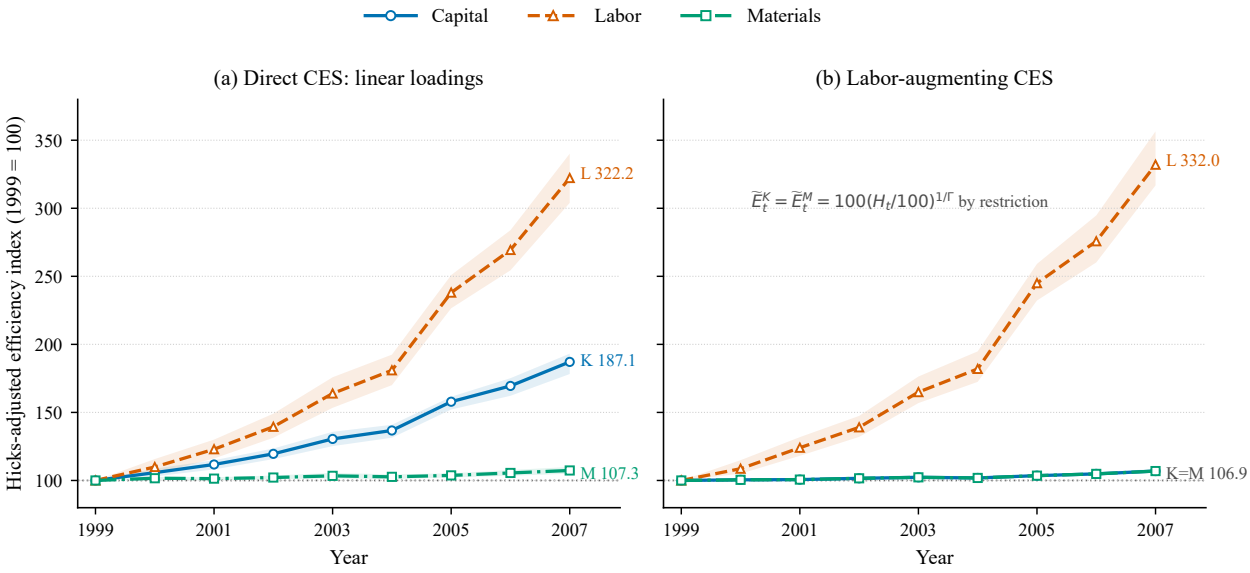
Across the four multi-factor-augmenting CES specifications—the three CES refits and the direct CES with linear factor-augmenting loadings—the 2007 Hicks-adjusted composite indexes range from 322.22–469.11 for labor, 187.12–228.96 for capital, and 89.14–107.34 for materials; the corresponding Hicks-neutral indexes range from 105.96 to 116.01. Thus, labor experienced the largest efficiency gains, capital efficiency also improved substantially, and materials efficiency was comparatively flat. Bootstrap uncertainty is largest for the quadratic-input CES refit, particularly for its labor index.

The labor-augmenting CES is excluded from these headline comparisons. It is reported

only as a deliberately restrictive diagnostic: because it fixes the capital- and materials-specific efficiency indexes at 100, their Hicks-adjusted indexes both equal $100(H_{2007}/100)^{1/\Gamma}$, and it cannot represent the distinct capital- and materials-specific paths permitted by the multi-factor specifications. Its contrast with the direct linear-loading CES illustrates potential restriction-induced misspecification bias, conditional on that maintained benchmark; it is neither a robustness check nor a preferred specification, and does not provide a formal or causal estimate of misspecification bias. These headline comparisons are conditional on each CES specification and its loading normalization: the refit rows are projections of flexible surfaces, the direct linear-loading row imposes CES during estimation.

Figure 1 uses the deliberately restrictive labor-augmenting CES as a diagnostic by contrasting its complete annual paths on a common vertical scale with those from the direct CES with linear factor-augmenting loadings. With linear factor-augmenting loadings, the 2007 adjusted capital, labor, and materials indexes are 187.12, 322.22, and 107.34, respectively. Under pure labor augmentation, the corresponding indexes are 106.91, 332.01, and 106.91. Because the latter specification fixes the unadjusted capital and materials indexes at 100, their adjusted paths both equal $100(H_t/100)^{1/\Gamma}$ and distinct capital- or materials-specific trends are ruled out. Relative to the linear-loading benchmark, the labor-only specification's 2007 adjusted capital index is therefore 80.20 points lower. The comparison illustrates potential restriction-induced misspecification bias: pure labor augmentation cannot represent the capital-specific path permitted by the richer loading structure. Because the linear-loading CES is itself a maintained model, these cross-panel gaps are best read as conditional diagnostics of the labor-only restriction rather than as formal or causal estimates of misspecification bias. The CES refits of the flexible production-function estimates provide complementary qualitative evidence: labor efficiency improved the most, capital efficiency also rose substantially, and materials efficiency remained comparatively flat.

Figure 1: Hicks-Adjusted Factor-Efficiency Paths under Alternative CES Restrictions



NOTE.—Panel (a) reports the direct CES estimator with linear factor-augmenting loadings, and Panel (b) reports the deliberately restrictive labor-augmenting CES diagnostic. Each panel plots the annual capital, labor, and materials efficiency indexes for 1999–2007 after incorporating the Hicks-neutral component. For factor $j \in \{k, l, m\}$, the plotted index is $\widetilde{E}_t^j = E_t^j (H_t/100)^{1/\Gamma}$, where Γ is the estimated CES returns to scale and E_t^j and H_t are each normalized to 100 in 1999. Lines and markers are point estimates. For each estimator separately, the shaded regions are pointwise 95% bootstrap percentile intervals from its 200 whole-firm bootstrap replications; the adjusted index is computed using each replication's estimated Γ, E_t^j, and H_t before the interval is computed. Draws are not paired across estimators. Both panels use the same vertical scale. Under pure labor augmentation, the unadjusted capital and materials indexes are fixed at 100, so their adjusted indexes both equal $100(H_t/100)^{1/\Gamma}$ in every year. Cross-panel differences describe sensitivity to the factor-loading restriction and are not causal bias estimates.

7 Conclusion

This paper develops a framework for identifying factor-biased productivity without assigning the latent productivity component an ex ante labor-, capital-, or materials-augmenting interpretation. Output is generated by a flexible, nonseparable production function of capital, labor, materials, and a scalar non-neutral productivity component, plus an additive Hicks-neutral component. The key restriction orders the non-neutral productivity component by the materials–labor relative marginal product at fixed inputs. Because the common cost multiplier and Hicks-neutral productivity cancel from the flexible-input first-order conditions, relative expenditures reveal that ordering. Combined with the model's timing and information structure, the nonseparable IV-quantile restriction identifies the non-neutral

productivity component.

The identification argument is constructive. The estimated elasticity-ratio field characterizes the directions of conditional isoquants, and a characteristic-flow index reduces the two flexible inputs to a single endogenous index. An additive NPIV step recovers a reduced form function G_t and the Hicks-neutral component. The structural production function is recovered only after composing G_t with the characteristic index; G_t is not itself the production function. This distinction clarifies the status of the factor-augmenting CES refit: it is a parsimonious projection of the estimated structural surface, useful for interpretation and comparison, but it supplies no additional primitive information.

The application to China's textile industry yields a consistent directional pattern across the three flexible production-function specifications. Materials have the largest local output elasticity, followed by labor and then capital; mean returns to scale range from 0.966 to 0.993. At a common fixed input vector, moving ω^B from its 1999 median to its 2007 median lowers the capital and labor elasticities and raises the materials elasticity. Under the maintained flexible-input first-order conditions, $\frac{w^M M}{w^L L} = \frac{\partial_m f_t}{\partial_\ell f_t}$. Thus, the same biased-productivity movement raises the model-implied materials–labor expenditure ratio and shifts the shares of combined labor-and-materials expenditure away from labor and toward materials. In the corresponding fixed-input output exercise, the three flexible specifications imply output gains of 5.2–8.5 percent from non-neutral productivity growth, 14.0–16.0 percent from Hicks-neutral productivity growth, and 22.0–25.8 percent from growth in both components. These exercises aim to identify the ceteris paribus effects of the productivity components and thus of technical change.

The four multi-factor-augmenting CES specifications provide a complementary, low-dimensional economic interpretation. Relative to 1999, their Hicks-adjusted 2007 composite indexes range from 322.22–469.11 for labor, 187.12–228.96 for capital, and 89.14–107.34 for materials; the Hicks-neutral index ranges from 105.96 to 116.01. Thus, labor experienced the largest efficiency gains, capital efficiency also improved substantially, and materials efficiency was comparatively flat. The deliberately restrictive labor-augmenting CES is reported only as a diagnostic of the labor-only restriction; it is neither a preferred specification nor a robustness check.

These findings show why relaxing labor augmentation is economically meaningful, rather than merely a more flexible parameterization. The robust flexible-model finding is an FOC-implied, fixed-input reallocation at the materials–labor margin: growth in the non-neutral productivity component raises the materials–labor expenditure ratio and shifts the model-

implied shares of combined labor-and-materials expenditure away from labor and toward materials. It identifies the direction of relative bias, but does not attribute the shift exclusively to labor augmentation. As a counterexample, in a nested-CES technology, purely capital-augmenting technical change can also raise the cost-minimizing materials–labor ratio when the capital–materials elasticity of substitution is lower than the elasticity of substitution between the capital–materials composite and labor. Imposing labor augmentation from the outset could therefore conflate the direction of relative bias at this margin with its factor-specific source. The two analyses are complementary: the flexible estimator identifies the direction of relative bias, while the CES projections provide a conditional factor-augmenting interpretation.

The nonparametric specification and multidimensional productivity structure of our model can make its production-function estimates less vulnerable to misspecification than parametric estimators or models that impose only a scalar productivity term. In this sense, our approach may provide greater reassurance in applications that rely on production-function estimates, such as the measurement of markups, as in De Loecker and Warzynski (2012) and Demirer (2020).

References

Acemoglu, D. (2002). Directed technical change. *The review of economic studies*, 69(4):781–809. 2, 11

Acemoglu, D. (2003). Labor-and capital-augmenting technical change. *Journal of the European Economic Association*, 1(1):1–37. 2

Ackerberg, D., Hahn, J., and Pan, Q. (2026). Nonparametric identification using timing and information set assumptions with an application to non-hicks neutral productivity shocks. *Working Paper*. 3, 15, 61

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451. 4, 34

Ackerberg, D. A. and De Loecker, J. (2026). Production functions under imperfect competition: A quality-equivalent cournot model. *AEA Papers and Proceedings*, 116:475–482. 1

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. 33

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261. 15, 60

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American economic review*, 102(6):2437–71. 43

Demirer, M. (2020). Production function estimation with factor-augmenting technology: An application to markups. Technical report, MIT working paper. 1, 2, 4, 9, 13, 14, 43

Doraszelski, U. and Jaumandreu, J. (2018). Measuring the bias of technological change. *Journal of Political Economy*, 126(3):1027–1084. 1, 2, 4, 5, 9, 11, 13

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. 3, 34, 62, 65

Hicks, J. R. (1932). *The Theory of Wages*. Macmillan, London. 2, 11

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448. 33

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2):317–341. 4, 14, 34

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578. 15, 26, 60

Oberfield, E. and Raval, D. (2021). Micro data and macro technology. *Econometrica*, 89(2):703–732. 4

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. 4, 14, 34

Pan, Q. (2026). Identification of gross output production functions with nonseparable productivity. Working paper. 11

Zhang, H. (2019). Non-neutral technology, firm heterogeneity, and labor demand. *Journal of Development Economics*, 140:145–168. 4, 5, 6

Zhao, S., Malikov, E., and Kumbhakar, S. C. (2024). On the estimation of nonneutral production technologies without homothetic separability. Working paper. 2

A Empirical Estimation Details

This appendix describes the finite-dimensional implementation of the eight estimators in Table 1: the baseline flexible estimator and its quadratic-input and proxy-first-stage extensions, their three Bernstein-loading CES refits, the direct CES with linear factor-augmenting loadings, and the labor-augmenting CES diagnostic. The replication labels P8 and P14 denote first stages with eight and 14 parameters, respectively; Q29 denotes the 29-parameter second stage used by all three flexible estimators. Each CES refit conditions on the productivity components and flexible surface from the corresponding point estimate or bootstrap draw, without re-estimating the flexible stages. The two directly estimated CES specifications jointly estimate their technology and both productivity components under their own mean-independence moments.

A.1 Analysis Sample and First-Stage Moments

The common starting sample is the CIC 171 sample described in Section 5: 42,612 observations from 13,793 firms after requiring the lagged city identifier used by the external wage instrument. The baseline flexible estimator and the proxy-first-stage extension exclude the same three observations under the predecessor/support screen and retain 42,609 observations. The quadratic-input extension applies a more restrictive whole-firm support rule and retains 36,647 observations from 11,958 firms.[9] Both directly estimated CES specifications use all 42,612 observations. Here, $i = (j, t)$ indexes a usable observation and $i- \equiv (j, t-1)$ its one-year lag.

All three flexible estimators use time-invariant elasticity-ratio maps, non-neutral productivity transitions, reduced production surfaces, and Hicks-neutral productivity transitions over 1999–2007. These are maintained implementation restrictions; period-specific identification does not impose them, and the productivity-component distributions may vary by year. Let $b_i \equiv \omega_i^B$ denote non-neutral productivity, o_{i-} the lagged ownership control, and z_{i-}^{wage} the lagged external city-year median wage. Write $x_i = (k_i, \ell_i, m_i)'$ and $x_i^c = x_i - \bar{x}$, where the point-estimate center is the arithmetic mean of current inputs,

$$\bar{x} = (9.0525894326, 5.1909322025, 10.1522329221)'. \tag{81}$$

The same center is used for current and lagged inputs and is recomputed from current inputs in every bootstrap draw. The relative-expenditure outcome χ_i^{ML} is defined in Section 5.

[9] The key results remain robust if we apply a less agressive support screen to the quadratic-input extension.

The three elasticity-ratio first stages use the same 14-column basis of primitive and generated instruments. It contains a constant, current capital, z_{i-}^{wage}, lagged output, the five lagged variables $(k, \ell, m, \chi^{ML}, o)$, and five interactions between a centered lagged auxiliary index and lagged (k, ℓ, m), z_{i-}^{wage}, and lagged output. The auxiliary index comes from the nonstructural linear model

$$\chi_i^{ML} = x_i^{c\prime}\beta^a + b_i^a, \qquad b_i^a = \alpha_0^a + \alpha_B^a b_{i-}^a + \alpha_o^a o_{i-} + u_i^a.$$

This index is used only to construct the instrument interactions. After removing the constant, we center and standardize the remaining 13 columns and apply symmetric Gram whitening. The generated auxiliary-index interactions are held fixed in the bootstrap, as discussed below.

For a candidate first-stage coefficient vector, let u_i^B denote the residual from the implemented linear transition for b_i. The reference innovation scale is its interquartile range divided by 1.349. The score grid uses empirical quantiles of the standardized reference innovation at

$$\mathscr{P} = \{.05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95\}. \tag{82}$$

We smooth the indicator with the integrated Epanechnikov kernel and bandwidth 0.20. The score blocks receive normalized trapezoidal weights over $\mathscr{P}$. Both the score and each instrument are centered, leaving the marginal location and CDF of u_i^B unrestricted. This gives $11 \times 13 = 143$ first-stage covariance moments. The reference scale and empirical cutoffs are fixed before each first-stage optimization and reconstructed under the specified rule for each bootstrap draw.

A.2 Baseline and Quadratic-Input Ratio First Stages

The baseline uses a normalized bilinear elasticity-ratio function,

$$\rho^{NBI}(x_i, b_i) = \beta' x_i^c + \{1 + \gamma' x_i^c\} b_i, \qquad b_i = \alpha_B b_{i-} + \alpha_o o_{i-} + u_i^B. \tag{83}$$

The eight parameters are the three elements of β, the three elements of γ, and (α_B, α_o). For each trial parameter vector, non-neutral productivity is recovered in closed form as

$$b_i = \frac{\chi_i^{ML} - \beta' x_i^c}{1 + \gamma' x_i^c}, \tag{84}$$

with the analogous expression for b_{i-}. The reference normalization $\rho^{NBI}(\bar{x}, b) = b$ holds exactly because $x^c = 0$ at $x = \bar{x}$. We impose $1+\gamma' x^c \geq 10^{-4}$ at all vertices of the box defined by the first and ninety-ninth percentiles of pooled current and lagged inputs. Because the loading is affine, these restrictions imply the same lower bound throughout the box. The baseline first stage uses no curvature penalty.

For the quadratic-input extension, define

$$q_2(x_i^c) = ((k_i^c)^2, (\ell_i^c)^2, (m_i^c)^2, k_i^c \ell_i^c, k_i^c m_i^c, \ell_i^c m_i^c)'.$$

Its elasticity-ratio function is

$$\rho^{NQBI}(x_i, b_i) = \beta' x_i^c + \delta' q_2(x_i^c) + \{1 + \gamma' x_i^c\} b_i, \qquad b_i = \alpha_B b_{i-} + \alpha_o o_{i-} + u_i^B. \tag{85}$$

The 14 parameters comprise three linear input terms, six quadratic input terms, three input–b interactions, and two transition coefficients. To recover b, replace the numerator of (84) with $\chi_i^{ML} - \beta' x_i^c - \delta' q_2(x_i^c)$. The additional quadratic terms preserve the reference normalization $\rho^{NQBI}(\bar{x}, b) = b$, and inversion remains available in closed form.

For this extension, the loading on b satisfies

$$1 + \gamma' x^c \geq 0.01 \tag{86}$$

throughout the retained current-and-lag input-support box. Eight vertex inequalities enforce this bound exactly because the loading remains affine in inputs. The quadratic terms enter the input intercept, while strict monotonicity in b is enforced by a hard constraint. Candidate inversions must also have positive slopes at every observed current and lagged input vector; the stronger 0.01 bound applies to the retained second-stage support.

Let N be the number of first-stage observations. We stabilize the six quadratic contributions with the penalty

$$\mathscr{R}_B(\delta) = \frac{1}{2N} \sum_{i=1}^{N} \left[\{\delta' q_2(x_i^c)\}^2 + \{\delta' q_2(x_{i-}^c)\}^2 \right]. \tag{87}$$

The first-stage objective adds $\lambda_B \mathscr{R}_B(\delta)$ with $\lambda_B = 0.001$. This value was selected to balance stability and fit on the previously evaluated homotopy grid $\{0, 0.0003, 0.001, 0.003, 0.01\}$ and is held fixed in bootstrap draws. The slope bound in (86) remains a hard identification-preserving restriction, separate from the curvature penalty.

A.3 Proxy-First-Stage Extension

The proxy-first-stage extension uses fitted output from an auxiliary regression in the second stage. It retains the relative-expenditure outcome χ_i^{ML} and the baseline elasticity-ratio map and inversion. Let $s_i = (k_i, \ell_i, m_i, \chi_i^{ML})'$ and let $B_3(\widetilde{s}_i)$ collect all monomials of total degree at most three in the four current standardized variables. This basis has 35 columns, including its constant. With T_i^O and T_i^I denoting the contemporaneous output- and input-tariff variables, respectively, we estimate by pooled unweighted OLS

$$q_i = X_i^P \pi + e_i, \qquad X_i^P = \left[B_3(\widetilde{s}_i),\ \widetilde{T}_i^O B_3(\widetilde{s}_i),\ \widetilde{T}_i^I B_3(\widetilde{s}_i) \right]. \tag{88}$$

The regression has 105 regressors and contains no firm or year fixed effects, tariff squares, or tariff cross-products. Current-sample means and scales standardize both the current regressors and their lagged counterparts. Applying the fitted coefficients to the latter gives $\widehat{q}_{i-}$. The auxiliary regression uses all 42,612 observations in the common sample and is re-estimated in every bootstrap draw.

The fitted values $\widehat{q}_i$ and $\widehat{q}_{i-}$ replace q_i and q_{i-} throughout the second stage, including its lagged-output instrument. Observed q_i enters the second stage only through the auxiliary OLS; it remains available for the unchanged ratio-first-stage instrument construction and audit checks. The observed relative-expenditure ratio χ_i^{ML}, ratio inversion, characteristic construction, second-stage sieve, and baseline support screen are otherwise unchanged. The extension therefore retains 42,609 observations, and its corresponding CES refit uses the resulting flexible production surface.

A.4 Second-Stage Samples

The sample rules are applied before fitting the production surface and do not use output residuals or second-stage elasticities. The baseline flexible estimator excludes three observations outside the admissible region for first-stage inversion and retains 42,609 observations from all 13,793 firms. The proxy-first-stage extension applies the same screen and uses the same second-stage sample, although its auxiliary output regression uses all 42,612 observations. Previously constructed lags are carried forward without being rebuilt after sample selection.

The quadratic-input extension uses a more conservative support rule. Among observations satisfying the 0.01 slope floor, we pool current and lagged values of each variable in (k, ℓ, m, b) and compute its first and ninety-ninth percentiles. An observation passes the initial screen if the current and lagged values of all four variables lie within their respective

bounds, including the endpoints, and both loadings on b satisfy (86). In the point-estimate sample, 3,451 of the 42,612 observations fail this screen. We then exclude all observations from any firm with at least one failed observation. This whole-firm rule removes another 2,514 otherwise admissible observations, leaving 36,647 observations from 11,958 firms. The support bounds and whole-firm screen are recomputed in every bootstrap draw, without rebuilding lags. The quadratic-input estimator and its CES refit use this same sample.

A.5 Characteristic Construction and the Second-Stage Surface

All three flexible estimators use the same form of reference rule for constructing the scalar flexible-input index. Set $d_i = m_i - \ell_i$ and standardize (k_i, b_i) by their pooled medians and interquartile ranges. The reference difference $d_0(k, b)$ is the pooled least-absolute-deviations (LAD) projection of d_i on

$$(1, \widetilde{k}, \widetilde{b}, \widetilde{k}^2, \widetilde{kb}, \widetilde{b}^2). \tag{89}$$

We assess leverage relative to its ninety-ninth percentile, but retain high-leverage observations in the quadratic LAD fit.

The characteristic is constructed from the estimated elasticity-ratio function. Let $M_i(s)$ solve

$$\frac{dM_i(s)}{ds} = \frac{1}{1 + \exp\{\widehat{\rho}(k_i, M_i(s) - s, M_i(s), b_i)\}}, \qquad M_i(d_i) = m_i, \tag{90}$$

from the observed difference d_i to $d_0(k_i, b_i)$. The scalar flexible-input index is the labor coordinate at this reference section,

$$H_i = M_i\{d_0(k_i, b_i)\} - d_0(k_i, b_i). \tag{91}$$

Equation (90) traces an isoquant because $dm/dd = f_\ell/(f_\ell + f_m) = 1/(1 + f_m/f_\ell)$. We solve the current and lagged characteristic equations with adaptive refinement and jointly integrate their variational equations. These equations differentiate the estimated ratio function, the initial condition, and the moving LAD reference, using analytic derivatives of the latter. They yield (H_k, H_ℓ, H_m) and their lagged counterparts. Numerical checks require the fitted relative-expenditure map to reconstruct χ^{ML} within 10^{-9} and each characteristic endpoint to agree within 10^{-8}. No additional observations are excluded on the basis of the fitted reference or path length.

The three flexible estimators share this characteristic construction and the Bernstein representation below, while their moment restrictions and regularization differ. Let $q_i^\star = q_i$ for the baseline and quadratic-input estimators and $q_i^\star = \widehat{q}_i$ for the proxy-first-stage extension, with the same convention for lagged output.

Let $B_{a,2}$ denote the degree-two Bernstein basis on $[0,1]$. We pool current and lagged values of (k, b, H), pad their coordinate ranges by 2 percent, and map each coordinate to $[0,1]$. The reduced production surface is

$$G(k,b,H) = \sum_{a=0}^{2}\sum_{c=0}^{2}\sum_{r=0}^{2} \theta^G_{acr} B_{a,2}(\widetilde{k}) B_{c,2}(\widetilde{b}) B_{r,2}(\widetilde{H}). \tag{92}$$

This representation has 27 Bernstein coefficients. Define the Hicks-neutral productivity component and its innovation as

$$\begin{aligned} \omega_i^H &= q_i^\star - G(k_i, b_i, H_i), \\ \xi_i^H &= \omega_i^H - a_H \omega_{i-}^H - b_H(o_{i-} - \bar{o}), \end{aligned} \tag{93}$$

where $a_H \in [0, 0.995]$. The two transition coefficients (a_H, b_H) bring the second-stage parameter count to 29. Ownership is centered and the transition has no intercept; the constant-instrument mean moment sets the empirical Hicks-neutral location convention.

The 42-column second-stage instrument basis consists of a degree-$(2,2,3)$ tensor Bernstein basis in $(k_i, b_i, z_{i-}^{\text{wage}})$, which gives 36 columns including a constant, and six standardized linear terms: $q_{i-}^\star$, k_{i-}, ℓ_{i-}, m_{i-}, χ_{i-}^{ML}, and o_{i-}. We add a scalar ridge to the instrument Gram matrix, choosing the smallest nonnegative value, subject to a numerical floor of 10^{-12}, that caps its condition number at 100,000.

The baseline and proxy-first-stage estimators use the 42 conditional-mean moments of ξ_i^H. The quadratic-input extension retains these moments and adds 451 full-independence moments: 11 centered CDF scores of ξ_i^H, smoothed with the integrated Epanechnikov kernel, interacted with the 41 nonconstant second-stage instruments. The bandwidth is 0.20, and the score probabilities and trapezoidal weights are those in (82). The innovation scale and cutoffs are recomputed from a mean-moment-only second-stage fit of the quadratic-input specification on the same sample and bootstrap draw, at the selected roughness value. The CDF-block weight is increased successively through $0.25, 0.50, 1.00$ and equals one in the reported estimator.

A.6 Second-Stage Shape and Roughness Regularization

All three flexible estimators add a normalized squared-second-difference penalty to the Bernstein coefficients along all three coordinates of G,

$$\lambda_G \mathscr{R}_G(\theta^G), \qquad \lambda_G = 10^{-5}. \tag{94}$$

This common value was selected from the archived roughness comparison and is held fixed in all bootstrap draws. We constrain adjacent Bernstein coefficient differences in the H direction so that

$$G_H(k, b, H) \geq 10^{-4} \tag{95}$$

throughout the padded sieve domain. Since the characteristic construction gives $H_\ell > 0$ and $H_m > 0$, this restriction ensures positive labor and materials output elasticities.

The quadratic-input extension also penalizes negative capital output elasticities. Let $f(k, \ell, m, b) = G(k, b, H(k, b, \ell, m))$ denote the composed production function. Its capital elasticity is

$$\partial_k f(k_{is}, \ell_{is}, m_{is}, b_{is}) = G_k(k_{is}, b_{is}, H_{is}) + G_H(k_{is}, b_{is}, H_{is}) H_{k,is}, \tag{96}$$

where $s \in \{t, t-1\}$ indexes the current or lagged value for observation i. Thus the penalty uses the full derivative of the composed production function, including the response of the flexible-input index to capital. For a candidate parameter vector ϑ, define

$$\mathscr{P}_K(\vartheta) = \frac{1}{2n_B} \sum_{i=1}^{n_B} \sum_{s \in \{t, t-1\}} \tau \log\left[1 + \exp\left\{-\frac{\partial_k f(k_{is}, \ell_{is}, m_{is}, b_{is}; \vartheta)}{\tau}\right\}\right], \tag{97}$$

where n_B is the number of retained observations and $\tau = 0.02$. This smooth penalty does not impose a hard nonnegativity constraint on $\partial_k f$. To scale the penalty grid, let J_0 and $\mathscr{P}_{K,0}$ be the criterion and penalty evaluated at the unpenalized fit on the same sample. A grid value γ_K corresponds to the absolute objective weight

$$w_{\gamma_K} = \gamma_K \frac{J_0}{\mathscr{P}_{K,0}}, \qquad \gamma_K \in \{0, .01, .03, .1, .3, 1, 3, 10\}. \tag{98}$$

The reported quadratic-input specification uses $\gamma_K = 3$, which gives the point-estimate

absolute weight $w_{\gamma_K} = 0.26724802036153944$. This absolute weight is held fixed in all bootstrap draws. The selected penalty substantially improves the capital-positivity diagnostics, although the resulting fit does not satisfy every earlier fit-preservation diagnostic relative to the unpenalized surface. We therefore treat this specification as a distinct, deliberately shape-regularized estimand and interpret its bootstrap distribution as conditional on the selected penalty.

The three restrictions serve different purposes. The hard condition (95) ensures a positive production response to the flexible-input index; the roughness penalty (94) stabilizes the finite Bernstein sieve; and the soft penalty (97) favors positive capital output elasticities without requiring them everywhere.

A.7 Bernstein-Loading CES Refits

For each flexible estimator, the corresponding Bernstein-loading CES refit projects its estimated production surface onto

$$f^{CES}(k,\ell,m,b) = \delta + \frac{\Gamma}{\rho}\log\left[s_K e^{\rho\{k+g_K(b)\}} + s_L e^{\rho\{\ell+g_L(b)\}} + s_M e^{\rho\{m+g_M(b)\}}\right], \tag{99}$$

with the continuous limit at $\rho = 0$. Here and in the direct CES specifications below, ρ denotes the scalar CES curvature parameter, as in the main text, rather than the elasticity-ratio function used above. We set $s_K = \alpha$, $s_M = \beta$, and $s_L = 1 - \alpha - \beta$, with each share bounded below by 0.001. Returns to scale satisfy $\Gamma \in [0.5, 1.5]$, curvature satisfies $\rho \in [-3, 0.95]$, and the elasticity of substitution is $\sigma = 1/(1-\rho)$.

Each factor loading is a degree-two Bernstein polynomial represented by its derivatives at the endpoints of the b domain. Let $[b_L, b_U]$ be the observed domain, b_0 the pooled median of b, $z = b - b_0$, $z_L = b_L - b_0$, and $\Delta = b_U - b_L$. For $j \in \{K, L, M\}$,

$$g_j(b) = d_{jL} z + (d_{jU} - d_{jL})\frac{z^2/2 - z_L z}{\Delta}, \qquad d_{jL}, d_{jU} \in [-2, 2]. \tag{100}$$

This representation imposes $g_j(b_0) = 0$, and $g_j'(b)$ linearly interpolates between the two endpoint derivatives. The loadings are unrestricted in monotonicity, and the CES refit adds no curvature or capital-elasticity penalty.

Let $\widehat{G}_i$ denote the flexible fitted structural log-output level before adding $\widehat{\omega}_i^H$. Assigning

equal total weight to each calendar year, the CES objective is

$$Q_G(\vartheta) = \sum_i w_i \left[\frac{\widehat{G}_i - f_i^{CES}(\vartheta)}{s_G} \right]^2,$$

$$Q_R(\vartheta) = \sum_i w_i \left[\frac{\chi_i^{ML} - \log(s_M/s_L) - \rho\{m_i - \ell_i + g_M(b_i) - g_L(b_i)\}}{s_R} \right]^2,$$

$$Q^{CES}(\vartheta) = Q_G(\vartheta) + \mu_R Q_R(\vartheta), \qquad \mu_R = 1. \tag{101}$$

The block scales are 1.4826 times the pooled median absolute deviation, with an interquartile-range fallback. The ratio block targets observed χ^{ML} through the identity implied by the flexible estimator; individual output elasticities from the flexible surface are not separate fitting targets. The intercept δ is profiled analytically from the level block.

For the baseline CES refit, the point-estimate reference, domain, and scales are

$$b_0 = 1.7740742176, \qquad [b_L, b_U] = [-5.9764938807, 10.4757488583],$$
$$(s_G, s_R) = (1.1345558111, 0.8023724412).$$

For the quadratic-input CES refit, they are

$$b_0 = 1.9731284897, \qquad [b_L, b_U] = [-1.1271144618, 5.2211045353],$$
$$(s_G, s_R) = (1.0499123559, 0.7573562382).$$

For the proxy-first-stage CES refit, they are

$$b_0 = 1.7740742176, \qquad [b_L, b_U] = [-5.9764938807, 10.4757488583],$$
$$(s_G, s_R) = (1.1548684802, 0.8023724412).$$

These normalizations are specific to each refit and do not impose equality across specifications. The numerical search uses 14 starting values: five structured starts, eight deterministic random starts, and one embedded linear-loading start. The non-neutral productivity reference point and domain, target scales, and year weights are recomputed in each bootstrap draw.

A.8 Direct CES Specifications

The direct CES with linear factor-augmenting loadings and the labor-augmenting CES diagnostic impose a three-factor CES technology within the estimation criterion. Both set the two nested curvature parameters equal. With common curvature ρ, the production equation is

$$q_i = \delta + \omega_i^H + \frac{\Gamma}{\rho} \log\left[s_K e^{\rho(k_i + \lambda_K \omega_i^B)} + s_L e^{\rho(\ell_i + \lambda_L \omega_i^B)} + s_M e^{\rho(m_i + \lambda_M \omega_i^B)} \right], \tag{102}$$

with the continuous limit at $\rho = 0$ and $(s_K, s_L, s_M) = (\alpha\beta, 1-\beta, \beta(1-\alpha))$. We estimate $\Gamma \in [0.5, 1.5]$ and $\rho \in [-12, 0.95]$, with $\sigma = 1/(1-\rho)$. The linear factor-augmenting loadings are parameterized as

$$\lambda = a_\lambda \mathbf{1} + \cos(\phi)\frac{(1,-1,0)'}{\sqrt{2}} + \sin(\phi)\frac{(1,1,-2)'}{\sqrt{6}}. \tag{103}$$

The centered loading component has zero sum and unit norm, while the raw loadings need not sum to zero. The direct linear-loading benchmark allows all three loadings to be signed and requires the CES-implied relative-expenditure map to be globally increasing in ω^B:

$$\frac{\partial \chi_i^{ML}}{\partial \omega_i^B} = \rho(\lambda_M - \lambda_L) \geq 10^{-4}. \tag{104}$$

For each candidate parameter vector, we invert the relative-expenditure ratio to recover ω_i^B and use the production equation to recover ω_i^H. Base-year mean normalizations fix the locations of both productivity components, with δ profiled to implement the Hicks-neutral normalization. Each component follows a time-invariant linear transition with a constant, its own lagged value, and lagged ownership. The criterion uses separate annual mean-independence blocks for the non-neutral and Hicks-neutral productivity innovations, equal total weight across years and components, and two-step fixed inverse-Gram weights. The non-neutral productivity instrument vector is

$$(1, k_i, z_{i-}^{\text{wage}}, q_{i-}, k_{i-}, \ell_{i-}, m_{i-}, \chi_{i-}^{ML}, o_{i-})',$$

and the Hicks-neutral productivity instrument vector is

$$(1, \omega_{i-}^H, o_{i-}, k_i, \omega_i^B, \omega_{i-}^B, z_{i-}^{\text{wage}})'.$$

The deliberately restrictive labor-augmenting CES diagnostic imposes $\lambda_K = \lambda_M = 0$ in (102). Under (103), this implies $|\lambda_L| = \sqrt{3/2}$. Both labor-loading orientations are considered subject to (104). Both directly estimated specifications use all 42,612 observations and estimate the CES structure without projecting a flexible surface.

A.9 Whole-Firm Bootstrap

Inference uses 200 whole-firm bootstrap replications. Firms are sampled with replacement, and repeated copies of a sampled firm receive distinct bootstrap identifiers. Each firm's observations are retained together with their previously constructed one-year lags; lags are not rebuilt after resampling.

For the baseline flexible estimator, each draw re-estimates the eight-parameter first stage, recovers ω^B, applies the support screen, and recomputes both current and lagged characteristics. It also rebuilds the 42-column second-stage instrument matrix and its Gram-matrix ridge and re-estimates the production surface using the mean moments. The corresponding CES refit then uses the flexible surface estimated in that draw.

For the quadratic-input extension, each draw re-estimates the regularized 14-parameter first stage, recomputes the inclusive support bounds and whole-firm screen, and rebuilds the quadratic LAD reference and characteristics. It also recomputes the mean-moment-only innovation reference for the full-independence block and re-estimates the production surface with the fixed absolute capital-penalty weight selected at $\gamma_K = 3$. Its CES refit uses the resulting flexible surface. For the proxy-first-stage extension, each draw re-estimates the 105-regressor auxiliary OLS, computes current and lagged fitted output, and re-estimates the baseline ratio first stage, second-stage surface, and corresponding CES refit.

For each direct CES specification, each draw resamples complete firm histories, preserves previously constructed lags, recenters current inputs and lagged ownership, and recomputes the base-year productivity normalizations, instrument standardization, and inverse-Gram weights. It then re-estimates the direct CES criterion. These bootstrap procedures are separate from those for the flexible estimators and their CES refits.

Input centers, support thresholds, LAD coefficients, Bernstein coordinate bounds, instrument transformations and Gram-matrix ridges, and CES non-neutral productivity domains, robust scales, and year weights are recomputed as specified for each estimator. The tuning choices $\lambda_B = 0.001$, $\lambda_G = 10^{-5}$, the score bandwidth and probability grid, the full-independence block weight, and the selected absolute capital-penalty weight are held fixed. The external city-wage series and generated auxiliary-index interaction columns are also held fixed as upstream inputs.

Reported bootstrap standard errors are sample standard deviations across the 200 replications, using the $B-1$ denominator. The archive also retains percentile and basic intervals. The bootstrap propagates uncertainty through each flexible estimator and its corresponding CES refit, and through each direct CES specification, conditional on the fixed tuning choices and upstream inputs. For the quadratic-input extension, inference is also conditional on the shape-regularized estimand selected at $\gamma_K = 3$; it does not account for penalty selection and is not robust to weak identification.

B Additional Derivations

B.1 Relative marginal products and input-ratio responses

This appendix relates the fixed-input bias condition used for identification to the Hicksian response of the materials–labor ratio. Work in logs, fix capital k, Hicks-neutral productivity ω^H, output $\bar{q}$, and the relative input price $p \equiv \log(w^M/w^L)$. To avoid a collision with the CES exponent ρ used below, write

$$\varphi_t(k,\ell,m,\omega) \equiv \log f_m(k,\ell,m,\omega) - \log f_\ell(k,\ell,m,\omega) \tag{105}$$

in this appendix, where ω abbreviates ω^B. The relative first-order condition and the fixed-output condition are

$$\varphi_t(k,\ell,m,\omega) - (m-\ell) = p, \tag{106}$$

$$f_t(k,\ell,m,\omega) = \bar{q} - \omega^H. \tag{107}$$

Differentiating (106)–(107) with respect to ω gives

$$(\varphi_\ell + 1)d\ell + (\varphi_m - 1)dm + \varphi_\omega d\omega = 0, \tag{108}$$

$$f_\ell d\ell + f_m dm + f_\omega d\omega = 0, \tag{109}$$

where all derivatives are evaluated at the same point. Solving this system for $dx/d\omega$, with $x = m - \ell$, yields

$$\frac{dx}{d\omega} = \frac{(f_\ell + f_m)\varphi_\omega - f_\omega(\varphi_\ell + \varphi_m)}{(f_\ell + f_m) + f_m\varphi_\ell - f_\ell\varphi_m}. \tag{110}$$

The denominator is positive under the standard second-order condition for the interior two-input cost-minimization problem. Moreover, differentiating the same system with respect to

the relative input price gives the local Hicksian elasticity of substitution

$$\sigma_{LM,t} \equiv -\left.\frac{\partial(m-\ell)}{\partial \log(w^M/w^L)}\right|_{\bar{q},k,\omega^B} = \frac{f_\ell + f_m}{(f_\ell + f_m) + f_m\varphi_\ell - f_\ell\varphi_m} > 0. \tag{111}$$

Substituting this expression into (110) gives (21), with $\varphi_t = \rho_t$. If the flexible-input tradeoff is locally homothetic conditional on capital, then $\varphi_\ell + \varphi_m = 0$, so the sign of the input-ratio response is the sign of the fixed-input bias φ_ω. Away from homotheticity, the scale-recomposition term can either reinforce or offset the direct bias term.

For the nested-CES example in (17)–(18), define

$$a \equiv \frac{\alpha(AK)^\nu}{\alpha(AK)^\nu + (1-\alpha)M^\nu}, \qquad u \equiv \frac{\beta X^\rho}{\beta X^\rho + (1-\beta)L^\rho}. \tag{112}$$

Here a is the capital share within the capital-materials nest and u is the share of the capital-materials composite in the outer nest. Direct differentiation gives

$$f_\omega = \gamma u a, \qquad f_\ell + f_m = \gamma(1-ua), \qquad \varphi_\ell + \varphi_m = a(\nu-\rho), \qquad \varphi_\omega = a(\rho-\nu), \tag{113}$$

where γ is the returns-to-scale parameter and ρ and ν are the outer and inner CES exponents. In this nested-CES case, the denominator in (110) equals

$$\gamma\left[(1-a)(1-\rho) + a(1-u)(1-\nu)\right], \tag{114}$$

which is strictly positive for $\rho, \nu < 1$. The common factor γ cancels from the Hicksian response. Therefore

$$\left.\frac{d(m-\ell)}{d\omega^B}\right|_{\bar{q},w,k} = \sigma_{LM,t}\frac{a(\rho-\nu)}{1-ua}. \tag{115}$$

Since $1-ua > 0$, the Hicksian response of M/L has the sign of $\rho - \nu$. Under the CES convention $\sigma_{KM} = 1/(1-\nu)$ and $\sigma_{XL} = 1/(1-\rho)$, this response is positive when $\sigma_{KM} < \sigma_{XL}$. Thus a capital-augmenting shock in this nested-CES technology can generate a rising cost-minimizing materials–labor ratio under the maintained fixed-output, fixed-capital, fixed-price comparison.

C Input Prices, Direct Instruments, and Lag Controls

This appendix clarifies what competitive input markets contribute to the identification argument, describes an alternative that instruments all current inputs directly, and explains why the baseline combines excluded price variation with lag controls. The distinction is between price taking, which is part of the firm's optimization problem, and stochastic exogeneity, which is a separate restriction on the joint distribution of prices and productivity.

C.1 Price taking is not price exogeneity

Let

$$P_{jt} \equiv \left(\log w_{jt}^{L}, \log w_{jt}^{M}\right) \tag{116}$$

collect the firm's observed input prices. Perfect competition in an input market makes an atomistic firm a price taker: when choosing an input, the firm treats its price as fixed rather than internalizing an upward-sloping marginal expenditure schedule. Together with homogeneous, correctly measured input quantities, this gives the structural exclusion used in (4)–(5): conditional on quantities and productivity, input prices do not enter the production function or ρ_t directly.

Price taking does not imply

$$P_{jt} \perp\!\!\!\perp \left(\omega_{jt}^{B}, \omega_{jt}^{H}\right). \tag{117}$$

Competitive prices clear markets and may therefore respond to market-wide labor or materials demand. A local productivity shock can raise both firms' input demands and the competitive market price even though no individual firm affects that price. Common industry shocks, spatial sorting, and the choice of input quality can generate the same statistical dependence. Consequently, (117), or an appropriate conditional version, must be justified separately. It is generally more credible for external supply-side price shifters than for firm-level average wages or materials unit values, which may incorporate productivity-related input composition or quality.

There is no formal contradiction in using an input price as an instrument in the relative-expenditure equation even though that price also enters

$$\chi_{jt}^{ML} = \log w_{jt}^{M} - \log w_{jt}^{L} + m_{jt} - \ell_{jt}. \tag{118}$$

At the exact first-order conditions, the structural equation is still $\chi_{jt}^{ML} = \rho_t(X_{jt}^B, \omega_{jt}^B)$, and the price is excluded from ρ_t after conditioning on inputs. This observation does not protect the moment condition against measurement error. If the same mismeasured price is used both to construct χ_{jt}^{ML} and as an instrument, its measurement error enters the left-hand side mechanically. An external price shifter or an independently measured market price is therefore preferable.

C.2 A direct-IV alternative

The productivity dynamics are not logically necessary if sufficiently rich external instruments are available for every endogenous argument. To state this alternative, let Z_{jt}^P contain exogenous labor and materials price shifters, let Z_{jt}^K contain excluded shifters for capital, and let Z_{jt}^ω contain excluded shifters for the recovered non-neutral productivity component in the production equation. Define

$$\widetilde{Z}_{jt}^B \equiv (Z_{jt}^K, Z_{jt}^P), \qquad \widetilde{Z}_{jt}^H \equiv (Z_{jt}^K, Z_{jt}^\omega, Z_{jt}^P). \tag{119}$$

For clarity, observed market and calendar-time controls are suppressed below; conditional analogues require the corresponding conditional rank and mean exclusions.

Because this alternative is period specific and does not link ω^B levels through a transition, it is convenient here to use a different ω^B coordinate. For this proposition only, transform the continuously distributed non-neutral productivity component to its period-specific marginal rank and write the transformed productivity component again as ω^B. Thus $\omega_{jt}^B \sim U(0,1)$ below. This appendix-specific rank scale replaces, rather than supplements, the baseline reference normalization in (28); the production function is expressed in the same rank coordinate.

Proposition C.1 (Direct IV)**.** *Maintain the flexible-input conditions of Lemma 2.1 and the globally increasing orientation supplied by the strengthened Assumption 2.1′. In the period-specific rank coordinate used here, maintain the normalization $\omega_{jt}^B \sim U(0,1)$. Suppose*

$$\widetilde{Z}_{jt}^B \perp\!\!\!\perp \omega_{jt}^B, \tag{120}$$

and suppose the IV-quantile restriction below has a unique solution for ρ_t in the maintained function class. Suppose further that

$$E[\omega_{jt}^H \mid \widetilde{Z}_{jt}^H] = 0, \tag{121}$$

and that the conditional-expectation operator associated with $X_{jt}^H \mid \widetilde{Z}_{jt}^H$ *is injective on the maintained class for* f_t*. Then* ρ_t*,* ω_{jt}^B*,* f_t*, and* ω_{jt}^H *are identified period by period without Assumption 2.2.*

Proof. For each $\tau \in (0,1)$, strict monotonicity and (120) imply

$$\begin{aligned} \Pr\left(\chi_{jt}^{ML} \leq \rho_t(X_{jt}^B, \tau) \mid \widetilde{Z}_{jt}^B\right) &= \Pr\left(\omega_{jt}^B \leq \tau \mid \widetilde{Z}_{jt}^B\right) \\ &= \tau. \end{aligned} \tag{122}$$

The maintained uniqueness condition therefore identifies ρ_t by the nonseparable IV-quantile argument of Chernozhukov and Hansen (2005). Applying the inverse in Lemma 2.1 recovers ω_{jt}^B. The production equation then gives

$$E\left[q_{jt} - f_t(X_{jt}^H) \mid \widetilde{Z}_{jt}^H\right] = 0. \tag{123}$$

Injectivity identifies f_t by the additive NPIV argument of Newey and Powell (2003); the production residual identifies ω_{jt}^H under its location normalization. □

Proposition C.1 makes clear what is required to remove the lags. Input-price shifters naturally move the flexible choices (ℓ_{jt}, m_{jt}), but they need not provide excluded variation for k_{jt}, which is predetermined, or for ω_{jt}^B, which is a productivity component rather than an input demand. If no Z_{jt}^K is available and capital is instead included as its own instrument in the first step, (120) becomes the stronger level restriction

$$(k_{jt}, Z_{jt}^P) \perp\!\!\!\perp \omega_{jt}^B. \tag{124}$$

Serially persistent productivity makes this restriction unattractive: current capital can be chosen before the current innovation while remaining correlated with the current productivity level through past productivity.

Likewise, absent separate instruments for capital and biased productivity in the second step, using them as their own instruments requires

$$E[\omega_{jt}^H \mid k_{jt}, \omega_{jt}^B, Z_{jt}^P] = 0. \tag{125}$$

This condition rules out conditional mean dependence between the levels of the two productivity components and rules out correlation between capital and the Hicks-neutral produc-

tivity level. Perfect competition in the labor and materials markets does not imply either (124) or (125).

Price relevance also remains a substantive condition. In the cost-minimization problem with output fixed, a common proportional change in w_{jt}^L and w_{jt}^M leaves the conditional input mix unchanged; the relative price $\log(w_{jt}^M/w_{jt}^L)$ is the direct shifter of that mix. Hence two recorded prices need not supply two independent dimensions of nonparametric variation. Their common component may affect scale through a separate output decision, but that channel is not supplied by the conditional cost-minimization problem alone. The completeness and support conditions in Proposition C.1 must therefore be assessed rather than inferred from the number of observed prices.

C.3 The lag-controlled hybrid

The baseline instead assigns different identifying roles to excluded prices and lag controls, following the hybrid logic in Ackerberg et al. (2026, Section 5.2). Set $Z_{jt}^{LM} = Z_{jt}^P$, or include the price shifters as components of Z_{jt}^{LM}. In the biased-component step, the relevant exclusion becomes

$$(Z_{jt}^P, k_{jt}, V_{j,t-1}^B) \perp\!\!\!\perp u_{jt}^B. \tag{126}$$

This condition does not require prices or capital to be independent of the level of ω_{jt}^B. Conditional on the inherited non-neutral productivity component summarized by $V_{j,t-1}^B$, it requires independence only from the new rank innovation u_{jt}^B. The price shifters provide excluded variation for current labor and materials, while the timing of capital makes k_{jt} valid relative to that innovation.

In the Hicks-neutral step, substituting the lagged production equation absorbs the predictable component of Hicks-neutral productivity into $a_t^H(V_{j,t-1}^H)$. The relevant moment with price instruments is

$$E[\xi_{jt}^H \mid k_{jt}, \omega_{jt}^B, Z_{jt}^P, V_{j,t-1}^H] = 0. \tag{127}$$

Here the price shifters provide excluded variation for the scalar flexible-input index H_{jt}^R. Under the maintained component-specific exclusion, capital can serve as its own instrument because it is chosen before ξ_{jt}^H. The recovered ω_{jt}^B can serve as its own instrument under the conditional cross-innovation independence in (57). The lags therefore do not replace excluded variation for H_{jt}^R. They change the exclusion target from persistent productivity levels to

current innovations and thereby address the otherwise missing instruments for capital and biased productivity.

The direct and hybrid routes are not ordered without further assumptions. The direct route avoids the controlled Markov restrictions, the initial-period loss, and the high-dimensional lag controls, but it requires valid excluded variation for all endogenous arguments or strong level exogeneity such as (125). The hybrid route permits input prices, capital, and the two productivity levels to share inherited determinants, but it imposes the controlled-Markov sufficiency for the two productivity components, timing, and cross-innovation restrictions used in the baseline. When credible price shifters are observed but separate instruments for capital and ω^B are not, using those shifters in Z_{jt}^{LM} while retaining the lag controls is the natural combination.

D Share-Based Identification under Perfect Competition

This appendix develops a more restrictive but simpler second-step identification argument based on flexible-input revenue shares, in the spirit of Gandhi et al. (2020). The baseline cost-minimization problem identifies the ratio of the two flexible-input elasticities without imposing any restriction on the output market. If firms also choose output optimally in a competitive output market, revenue shares identify the levels of both elasticities. This additional structure eliminates the need for excluded instruments for current labor and materials in the production equation.

Let p_{jt}^Q denote the output price, let $Q_{jt} = \exp(q_{jt})$, and define the labor and materials shares of gross output revenue as

$$S_{jt}^L \equiv \frac{w_{jt}^L L_{jt}}{p_{jt}^Q Q_{jt}}, \qquad S_{jt}^M \equiv \frac{w_{jt}^M M_{jt}}{p_{jt}^Q Q_{jt}}. \tag{128}$$

To connect these shares to the multiplier in (2), define the variable-cost function, holding $(k_{jt}, \omega_{jt}^B, \omega_{jt}^H)$ and input prices fixed,

$$\mathscr{C}_{jt}(\bar{q}) \equiv \min_{L,M} \left\{ w_{jt}^L L + w_{jt}^M M : \bar{q} = f_t(k_{jt}, \log L, \log M, \omega_{jt}^B) + \omega_{jt}^H \right\}. \tag{129}$$

Suppose the conditional cost minimizer is unique, regular, and interior, so that $\mathscr{C}_{jt}$ is differentiable at the firm's output choice. The envelope theorem then gives

$$\frac{\partial \mathscr{C}_{jt}(\bar{q})}{\partial \bar{q}} = \lambda_{jt}. \tag{130}$$

Under perfect competition in the output market, the firm's interior output choice solves

$$\max_{\bar{q}} \left\{ p_{jt}^{Q} \exp(\bar{q}) - \mathscr{C}_{jt}(\bar{q}) \right\}. \tag{131}$$

Its first-order condition and (130) imply

$$\lambda_{jt} = \frac{\partial \mathscr{C}_{jt}(q_{jt})}{\partial q} = p_{jt}^{Q} Q_{jt}. \tag{132}$$

Combining (132) with (4)–(5) yields

$$f_{t,\ell}(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B}) = S_{jt}^{L}, \qquad f_{t,m}(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^{B}) = S_{jt}^{M}. \tag{133}$$

No constant-returns-to-scale restriction is used in this derivation.

Proposition D.1 (Share-based identification under perfect competition)**.** *Suppose Proposition 3.1 holds. For a fixed period t, ρ_t and ω_{jt}^{B} are identified on the covered component. Suppose firms have regular interior profit-maximizing choices, are price takers in the output and flexible-input markets, and output revenue and flexible-input expenditures are correctly measured. Let f_t be twice continuously differentiable, with strictly positive flexible-input elasticities. For every (k, ω^{B}), suppose the conditional support $\mathscr{S}_{\ell m,t}(k, \omega^{B})$ of (ℓ, m) is piecewise-C^1 path connected.*

Then $f_{t,\ell}$ and $f_{t,m}$ are identified on the observed support. On each conditional support, they identify f_t up to an additive function $C_t(k, \omega^{B})$. If, in addition,

$$E\left[\xi_{jt}^{H} \mid k_{jt}, \omega_{jt}^{B}, V_{j,t-1}^{H}\right] = 0 \tag{134}$$

and the analogue of the additive-connectedness condition in Assumption 3.6 holds with X_{jt}^{H} replaced by $(k_{jt}, \omega_{jt}^{B})$, then C_t, f_t, and ω_{jt}^{H} are identified under the location normalization in Assumption 3.1.

Proof. Because ω_{jt}^{B} is identified on the fixed reported scale, (133) identifies both flexible-input derivative functions at every point on the observed support. The two share equations do not, however, provide two restrictions independent of the first step. Indeed,

$$\chi_{jt}^{ML} = \log \frac{S_{jt}^{M}}{S_{jt}^{L}} = \log f_{t,m}(X_{jt}^{H}) - \log f_{t,\ell}(X_{jt}^{H}) = \rho_t(X_{jt}^{H}). \tag{135}$$

This ratio restriction has already been used to identify ρ_t and ω_{jt}^B. Conditional on the identified ratio, one share level is sufficient:

$$f_{t,\ell}(X_{jt}^H) = S_{jt}^L, \qquad f_{t,m}(X_{jt}^H) = \exp\{\rho_t(X_{jt}^H)\} S_{jt}^L. \tag{136}$$

Thus, given the identified ratio, either share level identifies the common scale of both elasticities. The other share equation then supplies a cross-equation restriction in an estimator that does not impose the observed share ratio mechanically.

For each (k, ω^B), select a reference point $\big(\ell_{0t}(k,\omega^B), m_{0t}(k,\omega^B)\big)$ in $\mathscr{S}_{\ell m,t}(k,\omega^B)$. For any (ℓ, m) in the same conditional support, let $\gamma : [0,1] \to \mathscr{S}_{\ell m,t}(k,\omega^B)$ be a piecewise-C^1 path satisfying

$$\gamma(0) = \big(\ell_{0t}(k,\omega^B), m_{0t}(k,\omega^B)\big), \qquad \gamma(1) = (\ell, m),$$

and write its two coordinates as $\gamma_\ell(s)$ and $\gamma_m(s)$. Define

$$\begin{aligned} D_t(k,\ell,m,\omega^B) \equiv \int_0^1 \Big[& f_{t,\ell}\big(k, \gamma_\ell(s), \gamma_m(s), \omega^B\big)\dot{\gamma}_\ell(s) \\ & + f_{t,m}\big(k, \gamma_\ell(s), \gamma_m(s), \omega^B\big)\dot{\gamma}_m(s) \Big]\, ds. \end{aligned} \tag{137}$$

The integrand is the derivative of $f_t(k, \gamma_\ell(s), \gamma_m(s), \omega^B)$ along the path. Hence the fundamental theorem for line integrals gives

$$D_t(k,\ell,m,\omega^B) = f_t(k,\ell,m,\omega^B) - f_t\big(k, \ell_{0t}(k,\omega^B), m_{0t}(k,\omega^B), \omega^B\big)\,. \tag{138}$$

The value is therefore independent of the chosen admissible path. Defining

$$C_t(k,\omega^B) \equiv f_t\big(k, \ell_{0t}(k,\omega^B), m_{0t}(k,\omega^B), \omega^B\big) \tag{139}$$

gives

$$f_t(k,\ell,m,\omega^B) = D_t(k,\ell,m,\omega^B) + C_t(k,\omega^B). \tag{140}$$

For notational convenience, define the structural share functions

$$S_t^L(k,\ell,m,\omega^B) \equiv f_{t,\ell}(k,\ell,m,\omega^B), \qquad S_t^M(k,\ell,m,\omega^B) \equiv f_{t,m}(k,\ell,m,\omega^B). \tag{141}$$

On any region where $\mathscr{S}_{\ell m,t}(k,\omega^B)$ has nonempty interior, the two share functions must also

satisfy the local integrability restriction

$$\frac{\partial S_t^L(k,\ell,m,\omega^B)}{\partial m} = \frac{\partial S_t^M(k,\ell,m,\omega^B)}{\partial \ell}. \tag{142}$$

Together with zero circulation around admissible closed paths, this provides an overidentifying restriction on independently estimated share equations. On locally full-dimensional support and under the corresponding basis-rank condition, all parametric or sieve coefficients attached to terms that vary with ℓ or m, including interactions with k and ω^B, are identified from the shares. If the conditional support is lower dimensional, (138) still identifies production differences along admissible paths, but it does not identify an unrestricted extension of f_t away from that support or coefficients that require transverse variation.

Define transformed output by

$$\widetilde{q}_{jt} \equiv q_{jt} - D_t(k_{jt}, \ell_{jt}, m_{jt}, \omega_{jt}^B). \tag{143}$$

Equations (1), (27), and (140) imply

$$\widetilde{q}_{jt} = C_t(k_{jt}, \omega_{jt}^B) + a_t^H(V_{j,t-1}^H) + \xi_{jt}^H, \tag{144}$$

where a_t^H is the lag-control function defined in (51). Under (134),

$$E\big[\widetilde{q}_{jt} \mid k_{jt}, \omega_{jt}^B, V_{j,t-1}^H\big] = C_t(k_{jt}, \omega_{jt}^B) + a_t^H(V_{j,t-1}^H). \tag{145}$$

Additive connectedness identifies the two functions up to a period-specific constant. The normalization $E[\omega_{jt}^H] = 0$, together with $E[\xi_{jt}^H] = 0$, sets $E[a_t^H(V_{j,t-1}^H)] = 0$ and fixes this constant. Hence C_t, f_t, and $\omega_{jt}^H = q_{jt} - f_t(X_{jt}^H)$ are identified. □

Condition (134) follows from the baseline conditional cross-innovation independence in (57) and the component-specific exclusion used to obtain (58). In particular, the law of iterated expectations integrates Z_{jt}^{LM} out of (58). Thus k_{jt} and ω_{jt}^B can serve as their own instruments, and no second-step excluded instruments are required for ℓ_{jt} or m_{jt}. If conditional cross-innovation independence is not maintained, current ω_{jt}^B may remain correlated with ξ_{jt}^H and requires excluded variation or another control strategy.

The logic parallels Gandhi et al. (2020, Theorems 2–3). Their main analysis treats materials as the single flexible input and capital and labor as predetermined. The material revenue share identifies the material elasticity, which is integrated to recover the produc-

tion function up to a function of the remaining inputs; productivity dynamics and support variation then identify this constant of integration. They state the corresponding multiple-flexible-input extension in their Appendix O6. Their model also contains an ex-post shock ε_{jt} realized after input choices. If $\mathscr{E} \equiv E[\exp(\varepsilon_{jt})]$, their log-share equation contains $\log \mathscr{E} - \varepsilon_{jt}$, which they recover through the share regression before identifying the elasticity. Because the baseline model here excludes an ex-post shock, (133) holds exactly. The additional first-step inversion for ω^B changes the constant of integration from their function of nonflexible inputs to $C_t(k, \omega^B)$.

Finally, the share-level argument is more restrictive than the relative-share inversion. With an output markup μ_{jt} and otherwise undistorted flexible-input first-order conditions,

$$f_{t,\ell}(X_{jt}^H) = \mu_{jt} S_{jt}^L, \qquad f_{t,m}(X_{jt}^H) = \mu_{jt} S_{jt}^M. \tag{146}$$

The common markup cancels from the ratio and therefore does not invalidate the first-step identification of ω^B, but an unknown markup prevents the share levels from identifying the scale of the two elasticities. Measurement error in revenue or expenditures, optimization error, adjustment costs, and input-specific wedges likewise break the exact equalities in (133) and require an explicit error or wedge specification.

E Technical Details for the Elasticity-Ratio Index

Section 3.2 uses the elasticity-ratio field to reduce the flexible-input block in the second identification step. This appendix supplies the formal characteristic-flow argument and records its global continuation qualifications. Unlike the share-level argument in Appendix D, this construction does not require price-taking behavior in the output market. Write $z \equiv (k, \omega^B)$. On the supported domain, define

$$R_t(z, \ell, m) \equiv \exp\{\rho_t(z, \ell, m)\} = \frac{f_{t,m}(z, \ell, m)}{f_{t,\ell}(z, \ell, m)}. \tag{147}$$

Equation (146) shows that a common output markup cancels from this ratio even though it prevents either derivative level from being recovered. Thus R_t identifies the direction of the flexible-input gradient, or equivalently the conditional isoquants, but not the gradient's scale.

Assumption 3.4 embeds this identified field in the specified joint continuation (R_t^c, f_t^c) on $\mathscr{D}_t^c$, where $R_t^c = R_t$ and $f_t^c = f_t$ on the supported domain and $R_t^c = f_{t,m}^c / f_{t,\ell}^c$ throughout the continuation domain. This joint restriction matters only when a characteristic passes through

empirically unsupported points: extending slopes alone would not determine compatible output levels across the resulting path.

The ratio field can be converted into a scalar label for these isoquants. Under Assumption 3.4, the reference rule $m_{0t}(z)$ is fixed as part of the normalization. Every point in the target characteristic strip $\mathcal{U}_t^c$ lies on a unique solution to

$$\frac{\partial \Lambda_t(s; z, a)}{\partial s} = -R_t^c(z, \Lambda_t(s; z, a), s)\,, \qquad \Lambda_t(m_{0t}(z); z, a) = a. \tag{148}$$

Each solution intersects the fixed reference section exactly once. That section is

$$m = m_{0t}(z).$$

This characteristic-strip condition is global; the same construction is available locally under the stated smoothness conditions. For $(z, \ell, m) \in \mathcal{U}_t^c$, define

$$h_{R,t}(z, \ell, m) = a \quad \Longleftrightarrow \quad \Lambda_t(m; z, a) = \ell. \tag{149}$$

The normalization is $h_{R,t}(z, \ell, m_{0t}(z)) = \ell$: the index is the labor coordinate at which the characteristic through (ℓ, m) crosses the reference section.

Proposition E.1 (Ratio-index equivalence)**.** *Suppose Proposition 3.1 and Assumption 3.4 hold. Then the normalized $h_{R,t}$ is identified on supported characteristic components of $\mathcal{U}_t^c$ and is defined elsewhere on that strip conditional on the maintained continuation. It satisfies*

$$\frac{\partial_m h_{R,t}(z, \ell, m)}{\partial_\ell h_{R,t}(z, \ell, m)} = R_t^c(z, \ell, m). \tag{150}$$

Moreover, on $\mathcal{U}_t^c$, any continuously differentiable structural continuation with $f_{t,\ell}^c > 0$ satisfies

$$\frac{f_{t,m}^c}{f_{t,\ell}^c} = R_t^c \quad \Longleftrightarrow \quad f_t^c(z, \ell, m) = G_t^c(z, h_{R,t}(z, \ell, m)) \qquad \text{for } (z, \ell, m) \in \mathcal{U}_t^c \tag{151}$$

for an otherwise unrestricted continuously differentiable function G_t^c with $\partial_h G_t^c > 0$. A supported component here means that the complete path and its reference intersection lie in the support of the identified R. Conditional on the specified continuation, the representation imposes no additional restriction beyond the continued ratio relation and the reference normalization.

Proof. Differentiating the characteristic flow with respect to its initial label gives

$$\partial_a \Lambda_t(s; z, a) = \exp\left\{-\int_{m_{0t}(z)}^{s} \partial_\ell R_t^c(z, \Lambda_t(u; z, a), u)\, du\right\} > 0. \tag{152}$$

The flow is therefore invertible in a. Implicitly differentiating $\Lambda_t(m; z, h_{R,t}(z, \ell, m)) = \ell$ yields

$$\partial_\ell h_{R,t} = \frac{1}{\partial_a \Lambda_t}, \qquad \partial_m h_{R,t} = \frac{R_t^c}{\partial_a \Lambda_t}, \tag{153}$$

which proves (150).

For necessity, suppose $f_{t,m}^c = R_t^c f_{t,\ell}^c$. Along any characteristic,

$$\frac{d}{ds} f_t^c(z, \Lambda_t(s; z, a), s) = -R_t^c f_{t,\ell}^c + f_{t,m}^c = 0. \tag{154}$$

Hence the production function is constant along the same characteristics as $h_{R,t}$. Defining

$$G_t^c(z, a) \equiv f_t^c(z, a, m_{0t}(z)) \tag{155}$$

gives (151). Conversely, applying the chain rule to that representation gives $f_{t,m}^c / f_{t,\ell}^c = \partial_m h_{R,t} / \partial_\ell h_{R,t} = R_t^c$. □

The reference-section normalization above uniquely fixes $h_{R,t}$ on the characteristic strip. Without that normalization, the representation is invariant to regular increasing relabelings. In particular, if $\widetilde{h}_{R,t} = H_t(z, h_{R,t})$ with $\partial_h H_t > 0$, then the same production class is obtained after replacing $G_t(z, h)$ by $G_t(z, H_t^{-1}(z, h))$. This invariance is innocuous only when G_t is unrestricted in both arguments. Requiring an index that does not depend on z, or imposing additivity, linearity, a common coefficient, or another parametric restriction on G_t, adds structure beyond the ratio restriction. With two flexible inputs, no further local integrability condition on a smooth finite R_t^c is required. Globally, however, joining disconnected supported segments through an unsupported region requires the joint continuation in Assumption 3.4. The resulting off-support index and level comparisons are therefore maintained extrapolations, not nonparametrically identified counterfactuals.

The representation also clarifies what the ratio contributes to the second step. Define

$$z_{jt} \equiv (k_{jt}, \omega_{jt}^B), \qquad H_{jt}^R \equiv h_{R,t}(z_{jt}, \ell_{jt}, m_{jt}). \tag{156}$$

The production equation becomes

$$q_{jt} = G_t(k_{jt}, \omega^B_{jt}, H^R_{jt}) + a^H_t(V^H_{j,t-1}) + \xi^H_{jt}. \tag{157}$$

Thus the exact ratio restriction replaces the two-dimensional endogenous block (ℓ, m) by one endogenous composite H^R. Assumption 3.6 accordingly imposes injectivity over the smaller class of square-integrable additive differences

$$\mathscr{D}^R_t = \left\{\delta^R_t(z,h,v) = d_{G,t}(z,h) + d_{a,t}(v)\right\}, \tag{158}$$

A sufficiently rich excluded shifter for the scalar H^R may therefore support a weaker completeness condition than instruments for unrestricted movements in both flexible inputs. This dimensional reduction is not by itself an identification result: H^R_{jt} is chosen using current labor and materials and generally remains correlated with ξ^H_{jt}. Valid excluded variation and completeness are still required for this scalar direction. Finally, the ratio equation has already been used to recover ω^B in the first step. The index representation is therefore an exact maintained restriction on the second-step function class, not an additional moment.

F Correlated Productivity Innovations

The baseline uses Assumption 3.5 to make the identified non-neutral productivity component valid in the second-step instrument set. This appendix allows the two current productivity innovations to have unrestricted conditional-mean dependence. The period-specific second-step conditional-moment system developed below leaves a centered ω^B-specific location shift unresolved between the production function and Hicks-neutral productivity. It nevertheless identifies all three input elasticities and the pairwise bias of technical change. This is a reduced-moment result: jointly imposing the separate-Markov restrictions across periods may further restrict this location ambiguity under additional support and dynamic-separation conditions. Remark F.1 makes this distinction explicit.

Write $b_{jt} \equiv \omega^B_{jt}$ for the non-neutral productivity component on the reported reference scale. This component need not have a uniform marginal distribution. In the baseline model, the innovation rank u^B_{jt}, rather than b_{jt}, is Uniform$(0,1)$. Let $V_{j,t-1} \equiv V^H_{j,t-1}$ be the lag-only vector in (49). The inherited non-neutral productivity component and its transition control are measurable with respect to $V_{j,t-1}$. After the first step, b_{jt} and the normalized ratio index

$h_{R,t}$ are identified. Equation (54) gives

$$q_{jt} = G_t(k_{jt}, b_{jt}, H^R_{jt}) + a^H_t(V_{j,t-1}) + \xi^H_{jt}. \tag{159}$$

F.1 Conditional-Mean Residualization

Project the Hicks-neutral innovation directly on the lag controls and the identified non-neutral productivity component:

$$\lambda_t(V, b) \equiv E[\xi^H_{jt} \mid V_{j,t-1} = V, b_{jt} = b], \tag{160}$$

and define

$$\varepsilon^H_{jt} \equiv \xi^H_{jt} - \lambda_t(V_{j,t-1}, b_{jt}). \tag{161}$$

By construction,

$$E[\varepsilon^H_{jt} \mid V_{j,t-1}, b_{jt}] = 0. \tag{162}$$

This step permits unrestricted conditional-mean dependence between u^B_{jt} and ξ^H_{jt}. It asserts neither full independence nor independence of higher conditional moments.

Residualization does not establish validity of the remaining instruments. Let

$$W_{jt} \equiv (k_{jt}, Z^{LM}_{jt}) \tag{163}$$

and impose the separate exclusion

$$E[\xi^H_{jt} \mid W_{jt}, V_{j,t-1}, b_{jt}] = E[\xi^H_{jt} \mid V_{j,t-1}, b_{jt}]. \tag{164}$$

Equations (161) and (164) imply

$$E[\varepsilon^H_{jt} \mid W_{jt}, V_{j,t-1}, b_{jt}] = 0. \tag{165}$$

Thus no excluded instrument is required for b_{jt} itself. The exclusion in (164) still requires the instruments for current inputs to contain no additional information about ξ^H_{jt} once $(V_{j,t-1}, b_{jt})$ is fixed.

Define the combined predictable component as

$$A_t(V, b) \equiv a_t^H(V) + \lambda_t(V, b). \tag{166}$$

Then

$$q_{jt} = G_t(k_{jt}, b_{jt}, H_{jt}^R) + A_t(V_{j,t-1}, b_{jt}) + \varepsilon_{jt}^H. \tag{167}$$

The current-input arguments (k, H^R) appear only in G_t, whereas both G_t and A_t depend on b. This distinction determines what remains identified.

F.2 Identification of Input Elasticities

Proposition F.1 (Input elasticities with correlated innovations)**.** *Suppose Proposition 3.1 and Assumption 3.4 hold, so that b_{jt} and $h_{R,t}$ are identified under their stated normalizations. Suppose $G_t(k, b, h)$ and $h_{R,t}(k, b, \ell, m)$ are continuously differentiable on open neighborhoods of the relevant support, ξ_{jt}^H is integrable, and (164) holds.*

Let $\mathscr{D}_t^{\partial}$ be the class of square-integrable additive differences

$$\mathscr{D}_t^{\partial} = \{\delta_t(k, b, h, v) = d_{G,t}(k, b, h) + d_{A,t}(v, b)\}. \tag{168}$$

Assume the conditional-expectation operator is injective on this class:

$$\begin{aligned} &E[\delta_t(k_{jt}, b_{jt}, H_{jt}^R, V_{j,t-1}) \mid W_{jt}, V_{j,t-1}, b_{jt}] = 0 \\ \implies \quad &\delta_t(k_{jt}, b_{jt}, H_{jt}^R, V_{j,t-1}) = 0 \quad a.s. \end{aligned} \tag{169}$$

Finally, conditional on every b in the relevant support, suppose the support of (k, H^R, V) is additively connected and contains the open neighborhoods needed to identify derivatives with respect to (k, H^R).

Then the period-t conditional-moment system identifies the additive composite in (167). Its components are identified only up to a function of b:

$$G_t^*(k, b, h) = G_t(k, b, h) + c_t(b), \qquad A_t^*(v, b) = A_t(v, b) - c_t(b). \tag{170}$$

If the alternative decomposition is also required to satisfy the location normalization $E[\omega_{jt}^H] =$

0, *the admissible shifts are restricted to the centered class*

$$E[c_t(b_{jt})] = 0.$$

Here the expectation is with respect to the period-t distribution of the reference-scale non-neutral productivity component. Nevertheless, $G_{t,k}$ *and* $G_{t,h}$ *are uniquely identified, and therefore* $f_{t,k}$*,* $f_{t,\ell}$*, and* $f_{t,m}$ *are identified at interior points.*

Proof. Equation (165) gives the conditional moment

$$E\big[q_{jt} - G_t(k_{jt}, b_{jt}, H^R_{jt}) - A_t(V_{j,t-1}, b_{jt}) \,\big|\, W_{jt}, V_{j,t-1}, b_{jt}\big] = 0. \tag{171}$$

Injectivity in (169) identifies the additive composite. If (G_t, A_t) and $(\widetilde{G}_t, \widetilde{A}_t)$ generate the same composite, their differences satisfy

$$d_{G,t}(k, b, h) + d_{A,t}(v, b) = 0.$$

Fix b. At a common v, this equality makes $d_{G,t}$ equal at any two admissible (k, h) points. Conditional additive connectedness extends the equality throughout that slice, so

$$d_{G,t}(k, b, h) = c_t(b), \qquad d_{A,t}(v, b) = -c_t(b).$$

This proves (170). Moreover, (160), (166), and the Hicks-neutral transition imply

$$A_t(V_{j,t-1}, b_{jt}) = E[\omega^H_{jt} \mid V_{j,t-1}, b_{jt}].$$

Hence $E[\omega^H_{jt}] = 0$ implies $E[A_t(V_{j,t-1}, b_{jt})] = 0$. Requiring the same normalization for the alternative decomposition and using $A^*_t = A_t - c_t$ gives $E[c_t(b_{jt})] = 0$. No uniformity restriction is imposed on the reported reference-scale non-neutral productivity component b_{jt}; only the innovation rank u^B_{jt} is Uniform$(0, 1)$. Because $c_t(b)$ does not depend on k or h, $G_{t,k}$ and $G_{t,h}$ are invariant across all admissible decompositions.

Applying the chain rule to (43) gives

$$\begin{aligned} f_{t,\ell} &= G_{t,h}\, \partial_\ell h_{R,t}, & f_{t,m} &= G_{t,h}\, \partial_m h_{R,t}, \\ f_{t,k} &= G_{t,k} + G_{t,h}\, \partial_k h_{R,t}. \end{aligned} \tag{172}$$

Every object on the right-hand side is identified. □

Remark F.1 (Reduced moments versus the full dynamics)**.** Proposition F.1 concerns the period-t conditional moment (171). It does not assert that every centered c_t in (170) generates an alternative process satisfying the full separate-Markov restriction. The ratio-index representation maps this shift into

$$f_t^*(k,\ell,m,b) = f_t(k,\ell,m,b) + c_t(b), \qquad \omega_{jt}^{H*} = \omega_{jt}^H - c_t(b_{jt}).$$

Define the biased-productivity transition operator

$$(\mathscr{P}_{t+1}^B c)(b,o^B) \equiv E[c(b_{j,t+1}) \mid b_{jt} = b, o_{jt}^B = o^B] = \int_0^1 c\big(H_{t+1}^B(b,o^B,u)\big)\,du. \tag{173}$$

For the transformed non-neutral productivity component to satisfy even the conditional-mean part of the Hicks-neutral transition at $t+1$, there must exist a function g_{t+1}^{H*} such that

$$g_{t+1}^{H*}(x,o^H) = g_{t+1}^H\big(x + c_t(b), o^H\big) - (\mathscr{P}_{t+1}^B c_{t+1})(b,o^B) \tag{174}$$

jointly on the relevant support, and the left-hand side cannot depend on the remainder of the history. Suppose the support permits connected comparisons across values of b: for the compared b and admissible o^B values, (x,o^H) has a common nondegenerate x-interval at a common o^H. If, on that interval,

$$x \mapsto g_{t+1}^H(x+d,o^H) - g_{t+1}^H(x,o^H)$$

is nonconstant for every $d \neq 0$, then the right-hand side of (174) can be invariant to b only if c_t is constant on each connected b-support. Centering then gives $c_t \equiv 0$. Thus a nonlinear transition satisfying this no-translation-symmetry condition can restore level identification for an interior period.

By contrast, if

$$g_{t+1}^H(x,o^H) = r_{t+1}(o^H) + \beta_{t+1}x,$$

mean compatibility reduces to the operator recursion

$$(\mathscr{P}_{t+1}^B c_{t+1})(b,o^B) = \beta_{t+1}c_t(b) + \kappa_{t+1},$$

where centering implies $\kappa_{t+1} = 0$. Nonzero shifts can therefore survive the conditional-mean

restriction when this recursion has solutions. Satisfying (174) is not sufficient for the full model: Assumption 2.2 additionally requires

$$\mathscr{L}\left(\omega_{j,t+1}^{H} - c_{t+1}(b_{j,t+1}) \,\middle|\, \mathscr{I}_{j,t}\right)$$

to depend on $\mathscr{I}_{j,t}$ only through $(\omega_{jt}^{H} - c_t(b_{jt}), o_{jt}^{H})$. This requirement may fail through other elements of the history even when the conditional-mean equation holds.

There is no outgoing transition after the terminal period. Once $c_{T-1} = 0$, injectivity of $\mathscr{P}_T^B$ on centered functions is sufficient to eliminate c_T from the incoming conditional-mean restriction. Null-space shifts are not ruled out by that mean restriction alone; they are full-model alternatives only if the transformed conditional law also obeys the separate-Markov restriction. The maintained assumptions admit special transition laws with nonzero dynamically admissible sequences, so they do not by themselves establish level identification. This appendix does not impose or exploit the additional support, no-translation-symmetry, or transition-operator and full-kernel separation conditions needed to use these cross-period restrictions. Its identification claim is therefore relative to the period-specific second-step conditional-moment system.

The instruments must still generate relevant variation in the current directions k_{jt} and H_{jt}^{R} after conditioning on the lag controls and b_{jt}. The ratio restriction reduces the flexible-input block to one scalar, but it does not create the required completeness. Capital likewise requires valid own or excluded variation.

F.3 Identification of Technological Bias

Because inputs are measured in logs, output elasticities are the derivatives of f_t with respect to $\log K = k$, $\log L = \ell$, and $\log M = m$. At fixed inputs, define pairwise technological bias by

$$\mathscr{B}_t^{r:s}(k,\ell,m,b) \equiv \partial_b \log\left(\frac{\mathrm{MP}_{r,t}}{\mathrm{MP}_{s,t}}\right) = \partial_b \log\left(\frac{\partial f_t/\partial \log r}{\partial f_t/\partial \log s}\right), \qquad r,s \in \{K,L,M\}. \tag{175}$$

For materials relative to labor, the first step gives directly

$$\mathscr{B}_t^{M:L}(k,\ell,m,b) = \partial_b \log R_t(k,b,\ell,m). \tag{176}$$

This conclusion does not require second-stage cross-derivatives. If the conditions of Proposition F.1 hold jointly on a common open set as b varies, and if G_t and $h_{R,t}$ are twice

continuously differentiable there, then the b derivatives of the identified input elasticities are identified. Hence the pairwise bias measures involving capital are identified wherever the corresponding elasticities are positive.

Cardinal magnitudes in (175) depend on the first-step scale normalization for b. Under an increasing relabeling $\widetilde{b} = \phi_t(b)$,

$$\partial_{\widetilde{b}} \log \left(\frac{\partial f_t / \partial \log r}{\partial f_t / \partial \log s} \right) = \frac{1}{\phi_t'(b)} \partial_b \log \left(\frac{\partial f_t / \partial \log r}{\partial f_t / \partial \log s} \right). \tag{177}$$

The sign of the bias and the ordering of firms' productivity components are invariant to such a relabeling.

The limitation of the period-specific second-step conditional-moment system is a biased-productivity-specific location ambiguity. Under (170),

$$\begin{aligned} f_t^*(k, \ell, m, b) &= f_t(k, \ell, m, b) + c_t(b), \\ f_{t,b}^*(k, \ell, m, b) &= f_{t,b}(k, \ell, m, b) + c_t'(b). \end{aligned} \tag{178}$$

Thus, within this moment system and subject to $E[c_t(b_{jt})] = 0$, the b-specific level of f_t, its derivative $f_{t,b}$, the separate components G_t and A_t, the decomposition $A_t = a_t^H + \lambda_t$, and

$$\omega_{jt}^H = q_{jt} - f_t(k_{jt}, \ell_{jt}, m_{jt}, b_{jt})$$

are not separately identified. The extension does identify how biased productivity changes input elasticities and relative marginal products. Remark F.1 explains that jointly imposing the full cross-period dynamics may further restrict, and under additional conditions eliminate, the remaining location ambiguity.