Technology-Type-Heterogeneous Elasticities in Gross-Output Production Functions

Justin Doty[1] & Suyong Song[2]

[1]The MathWorks, Inc., Natick, MA, USA*

[2]Department of Economics & Finance, University of Iowa, Iowa City, IA, USA†

September 9, 2026

Abstract

We propose a new approach to nonparametrically identify and estimate production functions when elasticities are heterogeneous across firms and vary jointly with every input. We assume an unobserved, firm-specific technology type that enters the production function nonseparably alongside capital, labor, and materials. We show that under mild conditional independence restrictions on a two-period repeated-measurement system, the firm's flexible-input elasticity and its position in the type distribution are nonparametrically identified using a Kotlarski-type conditional deconvolution of the materials cost-share equation. The remaining elasticities can be recovered in the second stage, similar to Gandhi et al. (2020), where the fitted technology-types enter as an additional state variable to be conditioned on in a sieve-GMM procedure.

Keywords: Production functions, heterogeneous elasticities, unobserved heterogeneity

JEL Classification: C14, C33, D24

1 Introduction

Production function estimation is a long-standing empirical problem: it links firms' input choices to their output. Identification of the output elasticities and consequently the distribution of firm-level productivity is constrained by endogeneity issues. This is because

*Contact: jdoty@mathworks.com

†Contact: suyong-song@uiowa.edu

productivity is unobserved by the econometrician, but observed by the firm when making input decisions.

A popular approach to address this issue is to introduce a proxy variable for unobserved productivity, such as investment (Olley and Pakes (1996), henceforth OP) or an intermediate input (Levinsohn and Petrin (2003), henceforth LP; Ackerberg et al. (2015), henceforth ACF). These proxies are functions of state variables such as capital and unobserved productivity. These functions are assumed to be strictly increasing in their scalar unobserved productivity component. Inverting them controls for unobserved productivity, and the production function parameters can be estimated using a simple two-stage approach.

The standard Cobb-Douglas production function used in these procedures restricts the unobserved heterogeneity between firms to be additively separable from inputs. There has been substantial effort dedicated to extending these models to allow for unobserved heterogeneity in output elasticities. Recent evidence has found that common functional form assumptions, such as additive heterogeneity, are not compatible with the data. For example, Raval (2019) finds substantial variation in capital shares in the ready-mixed concrete industry, persistent across time and correlated with firm size. Heterogeneous coefficients are one way to reconcile these empirical findings by allowing for firm-specific production functions. The literature on heterogeneous production functions is small relative to the empirical research using the homogeneous coefficient model, even though many recent empirical studies have found heterogeneity in firm behavior and decisions.[1] This is because estimating the homogeneous coefficient model by itself is very difficult due to the issue of unobserved productivity and simultaneity bias.

An important downstream consequence of heterogeneous production function elasticities is an attenuated estimate of markup dispersion across firms. Flexible input elasticities and cost-share estimates are two objects required for the production-function-based approach to markup estimation. If elasticities are homogeneous, this suggests markups are invariant to firm-specific differences, which contradicts what the empirical literature finds on the cross-sectional dispersion and evolution of the markup distribution. Recent papers attribute changes in the markup distribution to labor-augmenting productivity (Demirer, 2020). Factor-augmenting productivity is a restriction on how heterogeneity is allowed to enter the production function: it rescales one input's effective quantity by a firm-specific shifter, leaving the underlying elasticity function common across firms. Therefore, hetero-

[1]Some notable examples are Kasahara et al. (2017), Balat et al. (2019), and Li and Sasaki (2017), to name a few. Gandhi et al. (2020) also estimate a nonparametric production function and obtain heterogeneous estimates.

geneity shows up only in the augmented input's own effective units and not in the other inputs' elasticities.

This paper offers an alternative, more general explanation. We allow an unobserved technology type to enter the production function nonseparably, so that the elasticities of materials, capital, and labor can all vary jointly with the same firm-specific type, rather than restricting heterogeneity to operate through a single input's effective units. Consistent with this distinction, our empirical results (Section 6) show all three elasticities moving systematically with the recovered technology type, a pattern a purely factor-augmenting specification cannot generate.

We show that the firm-level elasticity of output with respect to the flexible input, the distribution of the ex-post production shock, and the firm's rank in the distribution of an unobserved technology type are nonparametrically identified using a conditional deconvolution argument applied to a repeated measurement of the flexible input's cost share. This holds under conditional independence of the unobserved type across periods and mild regularity conditions on the relevant conditional characteristic functions. These conditions are the panel analogue of those used to identify measurement-error and random-coefficient models by conditional deconvolution, such as Hu et al. (2020) and Li and Vuong (1998). The conditional characteristic function of the bundled elasticity-and-shock term is identified in closed form via Kotlarski's identity, up to a location normalization; the ex-post shock's own characteristic function is identified as well, and hence its full distribution follows immediately. Identification of these characteristic functions gives identification of the corresponding conditional distribution functions by Fourier inversion, and hence identification of the flexible-input elasticity surface at every quantile of the technology-type distribution, due to the inverse relationship between quantiles and the CDF.

We propose a two-stage estimator motivated by this identification strategy. The first stage recovers each firm's ex-post shock and technology type from the deconvolution argument, and the second stage recovers the remaining (capital and labor) elasticities, together with productivity, using a sieve-GMM procedure resembling Gandhi et al. (2020)'s second stage. In our empirical application, we consider a widely used firm-level manufacturing dataset and estimate the full nonseparable production surface, including the materials elasticity across the recovered technology-type distribution and the capital and labor elasticities at each firm's own realized state. We show that heterogeneity in these estimates is systematic. Materials elasticities rise with a firm's recovered technology type in every industry we examine, while capital and labor elasticities tend to fall. This pattern has direct implications for how markups and productivity dispersion are measured.

The rest of the paper is organized as follows. Section 2 reviews the GNR procedure for estimating a gross-output production function and shows why it does not allow elasticities to vary with an unobserved technology type. Section 3 introduces the nonseparable production function and the technology type. Section 4 presents conditions under which the flexible-input elasticity, the ex-post shock, and the technology type are identified. Section 5 proposes a two-stage estimator and discusses its properties. Section 6 applies the two-stage estimator of Section 5 to a U.S. manufacturing dataset. Section 7 concludes with directions for future research.

2 Literature Review

We briefly review the procedure of Gandhi et al. (2020) for estimating a *gross-output* production function (in logs):

$$y_{it} = f_t(k_{it}, l_{it}, m_{it}) + \omega_{it} + \varepsilon_{it}, \tag{1}$$

where y_{it} denotes gross output for firm i at time t, k_{it} and l_{it} denote capital and labor, m_{it} is a flexible (variable) input such as materials, ω_{it} is Hicks-neutral productivity known to the firm when m_{it} is chosen, and ε_{it} denotes an ex-post shock to production realized only after m_{it} is chosen.

Gandhi et al. (2020) (henceforth GNR) show that applying the standard proxy-variable procedure of OP, LP, or ACF directly to a gross-output specification like this one runs into a fundamental identification problem. The flexible input's own optimal demand function, $m_{it} = m_{it}(k_{it}, l_{it}, \omega_{it})$, can be inverted to write ω_{it} as a function of the same state variables (k_{it}, l_{it}, m_{it}) that the first-stage proxy regression conditions on. Substituting this inverted function back into the production function collapses f_t and the (unknown) inverted productivity function into a single reduced-form function of (k_{it}, l_{it}, m_{it}) alone, so the materials elasticity cannot be separately recovered from the reduced form using only the information available in a gross-output first stage.

GNR's solution instead identifies the flexible input's own output elasticity directly from its static first-order condition, rather than from a proxy-variable inversion. Let

$$D_t(k_{it}, l_{it}, m_{it}) \equiv \frac{\partial f_t(k_{it}, l_{it}, m_{it})}{\partial m_{it}} \tag{2}$$

denote the output elasticity of materials. Because m_{it} is chosen to maximize expected profit after $(k_{it}, l_{it}, \omega_{it})$ are known, but before ε_{it} is realized, the firm's static first-order condition for m_{it} gives

$$s_{it} \equiv \ln(S_{it}) = \ln\big(D_t(k_{it}, l_{it}, m_{it})\big) + \ln \bar{\varepsilon} - \varepsilon_{it}, \qquad S_{it} \equiv \frac{P_{it}^M M_{it}}{P_{it} Y_{it}}, \qquad \bar{\varepsilon} \equiv \mathbb{E}[e^{\varepsilon_{it}}], \tag{3}$$

so that a nonparametric regression of the observed log materials share s_{it} on (k_{it}, l_{it}, m_{it}) identifies D_t directly. Given $\hat{D}_t$, integrating over materials while holding (k_{it}, l_{it}) fixed recovers f_t up to a constant of integration $C_t(k_{it}, l_{it})$: purging output of the identified materials contribution, $\check{y}_{it} \equiv y_{it} - \int \hat{D}_t(k_{it}, l_{it}, \sigma)\, d\ln\sigma = C_t(k_{it}, l_{it}) + \omega_{it}$. Approximating $C_t(k, l; \theta)$ with a sieve in (k, l), a trial θ inverts for trial productivity $\omega_{it}(\theta) = \check{y}_{it} - C_t(k_{it}, l_{it}; \theta)$; assuming productivity follows a Markov process and projecting $\omega_{it}(\theta)$ onto its own lag gives an innovation $\xi_{it}(\theta)$, and θ is estimated by GMM using the timing-based instruments k_{it} and l_{it-1}:

$$\mathbb{E}\big[\xi_{it}(\theta)\, k_{it}^j\, l_{it-1}^q\big] = 0 \qquad \text{for every } (j, q) \text{ in the sieve}, \tag{4}$$

which identifies the remaining capital and labor elasticities together with $C_t(k, l)$ itself.

GNR's materials elasticity $D_t(k, l, m)$ is a flexible function of the observed state variables only. Because inputs are chosen partly in response to the firm's own unobserved productivity ω_{it}, the realized elasticity $D_t(k_{it}, l_{it}, m_{it})$ varies across firms with ω_{it}, indirectly through firms' endogenous input choices. Any two firms that happen to realize the same input state (k_{it}, l_{it}, m_{it}), whatever their own productivity histories, are assigned exactly the same elasticity $D_t(k_{it}, l_{it}, m_{it})$ by construction. GNR's framework leaves no room for a further, independent source of unobserved heterogeneity that would let two firms at an identical input state nonetheless have different elasticities. That is, GNR allows output elasticities to vary with the unobservable ω_{it} only through its effect on observed input choices, not with a separate unobserved technology type that shifts a firm's elasticity independently of (k_{it}, l_{it}, m_{it}). The next section develops a model in which the materials elasticity depends on such a technology type U_{it} directly, in addition to (k_{it}, l_{it}, m_{it}), and Section 4 shows how this additional layer of heterogeneity can nonetheless be identified.

3 A Nonseparable Production Function in an Unobserved Technology Type

The purpose of this paper is identification and estimation of the following *gross-output* production function given (in logs) by:

$$y_{it} = f_t(k_{it}, l_{it}, m_{it}, U_{it}) + \omega_{it} + \varepsilon_{it}, \tag{5}$$

where k_{it}, l_{it}, and m_{it} are defined as earlier, ω_{it} is Hicks-neutral productivity, and ε_{it} is an ex-post shock to production realized after all inputs are chosen. No separability is imposed anywhere in f_t. The firm-specific technology type U_{it} is allowed to interact freely with capital, labor, and materials, so that every elasticity of f_t is heterogeneous and, in general, varies with the levels of all inputs, and the unobserved type. This is a key distinction from GNR, whose materials elasticity is a deterministic function of observed inputs alone, with no additional source of firm-level heterogeneity. We normalize U_{it} to be uniformly distributed on the unit interval, so that a firm's rank in the technology-type distribution is the object of interest throughout the paper.

In levels, $Y_{it} = e^{F(K_{it}, L_{it}, M_{it}, U_{it})} e^{\omega_{it} + \varepsilon_{it}} \equiv Y_{it}^0 e^{\varepsilon_{it}}$, where $Y_{it}^0 = e^{F(K_{it}, L_{it}, M_{it}, U_{it}) + \omega_{it}}$ is output net of the ex-post shock. Materials M_{it} is a static, flexible input: the firm chooses M_{it} after observing $(K_{it}, L_{it}, \omega_{it}, U_{it})$ but before ε_{it} is realized, so that ε_{it} is independent of the firm's information set $\mathcal{I}_{it} = \{K_{it}, L_{it}, \omega_{it}, U_{it}, P_{it}, P_{it}^M\}$ at the time M_{it} is chosen, with $E[e^{\varepsilon_{it}} | \mathcal{I}_{it}] = E[e^{\varepsilon_{it}}] \equiv \bar{\varepsilon}$ a free parameter. Define $\beta_m(k_{it}, l_{it}, m_{it}, U_{it}) \equiv \partial f_t(k_{it}, l_{it}, m_{it}, U_{it}) / \partial \ln m_{it}$, the materials log-elasticity, itself a function of every input and the unobserved technology type.

M_{it} maximizes expected profit conditional on $\mathcal{I}_{it}$,

$$\max_{M_{it}} \Big\{ P_{it}\, E[Y_{it} \mid \mathcal{I}_{it}] - P_{it}^M M_{it} \Big\} = \max_{M_{it}} \Big\{ P_{it}\, Y_{it}^0\, E[e^{\varepsilon_{it}} \mid \mathcal{I}_{it}] - P_{it}^M M_{it} \Big\}.$$

The first-order condition is

$$P_{it} \frac{\partial Y_{it}^0}{\partial M_{it}} E[e^{\varepsilon_{it}} \mid \mathcal{I}_{it}] - P_{it}^M = 0.$$

Since $\varepsilon_{it} \perp \mathcal{I}_{it}$, $E[e^{\varepsilon_{it}} \mid \mathcal{I}_{it}] = E[e^{\varepsilon_{it}}] \equiv \bar{\varepsilon}$. Because $Y_{it}^0 = e^{F(K_{it}, L_{it}, M_{it}, U_{it}) + \omega_{it}}$,

$$\frac{\partial Y_{it}^0}{\partial M_{it}} = \frac{\partial F}{\partial \ln M_{it}} \cdot \frac{Y_{it}^0}{M_{it}} = \beta_m(k_{it}, l_{it}, m_{it}, U_{it})\, \frac{Y_{it}^0}{M_{it}}.$$

Substituting and rearranging, this equates the elasticity, scaled by $\bar{\varepsilon}$, to the observed materials cost share times $e^{\varepsilon_{it}}$:

$$\beta_m(k_{it}, l_{it}, m_{it}, U_{it})\,\bar{\varepsilon} = \frac{P_{it}^M M_{it}}{P_{it} Y_{it}^0} = \frac{P_{it}^M M_{it}}{P_{it} Y_{it}}\, e^{\varepsilon_{it}} \equiv S_{it}\, e^{\varepsilon_{it}}, \qquad S_{it} \equiv \frac{P_{it}^M M_{it}}{P_{it} Y_{it}}. \tag{6}$$

Taking logs, we obtain the share equation

$$s_{it} \equiv \ln(S_{it}) = \ln\big(\beta_m(k_{it}, l_{it}, m_{it}, U_{it})\big) + \ln \bar{\varepsilon} - \varepsilon_{it}. \tag{7}$$

Since $\varepsilon_{it} \perp \mathcal{I}_{it}$, ε_{it} is independent of the technology type U_{it}: this is the timing assumption that allows the share equation (7) to separate variation in the materials elasticity driven by U_{it} from the transitory noise ε_{it}, which is exactly the deconvolution problem Section 4 solves.

Conditioning (7) on the firm's information set $\mathcal{I}_{it}$ is not enough to recover β_m at a specific technology type. Since M_{it} is itself chosen by the firm partly in response to U_{it}, any object built only from the conditional distribution of s_{it} given (k_{it}, l_{it}, m_{it}) recovers, at best, the elasticity averaged over U_{it}'s conditional distribution given materials usage. Recovering the full elasticity surface therefore requires exploiting the panel structure of the data, together with an additional timing assumption on ε_{it}, through the conditional deconvolution argument of the next section.

Once the materials elasticity surface and the firm's own technology type are recovered from (7), the remaining capital and labor elasticities are recovered by a second, GNR-style stage (Section 5) that treats $C(k, l, U) \equiv F(k, l, m_0, U)$, the production function net of the (now-identified) materials contribution, as a single nonseparable object of capital, labor, and the recovered type.

4 Identification

In this section we provide conditions under which the materials elasticity β_m, the distribution of the ex-post shock ε_{it}, and the firm's own technology type U_{it} are nonparametrically identified using conditional deconvolution arguments and other mild regularity conditions. To simplify notation, let $X_{it} = (k_{it}, l_{it}, m_{it})$ and

$$Z_{it} \equiv \ln\big(\beta_m(k_{it}, l_{it}, m_{it}, U_{it})\big) + \ln \bar{\varepsilon}, \tag{8}$$

so that the share equation (7) reads $s_{it} = Z_{it} - \varepsilon_{it}$. The log materials share is the bundled elasticity-and-constant term less the ex-post shock. For any random variable X and $\rho_\varepsilon \neq 0$, let $\tilde{X} = X/\rho_\varepsilon$. We assume the ex-post shocks follow an AR(1) process,

$$\varepsilon_{it} = \rho_\varepsilon \varepsilon_{it-1} + \zeta_{it}, \qquad \zeta_{it} \perp \varepsilon_{it-1} \mid X. \tag{9}$$

Writing out one more period conditional throughout on $X = X_{it}$,

$$\begin{aligned} s_{it} &= Z_{it} - \varepsilon_{it}, \\ s_{it+1} &= Z_{it+1} - \varepsilon_{it+1} = Z_{it+1} - \rho_\varepsilon \varepsilon_{it} - \zeta_{it+1}, \end{aligned} \tag{10}$$

gives two noisy measurements of the same ε_{it}, up to the scale ρ_ε on the second. We identify ρ_ε first, since it is needed to rescale the second measurement before the two can be fed into a deconvolution argument. With ρ_ε in hand, $\tilde{s}_{it+1} \equiv s_{it+1}/\rho_\varepsilon = \tilde{Z}_{it+1} - \varepsilon_{it} - \tilde{\zeta}_{it+1}$, so that $(s_{it}, \tilde{s}_{it+1})$ are now two noisy measurements of ε_{it} sharing a common component with mutually independent remaining pieces, which is the structure required by Kotlarski's identity.

Our goal is identification of the conditional distribution of Z_{it}, which is identified if the conditional characteristic function $\phi_{Z_{it}|X}(s|X)$ is identified, since the CDF then follows by Fourier (Gil-Pelaez) inversion,

$$F_{Z_{it}|X}(z|X) = \frac{1}{2} - \lim_{v \to \infty} \int_{-v}^{v} \frac{e^{-\mathrm{i}sz}}{2\pi \mathrm{i} s} \phi_{Z_{it}|X}(s|X) ds. \tag{11}$$

Since the quantile function is the inverse of the CDF, identification of $\phi_{Z_{it}|X}(s|X)$ implies identification of the elasticity surface $\beta_m(k, l, m; \tau)$ at every quantile τ of the technology-type distribution.[2] We utilize conditional deconvolution arguments to identify ρ_ε, the conditional characteristic functions of Z_{it} and ε_{it}, and the normalizing constant $\bar{\varepsilon}$ under the following assumptions:

Assumption 4.1 *Suppose the following set of assumptions*

(a) Random Sample: The random variables $(X_{it}, U_{it}, \varepsilon_{it})_{i=1}^N$ are independently and identically distributed on $\mathcal{X} \times \mathcal{U} \times \mathbb{R}$.

[2] Similar conditional deconvolution ideas have been used in measurement error models such as Li and Vuong (1998), Schennach (2004), Song et al. (2015), in panel data models such as Neumann (2007), and in a Kotlarski panel setting closely related to ours by Hu et al. (2020) and Evdokimov (2010).

(b) Shock Dynamics: The ex-post shock follows an AR(1) process, with innovation terms ζ_{it} satisfying $\mathbb{E}[\zeta_{it}\varepsilon_{it-1}|X] = 0$.

(c) Independent ex-post shocks: $\varepsilon_{it} \perp \mathcal{I}_{it}$, with the location normalization $\mathbb{E}[\varepsilon_{it} \mid X] = 0$.

(d) Conditional Independence: (i) $U_{it+1} \perp U_{it} \mid X$ (technology types are independent across periods, conditional on inputs); (ii) $U_{it} \perp \zeta_{it} \mid X$; (iii) $U_{it} \perp \varepsilon_{it} \mid X$.

(e) Characteristic Functions: The conditional characteristic functions $\phi_{Z_{it}|X}(s|X)$, $\phi_{\varepsilon_{it}|X}(s|X)$, and $\phi_{\zeta_{it+1}|X}(s|X)$ do not vanish for all $s \in \mathbb{R}$ and $X \in \mathcal{X}$.

Assumption 4.1(a) places restrictions on the data generating process. Assumption 4.1(b) is the panel AR(1) structure on the ex-post shock that lets a third period, $t+2$, pin down ρ_ε before deconvolution. Iterating the AR(1) process and taking conditional covariances of s_{it} with s_{it+1} and s_{it+2}, every term involving U_{it}, U_{it+1}, U_{it+2}, ζ_{it+1}, or ζ_{it+2} vanishes under Assumption 4.1(d) below, leaving only the autocovariance of ε_{it} with itself, so that ρ_ε is identified in closed form as the ratio

$$\rho_\varepsilon = \frac{\mathrm{Cov}(s_{it}, s_{it+2}|X)}{\mathrm{Cov}(s_{it}, s_{it+1}|X)}. \tag{12}$$

Proof: See Appendix A.

Assumption 4.1(c) gives the free parameter as in GNR, and its location normalization is what pins down $\phi_{Z_{it}|X}$ up to scale in (38) below. Assumption 4.1(d) is the conditional independence condition that makes the Kotlarski factorization possible. Because $X_{it} = (k_{it}, l_{it}, m_{it})$ is the full input vector, this rules out the unobserved technology type correlating across periods through any channel tied to input choices. Assumption 4.1(e) is a mild technical condition on characteristic functions satisfied by most standard families of distributions.

Theorem 4.1 *Under Assumption 4.1, the AR(1) coefficient ρ_ε, the conditional characteristic functions $\phi_{Z_{it}|X}(s|X)$ and $\phi_{\varepsilon_{it}|X}(s|X)$, the normalizing constant $\bar{\varepsilon}$, and hence the materials elasticity surface $\beta_m(k, l, m; \tau)$ at every quantile $\tau \in (0, 1)$ of the technology-type distribution, are nonparametrically identified. Given the recovered $\phi_{Z_{it}|X}(s|X)$ and $\phi_{\varepsilon_{it}|X}(s|X)$, each firm's own ex-post shock ε_{it} and technology type U_{it} are identified as functions of $(s_{it}, s_{it+1}, X_{it})$.*

Proof: See Appendix A.

Theorem 4.1 identifies the materials elasticity surface $\beta_m(k,l,m;\tau)$, the distribution of the ex-post shock ε_{it}, and, for each firm-period, the realized pair $(\varepsilon_{it}, U_{it})$. The elasticities of capital and labor remain to be identified, and we recover them by extending GNR's second-stage device to this setting, plugging in the objects Theorem 4.1 has just identified. Here, integrating the identified $\beta_m(k,l,m;\tau)$ over materials at each firm's own recovered rank $\hat{U}_{it}$, and subtracting the recovered ex-post shock $\hat{\varepsilon}_{it}$, purges output of both the flexible input's contribution and the transitory noise, leaving a nonseparable function of capital, labor, and the technology type:

$$\check{y}_{it} \equiv y_{it} - \hat{\varepsilon}_{it} - \int_{m_0}^{\widehat{m_{it}}} \hat{\beta}_m(k_{it}, l_{it}, \sigma; \hat{U}_{it})\, d\ln\sigma = C(k_{it}, l_{it}, U_{it}) + \omega_{it}, \tag{13}$$

with $\hat{U}_{it}$ entering as an additional state variable. Approximating $C(k,l,U;\theta)$ with a sieve, a trial θ inverts for trial productivity $\omega_{it}(\theta) = \check{y}_{it} - C(k_{it}, l_{it}, \hat{U}_{it}; \theta)$, and projecting $\omega_{it+1}(\theta)$ onto its own lag gives an innovation $\xi_{it+1}(\theta)$. Identification follows from an orthogonality condition on productivity innovations.

Assumption 4.2 *Productivity follows a first-order Markov process,*

$$\omega_{it+1} = h(\omega_{it}) + \xi_{it+1}, \qquad \mathbb{E}\left[\xi_{it+1} \mid \mathcal{I}_{it}\right] = 0, \tag{14}$$

where $\mathcal{I}_{it} = \{K_{it}, L_{it}, \omega_{it}, U_{it}, P_{it}, P^M_{it}\}$ is the same firm information set introduced in Section 3.

With this assumption in hand, and exactly as in Equation (4) but with $\hat{U}_{it}$ joining (k,l) as a third argument of both the sieve and the instrument set, θ is identified by the moment condition

$$\mathbb{E}\left[\xi_{it+1}(\theta)\, k^j_{it+1}\, l^q_{it}\, \hat{U}^p_{it}\right] = 0 \qquad \text{for every } (j,q,p) \text{ in the sieve}, \tag{15}$$

This moment condition identifies $C(k,l,U)$ and, by differentiation, the capital and labor elasticity surfaces, now indexed by U as well as (k,l). Section 5 develops the corresponding two-stage estimator: Stage 1 constructs $\hat{\beta}_m$, $\hat{\varepsilon}_{it}$, and $\hat{U}_{it}$ from Theorem 4.1's argument, and Stage 2 is the sample analogue of the extended GNR second stage just described, with these Stage 1 objects entering as generated regressors.

5 Econometric Procedure

In this section we introduce a two-stage estimator motivated by the identification argument of Section 4. From the repeated materials share equation, we recover the AR(1) coefficient ρ_ε, the conditional characteristic functions of Z_{it} and ε_{it}, the normalizing constant $\bar{\varepsilon}$, and each firm's own ex-post shock $\hat{\varepsilon}_{it}$ and technology type $\hat{U}_{it}$. The second stage takes $\hat{U}_{it}$ as a generated regressor and recovers the remaining capital and labor elasticities using a sieve-GMM procedure resembling Gandhi et al. (2020)'s second stage.

5.1 A Two-Stage Estimator

Stage 1(a): The AR(1) coefficient. We estimate the conditional means of materials shares $E[s_{it}|X_{it}]$, $E[s_{it+1}|X_{it}]$, and $E[s_{it+2}|X_{it}]$ using nonparametric kernel regression. Taking the residuals gives $\tilde{s}_{it}, \tilde{s}_{it+1}, \tilde{s}_{it+2}$; the sample analogue of the covariance-ratio identity in Equation (12) is

$$\hat{\rho}_\varepsilon = \frac{\sum_{i,t} \tilde{s}_{it}\tilde{s}_{it+2}}{\sum_{i,t} \tilde{s}_{it}\tilde{s}_{it+1}}. \tag{16}$$

This is similar to the covariance-differencing device used to identify serially correlated measurement error in the panel data models of Griliches and Hausman (1986).

Stage 1(b): Characteristic Functions and the Elasticity Surface. We represent every conditional density function by a common order-J Hermite-function sieve

$$f_{Z_{it}|X}(z|X) = \sum_{j=0}^{J} c_j(X)\psi_j(z), \tag{17}$$

where ψ_j is the (physicists') Hermite function of order j,

$$\psi_j(z) = \left(2^j j!\sqrt{\pi}\right)^{-1/2} H_j(z)\, e^{-z^2/2}, \tag{18}$$

H_j the physicists' Hermite polynomial of degree j. The $\{\psi_j\}_{j\geq 0}$ are a complete orthonormal basis of $L^2(\mathbb{R})$, so the coefficients $c_j(X)$ are the corresponding projections of the conditional density onto that basis,

$$c_j(X) = \int_{-\infty}^{\infty} f_{Z_{it}|X}(z|X)\, \psi_j(z)\, dz = \mathbb{E}\left[\psi_j(Z_{it}) \mid X\right]. \tag{19}$$

Because the ψ_j are eigenfunctions of the Fourier transform, the characteristic function

$$\phi_{Z_{it}|X}(\gamma|X) = \sqrt{2\pi} \sum_{j=0}^{J} i^j c_j(X) \psi_j(\gamma), \tag{20}$$

is available from the same coefficients $c_j(X)$ used for the density.

From this point, we make a departure from the traditional deconvolution estimation strategy, which relies on differentiation of a contour integral and numerical integration. Instead, we follow Kato et al. (2021) by showing that the true $\phi_{Z_{it}|X}$ satisfies, for every γ,

$$\mathbb{E}\Big[\big(\mathrm{i}\, s_{it}\, \phi_{Z_{it}|X}(\gamma|X) - \phi'_{Z_{it}|X}(\gamma|X)\big)\, e^{\mathrm{i}\gamma D_{it}} \,\Big|\, X\Big] = 0, \qquad D_{it} \equiv s_{it} - \tilde{s}_{it+1} \tag{21}$$

Proof: See Appendix B.

Substituting the Hermite-sieve representation $\phi_{Z_{it}|X}(\gamma|X)$ and its derivative makes this restriction linear in $c(X) = (c_0(X), \ldots, c_J(X))'$ at every γ, with only two data-based conditional expectations, $\mathbb{E}[s_{it}e^{\mathrm{i}\gamma D_{it}} \mid X]$ and $\mathbb{E}[e^{\mathrm{i}\gamma D_{it}} \mid X]$, each estimated directly by a nonparametric kernel regression. Stacking this restriction over a frequency grid $\gamma_1, \ldots, \gamma_L$ gives a homogeneous linear system $\hat{A}(X)c(X) \approx 0$ for a known, $(2L) \times (J+1)$ real weight matrix $\hat{A}(X)$ built from these two conditional expectations and the fixed, tabulated Hermite functions and their derivatives. Minimizing $\|\hat{A}(X)c\|^2$ subject to the density-integrates-to-one constraint $b'c(X) = 1$, for a fixed vector b closed-form in the Hermite basis has the closed-form solution

$$\hat{c}(X) = \frac{M(X)^{-1}b}{b'M(X)^{-1}b}, \qquad M(X) \equiv \hat{A}(X)'\hat{A}(X), \tag{22}$$

Proof: See Appendix B.

The coefficients and Hermite functions evaluated at γ give us our estimate of the conditional characteristic function, and hence the conditional density

$$\hat{f}_{Z_{it}|X}(z|X) = \sum_{j=0}^{J} \hat{c}_j(X)\psi_j(z), \tag{23}$$

Rather than obtaining the conditional CDF through numerical integration, and introducing another computational burden before getting to the quantile-object of interest, we show that the recursion relationship of Hermite functions yields an intuitive closed-form solution.

$$\hat{F}_{Z_{it}|X}(z|X) = \Phi(z) + e^{-z^2/2} \sum_{j=0}^{J} \hat{c}_j(X)\, q_j(z). \tag{24}$$

Proof: See Appendix B.

Here q_j is an explicit polynomial of degree $j-1$, itself a finite combination of the Hermite polynomials that generate the sieve basis,

$$\begin{aligned} q_j(z) &= -\sum_{k=0}^{\lceil j/2 \rceil - 1} \rho_k \sqrt{\frac{2}{j-2k}} \left(2^{j-2k-1}(j-2k-1)!\,\sqrt{\pi}\right)^{-1/2} H_{j-2k-1}(z) \\ \rho_k &\equiv \prod_{\ell=0}^{k-1} \sqrt{\frac{j-2\ell-1}{j-2\ell}} \quad (\rho_0 \equiv 1). \end{aligned} \tag{25}$$

The CDF is the standard normal CDF plus a polynomial correction, weighted by the same coefficients $\hat{c}_j(X)$ that fit the density, so departures of Z_{it}'s conditional distribution from normality show up as a low-order polynomial adjustment to $\Phi(z)$.

Evaluating $\phi_{s_{it}|X} = \phi_{Z_{it}|X}\phi_{\varepsilon_{it}|X}$ at the imaginary argument $\gamma = -\mathrm{i}$ delivers the plug-in estimator of the normalizing constant. Since $\bar{\varepsilon} = \mathbb{E}[e^{\varepsilon_{it}}]$ is the moment generating function of ε_{it} at 1, and the characteristic function analytically continues to the moment generating function via $M_\varepsilon(1) = \phi_{\varepsilon_{it}|X}(-\mathrm{i} \mid X)$ whenever $\bar{\varepsilon} < \infty$,

$$\hat{\bar{\varepsilon}} = \frac{1}{N} \sum_{i,t} \hat{\phi}_{\varepsilon_{it}|X}(-\mathrm{i} \mid X) = \frac{\hat{\phi}_{s_{it}|X}(\mathrm{i} \mid X)}{\hat{\phi}_{Z_{it}|X}(\mathrm{i} \mid X)}. \tag{26}$$

The numerator simplifies immediately, since $\phi_{s_{it}|X}(\mathrm{i} \mid X) = \mathbb{E}[e^{\mathrm{i}(\mathrm{i})s_{it}} \mid X] = \mathbb{E}[e^{-s_{it}} \mid X] = \mathbb{E}[1/S_{it} \mid X]$, which is the conditional mean of the reciprocal materials share, estimated directly. The denominator is Stage 1(b)'s own fitted $\hat{\phi}_{Z_{it}|X}$, evaluated at the imaginary point i. The elasticity surface then follows from (7) and the fitted CDF's inverse using a bisection algorithm.

$$\ln \hat{\beta}_m(k,l,m;\tau) = \hat{F}^{-1}_{Z_{it}|X}\big(\tau \mid X = (k,l,m)\big) - \ln \hat{\bar{\varepsilon}}, \qquad \tau \in (0,1). \tag{27}$$

Stage 1(c): Recovering $\hat{\varepsilon}_{it}$ and $\hat{U}_{it}$. Stage 1(b) estimates only the distribution of ε_{it}, not its realization for any given firm-period, yet Stage 2's generated regressor $\hat{U}_{it}$ and each firm's own materials elasticity $\hat{\beta}_m(k_{it}, l_{it}, m_{it}, U_{it}) = \exp(s_{it} + \hat{\varepsilon}_{it} - \ln \hat{\bar{\varepsilon}})$ from (7) both require the realization itself. The posterior mean of ε_{it} given the firm's own $(s_{it}, s_{it+1}, X_{it})$ weights every candidate value e by its likelihood given the observed data. A similar device is used by Bonhomme and Robin (2010) in an application that recovers transitory income shocks as opposed to the ex-post shocks considered in this paper. The posterior mean is given by

$$\hat{\varepsilon}_{it} = \frac{\displaystyle\int e\,\hat{f}_{\varepsilon_{it}|X}(e \mid X)\,\hat{f}_{Z_{it}|X}(s_{it} + e \mid X)\,\hat{f}_{Z_{it+1}-\zeta_{it+1}|X}\big(s_{it+1} + \hat{\rho}_{\varepsilon} e \mid X\big)\,de}{\displaystyle\int \hat{f}_{\varepsilon_{it}|X}(e \mid X)\,\hat{f}_{Z_{it}|X}(s_{it} + e \mid X)\,\hat{f}_{Z_{it+1}-\zeta_{it+1}|X}\big(s_{it+1} + \hat{\rho}_{\varepsilon} e \mid X\big)\,de}, \tag{28}$$

where the conditional densities $\hat{f}_{Z_{it}|X}$, $\hat{f}_{\varepsilon_{it}|X}$ are estimated from Stage 1(b). The third density $\hat{f}_{Z_{it+1}-\zeta_{it+1}|X}$ is shown to be recovered in Stage 1(b) as well, with full steps outlined in Appendix A.

Like any empirical-Bayes posterior mean, this estimator shrinks every firm's reconstructed ex-post shock toward its own conditional mean, which may understate true heterogeneity. This passes into the estimate of technology type, $\hat{U}_{it} = \hat{F}_{Z_{it}|X}(s_{it} + \hat{\varepsilon}_{it}|X)$, and therefore into the materials, capital, and labor elasticities as well.

Stage 2: Recovering Capital and Labor. Integrating the recovered materials elasticity over materials while holding $(k_{it}, l_{it}, \hat{U}_{it})$ fixed and purging $\hat{\varepsilon}_{it}$ from output leaves a nonseparable function C of capital, labor, and the recovered type:

$$\breve{y}_{it} = C(k_{it}, l_{it}, \hat{U}_{it}) + \omega_{it}. \tag{29}$$

We approximate C with a tensor-product sieve,

$$C(k, l, \hat{U}; \theta) \approx \sum_{j,q,p} \theta_{jqp}\, k^j l^q \hat{U}^p, \tag{30}$$

invert for a trial productivity,

$$\omega_{it}(\theta) = \breve{y}_{it} - C(k_{it}, l_{it}, \hat{U}_{it}; \theta), \tag{31}$$

and project $\omega_{it+1}(\theta)$ onto a sieve in $\omega_{it}(\theta)$ to construct the innovation $\xi_{it+1}(\theta)$, re-estimated at every trial θ. The sample counterpart to Equation (15) is a GMM criterion function, following Gandhi et al. (2020), whose minimizer $\hat{\theta}$ delivers firm-period capital and labor elasticities by differentiating the fitted sieve directly:

$$\hat{\beta}_k(k_{it}, l_{it}, \hat{U}_{it}) = \left.\frac{\partial C}{\partial \ln k}\right|_{\hat{\theta}}, \qquad \hat{\beta}_l(k_{it}, l_{it}, \hat{U}_{it}) = \left.\frac{\partial C}{\partial \ln l}\right|_{\hat{\theta}}. \tag{32}$$

We use a B-spline tensor sieve for C as opposed to raw polynomials as in Gandhi et al. (2020). This two-stage estimator is evaluated empirically in Section 6.

5.2 Large Sample Properties

The two-stage estimator has several tuning parameters: the Hermite sieve order J and the frequency grid $\{\gamma_1, \ldots, \gamma_L\}$ in Stage 1, the Nadaraya-Watson bandwidth h used throughout, the sieve degree and number of interior knots for $C(k, l, U; \theta)$ in Stage 2. Gandhi et al. (2020) itself derives no formal asymptotic theory for the sieve-GMM second stage this paper's own Stage 2 follows, and neither do we; formal rates and asymptotic normality for this class of two-stage sieve-deconvolution estimators, in the spirit of the sieve-GMM asymptotics of Ai and Chen (2003) and the two-step semiparametric GMM theory of Ackerberg et al. (2012) and Ackerberg et al. (2014) for a generated regressor estimated in a first stage, are left to future work, but the two stages likely differ in how tractable that theory would be to derive. Stage 1's materials elasticity comes from a linear system in the Hermite coefficients $c(X)$ with only two generated regressors entering it – the Nadaraya-Watson estimates $\widehat{\mathbb{E}}[s_{it}e^{\mathrm{i}\gamma D_{it}} \mid X]$ and $\widehat{\mathbb{E}}[e^{\mathrm{i}\gamma D_{it}} \mid X]$ – a structure close to standard semiparametric two-step estimation with generated regressors from a first-stage kernel regression, a setting whose asymptotics are comparatively well understood.

Stage 2 is more difficult: its capital and labor elasticities are sieve-GMM estimates that use $\hat{U}_{it}$, itself already a function of Stage 1's estimated coefficients and posterior means, as a regressor, so Stage 1's sampling error propagates into Stage 2's coefficients across two compounded stages, which coincides with the noisier firm-level estimates of $\hat{\beta}_k$, $\hat{\beta}_l$ in Section 6. We do not derive the resulting asymptotic distribution here. Instead, standard errors are reported from a nonparametric bootstrap clustered at the firm level. We leave the derivation of asymptotic properties, and validity of this bootstrap procedure, for future research.

6 Application

We apply the two-stage estimator of Section 5.1 to a firm-level manufacturing dataset from the U.S. to examine heterogeneity in production. For each manufacturing industry, we recover the materials elasticity surface $\hat{\beta}_m(k, l, m; \tau)$ across the recovered technology-type quantile τ, together with firm-period-level $\hat{\beta}_m$, $\hat{\beta}_k$, $\hat{\beta}_l$ at each firm's own realized state and recovered type $\hat{U}_{it}$. Every industry uses the same tuning parameters: a B-spline sieve of degree two with zero interior knots for Stage 2's $C(k, l, U; \theta)$. We also trim $\hat{U}_{it}$ to $[0.05, 0.95]$ before Stage 2 estimation. The tables and figures below report point estimates, dispersion across firm-periods, and dispersion across the recovered technology-type distribution.

6.1 U.S. Compustat

The U.S. manufacturing data come from Compustat, which covers publicly traded firms and contains data from their financial statements, giving measures of output, capital, labor, and material inputs. Manufacturing industries are classified by the two-digit NAICS codes 31, 32, and 33. NAICS 31 includes different types of food and beverage manufacturing as well as manufacturing of textiles and apparel. NAICS 32 includes manufacturing of wood, paper, chemicals, and plastics. NAICS 33 includes steel, mineral, equipment, and electronic manufacturing.

Before turning to the elasticity estimates, Table 1 reports the two industry-level structural parameters identified in Section 4: the AR(1) coefficient $\hat{\rho}_\varepsilon$ of the ex-post shock and the normalizing constant $\hat{\bar{\varepsilon}}$.

Table 1: Persistence and Intercept Estimates for U.S. Manufacturing Firms

NAICS	$\hat{\rho}_\varepsilon$	$\hat{\bar{\varepsilon}}$
31	0.952	1.370
	(0.003)	(0.034)
32	0.944	1.435
	(0.002)	(0.031)
33	0.952	1.460
	(0.002)	(0.017)

*$\hat{\rho}_\varepsilon$ is the ratio of residualized-share cross-moments (36); $\hat{\bar{\varepsilon}}$ is recovered from the same Kotlarski contour used for $\hat{\beta}_m$, evaluated at $\gamma = -i$.

The ex-post shock is highly persistent in every industry ($\hat{\rho}_\varepsilon$ between 0.94 and 0.95). Figure 1 shows both parameters directly: each firm-period's own denoised ex-post shock $\hat{\varepsilon}_{it}$ plotted against its one-period lag $\hat{\varepsilon}_{i,t-1}$.

Figure 1: Ex-Post Shock Persistence

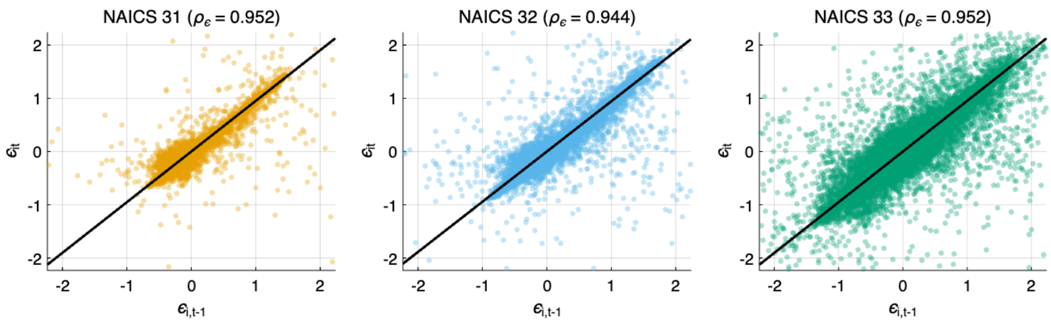


Table 2 reports the median of the firm-period-level materials, capital, and labor elasticities in each industry, together with median returns to scale and median capital intensity ($\hat{\beta}_k/\hat{\beta}_l$).

Table 2: Median Elasticity Estimates for U.S. Manufacturing Firms

	NAICS		
	31	32	33
Materials	0.267	0.202	0.356
	(0.011)	(0.007)	(0.005)
Capital	0.240	0.174	0.032
	(0.058)	(0.061)	(0.032)
Labor	0.236	0.468	0.755
	(0.077)	(0.076)	(0.049)
RTS	0.74	0.84	1.14
	(0.069)	(0.045)	(0.033)
Cap. Int.	1.02	0.37	0.04
	(4.83)	(0.21)	(0.05)

*Materials elasticities are Stage 1's $\hat{\beta}_m$ evaluated at each firm's own realized state and recovered type $\hat{U}_{it}$; capital and labor elasticities are Stage 2's $\hat{\beta}_k$, $\hat{\beta}_l$, both medians across firm-periods. RTS is the sum of the three median elasticities; capital intensity is the ratio of median capital to median labor elasticities.

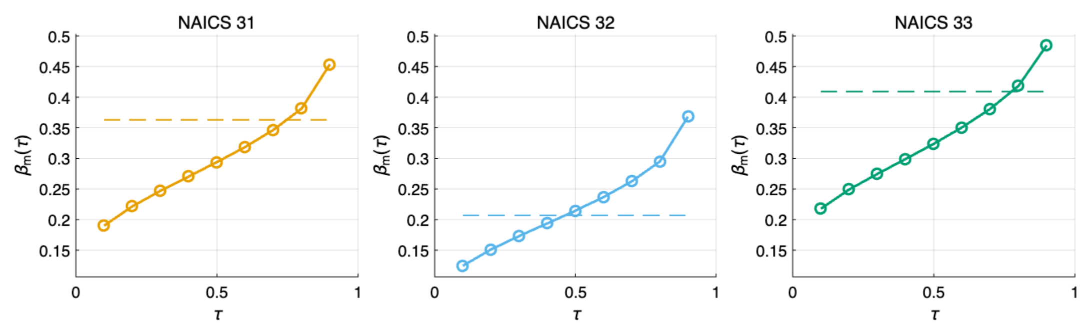
As a benchmark, Table 3 reports what the standard GNR estimator recovers using the default specification of their empirical application.

Table 3: GNR Elasticity Estimates for U.S. Manufacturing Firms

	NAICS		
	31	32	33
Materials	0.363	0.207	0.409
	(0.015)	(0.033)	(0.009)
Capital	0.303	0.138	0.077
	(0.091)	(0.103)	(0.029)
Labor	0.167	0.594	0.538
	(0.099)	(0.138)	(0.039)
RTS	0.83	0.94	1.02
	(0.055)	(0.057)	(0.024)
Cap. Int.	1.81	0.23	0.14
	(25.54)	(2.66)	(0.08)

One immediate conclusion when comparing these two results is that unobserved heterogeneity can change elasticity estimates drastically, even after controlling for heterogeneity in observables in the nonparametric estimator of Gandhi et al. (2020).

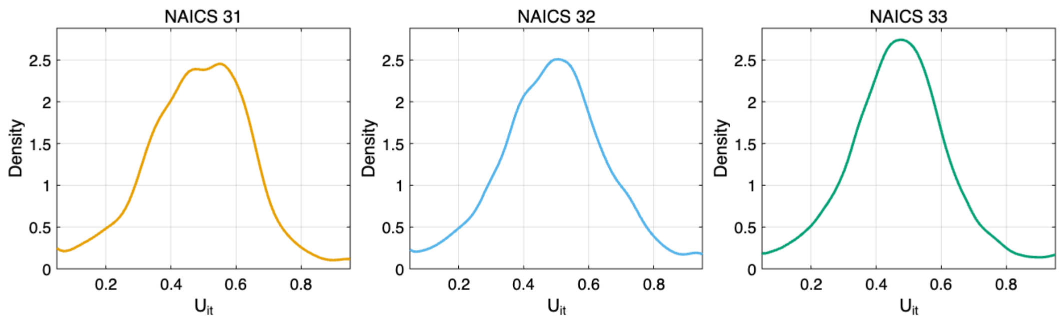
Figure 2: Materials Elasticity Across the Recovered Technology-Type Distribution



*Stage 1's materials elasticity surface $\hat{\beta}_m(\tau)$, averaged across firms at each quantile τ of the recovered technology-type distribution. The dashed horizontal line is the corresponding baseline GNR estimator.

Figure 2 plots the average materials elasticity across the recovered technology-type quantile $\tau \in \{0.05, 0.10, \ldots, 0.95\}$. The pattern is uniform across all three industries: the materials elasticity rises monotonically in τ, from 0.19 at $\tau = 0.10$ to 0.45 at $\tau = 0.90$ in NAICS 31, from 0.13 to 0.37 in NAICS 32, and from 0.22 to 0.49 in NAICS 33. Firms with a higher recovered technology type use materials substantially more intensively at the margin than firms with a lower type, holding capital, labor, and materials usage fixed. The GNR baseline's flat $\hat{\beta}_m$ (Table 3) lands inside this type-varying range in every industry.

Figure 3: Recovered Technology-Type Distribution: U.S.

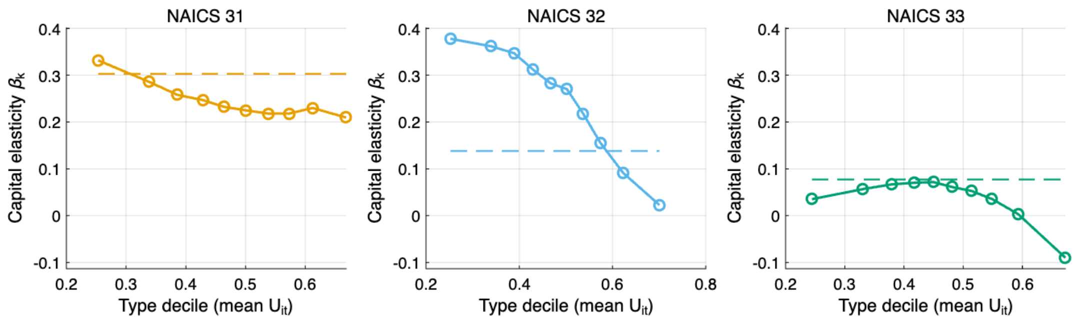


*Kernel density of the recovered technology type $\hat{U}_{it}$. The horizontal axis is truncated to $[0.05, 0.95]$ for legibility; the density is estimated on the full $[0, 1]$ support.

Figure 3 plots the distribution of technology types. Densities are unimodal and concentrated around $\hat{U}_{it} = 0.5$ in every industry, despite U_{it} being normalized to Uniform$(0, 1)$ by construction. This is consistent with the location normalization holding on average; however, the standard deviation (0.171-0.179) is only 59-62% of Uniform$(0, 1)$'s own $(1/\sqrt{12} \approx 0.289)$, and 76-80% of firm-periods fall in the middle two type quintiles $[0.3, 0.7]$, versus 40% under a true uniform.

This pattern is not a violation of Theorem 4.1, which identifies the population CDF of U_{it} exactly. Instead, it is the expected signature of estimating each firm's own $\hat{U}_{it}$ through the posterior *mean* $\hat{\varepsilon}_{it}$ of (43) rather than the true, unobserved ε_{it}. A posterior mean shrinks every firm's reconstructed $\hat{Z}_{it} = s_{it} + \hat{\varepsilon}_{it}$ toward its own conditional mean, a common feature of any empirical-Bayes posterior-mean estimator. Therefore, a firm truly near the tail of the type distribution is mapped by the fitted CDF to a less extreme $\hat{U}_{it}$ than its true rank. Pooled across firm-periods, this compresses the tails and inflates the center of the recovered distribution relative to the flat shape the identification argument implies for the population object.

Figure 4: Capital Elasticity by Technology-Type Decile

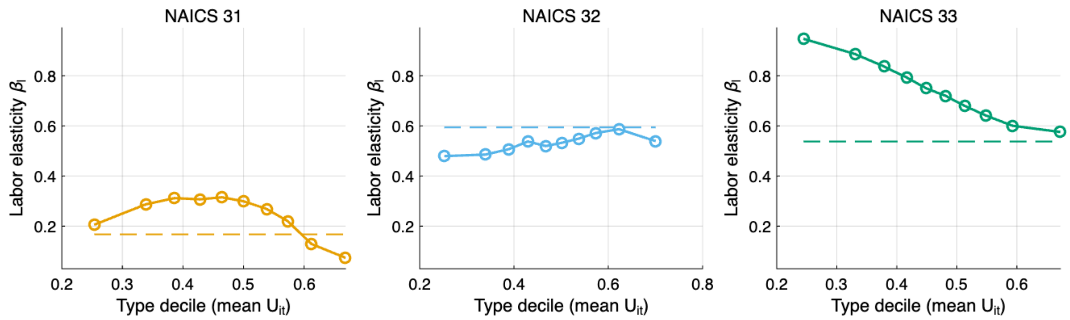


*Stage 2's $\hat{\beta}_k$, averaged within deciles of the matched recovered type $\hat{U}_{it}$. The dashed horizontal line is the corresponding median $\hat{\beta}_k$ from Table 3's baseline GNR estimator.

Unlike materials, capital elasticities *fall* across the recovered technology-type decile in every U.S. industry we examine: from 0.33 to 0.21 in NAICS 31, from 0.38 to 0.02 in NAICS 32, and, in NAICS 33, from 0.04 up to a mid-decile peak near 0.07 before falling to -0.09 at the highest decile. Labor's pattern is less uniform: it falls cleanly from 0.95 to 0.58 in NAICS 33, but is non-monotonic in NAICS 31 (rising from 0.21 in the lowest decile to a peak near 0.32 around the middle of the type distribution before falling to 0.07 at the top), and in NAICS 32 it rises over most of the range (from 0.48 to a peak near 0.59) before dipping slightly in the highest decile. Read together with Figure 2, the dominant pattern is that a higher recovered technology type substitutes toward materials intensity and away from capital intensity at the margin. This is consistent with the type indexing something closer to flexible-input-augmenting technology than to a uniform scale shift of every input's productivity.

This pattern has the flavor of factor-biased technical change, but is more general than the standard effective-units formulation (e.g., Acemoglu, 2002; Demirer, 2020), which rescales a single input's quantity by a firm-specific shifter before it enters an otherwise common elasticity function. This is equivalent to a horizontal relabeling of that input's own axis, with the elasticity function's underlying shape held fixed across firms. Because U_{it} enters the production function nonseparably rather than through such a rescaling, it can instead change the shape of the elasticity surface itself: materials', capital's, and labor's elasticities need not move together, or even monotonically, something a single input-augmenting shifter cannot generate. The pattern is consistent with firms differing in how much of their production process is embodied in purchased materials rather than in-house capital and labor, but the recovered type is not restricted, by construction, to operate only through that channel.

Figure 5: Labor Elasticity by Technology-Type Decile



*Stage 2's $\hat{\beta}_l$, averaged within deciles of the matched recovered type $\hat{U}_{it}$. The dashed horizontal line is the corresponding median $\hat{\beta}_l$ from Table 3's baseline GNR estimator.

6.2 Markups

In this section, we use the estimated flexible input elasticities (materials) to calculate markups using the production function approach. The first-order condition for M_{it} gives

$$\beta_m(k_{it}, l_{it}, m_{it}, U_{it})\,\bar{\varepsilon} = \mu_{it}\, S_{it}\, e^{\varepsilon_{it}}, \tag{33}$$

so that $\mu_{it} = \beta_m(k_{it}, l_{it}, m_{it}, U_{it})/S_{it}$. Under the maintained assumption that Stage 1's $\hat{\beta}_m$ is the firm's true, markup-free physical materials elasticity, and because Stage 1 already constructs $\hat{\beta}_m$ as $S_{it}\exp(\hat{\varepsilon}_{it})/\hat{\bar{\varepsilon}}$ (Section 5.1), this collapses algebraically to

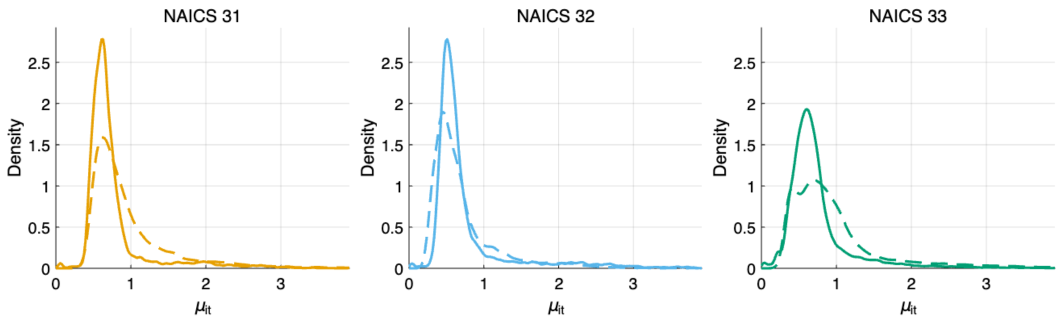
$$\hat{\mu}_{it} = \frac{\hat{\beta}_m(k_{it}, l_{it}, m_{it}, \hat{U}_{it})}{S_{it}} = \frac{\exp(\hat{\varepsilon}_{it})}{\hat{\bar{\varepsilon}}}, \tag{34}$$

computable from Stage 1's $\hat{\varepsilon}_{it}$ and $\hat{\bar{\varepsilon}}$. We emphasize that this is the single-flexible-input estimator, with materials as the only flexible input.

Table 4: Materials-Based Markup Estimates for U.S. Manufacturing Firms

NAICS	Median	Mean	[P5, P95]	Share $\hat{\mu}_{it} < 1$
31	0.644	0.822	[0.450, 1.960]	0.87
32	0.578	0.843	[0.386, 2.328]	0.85
33	0.641	0.837	[0.336, 1.777]	0.85

Figure 6: Materials-Based Markup Distributions



*Solid line is this paper's own markup estimate; dashed line is the markup estimated from the baseline GNR estimator.

Table 4 and Figure 6 show a distribution that is right-skewed and, at the median, below one in every industry: most firm-periods have $\hat{\mu}_{it} < 1$ (85-87% of the sample in each industry), while the mean is closer to, but still below, one (0.82-0.84) because of a long right tail of high-markup observations. We do not read the below-unity median as evidence that U.S. manufacturing firms systematically price below marginal cost. Two features of the construction can produce exactly this pattern even when the maintained assumption holds exactly in population. First, $\hat{\varepsilon}_{it}$ is a posterior *mean* and $\exp(\cdot)$ is convex, so $\exp(\hat{\varepsilon}_{it})$ understates $\mathbb{E}[\exp(\varepsilon_{it})]$ by Jensen's inequality. Second, and more fundamentally, a single flexible input gives no way to test the maintained assumption itself: with materials as the only flexible input, there is no independent check that $\hat{\beta}_m$ is free of market power. A second flexible input could supply exactly that test – checking whether materials- and labor-based markups agree – but we leave developing and implementing it to future work. In the next section, we estimate labor's own flexible elasticity independently and compare it, along with the persistence parameter and recovered technology type, against the main model's materials-based estimates.

6.3 Robustness Checks: Labor as a Flexible Input

Section 3 shows that the timing argument behind the materials share equation (7) applies equally to any input chosen after $(K_{it}, \omega_{it}, U_{it})$ are known but before ε_{it} is realized. Appendix C works out what follows if labor is treated as such a second flexible input: ρ_{ε}, $\bar{\varepsilon}$, and the technology-type surface can all be recovered a second time, independently, from labor's own share equation (49) (Corollary C.1), and comparing this independent recovery against the main model's materials-based estimates is a natural overidentification comparison, which we implement here.

We compare two sets of point estimates, which should be population-exact under the assumptions of Corollary C.1. This requires a labor revenue share S^l_{it}, and hence a labor-payments variable, which is reported in the U.S. Compustat data as wage bill, from which $s^l_{it} = \ln(\text{wage bill}_{it}) - \ln Y_{it}$. We re-run Stage 1 with labor in the flexible-input role (i.e., with s_{it} replaced by s^l_{it}) using the identical tuning parameters (Hermite order, frequency grid, bandwidth rule) as the materials-based estimator.

Table 5: Robustness Checks: Labor's Own Share Equation vs. Materials

NAICS	$\hat{\rho}^l_\varepsilon$	$\hat{\rho}_\varepsilon$	$\hat{\bar{\varepsilon}}^l$	$\hat{\bar{\varepsilon}}$	Corr$(\hat{U}^l_{it}, \hat{U}_{it})$	$\hat{\beta}_l$ Ratio
31	0.939	0.952	2.526	1.370	0.306	0.086 / 0.236
32	0.949	0.944	1.530	1.435	0.360	0.132 / 0.468
33	0.933	0.952	1.323	1.460	0.350	0.179 / 0.755

*$\hat{\rho}^l_\varepsilon$, $\hat{\bar{\varepsilon}}^l$: Corollary C.1's independent deconvolution of (49), with labor as the flexible input; $\hat{\rho}_\varepsilon$, $\hat{\bar{\varepsilon}}$: the main, materials-based Stage 1 of Table 1. Corr$(\hat{U}^l_{it}, \hat{U}_{it})$ is the Pearson correlation of the two independently recovered technology types. The $\hat{\beta}_l$ ratio is labor as a flexible input relative to labor as a predetermined input, estimated from the main model and obtained from Table 2.

The persistence parameter $\hat{\rho}^l_\varepsilon$ agrees closely with the materials-based $\hat{\rho}_\varepsilon$ in every industry (within 0.005-0.02), supporting the AR(1) persistence of the ex-post shock as a robust feature of the data, recoverable from either flexible input's own share equation. The normalizing constant $\hat{\bar{\varepsilon}}^l$ is farther from its materials-based counterpart, particularly in NAICS 31 (a factor of 1.8), and the recovered technology types $\hat{U}^l_{it}$ and $\hat{U}_{it}$ are positively but only moderately correlated (0.31-0.36), not the near-unity agreement that population-exactness under Assumption 4.1 would imply.

$\hat{\beta}_l$'s two independent estimates diverge more starkly: the direct, labor-based Stage 1 estimate is systematically *smaller* than when estimated in Stage 2 (roughly 2.8-4.2). What survives across the two estimates is the ranking across industries. NAICS 31 has the smallest labor elasticity and NAICS 33 has the largest. We read this pattern as an informative diagnostic and not necessarily a formal overidentification test for unobserved technology types. This comparison demonstrates that the timing assumptions on labor critically change the levels of the coefficient estimates; however, robust, real features of the data (e.g. persistence of ex-post shocks and ranking across industries) are preserved. A formal overidentification test is left for future work.

7 Conclusion

We propose a method for identifying and estimating a gross-output production function in which every elasticity is heterogeneous and interacts freely with an unobserved firm-specific technology type. Identification proceeds by a Kotlarski-type conditional deconvolution of the materials cost-share equation, exploiting a two-period repeated-measurement system and a set of conditional independence restrictions that are the panel analogue of those used elsewhere in the measurement-error and random-coefficients literature. The resulting two-stage estimator is computationally attractive relative to a fully joint estimation of the nonseparable surface. Stage 1 uses only kernel regression and a closed-form Hermite-sieve characteristic-function representation, and Stage 2 resembles the sieve-GMM second stage already standard in Gandhi et al. (2020). An application to a widely used U.S. manufacturing dataset shows that materials elasticities rise systematically with the recovered technology type in every industry we examine, while capital and labor elasticities tend to fall. These are patterns of heterogeneity that a homogeneous-coefficient production function cannot express.

This paper contributes to the growing literature on production functions with unobserved heterogeneity by identifying and estimating that heterogeneity directly from the flexible input's own first-order condition. Two important restrictions remain for future work. First, when labor is treated as a second flexible input dependent on the *same* technology type U_{it} as materials, Section 6.3's comparison between the two independently recovered objects is only qualitative; deriving a formal overidentification test, with a corresponding sampling distribution, is left to future work. Second, labor may instead depend on its own technology type, entirely separate from materials' U_{it}, a specification this paper does not consider; deriving the corresponding identification argument, together with a formal test of this alternative against the shared-type specification above, is a further direction for future work.

References

Acemoglu, D. (2002). Directed technical change. *The Review of Economic Studies*, 69(4):781–809.

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

Ackerberg, D., Chen, X., and Hahn, J. (2012). A practical asymptotic variance estimator for two-step semiparametric estimators. *Review of Economics and Statistics*, 94(2):481–498.

Ackerberg, D., Chen, X., Hahn, J., and Liao, Z. (2014). Asymptotic efficiency of semiparametric two-step GMM. *The Review of Economic Studies*, 81(3):919–943.

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843.

Balat, J., Brambilla, I., and Sasaki, Y. (2019). Heterogeneous firms: skilled-labor productivity and the destination of exports. Working paper.

Bonhomme, S. and Robin, J.-M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *Review of Economic Studies*, 77(2):491–533.

Demirer, M. (2020). Production function estimation with factor augmenting technology: An application to markups. Working paper.

Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Working paper.

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.

Griliches, Z. and Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31(1):93–118.

Hu, Y., Huang, G., and Sasaki, Y. (2020). Estimating production functions with robustness against errors in the proxy variables. *Journal of Econometrics*, 215(2):375–398.

Kasahara, H., Schrimpf, P., and Suzuki, M. (2017). Identification and estimation of production function with unobserved heterogeneity. Working paper.

Kato, K., Sasaki, Y., and Ura, T. (2021). Robust inference in deconvolution. *Quantitative Economics*, 12(1):109–142.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics*, 20(1):69–76.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–341.

Li, T. and Sasaki, Y. (2017). Constructive identification of heterogeneous elasticities in the cobb-douglas production function. Working paper.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65(2):139–165.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, 98(10):1955–1968.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263.

Raval, D. R. (2019). The micro elasticity of substitution and non-neutral technology. *The RAND Journal of Economics*, 50(1):147–167.

Schennach, S. M. (2004). Estimation of nonlinear models with measurement error. *Econometrica*, 72(1):33–75.

Song, S., Schennach, S. M., and White, H. (2015). Estimating nonseparable models with mismeasured endogenous variables. *Quantitative Economics*, 6(3):749–794.

Appendix

A Identification

Derivation of the share equation (7):
M_{it} maximizes expected profit conditional on $\mathcal{I}_{it} = \{K_{it}, L_{it}, \omega_{it}, U_{it}, P_{it}, P_{it}^M\}$,

$$\max_{M_{it}} \Big\{ P_{it}\, E[Y_{it} \mid \mathcal{I}_{it}] - P_{it}^M M_{it} \Big\} = \max_{M_{it}} \Big\{ P_{it}\, Y_{it}^0\, E[e^{\varepsilon_{it}} \mid \mathcal{I}_{it}] - P_{it}^M M_{it} \Big\},$$

using $Y_{it} = Y_{it}^0 e^{\varepsilon_{it}}$ with $Y_{it}^0 = e^{F(K_{it}, L_{it}, M_{it}, U_{it}) + \omega_{it}}$ nonrandom given $\mathcal{I}_{it}$. The first-order condition is

$$P_{it}\, \frac{\partial Y_{it}^0}{\partial M_{it}}\, E[e^{\varepsilon_{it}} \mid \mathcal{I}_{it}] - P_{it}^M = 0.$$

Since $\varepsilon_{it} \perp \mathcal{I}_{it}$, $E[e^{\varepsilon_{it}} \mid \mathcal{I}_{it}] = E[e^{\varepsilon_{it}}] \equiv \bar{\varepsilon}$. Because $Y_{it}^0 = e^{F(K_{it}, L_{it}, M_{it}, U_{it}) + \omega_{it}}$,

$$\frac{\partial Y_{it}^0}{\partial M_{it}} = \frac{\partial F}{\partial \ln M_{it}} \cdot \frac{Y_{it}^0}{M_{it}} = \beta_m(k_{it}, l_{it}, m_{it}, U_{it})\, \frac{Y_{it}^0}{M_{it}}.$$

Substituting into the first-order condition and rearranging,

$$\beta_m(k_{it}, l_{it}, m_{it}, U_{it})\, \bar{\varepsilon} = \frac{P_{it}^M M_{it}}{P_{it} Y_{it}^0} = \frac{P_{it}^M M_{it}}{P_{it} Y_{it}}\, e^{\varepsilon_{it}} \equiv S_{it}\, e^{\varepsilon_{it}}, \qquad S_{it} \equiv \frac{P_{it}^M M_{it}}{P_{it} Y_{it}}.$$

Taking logs gives (7). ■

Proof of Theorem 4.1:
Step 1: Identifying ρ_ε.
Iterating the AR(1) process by one more period,

$$\varepsilon_{it+2} = \rho_\varepsilon \varepsilon_{it+1} + \zeta_{it+2} = (\rho_\varepsilon)^2 \varepsilon_{it} + \rho_\varepsilon \zeta_{it+1} + \zeta_{it+2}, \qquad s_{it+2} = Z_{it+2} - (\rho_\varepsilon)^2 \varepsilon_{it} - \rho_\varepsilon \zeta_{it+1} - \zeta_{it+2}.$$

Under Assumption 4.1(d), $U_{it+1} \perp U_{it} \mid X$ and, applied one period further, $U_{it+2} \perp (U_{it}, U_{it+1}) \mid X$, together with $\varepsilon_{it} \perp (U_{it+1}, U_{it+2}) \mid X$ and $U_{it} \perp (\zeta_{it+1}, \zeta_{it+2}) \mid X$, every term in $\text{Cov}(s_{it}, s_{it+1} \mid X)$ and $\text{Cov}(s_{it}, s_{it+2} \mid X)$ involving Z_{it}, Z_{it+1}, Z_{it+2}, ζ_{it+1}, or ζ_{it+2} vanishes, leaving only the autocovariance of ε_{it} with itself:

$$\text{Cov}(s_{it}, s_{it+1} \mid X) = \rho_\varepsilon \text{Var}(\varepsilon_{it} \mid X), \qquad \text{Cov}(s_{it}, s_{it+2} \mid X) = (\rho_\varepsilon)^2 \text{Var}(\varepsilon_{it} \mid X). \tag{35}$$

Dividing the second relation in (35) by the first, $\text{Var}(\varepsilon_{it} \mid X)$ cancels and ρ_ε is identified in closed form, observed from the data, as

$$\rho_\varepsilon = \frac{\mathrm{Cov}(s_{it}, s_{it+2} \mid X)}{\mathrm{Cov}(s_{it}, s_{it+1} \mid X)}. \tag{36}$$

Step 2: The Kotlarski Factorization.
With ρ_ε identified, rescale the second period $\tilde{s}_{it+1} \equiv s_{it+1}/\rho_\varepsilon = \tilde{Z}_{it+1} - \varepsilon_{it} - \tilde{\zeta}_{it+1}$. The characteristic function of $(s_{it}, \tilde{s}_{it+1})$ conditional on X can be written as

$$\begin{aligned}
&\Psi_{(s_{it},\tilde{s}_{it+1})|X}(\gamma_1, \gamma_2|X) \\
=&\mathbb{E}\big[\exp\{\mathrm{i}[\gamma_1 Z_{it} + \gamma_2 \tilde{Z}_{it+1} - \gamma_2 \tilde{\zeta}_{it+1} - (\gamma_1+\gamma_2)\varepsilon_{it}]\} \mid X\big] \\
=&\mathbb{E}[\exp\{\mathrm{i}\gamma_2 \tilde{Z}_{it+1}\} \mid X]\,\mathbb{E}[\exp\{-\mathrm{i}\gamma_2 \tilde{\zeta}_{it+1}\} \mid X]\,\mathbb{E}[\exp\{\mathrm{i}\gamma_1 Z_{it}\} \mid X]\,\mathbb{E}[\exp\{-\mathrm{i}(\gamma_1+\gamma_2)\varepsilon_{it}\} \mid X] \\
=&\phi_{\tilde{Z}_{it+1}|X}(\gamma_2|X)\,\phi_{\tilde{\zeta}_{it+1}|X}(-\gamma_2|X)\,\phi_{Z_{it}|X}(\gamma_1|X)\,\phi_{\varepsilon_{it}|X}(-\gamma_1-\gamma_2|X),
\end{aligned} \tag{37}$$

where the second equality factors the joint expectation into four separate terms using Assumption 4.1(d)(i)-(iii): (d)(i) and (d)(iii) separate Z_{it} and ε_{it} from $(\tilde{Z}_{it+1}, \tilde{\zeta}_{it+1})$, and (d)(ii) further separates $\tilde{Z}_{it+1}$ from $\tilde{\zeta}_{it+1}$. Equation (37) is Kotlarski's identity (Kotlarski, 1967), brought into econometrics by Li and Vuong (1998) and adapted to this panel/AR(1) setting following Evdokimov (2010). Given two noisy measurements sharing a common component (ε_{it}) with mutually independent remaining components, the marginal characteristic function of Z_{it} is identified as

$$\phi_{Z_{it}|X}(\gamma|X) = \exp\left(\int_0^\gamma \frac{\mathrm{i}\,\mathbb{E}\big[s_{it}\exp\big(\mathrm{i}\alpha(s_{it}-\tilde{s}_{it+1})\big) \mid X\big]}{\phi_{s_{it}-\tilde{s}_{it+1}|X}(\alpha|X)}\,d\alpha \;+\; \mathrm{i}\gamma\,\mathbb{E}[\varepsilon_{it} \mid X]\right), \tag{38}$$

identified up to the location normalization $\mathbb{E}[\varepsilon_{it} \mid X] = 0$ of Assumption 4.1(c), and given that $\phi_{Z_{it}|X}$ and $\phi_{s_{it}-\tilde{s}_{it+1}|X}$ are non-vanishing according to Assumption 4.1(e).

Since $s_{it} = Z_{it} - \varepsilon_{it}$, $\phi_{s_{it}|X}(\gamma|X) = \phi_{Z_{it}|X}(\gamma|X)\,\phi_{\varepsilon_{it}|X}(-\gamma|X)$; dividing this out of the characteristic function of the data and evaluating at $-\gamma$,

$$\phi_{\varepsilon_{it}|X}(\gamma|X) = \frac{\phi_{s_{it}|X}(-\gamma|X)}{\phi_{Z_{it}|X}(-\gamma|X)}, \tag{39}$$

recovers the full distribution of the ex-post shock ε_{it}, given X. This proves identification of $\phi_{Z_{it}|X}(\gamma|X)$ and $\phi_{\varepsilon_{it}|X}(\gamma|X)$; the CDF $F_{Z_{it}|X}(z|X)$ then follows by Fourier (Gil-Pelaez) inversion of (38),

$$F_{Z_{it}|X}(z|X) = \frac{1}{2} - \lim_{\upsilon \to \infty} \int_{-\upsilon}^{\upsilon} \frac{e^{-\mathrm{i}sz}}{2\pi \mathrm{i}s} \phi_{Z_{it}|X}(s|X) ds. \tag{40}$$

Step 3: Identifying $\bar{\varepsilon}$.

$\bar{\varepsilon} = \mathbb{E}[e^{\varepsilon_{it}}]$ is the moment generating function of ε_{it} evaluated at 1. The characteristic function analytically continues to the moment generating function via $M_\varepsilon(1) = \phi_{\varepsilon_{it}|X}(-\mathrm{i} \mid X)$ whenever $\mathbb{E}[e^{\varepsilon_{it}}] < \infty$. Evaluating (39) at $\gamma = -\mathrm{i}$,

$$\bar{\varepsilon} = \phi_{\varepsilon_{it}|X}(-\mathrm{i} \mid X) = \frac{\phi_{s_{it}|X}(\mathrm{i} \mid X)}{\phi_{Z_{it}|X}(\mathrm{i} \mid X)}. \tag{41}$$

The numerator simplifies immediately since $\phi_{s_{it}|X}(\mathrm{i} \mid X) = \mathbb{E}[e^{\mathrm{i}(\mathrm{i})s_{it}} \mid X] = \mathbb{E}[e^{-s_{it}} \mid X] = \mathbb{E}[1/S_{it} \mid X]$, the conditional mean of the reciprocal level materials share. The denominator follows from evaluating (38) at $\gamma = \mathrm{i}$; the location normalization $\mathbb{E}[\varepsilon_{it} \mid X] = 0$ sets its second term to zero, leaving a contour integral from 0 to i along the imaginary axis, using the same integrand as (38). Combining this with (41) identifies $\bar{\varepsilon}$ from observed data alone.

Step 4: Identification of $\beta_m(k, l, m; \tau)$.

Since the quantile function is the inverse of the CDF (40), this fact, together with the results from Step 3, identifies, for every $\tau \in (0, 1)$,

$$\ln \beta_m(k, l, m; \tau) = F^{-1}_{Z_{it}|X}\big(\tau \mid X = (k, l, m)\big) - \ln \bar{\varepsilon}, \tag{42}$$

which is the materials elasticity surface at every observed level of (k, l, m) and every quantile τ of the (conditional) technology-type distribution.

Step 5: Identification of Firm Specific ε_{it} and U_{it}.

By the same argument as (37) with the roles of t and $t+1$ reversed, $\phi_{Z_{it+1}|X}$ and $\phi_{\zeta_{it+1}|X}$ are identified alongside $\phi_{Z_{it}|X}$ and $\phi_{\varepsilon_{it}|X}$; let $f_{Z_{it+1}-\zeta_{it+1}|X}$ denote the density corresponding to their convolution, which is itself identified since it is the Fourier transform of the (identified) product $\phi_{Z_{it+1}|X} \cdot \phi_{\zeta_{it+1}|X}$. Given the measurement equations (10), the posterior mean of ε_{it} given the firm's own observed (s_{it}, s_{it+1}, X) is

$$\hat{\varepsilon}_{it} = \mathbb{E}[\varepsilon_{it} \mid s_{it}, s_{it+1}, X] = \frac{\displaystyle\int e\, f_{\varepsilon_{it}|X}(e \mid X)\, f_{Z_{it}|X}(s_{it} + e \mid X)\, f_{Z_{it+1}-\zeta_{it+1}|X}\big(s_{it+1} + \rho_\varepsilon e \mid X\big)\, de}{\displaystyle\int f_{\varepsilon_{it}|X}(e \mid X)\, f_{Z_{it}|X}(s_{it} + e \mid X)\, f_{Z_{it+1}-\zeta_{it+1}|X}\big(s_{it+1} + \rho_\varepsilon e \mid X\big)\, de}, \tag{43}$$

a function only of (s_{it}, s_{it+1}, X) and the densities already shown to be identified in Steps 1-3, and hence itself identified. Applying the identified CDF $F_{Z_{it}|X}$ to the resulting $\hat{Z}_{it} = s_{it} + \hat{\varepsilon}_{it}$ identifies the firm's own position in the type distribution,

$$\hat{U}_{it} = F_{Z_{it}|X}\big(s_{it} + \hat{\varepsilon}_{it} \mid X\big). \qquad \blacksquare \tag{44}$$

An important conclusion of this theorem is that it nonparametrically identifies not only the population elasticity surface $\beta_m(k, l, m; \tau)$, but, through Step 5, each individual firm-period's own materials elasticity $\hat{\beta}_m(k_{it}, l_{it}, m_{it}, U_{it}) = \exp(s_{it} + \hat{\varepsilon}_{it} - \ln \bar{\varepsilon})$ directly from (7), which is what makes it possible in Section 5 to use $\hat{U}_{it}$ as a generated regressor in a second stage that recovers firm-period-level capital and labor elasticities as well.

B Estimation

This appendix collects the derivations behind Section 5's two-stage estimator.

The linear moment restriction used in Stage 1(b).
We verify the following claim using the same device Kato et al. (2021) use to prove their own moment restriction: the true $\phi_{Z_{it}|X}$ satisfies, for every γ,

$$\mathbb{E}\Big[\big(\mathrm{i}\, s_{it}\, \phi_{Z_{it}|X}(\gamma|X) - \phi'_{Z_{it}|X}(\gamma|X)\big)\, e^{\mathrm{i}\gamma D_{it}} \;\Big|\; X\Big] = 0, \qquad D_{it} \equiv s_{it} - \tilde{s}_{it+1}. \tag{45}$$

Verification of (45)*:*
Because $\phi_{Z_{it}|X}(\gamma|X)$ and $\phi'_{Z_{it}|X}(\gamma|X)$ are nonrandom given X, (45) is equivalent to

$$\mathrm{i}\, \phi_{Z_{it}|X}(\gamma|X)\, \mathbb{E}[s_{it} e^{\mathrm{i}\gamma D_{it}} \mid X] = \phi'_{Z_{it}|X}(\gamma|X)\, \mathbb{E}[e^{\mathrm{i}\gamma D_{it}} \mid X].$$

We verify this by differentiating (37) with respect to γ_1, holding γ_2 fixed, and then setting $\gamma_2 = -\gamma_1$.

Step 1: differentiate both sides of (37) *in* γ_1. The left side is $\mathbb{E}[e^{\mathrm{i}(\gamma_1 s_{it} + \gamma_2 \tilde{s}_{it+1})} \mid X]$ by definition, so

$$\partial_{\gamma_1} \mathrm{LHS} = \mathrm{i}\, \mathbb{E}\big[s_{it}\, e^{\mathrm{i}(\gamma_1 s_{it} + \gamma_2 \tilde{s}_{it+1})} \mid X\big].$$

On the right side, $\phi_{\tilde{Z}_{it+1}|X}(\gamma_2|X)$ and $\phi_{\tilde{\zeta}_{it+1}|X}(-\gamma_2|X)$ do not depend on γ_1; the product rule on the remaining two factors gives

$$\begin{aligned}\partial_{\gamma_1} \mathrm{RHS} = \phi_{\tilde{Z}_{it+1}|X}(\gamma_2|X)\, \phi_{\tilde{\zeta}_{it+1}|X}(-\gamma_2|X) \Big[&\phi'_{Z_{it}|X}(\gamma_1|X)\, \phi_{\varepsilon_{it}|X}(-\gamma_1 - \gamma_2|X) \\ &- \phi_{Z_{it}|X}(\gamma_1|X)\, \phi'_{\varepsilon_{it}|X}(-\gamma_1 - \gamma_2|X)\Big],\end{aligned}$$

using $\frac{d}{d\gamma_1}\phi_{\varepsilon_{it}|X}(-\gamma_1 - \gamma_2|X) = -\phi'_{\varepsilon_{it}|X}(-\gamma_1 - \gamma_2|X)$ by the chain rule.

Step 2: set $\gamma_2 = -\gamma_1$. The left side becomes $\mathrm{i}\, \mathbb{E}[s_{it} e^{\mathrm{i}\gamma_1 D_{it}} \mid X]$. On the right, $\phi_{\varepsilon_{it}|X}(-\gamma_1 - \gamma_2|X) \to \phi_{\varepsilon_{it}|X}(0|X) = 1$, and $\phi'_{\varepsilon_{it}|X}(-\gamma_1 - \gamma_2|X) \to \phi'_{\varepsilon_{it}|X}(0|X) = \mathrm{i}\, \mathbb{E}[\varepsilon_{it} \mid X] = 0$ by the location normalization (Assumption 4.1(c)), so the second bracketed term vanishes, leaving

$$\partial_{\gamma_1} \mathrm{RHS}\big|_{\gamma_2 = -\gamma_1} = \phi_{\tilde{Z}_{it+1}|X}(-\gamma_1|X)\, \phi_{\tilde{\zeta}_{it+1}|X}(\gamma_1|X)\, \phi'_{Z_{it}|X}(\gamma_1|X).$$

Step 3: eliminate $\phi_{\tilde{Z}_{it+1}|X}\phi_{\tilde{\zeta}_{it+1}|X}$. Setting $\gamma_2 = -\gamma_1$ in (37) gives $\phi_{s_{it} - \tilde{s}_{it+1}|X}(\gamma_1|X) = \phi_{\tilde{Z}_{it+1}|X}(-\gamma_1|X)\, \phi_{\tilde{\zeta}_{it+1}|X}(\gamma_1|X)\, \phi_{Z_{it}|X}(\gamma_1|X)$, so, dividing by the non-vanishing $\phi_{Z_{it}|X}(\gamma_1|X)$

(Assumption 4.1(e)),

$$\phi_{\tilde{Z}_{it+1}|X}(-\gamma_1|X)\,\phi_{\tilde{\zeta}_{it+1}|X}(\gamma_1|X) = \frac{\phi_{s_{it}-\tilde{s}_{it+1}|X}(\gamma_1|X)}{\phi_{Z_{it}|X}(\gamma_1|X)}.$$

Step 4: combine. Substituting Step 3 into Step 2, equating to the differentiated left side, and writing γ for γ_1,

$$\mathrm{i}\,\phi_{Z_{it}|X}(\gamma|X)\,\mathbb{E}[s_{it}e^{\mathrm{i}\gamma D_{it}} \mid X] = \phi_{s_{it}-\tilde{s}_{it+1}|X}(\gamma|X)\,\phi'_{Z_{it}|X}(\gamma|X) = \mathbb{E}[e^{\mathrm{i}\gamma D_{it}} \mid X]\,\phi'_{Z_{it}|X}(\gamma|X),$$

using $\phi_{s_{it}-\tilde{s}_{it+1}|X}(\gamma|X) = \mathbb{E}[e^{\mathrm{i}\gamma D_{it}} \mid X]$ by definition, which is the form of (45) stated above. ■

Equation (45) holds without differentiating (38) and without the non-vanishing-characteristic-function condition that the formula requires. This is the same property Kato et al. (2021) establish for their own moment restriction by a similar argument. Section 5.1's Stage 1(b) uses (45) directly for estimation.

The Closed-Form Coefficient Solution.

Stacking (45) over the frequency grid $\gamma_1, \ldots, \gamma_L$ gives the homogeneous linear system $\hat{A}(X)c(X) \approx 0$ of Section 5.1. Estimating $c(X)$ minimizes $\|\hat{A}(X)c\|^2$ subject to the density-integrates-to-one constraint $b'c(X) = 1$, which pins down scale and rules out the trivial solution $c(X) = 0$. We derive the closed-form minimizer explicitly.

Step 1: Set Up the Lagrangian. Writing the objective as $\|\hat{A}(X)c\|^2 = c'\hat{A}(X)'\hat{A}(X)c = c'M(X)c$ for $M(X) \equiv \hat{A}(X)'\hat{A}(X)$, the constrained minimization problem has Lagrangian

$$\mathcal{L}(c,\lambda) = c'M(X)c - \lambda\big(b'c(X) - 1\big).$$

The two first-order conditions of $\mathcal{L}$ deliver the estimator directly. Since $M(X)$ is symmetric, $\partial\big(c'M(X)c\big)/\partial c = 2M(X)c$, so the first-order condition in c is $\partial\mathcal{L}/\partial c = 2M(X)c(X) - \lambda b = 0$, giving $c(X) = \frac{\lambda}{2}M(X)^{-1}b$ (with $M(X)$ invertible provided $\hat{A}(X)$ has full column rank); the first-order condition in λ, $\partial\mathcal{L}/\partial\lambda = -\big(b'c(X) - 1\big) = 0$, just returns the constraint $b'c(X) = 1$. Substituting the first expression into the second pins down the multiplier,

$$b'\left(\frac{\lambda}{2}M(X)^{-1}b\right) = 1 \quad \Longrightarrow \quad \frac{\lambda}{2} = \frac{1}{b'M(X)^{-1}b},$$

and substituting back into $c(X) = \frac{\lambda}{2} M(X)^{-1} b$ gives the closed form stated in Section 5.1,

$$\hat{c}(X) = \frac{M(X)^{-1} b}{b' M(X)^{-1} b}. \tag{46}$$

Because $M(X) = \hat{A}(X)' \hat{A}(X)$ is positive semidefinite, $c' M(X) c$ is convex in c, so the stationary point identified above is the global minimizer. ■

The CDF Recursion Used in Stage 1(b). Since Stage 1(b) fits $\phi_{Z_{it}|X}$ in the Hermite-sieve basis, the same coefficients $\hat{c}(X)$ give the CDF as the closed finite sum used in Section 5.1,

$$\hat{F}_{Z_{it}|X}(z|X) = \sum_{j=0}^{J} \hat{c}_j(X) \Psi_j(z), \qquad \Psi_j(z) \equiv \int_{-\infty}^{z} \psi_j(v)\, dv, \tag{47}$$

where Ψ_j satisfies the two-term recursion

$$\Psi_j(z) = -\sqrt{\frac{2}{j}}\, \psi_{j-1}(z) + \sqrt{\frac{j-1}{j}}\, \Psi_{j-2}(z), \qquad j \geq 1, \qquad \Psi_{-1}(z) \equiv 0, \tag{48}$$

bottoming out at $\Psi_0(z) = \sqrt{2}\, \pi^{1/4} \Phi(z)$, where Φ is the standard normal CDF. Because the recursion steps down by 2 in j, every even-order Ψ_j embeds one evaluation of $\Phi(z)$, while every odd-order Ψ_j bottoms out at the elementary $\Psi_1(z) = -\sqrt{2}\, \psi_0(z)$. Φ is evaluated by a standard library routine, so (48) gives every Ψ_j exactly and cheaply.

Derivation of (48)*:* Let $I_j(z) \equiv \int_{-\infty}^{z} H_j(v) e^{-v^2/2}\, dv$, so that $\Psi_j(z) = (2^j j! \sqrt{\pi})^{-1/2} I_j(z)$. The Hermite polynomials satisfy the derivative relation $H'_{j-1}(z) = 2(j-1) H_{j-2}(z)$ and the three-term recurrence $H_j(z) = 2z H_{j-1}(z) - 2(j-1) H_{j-2}(z)$, the latter rearranged as $z H_{j-1}(z) = \frac{1}{2} H_j(z) + (j-1) H_{j-2}(z)$. Differentiating $-H_{j-1}(z) e^{-z^2/2}$ and substituting these two relations,

$$\begin{aligned}
\frac{d}{dz}\big[-H_{j-1}(z) e^{-z^2/2}\big] &= -H'_{j-1}(z) e^{-z^2/2} + z H_{j-1}(z) e^{-z^2/2} \\
&= -2(j-1) H_{j-2}(z) e^{-z^2/2} + \Big[\tfrac{1}{2} H_j(z) + (j-1) H_{j-2}(z)\Big] e^{-z^2/2} \\
&= \tfrac{1}{2} H_j(z) e^{-z^2/2} - (j-1) H_{j-2}(z) e^{-z^2/2},
\end{aligned}$$

which rearranges to

$$H_j(z) e^{-z^2/2} = 2 \frac{d}{dz}\big[-H_{j-1}(z) e^{-z^2/2}\big] + 2(j-1) H_{j-2}(z) e^{-z^2/2}.$$

Integrating from $-\infty$ to z, the boundary term vanishes because $H_{j-1}(v)e^{-v^2/2} \to 0$ as $v \to -\infty$, leaving

$$I_j(z) = -2H_{j-1}(z)e^{-z^2/2} + 2(j-1)I_{j-2}(z), \qquad j \geq 2,$$

with base cases, by direct integration, $I_0(z) = \int_{-\infty}^{z} e^{-v^2/2}dv = \sqrt{2\pi}\,\Phi(z)$ and $I_1(z) = \int_{-\infty}^{z} 2ve^{-v^2/2}dv = -2e^{-z^2/2}$. Substituting $H_j(z)e^{-z^2/2} = (2^j j!\sqrt{\pi})^{1/2}\psi_j(z)$ and $I_j(z) = (2^j j!\sqrt{\pi})^{1/2}\Psi_j(z)$ into the recursion for I_j and dividing by $(2^j j!\sqrt{\pi})^{1/2}$,

$$\Psi_j(z) = -\frac{2\big(2^{j-1}(j-1)!\sqrt{\pi}\big)^{1/2}}{\big(2^j j!\sqrt{\pi}\big)^{1/2}}\,\psi_{j-1}(z) + \frac{2(j-1)\big(2^{j-2}(j-2)!\sqrt{\pi}\big)^{1/2}}{\big(2^j j!\sqrt{\pi}\big)^{1/2}}\,\Psi_{j-2}(z),$$

and simplifying each coefficient. The $\sqrt{\pi}$ factors cancel, leaving $2\big(2^{j-1}(j-1)!\big)^{1/2}/\big(2^j j!\big)^{1/2} = \sqrt{2/j}$ and $2(j-1)\big(2^{j-2}(j-2)!\big)^{1/2}/\big(2^j j!\big)^{1/2} = \sqrt{(j-1)/j}$. This gives (48) for $j \geq 2$. Extending to $j = 1$ with the convention $\Psi_{-1}(z) \equiv 0$ gives $\Psi_1(z) = -\sqrt{2}\,\psi_0(z)$, which is equivalent to $I_1(z) = -2e^{-z^2/2}$. ■

C Extension: A Second Flexible Input

The baseline model of Sections 3-5 treats materials as the only flexible input: capital and labor are recovered jointly, in Stage 2, as a single nonseparable object $C(k, l, U)$. In many applications labor is chosen with as much within-year flexibility as materials – hired and laid off period by period, well before the ex-post shock ε_{it} is realized – so it is natural to ask what changes if labor is treated as a second flexible input rather than folded into Stage 2. This section works out that extension in theory: it shows that both flexible inputs' elasticities are recovered in Stage 1, using a ratio identity that turns out to be robust to the ex-post shock, and that Stage 2 then reduces to capital alone. The reduced Stage 2 itself has no accompanying empirical implementation here; Section 6.3 of the main text, however, does implement and apply this section's overidentification check (Corollary C.1) to the U.S. application.

C.1 Labor as a Second Flexible Input

Suppose labor L_{it}, like materials M_{it}, is chosen after $(K_{it}, \omega_{it}, U_{it})$ are known but before ε_{it} is realized, so that the same timing argument behind (7) applies to labor directly. Define the labor log-elasticity $\beta_l(k_{it}, l_{it}, m_{it}, U_{it}) \equiv \partial f_t(k_{it}, l_{it}, m_{it}, U_{it})/\partial \ln l_{it}$. The static first-order-condition argument (Appendix A) applied to L_{it} instead of M_{it} gives a second share equation,

$$s^l_{it} \equiv \ln(S^l_{it}) = \ln\big(\beta_l(k_{it}, l_{it}, m_{it}, U_{it})\big) + \ln\bar{\varepsilon} - \varepsilon_{it}, \qquad S^l_{it} \equiv \frac{P^L_{it} L_{it}}{P_{it} Y_{it}}, \tag{49}$$

with the same normalizing constant $\bar{\varepsilon}$ and ex-post shock ε_{it} as (7), since both are shocks to the same realized output Y_{it}.

Corollary C.1 *Under Assumption 4.1, with M_{it} replaced by L_{it}, the objects ρ_ε, $\phi_{Z^l_{it}|X}(s|X)$ (the characteristic function of $Z^l_{it} \equiv \ln\beta_l(\cdot) + \ln\bar{\varepsilon}$), $\phi_{\varepsilon_{it}|X}(s|X)$, $\bar{\varepsilon}$, and the labor elasticity surface $\beta_l(k, l, m; \tau)$ are nonparametrically identified by the argument of Theorem 4.1, applied verbatim to (49).*

Corollary C.1 means labor's elasticity surface, in theory, could be recovered by an entirely independent second deconvolution, run on (49) exactly as Stage 1 already runs on (7). But because ε_{it} and $\bar{\varepsilon}$ enter (7) and (49) identically, a second deconvolution is unnecessary for recovering labor's elasticity. Differencing the two share equations cancels both terms exactly.

Lemma C.1 (Ratio identity) *Under the timing assumption of this section, for every firm-period with* $S^m_{it} > 0$ *and* $S^l_{it} > 0$*,*

$$\ln\left(\frac{\beta_l(k_{it}, l_{it}, m_{it}, U_{it})}{\beta_m(k_{it}, l_{it}, m_{it}, U_{it})}\right) = s^l_{it} - s_{it} = \ln\left(\frac{S^l_{it}}{S^m_{it}}\right). \tag{50}$$

Proof: Subtract (7) from (49): $s^l_{it} - s_{it} = \left[\ln\beta_l + \ln\bar{\varepsilon} - \varepsilon_{it}\right] - \left[\ln\beta_m + \ln\bar{\varepsilon} - \varepsilon_{it}\right] = \ln\beta_l - \ln\beta_m$. ■

Given $\hat{\varepsilon}_{it}$ and $\hat{U}_{it}$ from Stage 1's materials-only deconvolution (Section 5.1), Lemma C.1 recovers labor's own firm-period elasticity directly from the two observed revenue shares,

$$\hat{\beta}_l(k_{it}, l_{it}, m_{it}, \hat{U}_{it}) = \hat{\beta}_m(k_{it}, l_{it}, m_{it}, \hat{U}_{it}) \cdot \frac{S^l_{it}}{S^m_{it}}, \tag{51}$$

so that both flexible inputs' elasticities are available from Stage 1: $\hat{\beta}_m$ from the deconvolution as before, $\hat{\beta}_l$ from (51). Running the independent deconvolution of Corollary C.1 on (49) could, however, serve as a natural overidentification test on the model by checking that $\hat{\rho}_\varepsilon$, $\hat{\bar{\varepsilon}}$, and $\hat{\beta}_l$ agree with (51).

C.2 A Reduced Second Stage: Capital Alone

With both flexible inputs' elasticity surfaces in hand, Stage 2 no longer needs to recover labor. Define $C(k, U) \equiv F(k, l_0, m_0, U)$, the production function net of both flexible inputs' contributions, and purge output of the integrated contribution of M_{it} and L_{it} along any path from (l_0, m_0) to (l_{it}, m_{it}) holding (k_{it}, U_{it}) constant.

$$\check{y}_{it} \equiv y_{it} - \hat{\varepsilon}_{it} - \int_{(l_0, m_0)}^{\widehat{(l_{it}, m_{it})}} \left[\hat{\beta}_m(k_{it}, \sigma_l, \sigma_m, \hat{U}_{it})\, d\ln\sigma_m + \hat{\beta}_l(k_{it}, \sigma_l, \sigma_m, \hat{U}_{it})\, d\ln\sigma_l\right] = C(k_{it}, \hat{U}_{it}) + \omega_{it}. \tag{52}$$

Approximating $C(k, U; \theta)$ with a sieve in (k, U), the GMM construction of Section 5.1 applies directly, with the instrument vector reduced to $k^j_{it+1}\hat{U}^p_{it}$ and delivers $\hat{\beta}_k(k_{it}, \hat{U}_{it}) = \partial C / \partial \ln k|_{\hat{\theta}}$ by differentiating the fitted sieve as before.