Nonparametric Identification and Estimation of Production Functions Invariant to Productivity Dynamics*

Rentaro Utamaru†

April 2, 2026

Click here for the latest version

Abstract

I identify the production function and the distribution of productivity without assuming that productivity follows a Markov process. Existing proxy variable estimators require this untestable assumption; I replace it with conditional independence across intermediate input demands, a static condition with transparent microfoundations in input market segmentation. Three intermediate inputs (materials, energy, water) serve as noisy measurements of the same latent variable, enabling nonparametric identification from a single cross-section. I characterize the residual indeterminacy completely and close it by two routes, neither requiring dynamic assumptions; I develop a GMM estimator and establish consistency and asymptotic normality. In Monte Carlo simulations where the Markov assumption is systematically varied, the proposed estimator is unbiased while Ackerberg, Caves, and Frazer (2015) exhibits persistent bias. Applying the estimator to 502 Japanese manufacturing industries, the proposed method yields industry-level markups with a median of 0.93, while the standard method produces a median of 1.03; the gap is consistent with the upward Markov misspecification bias documented in the simulations. The recovered productivity measures show substantially stronger associations with economic fundamentals than those from the standard method, consistent with a higher signal-to-noise ratio from separating input-specific demand shocks.

Keywords: Production Function, Productivity, Nonparametric Identification, Measurement Error, GMM
JEL Classification Codes: C13, C14, D24, L11

*I am grateful to Yasutora Watanabe, Yuta Toyama, Shosei Sakaguchi, Hidehiko Ichimura, and Satoshi Imahie for their insightful comments and detailed discussions. I also thank Takanori Adachi, Daiya Isogawa, Yuta Kikuchi, Toshifumi Kuroda, Yusuke Matsuki, Masato Nishiwaki, Tatsushi Oka, Ryo Okui, Hidenori Takahashi, and Naoki Wakamori for helpful comments, as well as participants at the Japan Empirical Industrial Organization Workshop and the Kansai Econometric Society Meeting. This research was financially supported by the Project Research Program of the Joint Usage/Research Center Programs at the Institute of Economic Research, Hitotsubashi University (Grant Number: IERPK2437); the JST SPRING fellowship; and a Grant-in-Aid for JSPS Fellows (Grant Number: 25KJ0910). This research was conducted under approval number 20240708-stat-No1 dated July 8, 2024, by the Statistics Bureau, Ministry of Internal Affairs and Communications. I utilized microdata from the *Census of Manufactures* (Ministry of Economy, Trade and Industry) and the *Economic Census for Business Activity* (Ministry of Internal Affairs and Communications; Ministry of Economy, Trade and Industry). The views expressed in this paper are those of the author and do not necessarily reflect the views of the Japanese government or the ministries. All remaining errors are my own.

†Graduate School of Public Policy, University of Tokyo, 7-3-1 Hongo, Tokyo, Japan. Email: rentaro.utamaru@gmail.com

1 Introduction

I identify the production function and the distribution of productivity from a single cross-section, replacing the first-order Markov assumption on productivity with a static condition: mutual independence of demand shocks across three intermediate inputs. The Markov assumption is not a technical convenience but the source of identification in both proxy variable methods (Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg, Caves, and Frazer 2015) and dynamic panel methods (Arellano and Bond 1991; Blundell and Bond 1998). This paper shows it is unnecessary.

Three flexible intermediate inputs (raw materials, energy, water) whose demands are jointly determined by productivity serve as three noisy measurements of the same latent variable. I recover the productivity distribution from these conditionally independent signals using the spectral decomposition of Hu and Schennach (2008) (hereafter HS08), without any restriction on the law of motion for productivity. The key departure from prior applications of HS08 to production functions (Brand 2020; Hu, Huang, and Sasaki 2020; Doty 2022) is the use of three contemporaneous input demands as measurements, rather than lagged variables as instruments, which eliminates the dynamic requirement entirely.

The Markov restriction has two costs. First, when productivity evolves endogenously (through R&D, learning, or managerial turnover), omitting the relevant state variables from the transition equation generates a transmission bias (De Loecker 2007; De Loecker 2013; Doraszelski and Jaumandreu 2013). Second, the assumption presupposes stationary dynamics and conflicts with the potential outcomes framework: if a treatment alters the transition path of productivity, the Markov property assumed during estimation is violated (Chen, Liao, and Schurter 2024). I resolve this inconsistency in Section 2.5.

Both costs are consequences of the same feature: the Markov assumption places a restriction on the data-generating process that is untestable within the proxy variable framework. By contrast, the conditional independence assumption I substitute has transparent microfoundations in input market segmentation and yields a pairwise convergence diagnostic (Remark 1): under the joint null, the ratio of demand slopes on each primary input to the corresponding productivity loading must be equal across the three intermediate inputs. In 502 industries, this diagnostic converges to zero for capital but not for labor, providing direct evidence that the exclusion restriction holds for capital but fails for labor across the manufacturing sector. When conditional independence is violated through a common utility shock to electricity and water—the most economically salient threat—the bias in $\hat{\beta}_m$ is *upward*, the same direction as the Markov misspecification bias in existing methods; hence the empirical gap between the proposed and ACF estimates cannot be attributed to CI violation (Section 4, Appendix J). No analogous diagnostic exists for the Markov assumption: within the proxy variable framework, there is no restriction that distinguishes a correctly specified AR(1) from an AR(2) or a potential outcome process.

Several recent approaches weaken or circumvent the Markov restriction. One strand applies the HS08 framework to production functions (Brand 2020; Hu, Huang, and Sasaki 2020; Doty 2022), but uses lagged variables as instruments and therefore retains the Markov assumption. A second strand, led by Gandhi, Navarro, and Rivers (2020) (hereafter GNR), exploits static first-order conditions; this resolves the within-period identification problem but requires competitive input markets with common prices and remains dependent on the Markov assumption for capital and labor elasticities. Table 1 provides a systematic comparison.

I make three contributions. First, I establish nonparametric identification of the production func-

Table 1: Comparison with other studies

	Identification Method	Req. Markov	Req. Scalar Unobs.	Nonpara Non-Hicks	Function Type	Proxy or Control
Proposed method	Hu et al.			✓	Gross	e_{jt}, w_{jt}
Gandhi, Navarro, and Rivers (2020)	FOC + Markov	✓	✓		Gross	s_{jt} (share)
Doty (2022)	Hu et al.	✓	✓	✓	Gross	I_{jt}, y_{jt+1}
Hu, Huang, and Sasaki (2020)	Hu et al.	✓	✓		Gross	I_{jt}, m_{jt+1}
Brand (2020)	Hu et al.	✓	✓		Gross	y_{jt-1}, y_{jt+1}
Zeng (2023)	Matzkin Imbens et al.	✓	✓	✓	Value	K_{jt-1}, I_{jt-1}
Ackerberg, Hahn, and Pan (2022)	Matzkin Imbens et al.	✓	✓	✓	Gross	$\{y_{j\tau}, x_{j\tau}\}_{\tau=t-M}^{t-1}$
Navarro and Rivers (2018)	Matzkin Imbens et al.	✓	✓	✓	Gross	$x_{jt-1}, \mathcal{Y}_{jt-1}$
Pan (2022)	Matzkin Imbens et al.	✓	✓	✓	Gross	$\{y_{j\tau}, x_{j\tau}\}_{\tau=t-M}^{t-1}$

Notes: "Req. Markov" indicates whether the method requires a Markov assumption on productivity; a blank cell indicates the method does not. "Req. Scalar Unobs." indicates whether the method requires scalar unobservability (productivity as the sole unobservable in input demand); a blank cell indicates that the method permits input-specific demand shocks. "Nonpara Non-Hicks" indicates nonparametric identification under non-Hicks-neutral specifications. "Function Type" distinguishes gross output from value-added production functions. "Proxy or Control" lists the proxy variables or control variables used for identification. For the proposed method, the "Nonpara Non-Hicks" checkmark refers to the identification result in Appendix C; the implemented estimator is Hicks-neutral (equation (12)). The proposed method requires conditional independence of input-specific demand shocks (Assumption 2) in place of the Markov and scalar unobservability conditions; both blank cells in its row reflect this substitution, not an absence of identifying assumptions.

tion and the distribution of productivity from static data alone. This is a substitution, not a relaxation, of identifying assumptions: the conditional covariance structure across three contemporaneous flexible inputs fully replaces the time-series restriction, delivering point identification without any assumption on the productivity process. I characterize the residual indeterminacy completely: any two observationally equivalent structures differ only by a location shift $\Delta(k, l)$ applied to productivity, ruling out nonlinear transformations (Theorem 2). This indeterminacy is eliminated nonparametrically through exclusion restrictions (Corollary 1) or parametrically through a homothetic regularity condition on $\mathbb{E}[\omega \mid k, l]$ (Theorem 3); I develop a GMM estimator and establish consistency and asymptotic normality (Theorem 4).

Second, I show that the conditional independence assumption admits a testable diagnostic with no analogue under the Markov assumption. The GMM system yields a pairwise discrepancy statistic (Remark 1) that tests whether demand coefficients on primary inputs are proportional across intermediate inputs. Its standard error is obtained by the delta method from the GMM variance-covariance matrix, so no generated-regressors correction is needed. In 502 Japanese manufacturing industries, the statistic converges to zero for capital but not for labor, revealing differential applicability of the exclusion restriction across input types.

Third, I provide Monte Carlo evidence that the Markov assumption generates a consistency failure, not merely finite-sample bias. Under three DGPs that progressively violate the first-order Markov assumption, ACF's bias in $\hat{\beta}_m$ does not vanish as T increases: it is $+0.026$ under AR(2) dynamics and $+0.266$ under potential outcome dynamics, remaining constant as sample size grows. The proposed estimator is unbiased across all specifications. In 502 Japanese manufacturing industries, this consistency gap translates into systematically different markup distributions and productivity measures with substantially stronger associations with economic fundamentals (Section 5).

Related literature. Table 1 positions my identification strategy within the recent literature. The most closely related work is Gandhi, Navarro, and Rivers (2020) (GNR). GNR's Theorem 1 establishes that proxy variable methods alone cannot identify the gross output production function; additional within-period, cross-sectional information is required. Both approaches supply such information: GNR through the structural link between the production function and the firm's first-order condition, yielding a nonparametric share regression that directly identifies the flexible input elasticity; my approach through the measurement error structure of HS08, using conditional independence across intermediate inputs to recover the distribution of unobserved productivity.

The two approaches rest on different assumptions regarding input markets. GNR's first-order condition requires competitive input markets with common prices and that any unobserved component in the share equation is non-persistent (their Appendix O6, Assumption 7); when input-specific markdowns or procurement frictions are persistent, the FOC-based estimation equation does not hold and the share regression is misspecified. My framework permits persistent, input-specific demand shocks arising from procurement relationships, supply contracts, or input-specific markdowns; identification requires only mutual independence across inputs at each time point, accommodating arbitrary serial dependence within each shock. GNR's second stage recovers capital and labor elasticities using the Markov structure; my approach requires no dynamic assumption at any stage. The scalar unobservability case is a special case of my model, obtained when the input-specific shocks are degenerate (Section 4).

Additional related work is summarized in Table 1. Zeng (2023) avoid the Markov restriction at the estimation stage but presuppose it for the investment policy function. A growing literature on non-Hicks-neutral identification (Navarro and Rivers 2018; Ackerberg, Hahn, and Pan 2022; Pan 2022; Kasahara, Schrimpf, and Suzuki 2023; Doty 2022), including factor-augmenting approaches (Doraszelski and Jaumandreu 2018; Demirer 2022; Raval 2019), retains first- or higher-order Markov assumptions; my identification results extend to these models without dynamic restrictions (Appendix C), though the implemented estimator and all empirical results use the Hicks-neutral Cobb–Douglas specialization (Section 3.1.2).

The remainder of the paper is organized as follows. Section 2 presents the model and the nonparametric identification results. Section 3 develops the GMM estimator. Section 4 presents Monte Carlo evidence. Section 5 applies the estimator to 502 Japanese manufacturing industries. Section 6 concludes.

2 Model and Identification

This section establishes the identification strategy in three steps. First, I show that three conditionally independent input demands identify the joint distribution of productivity and inputs within each capital-labor cell (Theorem 1); any two observationally equivalent structures differ only by a location shift $\Delta(k, l)$ (Theorem 2). Second, I provide two conditions that eliminate this indeterminacy: an exclusion restriction (Corollary 1) and a homothetic regularity condition (Theorem 3). The exclusion restriction carries a testable implication (Remark 1). The formal statement of density identification (Theorem 1) and the technical regularity conditions (Assumptions A.1–A.3) are in Appendix A.

2.1 Model Setup

I define the general gross output production function for firm j at time t as follows:

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

Here, y_{jt} is the logarithm of output, k_{jt} and l_{jt} are the logarithms of capital and labor. Following the production function literature (Olley and Pakes 1996; Ackerberg, Caves, and Frazer 2015; Bond and Söderbom 2005), capital and labor are treated as dynamic or quasi-fixed inputs whose current values are predetermined relative to intermediate input decisions. The model requires at least three distinct intermediate inputs: m_{jt} (raw materials), e_{jt} (electricity), and w_{jt} (industrial water). Three inputs are the minimum required by the Hu and Schennach (2008) spectral decomposition: it identifies the latent productivity distribution from three mutually independent measurements of a common latent variable; two measurements do not suffice for nonparametric identification without additional restrictions.[1] ω_{jt} is the firm's productivity, unobserved by the econometrician but known to the firm when making input decisions. ε_{jt} denotes ex-post production shocks (measurement error or unexpected disruptions), unobserved by both the firm and the econometrician at the time of input choice.

The state variable vector $x_{jt} = (k_{jt}, l_{jt}, z_{jt})$ determines input demand. Here, k_{jt} and l_{jt} are the primary inputs, while z_{jt} represents additional firm-specific state variables such as inventory levels, input prices, or market conditions that do not directly enter the production function but influence input demand. Given x_{jt}, the demand for each intermediate input is determined as follows:

$$m_{jt} = g_m(x_{jt}, \omega_{jt}, \tau_{jt}) \tag{2}$$

$$e_{jt} = g_e(x_{jt}, \omega_{jt}, \nu_{jt}) \tag{3}$$

$$w_{jt} = g_w(x_{jt}, \omega_{jt}, \eta_{jt}) \tag{4}$$

The functions $g_m(\cdot)$, $g_e(\cdot)$, and $g_w(\cdot)$ are unknown and potentially nonlinear. τ_{jt}, ν_{jt}, and η_{jt} are unobserved shock terms specific to each input demand, following Hu, Huang, and Sasaki (2020), Brand (2020), and Doty (2022). These shocks capture optimization errors, supply disruptions, and adjustment frictions not explained by productivity and state variables. Appendix B derives the demand system from a cost-minimization problem under imperfect input markets and shows that these shocks correspond to input-specific markdowns, prices, and wedges—specifically, the components of markdowns and wedges orthogonal to observable state variables.

The presence of input-specific shocks represents a departure from the scalar unobservability assumption maintained in Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015), GNR, and others, which requires productivity to be the sole unobservable affecting input demand. When scalar unobservability fails because firm-level input prices, markdowns, or wedges are unobserved, standard proxy variable estimators are inconsistent (Jaumandreu and Doraszelski 2021; Doraszelski and Li 2025). In my framework, all unobserved firm-specific heterogeneity beyond productivity is absorbed into $\tau_{jt}, \nu_{jt}, \eta_{jt}$, and identification requires only that these shocks be mutually independent across inputs, not that they be absent. Scalar unobservability is nested as the special case $\tau = \nu = \eta = 0$.

This formulation also addresses the collinearity problem identified by Gandhi, Navarro, and Rivers (2020): under scalar unobservability, flexible inputs determined by static optimization lack sufficient

[1] When labor adjusts rapidly to current productivity, it may serve as a third measurement of ω_{jt}, reducing the required number of flexible intermediate inputs from three to two; see Footnote 6 for details.

residual variation to identify the gross production function (Ackerberg, Caves, and Frazer 2015; Bond and Söderbom 2005). GNR resolve this problem by exploiting the first-order condition for the flexible input, which identifies its output elasticity from the revenue share. My approach resolves the collinearity through independent input-specific shocks, which supply the cross-sectional variation needed for identification via the measurement error structure of HS08, without relying on the first-order condition or dynamic moment conditions. The practical difference is that the share regression requires the first-order condition to hold with common input prices, whereas my approach permits firm-specific input prices and markdowns (Appendix B).

2.2 Assumptions for Identification

The identification theory rests on two substantive assumptions stated here, together with three regularity conditions (Assumptions A.1–A.3) collected in Appendix A.

Assumption 1 (Additive Error Structure)**.** *The production function has an additive error structure:*

$$y_{jt} = f_t(k_{jt}, l_{jt}, m_{jt}, e_{jt}, w_{jt}, \omega_{jt}) + \varepsilon_{jt}, \tag{5}$$

where the ex-post shock ε_{jt} satisfies

$$\mathbb{E}\big[\varepsilon_{jt} \mid k_{jt}, l_{jt}, m_{jt}, e_{jt}, w_{jt}, \omega_{jt}\big] = 0.$$

Role and economic content. This is standard in the production function literature (Olley and Pakes 1996; Ackerberg, Caves, and Frazer 2015). The shock ε_{jt} captures ex-post deviations (measurement error, unexpected disruptions) that are realized after input choices are made and are therefore uncorrelated with all inputs and productivity. It acts as classical measurement error in the dependent variable and inflates standard errors but does not bias the production function estimates (Theorem 4).

Assumption 2 (Conditional Independence)**.** *The demand shocks $(\tau_{jt}, \nu_{jt}, \eta_{jt})$ for the three intermediate inputs are mutually independent, conditional on productivity ω_{jt} and state variables $x_{jt} = (k_{jt}, l_{jt}, z_{jt})$:*

$$f_{\tau,\nu,\eta|\omega,x} = f_{\tau|\omega,x} \cdot f_{\nu|\omega,x} \cdot f_{\eta|\omega,x}.$$

Mutual independence is required; pairwise independence does not suffice for the spectral decomposition of HS08.

Role. This is the substantive identifying condition. Together with the regularity conditions in Appendix A (Assumptions A.1–A.3), it enables the unique spectral decomposition of the integral equation (39). Conditional independence is the economically substantive condition; it restricts the data generating process rather than regularity of the operators.

Economic content. The assumption posits that, for a firm with given state variables and productivity level, an unexpected shock to raw material demand (e.g., a supply chain disruption) is independent of a shock to electricity demand (e.g., an unscheduled rate surcharge). This is natural when input markets are segmented: raw materials, electricity, and water are procured through distinct channels, under separate contracts, with different suppliers. The common components of demand variation (product demand fluctuations, aggregate markup changes) are captured by x_{jt}; $\tau_{jt}, \nu_{jt}, \eta_{jt}$ represent the residual, input-specific components. The microfoundations in Appendix B make this structure precise.

2.3 Interpretation and Robustness of the Conditional Independence Assumption

The general principle is as follows. Common shocks that affect all three input demands (product demand fluctuations, markup variation, aggregate input price movements) can be absorbed by projecting onto observable control variables z_{jt}; the shock terms $\tau_{jt}, \nu_{jt}, \eta_{jt}$ are then defined as the orthogonal residuals of this projection (Appendix B). The independence assumption therefore requires only that the *residual*, input-specific components of demand variation are mutually independent.

Several potential threats illustrate this principle. *Unobserved demand shocks* generate common variation across all inputs, but can be proxied by inventory fluctuations (Kumar and Zhang 2019) or recovered from revenue data (Kasahara and Sugita 2020), included in z_{jt}. *Product market power* affects all input demands through marginal revenue; following Ackerberg and De Loecker (2024) and Jaumandreu (2025), low-dimensional sufficient statistics for the markup (e.g., competitors' output, average variable cost) can be included in z_{jt}.[2] *Input market power* (markdowns) may generate common bargaining advantages across inputs, but the common component depends on firm attributes (size, liquidity) captured by (k_{jt}, l_{jt}, z_{jt}); what remains in the shock terms are idiosyncratic variations from individual supplier relationships. It is economically reasonable that the outcome of negotiations with raw material suppliers is independent of electricity rate negotiations, conditional on firm size and other observables.[3] *Common input price shocks* (e.g., oil price hikes) affect multiple inputs symmetrically and are controlled by time fixed effects or industry-specific deflators in z_{jt}. Firm-specific price variations are absorbed as part of the structural shock terms and need only be independent across inputs.

Structural proportionality under Cobb–Douglas. Beyond the observable controls z_{jt}, the Cobb–Douglas specification provides a structural defense against common shocks. Under Cobb–Douglas, the first-order conditions for all three intermediate inputs share a common dependence on productivity ω_{jt} and the common market factor $\ln(P_{jt}/\mu_{jt})$: both enter each equation with identical reduced-form coefficients (Appendix B), because the log marginal product $\ln(\partial F/\partial h)$ depends on ω only through $y - h$, which loads ω identically across inputs. Any unobserved common component ξ_{jt} that enters the first-order conditions symmetrically is therefore absorbed into an effective productivity measure $\tilde{\omega}_{jt} \equiv \omega_{jt} + \xi_{jt}$, with the input-specific residuals $(\tau^*_{jt}, \nu^*_{jt}, \eta^*_{jt})$ remaining mutually independent. The demand-side identification (Theorems 1–2) proceeds with $\tilde{\omega}$ in place of ω, and the moment conditions that eliminate productivity across demand pairs (u_1, u_2 in Block A and the demand–demand covariances in Block B) are exactly invariant to ξ_{jt}.[4] The conditional independence assumption therefore restricts only the input-specific residual components *after* this structural absorption—a weaker condition than independence of the raw demand shocks.

2.4 Identification of the Production Function

The identification proceeds in two stages: first, I recover the production function and productivity distribution within each capital-labor cell (k_0, l_0); second, I characterize and resolve the residual

[2] Under Cournot competition, Ackerberg and De Loecker (2024) show that the total output of competitors serves as a sufficient statistic.

[3] When an intermediate input is traded on competitive commodity markets, the firm is a price-taker and the markdown on that input vanishes. Avignon and Guigue (2025) exploit this property for globally traded dairy commodities to separately identify markups and markdowns on other inputs.

[4] The production–demand cross moments (u_3 and the y–demand covariances in Block B) are not invariant when ξ_{jt} does not enter the production function. Under revenue data, the common market factor is absorbed into revenue productivity, so the cross moments are approximately invariant. Appendix J characterizes the direction of bias in $\hat{\beta}_m$ under residual violations.

indeterminacy that arises when linking these cell-specific results across different values of (k, l).[5]

2.4.1 Identification within Each (k, l)

The foundational identification result applies the spectral decomposition of HS08, whose conditions I verify under the present assumptions.

Theorem 1 (Identification of Densities)**.** *Under Assumptions 1–2 and Assumptions A.1–A.3, the observable conditional joint density $f_{m_{jt},e_{jt}|x_{jt},w_{jt}}$ uniquely identifies the three unknown conditional density functions: $f_{m_{jt}|\omega_{jt},x_{jt}}$, $f_{e_{jt}|\omega_{jt},x_{jt}}$, and $f_{\omega_{jt}|x_{jt},w_{jt}}$.*

The proof, which verifies the conditions of HS08's Theorem 1 for the integral equation (39), is in Appendix A.2.

As a consequence of Theorem 1 and equation (41), for each fixed (k_0, l_0), the following are non-parametrically identified: the conditional densities $f_{m|\omega,k_0,l_0}$, $f_{e|\omega,k_0,l_0}$, $f_{w|\omega,k_0,l_0}$, and $f_{\omega|k_0,l_0,m,e,w}$.

Using these identification results, I recover the structure of f_t as a function of (m, e, w, ω). I focus on the Hicks-neutral specification, widely adopted in the empirical literature, and defer the general case to Appendix C. Under this specification $y = g_t(k, l, m, e, w) + \omega + \varepsilon$, Assumption 1 implies

$$g_t(k_0, l_0, m, e, w) = \mathbb{E}[y \mid x, m, e, w] - \mathbb{E}[\omega \mid x, m, e, w]. \tag{6}$$

Here g_t represents the component of the production technology that depends on intermediate inputs, with productivity ω separated out. The first term on the right-hand side is a conditional expectation identified directly from the data, and the second is computable from the posterior density in equation (41). Thus g_t is identified as a function of (m, e, w) without additional assumptions. For the general non-Hicks-neutral model, $f_t(k_0, l_0, m, e, w, \omega)$ is identified as a function of (m, e, w, ω) under additional regularity conditions on the distribution of ε; see Appendix C for details.

For each fixed (k_0, l_0), the conditional distribution $f_{\omega|k_0,l_0,m,e,w}$ is fully characterized, and the conditional expectation

$$\hat{\omega}_{jt} \equiv \mathbb{E}[\omega_{jt} \mid x_{jt}, m_{jt}, e_{jt}, w_{jt}] = \int \omega\, f_{\omega|x,m,e,w}(\omega \mid \cdot)\, d\omega \tag{7}$$

provides a firm-level productivity measure for each firm j and period t. The empirical applications of this within-(k, l) identification, including markup estimation and policy evaluation, are developed in Section 2.5 after the identification theory is completed.

However, to identify f_t as a function of (k, l) as well, additional structure is needed. (When labor adjusts rapidly to current productivity, it serves as an additional measurement, reducing the required intermediate inputs from three to two.[6]) ω must be defined on a common scale across different values of (k, l). Since Theorem 1 applies the HS08 procedure independently for each (k, l), there is no automatic correspondence between the ω values identified at (k_1, l_1) and those identified at (k_2, l_2). I now formalize this problem.

[5] In the following, firm subscripts j are suppressed as I discuss population-level arguments. The time subscript t is retained only to indicate time-variation in the production function f_t.

[6] When labor adjusts within the production period, it serves as a third measurement of ω_{jt}, and the HS08 identification procedure (Theorem 1) applies to the triple (l_{jt}, m_{jt}, e_{jt}), reducing the required flexible intermediate inputs from three to two. This extension applies when adjustment costs are small enough that l_{jt} responds to within-period productivity innovations; industries with high turnover or temporary staffing (e.g., food processing, garment manufacturing) are natural candidates. When labor is quasi-fixed, l_{jt} reflects past rather than current productivity, and the conditional independence conditions for (l, m, e) do not hold. See Appendix C for details.

2.4.2 Observational Equivalence and Limits of Identification

Theorem 1 identifies the production function within each (k_0, l_0), but a practitioner needs parameters that are comparable across different capital-labor combinations. The next result shows exactly what remains unresolved and rules out the possibility that the indeterminacy takes a nonlinear form.

Under Assumptions 1–2 and the regularity conditions in Appendix A (Assumptions A.1–A.3), the conditional densities $f_{m|\omega}$, $f_{e|\omega}$, $f_{w|\omega}$ and the marginal density f_ω are nonparametrically identified from the joint density of (m, e, w) conditional on (k, l, z) (Theorem 1). This pins down the shape of each conditional distribution but leaves a common location shift $\Delta(k, l)$ applied to the latent variable unresolved. The following theorem characterizes this residual indeterminacy completely.

Theorem 2 (Complete Characterization of Observational Equivalence)**.** *Under Assumptions 1–2 and Assumptions A.1–A.3, a necessary and sufficient condition for two structures (f_t, ω) and $(\tilde{f}_t, \tilde{\omega})$ to generate the same joint distribution of observables is that there exists a continuous function $\Delta(k, l)$ such that*

$$\tilde{\omega} = \omega + \Delta(k, l), \qquad \tilde{f}_t(k, l, m, e, w, \tilde{\omega}) = f_t(k, l, m, e, w, \tilde{\omega} - \Delta(k, l)). \tag{8}$$

The proof is given in Appendix D; the key steps are as follows. The HS08 eigenvalue-eigenfunction decomposition uniquely determines the functional form of each conditional density within each (k_0, l_0), ruling out nonlinear transformations of ω. Any remaining degree of freedom must therefore be a location shift that varies across (k, l), yielding (8). The continuity of $\Delta(k, l)$ follows from the continuous dependence of $f_{m|\omega,k,l}$ on (k, l) (stated after Assumption A.3) together with the perturbation theory of compact operators under simple eigenvalues (Assumption A.2; see Appendix D for details).

Nonlinear transformations (including scale transformations) are ruled out because the eigenvalue–eigenfunction decomposition in HS08 uniquely determines the functional form of each conditional density within each (k_0, l_0). Second, the $\Delta(k, l)$ indeterminacy arises inherently from the fact that Theorem 1 applies the HS08 procedure independently for each (k, l). Within each (k_0, l_0), Assumption A.3 fixes the level of ω, but the reference point of this normalization may depend on (k_0, l_0). The data on conditional distributions of intermediate input demands do not contain information to unify ω levels across different (k, l).[7]

Economically, the $\Delta(k, l)$ indeterminacy means that the effect of (k, l) on the production function and $\mathbb{E}[\omega \mid k, l]$ cannot be separated without additional restrictions. As a direct consequence, f_t is identified up to the specification of $\mathbb{E}[\omega \mid k, l]$: fixing $\mathbb{E}[\omega \mid k, l]$ pins down $\Delta = 0$ (Theorem A.1, Appendix A).

The $\Delta(k, l)$ indeterminacy also arises in the existing literature: Gandhi, Navarro, and Rivers (2020) resolve it in the Hicks-neutral setting by combining first-order conditions with a Markov assumption, which reduces $\Delta(k, l)$ to a constant; for non-Hicks-neutral models, this strategy fails because ω cannot be separated from the first-order condition.[8]

I provide two alternative methods that close the identification gap without dynamic assumptions. Theorem 2 guarantees that $\Delta(k, l)$ is a continuous function of (k, l) alone, which both methods exploit.

[7] Hahn, Liao, and Ridder (2023) show that in dynamic approaches such as Olley and Pakes (1996), the identification of dynamic input elasticities relies on an index restriction that collapses state variables into a one-dimensional scalar. Theorem 1 does not provide such an index restriction, and hence the indeterminacy with respect to the dynamic elasticities persists.

[8] In the Hicks-neutral model, $\Delta(k, l)$ shifts the (k, l) component of the production function: $\tilde{g}_t(k, l, \cdot) = g_t(k, l, \cdot) - \Delta(k, l)$. In non-Hicks-neutral models, the FOC $P_t \cdot (\partial f_t / \partial M) = \rho_t$ retains ω on the left-hand side, precluding a share regression. Li and Sasaki (2024) show that heterogeneous output elasticities with respect to flexible inputs remain identifiable under a scalar unobservable assumption on the proxy variable.

Section 2.4.4 imposes exclusion restrictions on intermediate input demands that directly constrain the functional form of $\Delta(k, l)$, achieving nonparametric point identification. Section 2.4.5 parametrically specifies the (k, l) component and introduces a regularity condition on the shape of $\mathbb{E}[\omega \mid k, l]$, achieving parametric identification through the non-constant curvature of the homothetic transformation.

2.4.3 Closing the Identification Gap

The $\Delta(k, l)$ indeterminacy is the cost of dispensing with the Markov assumption. I now show this cost is payable: two conditions, each operating without dynamic restrictions, eliminate the indeterminacy and deliver point identification.

2.4.4 Nonparametric Identification via Exclusion Restrictions

The $\Delta(k, l)$ indeterminacy arises because the location normalization in Assumption A.3 is applied independently for each (k, l) (Theorem 2). If the HS08 location normalization $M[f_{m|\omega,k,l}(\cdot \mid \omega)] = \omega$ could be applied *uniformly* across all (k, l), then $\Delta(k, l) = 0$ would follow immediately. However, for this uniform normalization to hold, $M[f_{m|\omega,k,l}(\cdot \mid \omega)]$ must not depend on (k, l); that is, the conditional demand for the intermediate input, given ω, must be independent of (k, l). This observation suggests that exclusion restrictions on intermediate input demands directly constrain $\Delta(k, l)$.

Corollary 1 (Identification via Exclusion Restrictions)**.** *In addition to Assumptions 1–2 and A.1–A.3, suppose one of the following conditions holds:*

(i) The demand for some intermediate input (e.g., w) does not depend on (k, l): $f_{w|\omega,k,l} = f_{w|\omega}$.

(ii) The demand for one input (e.g., m) does not depend on k, and the demand for another (e.g., e) does not depend on l: $f_{m|\omega,k,l} = f_{m|\omega,l}$ and $f_{e|\omega,k,l} = f_{e|\omega,k}$.

Then, under the normalization $\mathbb{E}[\omega] = 0$, the production function f_t is nonparametrically point-identified. Condition (i) is a special case of condition (ii).

Proof. By Theorem 2, observationally equivalent structures are parameterized by $\tilde{\omega} = \omega + \Delta(k, l)$. Requiring that the exclusion restriction be maintained in the alternative structure:

Condition (i): $f_{w|\omega,k,l} = f_{w|\omega}$ implies that in the alternative structure, $f_{w|\tilde{\omega},k,l}(w \mid \tilde{\omega}) = f_{w|\omega}(w \mid \tilde{\omega} - \Delta(k, l))$, which is independent of (k, l) only if $\Delta(k, l)$ is constant.

Condition (ii): $f_{m|\omega,k,l} = f_{m|\omega,l}$ implies Δ does not depend on k. $f_{e|\omega,k,l} = f_{e|\omega,k}$ implies Δ does not depend on l. Together, Δ is constant.

In both cases, $\mathbb{E}[\omega] = 0$ pins down $\Delta = 0$. □

Economically, condition (i) requires that the demand for some intermediate input (e.g., electricity) depends on productivity alone and not on capital or labor intensity; this may hold in energy-intensive industries where electricity consumption is driven by production volume rather than by the composition of capital equipment. Condition (ii) requires that different inputs exclude different primary inputs from their demand: for example, raw material demand does not depend on labor intensity, and fuel demand does not depend on capital intensity. These exclusion restrictions limit the scope of application to industries where institutional knowledge supports them. For settings where such restrictions cannot be justified, I provide a parametric alternative in the next subsection.

Remark 1 (Testability of the Exclusion Restriction)**.** *Under the linear demand specification* (13)–(15)*, let* a_k^h*,* a_l^h*, and* a_ω^h *denote the slope coefficients on* k*,* l*, and* ω *in the demand for input* h*:* $(a_k^m, a_l^m, a_\omega^m) = (\gamma_k, \gamma_l, \gamma_\omega)$*,* $(a_k^e, a_l^e, a_\omega^e) = (\delta_k, \delta_l, \delta_\omega)$*,* $(a_k^w, a_l^w, a_\omega^w) = (\zeta_k, \zeta_l, \zeta_\omega)$*. The exclusion restriction of Corollary 1 for a single input* h *is not separately testable from Block A+B estimates. Under the normalization* $\beta_k = \beta_l = 0$ *(Section 3.1), the estimated demand coefficient* $\hat{a}_k^{h*}$ *converges to* $a_k^h - a_\omega^h \beta_k$*, confounding the structural exclusion parameter* a_k^h *with the indeterminacy* $a_\omega^h \beta_k$ *from Theorem 2.*

The joint *restriction across inputs, however, yields a diagnostic test—a necessary condition for consistency with the exclusion restriction, but not a sufficient one. Under Proposition A.1, the OLS estimate* $\hat{\beta}_k^{(h)}$ *from input* h *converges to* $\beta_k - a_k^h / a_\omega^h$*. Define the pairwise discrepancy*

$$d_k^{(h_1,h_2)} \equiv \frac{\hat{a}_k^{h_2*}}{\hat{a}_\omega^{h_2}} - \frac{\hat{a}_k^{h_1*}}{\hat{a}_\omega^{h_1}} \xrightarrow{p} \frac{a_k^{h_2}}{a_\omega^{h_2}} - \frac{a_k^{h_1}}{a_\omega^{h_1}}, \tag{9}$$

which is free of the $\Delta(k,l)$ *indeterminacy since* β_k *cancels in the difference. Under the joint exclusion restriction* $a_k^{h_1} = a_k^{h_2} = 0$*,* $d_k = 0$*; the converse does not hold. The test statistic* $d_k = 0$ *is a* necessary condition *for the full joint exclusion restriction, not a sufficient one:* $d_k = 0$ *also obtains in the knife-edge case where* a_k^h / a_ω^h *is equal across inputs but nonzero. This configuration has no structural basis when the three inputs involve distinct procurement channels, but the possibility cannot be ruled out on the basis of* d_k *alone (Appendix H.1). The test is therefore best interpreted as a* diagnostic*: a rejection of* $d_k = 0$ *is evidence against the exclusion restriction, while non-rejection is consistent with—but does not establish—it. Since* d_k *is a smooth function of the Block A+B parameters, its standard error is obtained by the delta method from the GMM variance-covariance matrix, yielding a Wald test without the generated-regressors problem that would arise from testing OLS estimates directly. With three inputs, the formal test has two degrees of freedom* $(d_k = d_l = 0)$ *(Appendix H.1). I apply this test in Section 5.3.*

The formal statement and proof are given in Proposition A.1 (Appendix A).[9]

2.4.5 Parametric Identification via Homothetic Regularity

As an alternative when exclusion restrictions cannot be justified, I parametrically specify the (k,l) component and introduce a regularity condition on $\mathbb{E}[\omega \mid k,l]$. Consider the additively separable model

$$y = g(k,l;\theta) + q(m,e,w) + \omega + \varepsilon, \tag{10}$$

where g is parametric with known functional form and q is nonparametric. From Section 2.4.1, q is nonparametrically recoverable for each fixed (k_0, l_0, ω_0).

Specializing to $g(k,l;\theta) = \beta_k k + \beta_l l$, Theorem 2 reduces the identification indeterminacy to

$$\Delta(k,l) = c_k k + c_l l, \quad (c_k, c_l) \in \mathbb{R}^2. \tag{11}$$

To eliminate this two-dimensional indeterminacy, I introduce the following regularity condition.

Assumption 3 (Homothetic Weak Separability)**.** *The conditional expectation of TFP in the cross-section has a homothetic structure: there exist continuously differentiable functions* $h\colon \mathbb{R} \to \mathbb{R}$ *and* $v\colon \mathbb{R}^2 \to \mathbb{R}$ *such that*

$$\bar{\omega}(k,l) \equiv \mathbb{E}[\omega \mid k,l] = h(v(k,l)),$$

[9] Replacing the linear subtraction of $\hat{\omega}^h$ in Proposition A.1 with a polynomial regression is not consistent in general; see Appendix H.3 for details.

where:

(A) Nonlinear transformation: h' is not a constant function.

(B) Translation homogeneity: v satisfies $v(k+c, l+c) = v(k,l) + c$ for all $c \in \mathbb{R}$.

(C) Imperfect substitutability: The isoquants of v are strictly convex, and the marginal rate of substitution v_k/v_l is not constant on (k,l).

All three conditions are necessary for Theorem 3: (A) prevents observational equivalence with linear functions; (B) ensures the counterfactual index is also translation homogeneous, so that the MRS of $\tilde{v}$ is translation invariant; (C) excludes Cobb–Douglas, where v_k/v_l is constant and a one-dimensional indeterminacy persists. Economically, (A) requires nonlinear returns to the input bundle, (B) corresponds to constant returns to scale in the level variables (since translation homogeneity on the log scale is equivalent to degree-one homogeneity in levels), and (C) requires a finite and non-unit elasticity of substitution, satisfied by CES, translog, and normalized quadratic forms. Assumption 3 can be checked from Blocks A and B alone (Section 3.1.7); detailed necessity arguments and testability procedures are in Appendix H.5.

To illustrate, consider the CES specification where $v(k,l) = \frac{1}{\rho_v}\log\left(\alpha e^{\rho_v k} + (1-\alpha)e^{\rho_v l}\right)$ is translation homogeneous on the log scale: $v(k+c, l+c) = v(k,l) + c$. With $h(v) = \gamma v$ (for $\gamma \neq 0$ and higher-order terms $\rho_2 v^2 + \rho_3 v^3$ with $\rho_2 \neq 0$ or $\rho_3 \neq 0$), h' is non-constant (satisfying (A)), v is translation homogeneous (satisfying (B)), and the MRS $v_k/v_l = [\alpha/(1-\alpha)]e^{\rho_v(k-l)}$ is non-constant for $\rho_v \neq 0$ (satisfying (C)). The Cobb–Douglas case ($\rho_v \to 0$, so $v \to \alpha k + (1-\alpha)l$) yields a linear v and a constant MRS, violating conditions (A) and (C) simultaneously; the rank condition in Theorem 3 fails, and (β_k, β_l) cannot be separately identified. More generally, when ρ_v is close to zero, identification of (β_k, β_l) through Block C becomes weak: the marginal rate of substitution v_k/v_l approaches a constant as $\rho_v \to 0$, so the cross-sectional variation in (k_{jt}, l_{jt}) provides little leverage on the curvature parameters. In the empirical analysis, the t-statistics for $\hat{\rho}_2$ and $\hat{\rho}_3$ (Section 3.1.7) provide a direct diagnostic for this failure; industries where both are statistically insignificant should not be relied upon for separate identification of β_k and β_l through Block C alone. When $\rho_v = 0$, the exclusion restriction of Corollary 1 provides an alternative identification route.

Theorem 3 (Static Parametric Identification)**.** *Under Assumptions 1–2, A.1–A.3, and 3, the structural parameters β_k and β_l in model (10) are point-identified from static data alone.*

Proof. By contradiction. Suppose an observationally equivalent $\tilde{\beta}_i = \beta_i + c_i$ $(i = k, l)$ exists with $(c_k, c_l) \neq (0,0)$. By (11), the alternative TFP function satisfies $\mathbb{E}[\tilde{\omega} \mid k, l] = h(v(k,l)) - c_k k - c_l l$. Requiring that $\mathbb{E}[\tilde{\omega} \mid k, l] = \tilde{h}(\tilde{v}(k,l))$ for some translation homogeneous $\tilde{v}$ and differentiable $\tilde{h}$, the translation invariance of the marginal rate of substitution of $\tilde{v}$ requires

$$(c_l\, v_k - c_k\, v_l)\big(h'(v+c) - h'(v)\big) = 0$$

for all $(k,l) \in \mathbb{R}^2$ and $c \in \mathbb{R}$. By condition (A), h' is non-constant, so the second factor is nonzero for some (v_0, c_0). Hence $c_l\, v_k - c_k\, v_l = 0$ everywhere, so v_k/v_l equals the constant c_k/c_l. Under translation homogeneity, a constant MRS forces $v(k,l) = \alpha k + (1-\alpha)l$, which is linear in (k,l), contradicting condition (C). Therefore $(c_k, c_l) = (0,0)$.

Condition (B) (translation homogeneity) enters the argument through the translation invariance of the MRS of $\tilde{v}$: since $\tilde{v}_k + \tilde{v}_l = 1$ (implied by translation homogeneity), without it $\tilde{v}$ need not be translation homogeneous, and the equality $c_l\, v_k - c_k\, v_l = 0$ does not follow. ☐

Theorem 3 is stated and proved for the CES specification of $v(k,l)$; the argument extends to other parametric forms (e.g., translog) subject to verifying the rank condition specific to each functional form.[10]

2.5 Implications for Empirical Applications

The within-(k_0, l_0) identification results have direct empirical applications that differ in what they require. Markup estimation requires only β_m, which is identified by Blocks A and B alone. Event studies and difference-in-differences designs similarly require only Block A+B: because the estimator uses no transition equation for ω, the recovered $\hat{\omega}_{jt}$ is valid under any productivity dynamics, including treatment-induced non-Markov paths (Proposition A.2, Appendix A). Full productivity-level analysis—including the identification of β_k and β_l—requires Block C in addition.

Applications. Because estimation does not employ a transition process for ω, the estimates are invariant to how a policy D_{jt} affects productivity dynamics (Proposition A.2, Appendix A). For markup estimation, the within-(k_0, l_0) results suffice: output elasticities $\partial f_t / \partial m$ are identified for each fixed (k_0, l_0), which recovers markups as the ratio of the output elasticity to the revenue share (De Loecker and Warzynski 2012).

Remark 2 (Functional Form Generality)**.** *The identification results of this paper rest on the conditional independence of intermediate input demands (Assumption 2), not on the functional form of production. Theorems 1 and 2 establish nonparametric identification via the HS08 spectral decomposition for any production function satisfying Assumptions 1–2. The GMM estimator of Section 3 implements this under Cobb–Douglas, where input demands are linear in productivity and the moment conditions take a tractable linear form. Appendix K shows that the same identification source—conditional independence—yields nonlinear moment conditions under translog production. The empirical implementation focuses on Cobb–Douglas to maintain computational tractability and to isolate the effect of relaxing the Markov assumption from functional form complexities.*

3 Estimation Methods

The nonparametric identification results of Section 2 establish that the production function and productivity distribution are identified from the joint density of intermediate inputs; nonparametric sieve estimation could in principle implement this directly, but the high-dimensional numerical integration required is computationally prohibitive for census-scale panels spanning hundreds of industries. I therefore develop a GMM estimator that specializes to a linear production function and linear demand functions. Under this parametric restriction, the observational equivalence class of Theorem 2 reduces to a two-dimensional indeterminacy (c_k, c_l) (equation (11)), and the identification results of Corollary 1 and Theorem 3 carry through directly.

3.1 Estimation Based on the Generalized Method of Moments

As noted in Remark 2, the identification results of Section 2 apply to general production functions. The parametric implementation below specializes to the Cobb–Douglas case, where input demand functions

[10] For instance, with a translog specification $g = \beta_k k + \beta_l l + \beta_{kk} k^2 + \beta_{ll} l^2 + \beta_{kl} kl$, Δ is restricted to the corresponding polynomial class and the homothetic regularity condition eliminates the indeterminacy by a similar argument, but the conditions on the MRS differ from the CES case.

are linear in productivity (Appendix B). This linearity yields the tractable linear GMM system of Blocks A–B. Extension to flexible functional forms such as translog is developed in Appendix K; the identification source remains the conditional independence of demand shocks.

3.1.1 Overview

The GMM estimator jointly recovers the production function and demand parameters from three blocks of moment conditions:

(i) **Block A** (Proxy moments): orthogonality conditions derived from eliminating ω_{jt} across pairs of demand residuals and the production residual, using an asymmetric instrument strategy;

(ii) **Block B** (Covariance moments): cross-covariance restrictions among demand and production residuals, exploiting the mutual independence of demand shocks;

(iii) **Block C** (Curvature moments): conditional moment restrictions derived from the homothetic regularity condition on $\mathbb{E}[\omega_{jt} \mid k_{jt}, l_{jt}]$ (Assumption 3), which closes the $\Delta(k, l)$ identification gap characterized in Theorem 2.

Blocks A and B identify the intermediate input elasticities $(\beta_m, \beta_e, \beta_w)$, the demand function parameters (θ_g, ψ_ω), and certain composite functions of (β_k, β_l) and the demand slopes. However, as shown in Section 2.4.2, these blocks alone cannot separate β_k and β_l from the demand function slopes on (k, l) due to the $\Delta(k, l)$ observational equivalence (Theorem 2). Block C resolves this indeterminacy through the nonlinear curvature of $\mathbb{E}[\omega \mid k, l]$ imposed by Assumption 3, thereby achieving point identification of all structural parameters (Theorem 3). When its identifying conditions are weak, the exclusion restriction of Corollary 1 provides an alternative route. Figure 17 (Appendix L) provides a visual overview of the full estimation and inference pipeline, including the diagnostic branches that determine which identification route applies.

3.1.2 Model Specification and Parameters

The parametric specialization below implements the identification results of Section 2 under additive separability; this restriction reduces the nonparametric problem to a finite-dimensional GMM system while preserving all theoretical properties of Theorems 2–3. To apply GMM, I impose additive separability on both the production and demand functions.

Production function. Following the parametric model of Section 2.4.5, the production function is specified as:

$$y_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \beta_e e_{jt} + \beta_w w_{jt} + \omega_{jt} + \varepsilon_{jt}. \tag{12}$$

Here $g(k, l; \theta) = \beta_k k + \beta_l l$ is the parametric (k, l) component and $q(m, e, w) = \beta_m m + \beta_e e + \beta_w w$ is the (linear) intermediate input component, corresponding to the additively separable model (10).

Demand functions. The intermediate input demands take the additively separable form:

$$m_{jt} = \gamma_k\, k_{jt} + \gamma_l\, l_{jt} + h_m(z_{jt}) + \gamma_\omega\, \omega_{jt} + \tau_{jt}, \tag{13}$$

$$e_{jt} = \delta_k\, k_{jt} + \delta_l\, l_{jt} + h_e(z_{jt}) + \delta_\omega\, \omega_{jt} + \nu_{jt}, \tag{14}$$

$$w_{jt} = \zeta_k\, k_{jt} + \zeta_l\, l_{jt} + h_w(z_{jt}) + \zeta_\omega\, \omega_{jt} + \eta_{jt}, \tag{15}$$

where the functions h_m, h_e, h_w are left unrestricted and $\psi_\omega = (\gamma_\omega, \delta_\omega, \zeta_\omega)$ are the productivity loading coefficients.[11] The demand slope parameters $\theta_g = (\gamma_k, \gamma_l, \delta_k, \delta_l, \zeta_k, \zeta_l)$ and the productivity loadings ψ_ω are estimated jointly by GMM together with the $3\,d_z$ coefficients of h_m, h_e, h_w on the polynomial basis in z.

Homothetic structure of $\mathbb{E}[\omega \mid k, l]$. Under Assumption 3 (Homothetic Weak Separability), the conditional expectation of productivity admits the representation $\mathbb{E}[\omega_{jt} \mid k_{jt}, l_{jt}] = h(v(k_{jt}, l_{jt}))$. The economic motivation is discussed in Section 2.4.5. I parametrize the index function using a CES aggregator:

$$v_{jt}(\alpha, \rho_v) = \frac{1}{\rho_v} \log\!\big(\alpha\, e^{\rho_v\, k_{jt}} + (1-\alpha)\, e^{\rho_v\, l_{jt}}\big), \tag{16}$$

which, in levels, corresponds to the CES aggregator $V = \big(\alpha\, K^{\rho_v} + (1-\alpha)\, L^{\rho_v}\big)^{1/\rho_v}$. This nests the Cobb–Douglas case ($\rho_v \to 0$, where $v \to \alpha\, k + (1-\alpha)\, l$) as a special case and satisfies the degree-one homogeneity requirement (Assumption 3(B)) and the strict convexity of isoquants (Assumption 3(C)) for $\alpha \in (0, 1)$ and any $\rho_v \neq 0$. The transformation function h is approximated by a cubic polynomial:

$$h(v;\, \rho) = \rho_1\, v + \rho_2\, v^2 + \rho_3\, v^3, \tag{17}$$

where the constant ρ_0 is absorbed by de-meaning prior to estimation. Under the normalization $\mathbb{E}[\omega] = 0$, the constant satisfies $\rho_0 = -\mathbb{E}[\rho_1 v + \rho_2 v^2 + \rho_3 v^3]$; this constant is not separately identified from the production function intercept and is recovered post-estimation. Condition (A) of Assumption 3 (h' non-constant) requires $\rho_2 \neq 0$ or $\rho_3 \neq 0$; this is a necessary condition for the identification of β_k and β_l (Theorem 3). I report results for polynomial orders 3 through 5 as a robustness check; computational details including the parametrization of h are in Appendix I.4.

Parameter classification. The full parameter vector is $\Theta = (\theta_1', \theta_2')'$, where:

$$\begin{aligned} \theta_1 &= (\beta_m,\, \beta_e,\, \beta_w,\, \theta_g,\, \psi_\omega) && \text{(intermediate input and demand parameters)}, \\ \theta_2 &= (\beta_k,\, \beta_l,\, \alpha,\, \rho_1,\, \rho_2,\, \rho_3) && \text{(primary input and homothetic parameters)}. \end{aligned}$$

Residuals. Define the observable residuals, where the nuisance functions $h_m(z), h_e(z), h_w(z)$ are estimated jointly as described below:

$$\tilde{m}_{jt} \equiv m_{jt} - \gamma_k\, k_{jt} - \gamma_l\, l_{jt} - h_m(z_{jt}) = \gamma_\omega\, \omega_{jt} + \tau_{jt}, \tag{18}$$

$$\tilde{e}_{jt} \equiv e_{jt} - \delta_k\, k_{jt} - \delta_l\, l_{jt} - h_e(z_{jt}) = \delta_\omega\, \omega_{jt} + \nu_{jt}, \tag{19}$$

$$\tilde{w}_{jt} \equiv w_{jt} - \zeta_k\, k_{jt} - \zeta_l\, l_{jt} - h_w(z_{jt}) = \zeta_\omega\, \omega_{jt} + \eta_{jt}, \tag{20}$$

$$\tilde{y}_{jt} \equiv y_{jt} - \beta_m m_{jt} - \beta_e e_{jt} - \beta_w w_{jt} = \beta_k k + \beta_l l + \omega_{jt} + \varepsilon_{jt}. \tag{21}$$

The equalities following the definition signs hold at the true parameter values. The nuisance functions $h_m(z), h_e(z), h_w(z)$ are approximated by second-degree polynomials in z and estimated jointly with the structural parameters; details are in Appendix I.3.

[11] The Cobb–Douglas first-order condition (Appendix B) structurally constrains the demand function to be linear in (k, l, ω), but imposes no restriction on the functional form of the dependence on z. The state variables z_{jt} enter through input prices $\ln P_{h,jt}$, the common market factor $\ln(P_{jt}/\mu_{jt})$, markdowns $\ln \psi_{h,jt}$, and wedges $\ln \Upsilon_{h,jt}$ (equation (52)), each of which may depend nonlinearly on z.

3.1.3 Moment Conditions

Block A: Proxy Moments. [12] By eliminating ω_{jt} across pairs of residuals, I construct three error terms that depend only on the structural shocks:

$$u_{1,jt} \equiv \delta_\omega \,\tilde{m}_{jt} - \gamma_\omega \,\tilde{e}_{jt} = \delta_\omega \,\tau_{jt} - \gamma_\omega \,\nu_{jt}, \tag{22}$$

$$u_{2,jt} \equiv \zeta_\omega \,\tilde{m}_{jt} - \gamma_\omega \,\tilde{w}_{jt} = \zeta_\omega \,\tau_{jt} - \gamma_\omega \,\eta_{jt}, \tag{23}$$

$$u_{3,jt} \equiv \gamma_\omega \,\tilde{y}_{jt} - \tilde{m}_{jt} = \gamma_\omega \,\varepsilon_{jt} - \tau_{jt}. \tag{24}$$

An asymmetric instrument strategy assigns different instruments to each error based on the shock composition. Since $u_{i,jt}$ excludes certain shocks, the corresponding intermediate inputs serve as valid additional instruments (Appendix I.1):

$$\mathbb{E}\big[(Z_{\text{base}},\, w_{jt}) \otimes u_{1,jt}(\Theta)\big] = \mathbf{0}, \tag{25}$$

$$\mathbb{E}\big[(Z_{\text{base}},\, e_{jt}) \otimes u_{2,jt}(\Theta)\big] = \mathbf{0}, \tag{26}$$

$$\mathbb{E}\big[(Z_{\text{base}},\, e_{jt},\, w_{jt}) \otimes u_{3,jt}(\Theta)\big] = \mathbf{0}, \tag{27}$$

where $Z_{\text{base},jt} = (k_{jt},\, l_{jt},\, z_{jt})$. Block A is invariant to the $\Delta(k,l)$ transformation of Theorem 2 and therefore cannot separately identify β_k from the demand slopes on (k,l) (Appendix I.1).

Block B: Covariance Moments. Let ϕ_h denote the productivity loading of residual $\tilde{h}$: $\phi_m \equiv \gamma_\omega$, $\phi_e \equiv \delta_\omega$, $\phi_w \equiv \zeta_\omega$, and $\phi_y \equiv 1$. The mutual exogeneity of shocks (Assumption A.4(3)) implies that $\mathrm{Cov}(\tilde{h}_1, \tilde{h}_2) = \phi_{h_1}\,\phi_{h_2}\,\mathrm{Var}(\omega)$ for each pair $(h_1, h_2) \in \{y, m, e, w\}$, $h_1 \neq h_2$. Eliminating $\mathrm{Var}(\omega)$ across the six distinct pairs yields six covariance relations of the form

$$\mathbb{E}\big[\tilde{h}_1\,\tilde{h}_2 - \phi_{h_2}\,\tilde{y}\,\tilde{h}_1\big] = 0, \tag{28}$$

for each pair (Appendix I.2 lists the individual conditions). Of these six relations, four are algebraically implied by the Block A instrumental variable moments: the conditions involving cross-products of the demand residuals $\tilde{e}$ and $\tilde{w}$ with the proxy equation errors are already encoded in the Block A moment conditions through the instruments $Z_3 = (k, l, \tilde{e}, \tilde{w})$. Consequently, Block B contributes only two independent moment conditions beyond Block A, and the combined Block A+B system is just-identified. The concentrated covariance-ratio formulas derived in Appendix I.2 remain useful for obtaining closed-form scale parameter estimates, improving computational efficiency. As with Block A, Block B is invariant to the $\Delta(k,l)$ transformation.

Block C: Curvature Moments. Block C resolves the $\Delta(k,l)$ indeterminacy by implementing the homothetic regularity condition (Assumption 3, Theorem 3).

Define the net output residual $\tilde{y}_{jt}(\theta_1) = y_{jt} - \beta_m m_{jt} - \beta_e e_{jt} - \beta_w w_{jt}$. Evaluating at the true parameter vector Θ_0, the production function (12) gives $\tilde{y}_{jt} = \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \varepsilon_{jt}$. Taking the conditional expectation with respect to (k_{jt}, l_{jt}):

$$\mathbb{E}[\tilde{y}_{jt} \mid k, l] = \beta_k\, k + \beta_l\, l + h(v(k,l)). \tag{29}$$

[12] The moment conditions require Assumption A.4 (Appendix A), which is implied by the zero conditional mean condition together with Assumption 2.

The first step uses $\mathbb{E}[\varepsilon_{jt} \mid k, l] = 0$, which follows from Assumption 1 by the law of iterated expectations. The second step uses Assumption 3. No structural decomposition of ω_{jt} is postulated; equation (29) follows entirely from the definition of conditional expectation and the regularity condition on its functional form.

Define the structural error:

$$u_{jt}(\Theta) \equiv \tilde{y}_{jt}(\theta_1) - \beta_k\, k_{jt} - \beta_l\, l_{jt} - h\big(v(k_{jt}, l_{jt};\, \alpha);\, \rho\big). \tag{30}$$

Equation (29) implies $\mathbb{E}[u_{jt} \mid k_{jt}, l_{jt}] = 0$ at Θ_0, which yields valid moment conditions with any function of (k, l) as instruments. I use the polynomial instrument vector:

$$Z_{2,jt} = \big(k_{jt},\ l_{jt},\ k_{jt}^2,\ l_{jt}^2,\ k_{jt}\, l_{jt},\ k_{jt}^3,\ l_{jt}^3,\ k_{jt}^2\, l_{jt},\ k_{jt}\, l_{jt}^2\big)', \tag{31}$$

giving the moment conditions:

$$\mathbb{E}\big[Z_{2,jt} \cdot u_{jt}(\Theta)\big] = \mathbf{0}. \tag{32}$$

As with Block A, the constant term is excluded from $Z_{2,jt}$ and ρ_0 (the intercept of h) is recovered post-estimation from the de-meaned residuals.

Identification mechanism. The structural error u_{jt} depends on $\theta_2 = (\beta_k, \beta_l, \alpha, \rho_1, \rho_2, \rho_3)$. Theorem 3 establishes that under Assumption 3, the $\Delta(k, l) = c_k k + c_l l$ transformation is incompatible with the homothetic structure unless $(c_k, c_l) = (0, 0)$. Operationally, this identification works through the higher-order instruments in $Z_{2,jt}$: the nonlinear terms v^2 and v^3 in h interact with the homogeneity of v in a manner that uniquely pins down β_k and β_l.

If $\rho_2 = \rho_3 = 0$ (i.e., h is linear), then β_k and $\rho_1\alpha$ are linearly confounded and identification fails. The significance of $\hat{\rho}_2$ and/or $\hat{\rho}_3$ therefore serves as a diagnostic for the strength of identification. I report estimates and standard errors of these parameters in both the simulation and the empirical analysis.[13]

3.1.4 De-Meaning, Intercepts, and Estimation Procedure

De-meaning and estimation procedure. All variables are de-meaned prior to estimation and the constant is excluded from all instrument vectors. All parameters $\Theta = (\theta_1, \theta_2)$ are estimated simultaneously by two-step GMM:

$$\hat{\Theta} = \arg\min_{\Theta}\ g_N(\Theta)'\, \hat{W}\, g_N(\Theta), \qquad g_N(\Theta) = \frac{1}{N}\sum_{j=1}^{N} \bar{g}_j(\Theta), \tag{33}$$

where $\bar{g}_j(\Theta) = T^{-1}\sum_{t=1}^{T} g_{jt}(\Theta)$ stacks all moment conditions, and $\hat{W}$ is the optimal weighting matrix estimated from a first-step identity-weighted GMM. Post-estimation intercepts and further implementation details are in Appendix I.3.

[13] In practice, even when ρ_2 and ρ_3 are nonzero, the near-collinearity between $\rho_1 v(k, l)$ and $(\beta_k k, \beta_l l)$ can impede numerical optimization. I orthogonalize the polynomial basis (v, v^2, v^3) against the linear span of $(1, k, l)$ before constructing h, so that only the nonlinear component of $h(v)$—the source of identification (Theorem 3)—enters the Block C moment conditions. This is a reparametrization: the structural parameters $(\beta_k, \beta_l, \alpha)$ are invariant, while the polynomial coefficients (ρ_1, ρ_2, ρ_3) are redefined as loadings on the orthogonalized basis.

3.1.5 Recovering Productivity

Given the estimated parameters $\hat{\Theta}$, the firm-level productivity measure is computed as:

$$\hat{\omega}_{jt} = y_{jt} - \hat{\beta}_k\, k_{jt} - \hat{\beta}_l\, l_{jt} - \hat{\beta}_m\, m_{jt} - \hat{\beta}_e\, e_{jt} - \hat{\beta}_w\, w_{jt}. \tag{34}$$

If $\varepsilon_{jt} = 0$, this equals ω_{jt}. Otherwise, $\hat{\omega}_{jt} = \omega_{jt} + \varepsilon_{jt}$; the ex-post shock acts as classical measurement error when $\hat{\omega}_{jt}$ is used in subsequent regressions (Theorem 4).

Practical treatment of the $\Delta(k, l)$ indeterminacy. When Block C is not imposed, $\hat{\omega}_{jt}$ includes a location shift $c(k_{jt}, l_{jt})$ (Theorem 2). Since c depends only on (k, l), flexible controls in (k, l) absorb this shift in regression analysis; in difference-in-differences designs with parallel (k, l) trends, c is automatically differenced out. When the identifying restrictions of Section 2.4.3 are imposed, $\Delta(k, l)$ reduces to a constant absorbed by fixed effects. The proposed estimator therefore supports event studies and productivity regressions without requiring Block C: the $\Delta(k, l)$ component is controlled via polynomial (k, l) regressors in all subsequent regressions (Section 5). The formal justification is provided by Proposition A.2 and the ATT identification result in Appendix H.4.

3.1.6 Asymptotic Properties

Under standard regularity conditions (Appendix I.5), the GMM estimator satisfies:

Theorem 4 (Asymptotic Properties of the GMM Estimator)**.** *As $N \to \infty$ with T fixed: (a) $\hat{\Theta} \xrightarrow{p} \Theta_0$; and (b) $\sqrt{N}\,(\hat{\Theta} - \Theta_0) \xrightarrow{d} N(0, V)$, where*

$$V = (G'WG)^{-1}\, G'W\Sigma WG\, (G'WG)^{-1} \tag{35}$$

with $\Sigma = \mathbb{E}[\bar{g}_j(\Theta_0)\, \bar{g}_j(\Theta_0)']$ and $G = \mathbb{E}[\nabla_\Theta \bar{g}_j(\Theta_0)]$.

Standard errors are clustered at the firm level to accommodate arbitrary within-firm serial dependence. The proof and regularity conditions are in Appendix I.5.

Computational details are in Appendix I.4.

3.1.7 Specification Testing and Diagnostics

Identification count. The combined Block A+B system is just-identified: Block A contributes 10 moment conditions, and Block B contributes exactly two independent moment conditions beyond Block A (four of the six Block B covariance relations are algebraically redundant with Block A; Section 3.1), giving 12 moment conditions matching the 12 free parameters in θ_1. The scale parameters $(\gamma_\omega, \delta_\omega, \zeta_\omega)$ are estimated via closed-form covariance ratios for computational efficiency.

Strength of identification for β_k, β_l. As discussed in Section 3.1.3, the identification of β_k and β_l relies on the nonlinearity of h ($\rho_2 \neq 0$ or $\rho_3 \neq 0$). I report the estimates and t-statistics of $\hat{\rho}_2$ and $\hat{\rho}_3$ as diagnostics. If both are insignificant, the identification of primary input elasticities may be weak, and the researcher should interpret β_k and β_l with caution or consider exclusion restrictions (Corollary 1) as an alternative identification strategy.

Reduced-form check of Assumption 3. As a pre-estimation diagnostic, one may estimate θ_1 from Blocks A and B alone (which does not require Assumption 3), construct $\tilde{y}_{jt}(\hat{\theta}_1)$, and examine whether $\mathbb{E}[\tilde{y} \mid k, l]$ exhibits a homothetic structure via nonparametric regression. A visual departure from homotheticity would indicate a violation of the identifying assumption.

Polynomial degree selection. The cubic specification of h can be extended to higher-order polynomials. I recommend reporting results for polynomial orders 3 through 5 and selecting via information criteria.

Appendix G.6 reports the full-sample Block C recovery results for (β_k, β_l) across all 502 industries, comparing the homothetic approach with the exclusion restriction and ACF estimators.

4 Monte Carlo Simulation

This section evaluates the proposed estimator through Monte Carlo simulations. The primary objective is to assess whether the estimator remains consistent when the first-order Markov assumption is violated, and to compare its performance against existing methods.

4.1 Data Generating Process (DGP)

All DGPs share a common structure for the production function, demand functions, and dynamic input decisions, differing only in the productivity process. Detailed parameter settings are in Appendix E.

4.1.1 Basic Structure

The firm's production function is Cobb–Douglas in all inputs:[14]

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \beta_e e_{jt} + \beta_w w_{jt} + \omega_{jt} + \varepsilon_{jt}, \tag{36}$$

with true parameter values $(\beta_0, \beta_k, \beta_l, \beta_m, \beta_e, \beta_w) = (0.1, 0.2, 0.3, 0.3, 0.15, 0.1)$ and $\varepsilon_{jt} \sim$ i.i.d. $N(0, 0.05^2)$.

Intermediate input demands are log-linear in (k, l, ω) with input-specific demand shocks following independent AR(1) processes ($\rho = 0.5$, $\sigma = 0.15$). The demand function coefficients are calibrated from the first-order conditions of cost minimization under input-specific markdowns (Appendix B); the productivity loading coefficients $(\gamma_\omega, \delta_\omega, \zeta_\omega) = (2.2, 2.0, 1.8)$ differ across inputs, reflecting heterogeneous markdowns.[15] Conditional independence (Assumption 2) is a cross-sectional condition requiring mutual independence across inputs at each point in time; it is unaffected by the serial correlation of individual shocks, since each AR(1) has mutually independent innovations.

Primary inputs are endogenously determined. Capital accumulates through dynamic investment, and labor is chosen based on forecasted productivity from an AR(1) model. The labor demand function is structured so that Assumption 3 holds: the conditional expectation $\mathbb{E}[\omega_{jt} \mid k_{jt}, l_{jt}]$ is a function of a CES aggregator with $(\alpha, \rho) = (0.4, 0.3)$. Full parameter details are provided in Appendix E.

To test the robustness of the proposed method, I generate productivity under three scenarios:

[14] The Cobb–Douglas specification is standard in Monte Carlo studies of production function estimators (Ackerberg, Caves, and Frazer 2015; Gandhi, Navarro, and Rivers 2020). Evaluating the proposed method under more flexible production functions (e.g., translog) is left for future work; the identification results (Theorems 1–3) do not require Cobb–Douglas.

[15] Under perfect competition with a Cobb–Douglas production function, the first-order condition implies $\gamma_\omega = 1/(1 - \beta_m) \approx 1.43$; the larger values incorporate input-specific markdowns and procurement frictions (see Appendix B).

1. **DGP1: AR(1) Markov Process (Baseline).** The standard case where existing methods are correctly specified. $\omega_{jt} = 0.8\,\omega_{j,t-1} + \xi_{jt}$, with $\sigma_\xi = 0.2$.

2. **DGP2: AR(2) Process.** Productivity depends on its own two-period history, as when R&D investments require two years to affect efficiency. The first-order Markov assumption is violated. $\omega_{jt} = 0.6\,\omega_{j,t-1} + 0.3\,\omega_{j,t-2} + \xi_{jt}$, $\sigma_\xi = 0.15$.

3. **DGP3: Potential Outcome Model.** A firm's realized productivity is determined by an endogenous binary treatment D_{jt}, generating a potential outcome process incompatible with the first-order Markov assumption. Following the diagonal reference model of Chen, Liao, and Schurter (2024), untreated productivity follows $\omega^0_{jt} = 0.8\,\omega^0_{j,t-1} + \xi_{0,jt}$ ($\sigma_{\xi 0} = 0.2$) and treated productivity follows $\omega^1_{jt} = 0.5\,\omega^1_{j,t-1} + 0.15 + \xi_{1,jt}$ ($\sigma_{\xi 1} = 0.25$). Observed productivity is $\omega_{jt} = (1 - D_{jt})\omega^0_{jt} + D_{jt}\omega^1_{jt}$. Treatment is reversible and endogenous: $D_{jt} = \mathbb{I}(\omega^0_{jt} > 0)$, so firms enter and exit treatment as their untreated potential productivity fluctuates above and below zero. Full parameter settings, including the capital accumulation and labor decision rules common to all DGPs, are provided in Appendix E.

4. **DGP4: Conditional Independence Violation.** The productivity process is AR(1) as in DGP1, but the electricity demand shock ν_{jt} and the water demand shock η_{jt} are correlated via a common factor: $\nu_{jt} = \sqrt{1-\rho^2}\,\sigma_\nu\,\epsilon_\nu + \rho\,\sigma_\nu\,\epsilon_{\text{common}}$ and similarly for η_{jt}, where $\epsilon_{\text{common}} \sim \mathcal{N}(0,1)$ is independent of ω_{jt}; the materials shock τ_{jt} remains independent. This generates $\text{Corr}(\nu_{jt}, \eta_{jt}) = \rho \in \{0, 0.05, 0.10, 0.20, 0.30\}$, directly violating Assumption 2 when $\rho > 0$. A common energy price shock or seasonal supply constraint that simultaneously raises both electricity and water costs is one economic interpretation—arguably the most salient threat to conditional independence, since both are utility services subject to common regulatory and infrastructure conditions. This DGP tests the robustness of the proposed method to violations of the conditional independence assumption.

4.2 Estimation Methods Compared

Using the generated data, I organize the estimation into two parts to isolate the contributions of each block of moment conditions.

Part 1: Flexible input parameters. I estimate the intermediate input elasticities $(\beta_m, \beta_e, \beta_w)$ and compare four estimators.[16] The four estimators are:

1. **Proposed (Block A+B):** The GMM estimator of Section 3.1, using the intermediate input moment conditions (Block A) and the covariance moments (Block B). This part does not identify β_k and β_l, which remain subject to the $\Delta(k,l)$ indeterminacy (Theorem 2).

2. **Standard ACF:** The two-step GMM estimator of Ackerberg, Caves, and Frazer (2015), assuming a first-order Markov process for productivity.

3. **Modified ACF (ACF-Mod):** A variant of the ACF estimator in which the demand shocks $(\tau_{jt}, \nu_{jt}, \eta_{jt})$ are treated as observed and included as controls in the first stage. This ensures scalar unobservability by construction. Any remaining bias in ACF-Mod can therefore be attributed solely to the violation of the Markov assumption, isolating the dynamic misspecification channel.

[16] The main text figures report two of the four estimators (ACF and Proposed). ACF-Mod results are in Appendix F; GNR results are also reported there.

4. **GNR:** The estimator of Gandhi, Navarro, and Rivers (2020), implemented with a polynomial share regression of degree 2 and a degree-3 polynomial for $g(\omega_{t-1})$, with common prices ($P = r = 1$). In the DGP, persistent input-specific demand shocks ($\rho = 0.5$) violate both the FOC premise and the non-persistence condition of GNR (their Appendix O6, Assumption 7), so GNR tests the share regression approach under persistent input market imperfections. GNR is included in Part 1 only, as its second stage is structurally identical to ACF.

Part 2: Fixed input parameters. I additionally estimate (β_k, β_l) by adding the homothetic regularity condition (Block C) to the proposed estimator:

1. **Proposed (Block A+B+C):** The full GMM estimator using all three blocks, with the CES aggregator $v(k,l) = \frac{1}{\rho_v}\log\big(\alpha e^{\rho_v k} + (1-\alpha)e^{\rho_v l}\big)$ evaluated at the true DGP values $(\rho_v, \alpha) = (0.3, 0.4)$.[17] The comparison with Part 1 isolates the contribution of Block C.

2. **Standard ACF** and **ACF-Mod:** Same as above, now evaluated on (β_k, β_l) as well.

4.3 Evaluation Metrics

I report bias, $\text{Bias}(\hat{\beta}) = \mathbb{E}_R[\hat{\beta}^{(r)}] - \beta_{\text{true}}$, and RMSE, $\text{RMSE}(\hat{\beta}) = \sqrt{\mathbb{E}_R[(\hat{\beta}^{(r)} - \beta_{\text{true}})^2]}$, averaged over R Monte Carlo repetitions.

4.4 Simulation Execution

For Part 1, I run $R = 100$ replications for each combination of DGP and estimation method, varying the number of firms $N \in \{50, 200, 500\}$ and the observation period $T \in \{10, 20, 50\}$ to examine the impact of sample size. For Part 2, I run $R = 100$ replications at $(N, T) = (200, 50)$. The parameter estimates obtained in each repetition are collected, and mean bias and RMSE are calculated for comparison. With $R = 100$, the simulation standard error of the estimated bias is approximately $\text{SD}/\sqrt{R}$; for the typical standard deviation of $\hat{\beta}_m$ (≈ 0.005), this yields a simulation uncertainty of ≈ 0.0005, which is small relative to the reported biases.

4.5 Results

I report the Part 1 results in Figures 1 and 2 and the Part 2 results in Figure 3.[18] These results confirm that the proposed estimator performs well under general conditions and illustrate the sensitivity of the ACF framework to violations of the Markov assumption.

Remark on GNR. GNR shares the static identification strategy of the proposed method—both recover β_m from within-period variation without a Markov assumption—but requires competitive input markets with non-persistent demand shocks (their Appendix O6, Assumption 7). The present DGP, which features persistent input-specific shocks ($\rho_\tau = 0.5$), is therefore outside GNR's maintained assumptions by design: the DGP is calibrated to the proposed method's setting, not GNR's. Under GNR's own assumptions ($\tau = \nu = \eta = 0$), the share regression recovers β_m consistently regardless of the productivity process. The simulation results for GNR (Appendix F) should accordingly be read as

[17] These parameters are known by construction in the simulation; the empirical application treats them as unknown and estimates them by profile GMM (Section 5.2).

[18] Additional summary tables, including GNR results, are provided in Appendix F. RMSE convergence plots are in Figure 7.

illustrating the sensitivity of the FOC-based approach to input market imperfections, not as a general performance comparison.

DGP1 (AR(1) Baseline): Under DGP1, where the Markov assumption holds, all three estimators (ACF, ACF-Mod, and Proposed) are consistent. The bias for each method decays toward zero as T increases (Figure 1). The boxplots in Figure 2 corroborate this finding; the proposed method remains centered on the true values. The ACF and ACF-Mod estimators show small positive finite-sample bias that diminishes with sample size (see Appendix F for detailed tables). However, the proposed estimator exhibits larger variance than the ACF estimator under DGP1, resulting in higher RMSE when the Markov assumption is correctly specified (Appendix Table 8). This is the efficiency cost of the static approach: the proposed method trades time-series information for robustness to dynamic misspecification. Under DGP2 and DGP3, this ranking reverses: ACF's bias dominates its variance advantage, yielding larger mean squared error. In applied settings, non-Markov productivity dynamics are the norm rather than the exception: R&D, learning-by-doing, export entry, and policy interventions all generate history-dependent processes. Even in the rare case where the first-order Markov assumption holds (DGP1), the proposed estimator remains unbiased; the only cost is precision. The static identification strategy is also the only approach in this literature that permits event study and difference-in-differences designs, where the treatment itself violates the Markov assumption (Section 5.6).

DGP2 (AR(2)) and DGP3 (Potential Outcome): Under the scenarios where the first-order Markov assumption is violated, Figure 1 reveals a clear divergence in performance. Under DGP2 and DGP3, the ACF estimator exhibits positive bias in $\hat{\beta}_m$ that does not vanish with increasing T, indicating inconsistency. This upward bias is consistent with the transmission of omitted productivity persistence into the materials elasticity. An infeasible oracle benchmark (ACF-Mod) that removes scalar unobservability by treating demand shocks as observed shows comparable bias under both DGP2 and DGP3, confirming that the source is Markov misspecification rather than demand shock contamination (Appendix F).

The proposed method, by contrast, exhibits negligible bias across these specifications. The bias remains close to zero for all values of T under both DGP2 and DGP3. Because the estimator relies solely on static conditional independence, it remains invariant to the underlying productivity dynamics. The main text figures report results for $N = 500$; increasing N reduces variance for all estimators but does not mitigate ACF's asymptotic bias under DGP2 or DGP3 (Appendix Figure 13), confirming that the bias is asymptotic rather than finite-sample.

Block A+B vs. Block A+B+C (Part 2): Part 2 supplements Part 1 by adding Block C to recover (β_k, β_l). I use the design point $(N, T) = (200, 50)$, which matches the Part 1 baseline, to examine whether Block C disturbs the Block A+B parameters. Figure 3 presents the results, where Block C is added to identify (β_k, β_l). In the baseline DGP1, the proposed method recovers both parameters with negligible bias. Under DGP3, where ACF estimates of β_k and β_l collapse toward zero (RMSE $\approx$ 0.20–0.30), the proposed method achieves substantially lower error (RMSE $\approx$ 0.02). The intermediate input elasticities $(\beta_m, \beta_e, \beta_w)$ remain stable between Part 1 and Part 2, confirming that the addition of Block C moments does not contaminate the well-identified flexible input parameters. This stability confirms in finite samples that the joint GMM system does not transmit Block C misspecification into the flexible input estimates: the intermediate input elasticities are identified by Blocks A and B alone

(Theorem 1, specialized to the Cobb–Douglas parametric model of Section 3.1), so any misspecification in Block C affects only (β_k, β_l). Because markups depend solely on β_m (equation (37)), the primary empirical application is insulated from Block C specification.

DGP4 (Conditional Independence Violation): DGP4 examines the cost of violating Assumption 2 by introducing correlation between the electricity demand shock ν_{jt} and the water demand shock η_{jt}—arguably the most economically salient threat to conditional independence, since both are utility services subject to common energy prices and infrastructure constraints. The materials shock τ_{jt} remains independent. The correlation $\rho \equiv \text{Corr}(\nu_{jt}, \eta_{jt})$ varies from 0 to 0.30.

The bias mechanism operates through the scale parameter ζ_ω. Positive $\text{Cov}(\nu, \eta)$ inflates $\text{Cov}(\tilde{e}, \tilde{w})$, causing the concentrated scale estimator $\hat{\zeta}_\omega = \text{Cov}(\tilde{e}, \tilde{w})/\text{Cov}(\tilde{y}, \tilde{e})$ to overestimate ζ_ω (Appendix J). The overestimated $\hat{\zeta}_\omega$ introduces a positive productivity component into the Block A residual $u_2 = \hat{\zeta}_\omega \tilde{m} - \hat{\gamma}_\omega \tilde{w}$, which the GMM compensates by *increasing* $\hat{\beta}_m$—yielding an *upward* bias.

Table 14 (Appendix Figure 15) reports the results. When $\rho = 0$, the proposed method is approximately unbiased. As ρ increases, $\hat{\beta}_m$ exhibits increasing upward bias. The magnitudes suggest that the estimator is robust to moderate violations. Importantly, the bias direction is the *same* as the Markov misspecification bias documented in DGPs 2 and 3 for ACF: both push $\hat{\beta}_m$ upward. Therefore, the empirical finding that the proposed estimator yields *lower* $\hat{\beta}_m$ than ACF (Section 5.4) cannot be attributed to CI violation; it must reflect Markov misspecification bias in ACF.[19]

Table 2 summarizes the bias properties across DGPs 1–3. The proposed method is unbiased across all three specifications, while ACF exhibits positive bias under Markov misspecification.

Table 2: Monte Carlo Summary: Bias Properties across DGPs ($N = 200$, $T = 50$)

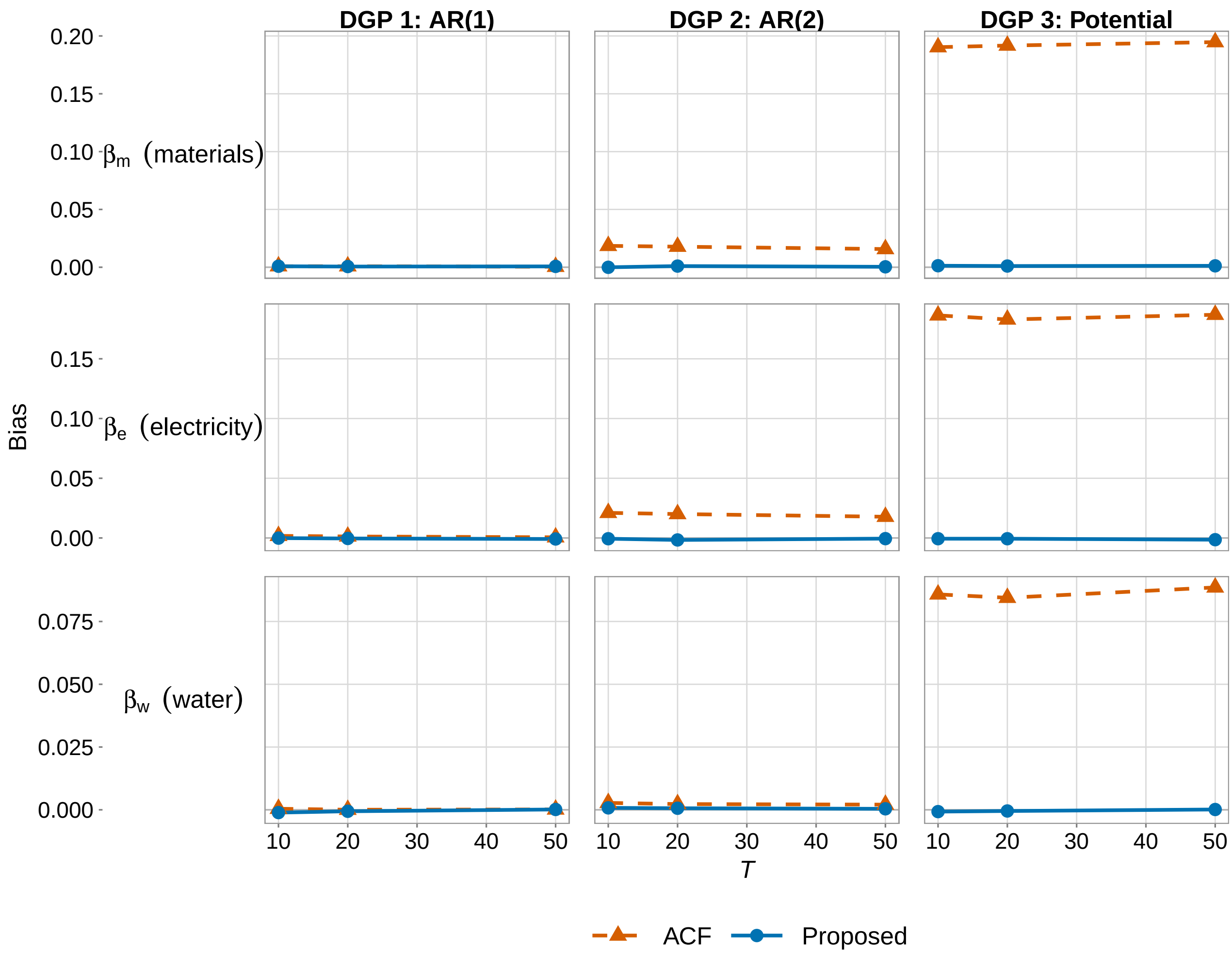
	DGP 1 (AR1)	DGP 2 (AR2)	DGP 3 (PO)
Proposed	Unbiased (+0.002)	Unbiased (−0.001)	Unbiased (+0.000)
ACF	Unbiased (+0.001)	Biased (+0.026)	Biased (+0.266)
GNR	Biased (+0.589)	Biased (+0.589)	Biased (+0.592)

Notes: "Biased (+)" indicates positive asymptotic bias in $\hat{\beta}_m$ that does not diminish with sample size. See Figures 1–3 for detailed convergence plots and Tables 8–13 (Appendix F) for full RMSE and SD by method and DGP.

Taken together, the Monte Carlo simulations confirm that the proposed method consistently recovers production function parameters without imposing restrictions on the productivity process. In contrast, standard methods exhibit substantial positive bias in $\hat{\beta}_m$ when the assumed law of motion for productivity does not match the true data generating process. The simulations establish that Markov misspecification generates a detectable and economically meaningful bias. The empirical application then examines whether these patterns hold in Japanese manufacturing data.

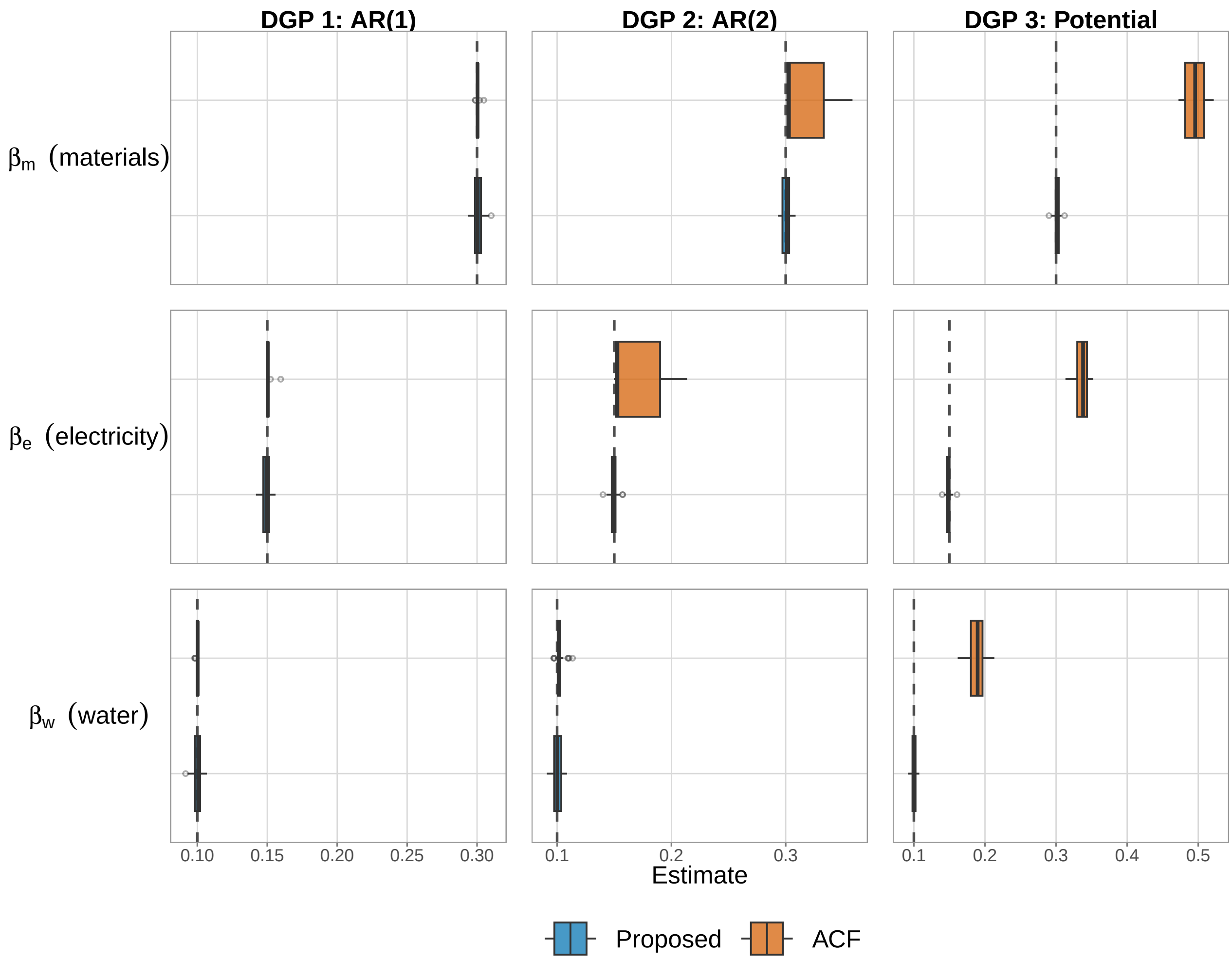
[19] ACF uses only the materials demand proxy and does not exploit cross-shock variation, so it is unaffected by $\text{Corr}(\tau, \eta)$.

Figure 1: Part 1: Mean Bias Convergence ($N = 500$)



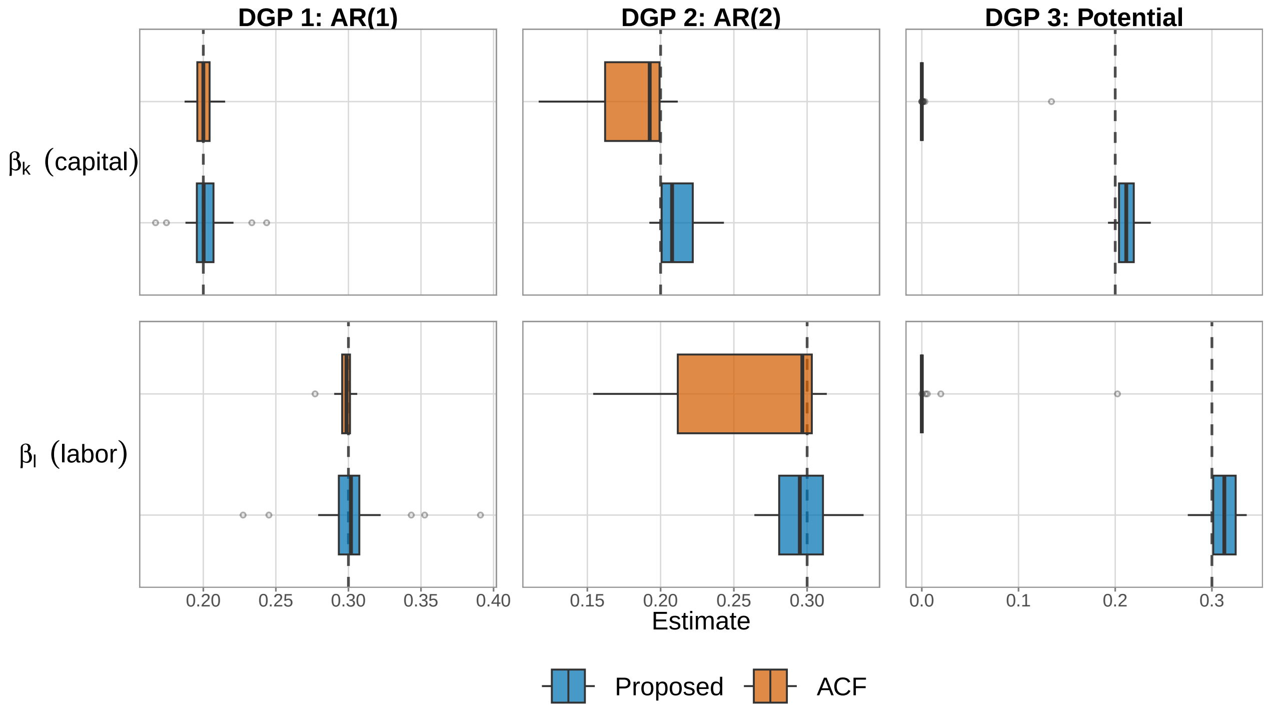
Notes: Mean bias of $(\hat{\beta}_m, \hat{\beta}_e, \hat{\beta}_w)$ *as a function of* T *for three DGPs* ($N = 500$, $R = 100$)*. Under DGP 1 (baseline AR(1)), both methods are approximately unbiased. Under DGPs 2 and 3, where the first-order Markov assumption is violated, ACF exhibits persistent bias while the proposed method remains centered at zero. Three-method comparison including ACF-Mod is in Figure 12; GNR results are in Appendix F.*

Figure 2: Part 1: Distribution of Estimates ($N = 500$, $T = 50$)



Notes: Distribution of $(\hat{\beta}_m, \hat{\beta}_e, \hat{\beta}_w)$ *across* $R = 100$ *replications for* $N = 500$, $T = 50$. *Dashed lines indicate true values. The proposed method remains centered on the true values across all DGPs. Under DGP 2 and DGP 3, ACF distributions are shifted rightward, consistent with the positive Markov misspecification bias. A four-method comparison including ACF-Mod and GNR is in Figure 11.*

Figure 3: Part 2: Distribution of $\hat{\beta}_k$ and $\hat{\beta}_l$ ($N = 200$, $T = 50$)



Notes: Distribution of $(\hat{\beta}_k, \hat{\beta}_l)$ *from Block A+B+C estimation* ($R = 100$)*. The proposed method identifies* (β_k, β_l) *with moderate accuracy across all DGPs. Under DGP 3, the proposed method achieves substantially lower RMSE* (≈ 0.02)*. ACF estimates of* β_k *collapse to near zero under DGP 3 (mean* $\hat{\beta}_k \approx 0.003$*, true value* 0.20*).*

5 Empirical Analysis

The empirical analysis has two objectives: to test whether the conditional independence framework produces economically plausible estimates across the manufacturing sector, and to assess the relative plausibility of the static and dynamic identifying assumptions through the convergence diagnostic of Remark 1. I estimate the production function for all 502 manufacturing industries using Block A+B, reporting analytical standard errors. A practical consequence of this block structure: the markup estimates, productivity determinants, and convergence diagnostics reported below require only Blocks A and B. These results do not depend on the resolution of the $\Delta(k, l)$ indeterminacy and are available for all 502 industries. Block A+B+C is used for a subset of industries where (β_k, β_l) recovery is needed for productivity level analysis.

The section is organized as follows. Sections 5.1 and 5.2 describe the data and estimation specifications. Section 5.3 presents specification the exclusion restriction diagnostic. Section 5.4 reports production function parameters and markup estimates. Section 5.5 examines productivity determinants. Section 5.6 presents an event study application exploiting the non-Markov validity of the estimator. Section 5.7 reports estimates of (β_k, β_l) from two independent identification routes.

5.1 Data and Analytical Framework

I apply the proposed method to the Japanese Census of Manufactures and the Economic Census for Business Activity. I estimate the production function for all manufacturing industries with at least 50 firm-year observations in the extended panel (2003–2020), yielding Block A+B estimates for 502 industries covering 559,381 firm-year observations.[20] These estimates provide markup distributions and productivity determinants at the level of the entire manufacturing sector. Four industries—food processing (Bread, industry code 971), paper products (Corrugated board boxes, code 1453), chemicals (Plastic film, code 1821), and machinery (Industrial robots, code 2694)—serve as representative cases for the time-varying parameter analysis in Appendix G.5, covering major manufacturing sectors (food, paper, chemicals, machinery). Analytical standard errors from the GMM sandwich formula are reported for both the proposed method and the ACF benchmark.

The core variables include the logarithm of real output, y_{jt}, the logarithm of real capital stock, k_{jt}, and the logarithm of labor input, l_{jt}.

I map the theoretical input triplet to observable data as follows. I designate the real value of primary raw materials as m_{jt}, the quantity of electricity as e_{jt}, and the quantity of industrial water as w_{jt}. This selection exploits the fact that industrial water and electricity prices are typically regulated, limiting firm-specific bargaining. This institutional feature reduces the risk of unobserved common price shocks inducing correlation between ν_{jt} and η_{jt}, thereby supporting the validity of the conditional independence assumption ($\tau_{jt} \perp \nu_{jt} \perp \eta_{jt} \mid (\omega_{jt}, x_{jt})$). The principal remaining threat is commodity price shocks that jointly affect raw materials costs and electricity generation costs. Two features mitigate this concern: (i) industrial electricity prices exhibit less high-frequency variation than raw materials procurement costs, as the fuel cost adjustment mechanism smooths commodity price pass-through on a quarterly basis; and (ii) even if a residual common utility shock induces positive $\text{Corr}(\nu_{jt}, \eta_{jt})$, the resulting bias in $\hat{\beta}_m$ is *upward*—the same direction as ACF's Markov bias—so the empirical gap between methods cannot be attributed to CI violation (Section 4, Appendix J).

I augment x_{jt} with control variables z_{jt} consisting of beginning-of-period total inventory ($z_{1,jt}$), its square ($z_{2,jt} \equiv z_{1,jt}^2$), plant fixed effects, and year fixed effects. These controls directly implement the conditioning strategy of Section 2.3, where common shocks are absorbed by z_{jt} so that the residual shock terms $\tau_{jt}, \nu_{jt}, \eta_{jt}$ satisfy the conditional independence assumption. Inventory proxies for unobserved product demand fluctuations (Kumar and Zhang 2019): a firm anticipating high demand accumulates more stock in advance, so inventory captures the common demand component that would otherwise enter all three input demands simultaneously (Section 2.3). The quadratic term $z_{2,jt}$ accommodates a nonlinear relationship between inventory and unobserved demand, consistent with the structural decomposition in equation (52) where demand-related terms enter input prices nonlinearly. Year fixed effects absorb common input price shocks (e.g., energy price movements) that affect all inputs simultaneously, as discussed in Section 2.3. Plant fixed effects absorb time-invariant plant-level heterogeneity in input prices and buyer–supplier relationships, capturing the firm-attribute component of input market power noted in Section 2.3.

5.2 Specification of Estimation Methods

I contrast the results of my approach with those obtained from the standard ACF framework.

First, I implement the **proposed method** using the GMM estimator derived in Section 3.1. The estimator jointly recovers the production function and demand parameters as described in Section 3.1. The CES aggregator parameters (ρ_v, α) are selected via profile GMM: for a grid of (ρ_v, α) values, the remaining parameters are estimated by minimizing the GMM objective, and the pair yielding the

[20] The identification results of Section 2 require only the joint distribution of $(m_{jt}, e_{jt}, w_{jt}, k_{jt}, l_{jt})$ at a single point in time; no assumption on the time-series dynamics of ω_{jt} is needed (Appendix G.5). The panel dimension is exploited solely to improve estimation efficiency by time-averaging the sample moment conditions, $\bar{g}_j(\Theta) = T^{-1} \sum_t g_{jt}(\Theta)$, which reduces finite-sample variance without affecting consistency.

smallest J-statistic is selected.[21] The nuisance functions h_m, h_e, h_w are approximated by second-degree polynomials in $(z_{1,jt}, z_{2,jt})$, giving a polynomial basis of dimension $d_z = 2$ and thus $\dim \Theta = 24$ where Block A+B is just-identified (Section 3.1.7). I report the estimates of $\hat{\rho}_2$ and $\hat{\rho}_3$ as diagnostics for the strength of identification of β_k and β_l (Section 3.1.7).

As a benchmark, I estimate the ACF two-step GMM with the same control variables z_{jt} to ensure comparability. Analytical standard errors from the GMM sandwich formula are reported for both methods.

Identifying assumptions in practice. The ACF framework requires scalar unobservability (productivity as the sole unobservable in input demand) and a first-order Markov process for productivity. GNR requires scalar unobservability and competitive input markets. The proposed method requires conditional independence of input-specific demand shocks. Scalar unobservability rules out procurement relationships, supply contracts, and input-specific markdowns; the proposed method permits these. The GNR competitive input market assumption precludes markup estimation, since the identifying condition coincides with the object of interest. The conditional part of the independence assumption depends on the adequacy of the control variables z_{jt}, but this dependence is shared by the ACF proxy equation.

5.3 Specification Diagnostics

Two diagnostics probe different layers of the identification strategy before any structural results are interpreted:

(i) **Exclusion restriction diagnostic**: tests whether the pairwise discrepancy $d_k = d_l = 0$ (equation (9)), a necessary condition for the exclusion restriction of Corollary 1 that resolves the $\Delta(k, l)$ indeterminacy nonparametrically.

(ii) **Block C diagnostic**: assesses the strength of the CES curvature ($\rho_v \neq 0$), the identifying condition for separate recovery of (β_k, β_l) via Theorem 3 (Appendix G, Table 20).

Table 3 summarizes the three diagnostics and their empirical outcomes.

Table 3: Identification Roadmap and Specification Diagnostics

Diagnostic	Assumption tested	Enables	Null hypothesis	Outcome
Exclusion diagnostic	Excl. restriction (Cor. 1)	Check on $\hat{\beta}_k$	$d_k = d_l = 0$	Capital only
Block C diagnostic	Homotheticity + CES curvature (Thm. 3)	$\hat{\beta}_k$, $\hat{\beta}_l$	$\rho_v \neq 0$	Section 5.7

Notes: The two diagnostics probe successive layers of the identification strategy. Row 1 is tested using Blocks A and B alone and requires only Cobb–Douglas and conditional independence; results are reported in Sections 5.3–5.4. Row 2 additionally invokes Assumption 3 (Homothetic Weak Separability); results are reported in Section 5.7. The exclusion diagnostic ($d_k = d_l = 0$) provides a Wald test with 2 degrees of freedom. Block C diagnostic details are in Table 20.

[21] Under strong identification of (ρ_v, α), the profile GMM procedure yields a J-statistic with the standard χ^2 distribution asymptotically (Newey and McFadden 1994). When identification of these parameters is weak, the minimum-J selection may bias the test toward under-rejection, making the test conservative. The block bootstrap standard errors reported below account for the uncertainty in (ρ_v, α) selection by re-running the profile grid search within each bootstrap replication.

Exclusion restriction diagnostic. Figure 4 applies the exclusion-based OLS recovery of Proposition A.1 to all 502 manufacturing industries, plotting $\hat{\beta}_k^{(m)}$ against $\hat{\beta}_k^{(e)}$ (panel a) and $\hat{\beta}_l^{(m)}$ against $\hat{\beta}_l^{(e)}$ (panel b). Under the exclusion restriction, both panels should cluster along the 45-degree line. Panel (a) confirms this for capital: points concentrate tightly around the diagonal, consistent with $a_k^h = 0$ across industries. Panel (b) reveals the opposite for labor: points scatter widely, indicating that different proxy equations yield systematically different $\hat{\beta}_l$ values.

The asymmetry between capital and labor is the central diagnostic finding. Capital is quasi-fixed within the production period and does not directly influence short-run intermediate input procurement, so $a_k^h = 0$ is economically plausible. The systematic failure for labor is consistent with labor affecting production scheduling, shift patterns, and input utilization through input-specific channels (Appendix B). The formal Wald test of $d_k = d_l = 0$ is rejected for 37% of industries at the 5% level, while the labor-only Wald test ($d_l = 0$) is rejected for 28% of industries, confirming that the labor exclusion restriction is violated for a substantial share of the sample while capital passes in most cases.

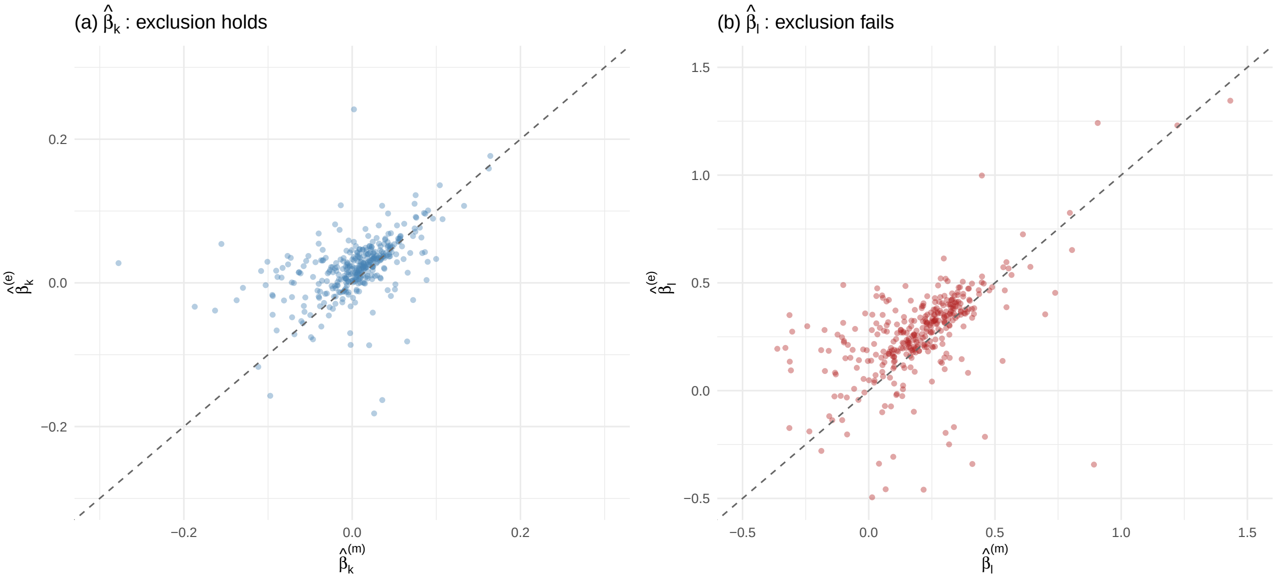


Figure 4: Recovery of (β_k, β_l) via the Exclusion Restriction

Notes: Each industry's β_k and β_l are recovered via OLS from each proxy equation using Proposition A.1. Panel (a): $\hat{\beta}_k^{(m)}$ versus $\hat{\beta}_k^{(e)}$. Panel (b): $\hat{\beta}_l^{(m)}$ versus $\hat{\beta}_l^{(e)}$. Dashed lines are the 45-degree reference. Under the exclusion restriction, both panels should cluster along the diagonal. Outliers $|\hat{\beta}| > 2$ are trimmed for readability; the full distribution is reported in Table 7.

5.4 Production Function Parameters and Markups

The intermediate input elasticities $(\beta_m, \beta_e, \beta_w)$ are identified by Blocks A and B alone (Theorem 1, specialized to the Cobb–Douglas parametric model of Section 3.1), without requiring Block C or Assumption 3. The ACF estimates of $\hat{\beta}_m$ are systematically higher than those from the proposed method (Table 15 in Appendix G), consistent with the Markov misspecification bias documented in the Monte Carlo simulations (Section 4). The cross-industry distribution of all Block A+B and Block C parameter estimates is reported in Table 15 in Appendix G.

Electricity and water elasticities are small across industries (median $\hat{\beta}_e = 0.001$ and $\hat{\beta}_w = 0.006$, respectively), consistent with these inputs serving auxiliary rather than central production roles in Japanese manufacturing. Their demand shocks nevertheless remain valid exclusion restrictions for identifying β_m in the proposed GMM.

Under perfect competition, β_m equals the revenue share, which is the basis of GNR's share regression. My estimator identifies β_m independently of the first-order condition, permitting imperfect competition in both product and input markets.

Markups. Markups are computed following De Loecker and Warzynski (2012). Under the Cobb-Douglas specification maintained throughout, the markup formula simplifies to

$$\hat{\mu}_{jt} = \frac{\hat{\beta}_m}{s_{m,jt}}, \tag{37}$$

where $s_{m,jt}$ denotes the expenditure share of raw materials in total revenue. Unlike the standard production approach, in which Hicks-neutral productivity and scalar unobservability jointly imply $\hat{\beta}_h/s_{h,jt} = \mu_{jt}$ for every variable input h—so that materials, labor, and energy serve as interchangeable markup proxies—this paper allows input-specific markdowns $\psi_{h,jt}$ for each static input $h \in \{m, e, w\}$, captured by the demand shocks $(\tau_{jt}, \nu_{jt}, \eta_{jt})$ (Appendix B). Consequently, $\hat{\beta}_h/s_{h,jt}$ will generally differ across inputs by design; this divergence reflects the richer structure of the framework, not an overidentification failure. Raw materials are selected for markup computation because competitive commodity markets support the absence of buyer-side market power ($\psi_{m,jt} \approx 1$; Avignon and Guigue (2025)), giving $\hat{\beta}_m/s_{m,jt} \approx \mu_{jt}$; this is a maintained assumption.[22] These estimates require only Blocks A and B and are invariant to the $\Delta(k, l)$ indeterminacy (Theorem 2), since equation (37) depends only on $\hat{\beta}_m$ and the observable cost share. I restrict the comparison to industries with at least 50 firms ($N_{\text{firms}} \geq 50$), which removes industries where the lower bound $\hat{\beta}_m \approx 0$ reflects identification failure rather than true low input elasticities.[23]

Comparison with ACF. Figure 5 plots the empirical CDF of industry-level median markups under the proposed method and the ACF benchmark for the $N_{\text{firms}} \geq 50$ subsample. The two distributions are stochastically ordered: the ACF CDF lies strictly to the right of the proposed CDF at every percentile (Table 4). The proposed method yields a median markup of 0.926, while ACF yields 1.027—a gap of 0.101 at the median. At the 90th percentile the gap widens to approximately 0.15. Under the proposed method, 37% of industries show markups above unity, compared with 54% under ACF.

The Monte Carlo evidence in Section 4 provides a structural interpretation. Under DGP 3 (potential-outcome dynamics, Table 10), ACF incurs a bias of +0.19 in $\hat{\beta}_m$ (true value 0.30), a 63% relative overestimate, while the proposed estimator is essentially unbiased (bias = 0.001). The empirical gap of +0.10 at the median corresponds to a relative overestimate of roughly 11% in $\hat{\beta}_m$, well within the range predicted by the DGP 3 calibration. The evidence is therefore consistent with the theoretical prediction that ACF overestimates $\hat{\beta}_m$ when productivity dynamics deviate from the Markov assumption.

[22] The empirical specification imposes Hicks-neutral Cobb-Douglas production; if factor-augmenting productivities differ across inputs, $\hat{\beta}_m$ may absorb non-neutral components and bias the markup estimate (Raval 2023). The identification theory accommodates non-Hicks-neutral production (Appendix C), but the implemented GMM does not exploit this generality.

[23] The value-added markup $\mu^{VA} = \beta_l^{VA}/s_l^{VA}$ can differ substantially from the gross output markup when the materials share is large. Gandhi, Navarro, and Rivers (2017) document that gross output and value-added specifications yield fundamentally different productivity estimates. I report gross output markups throughout.

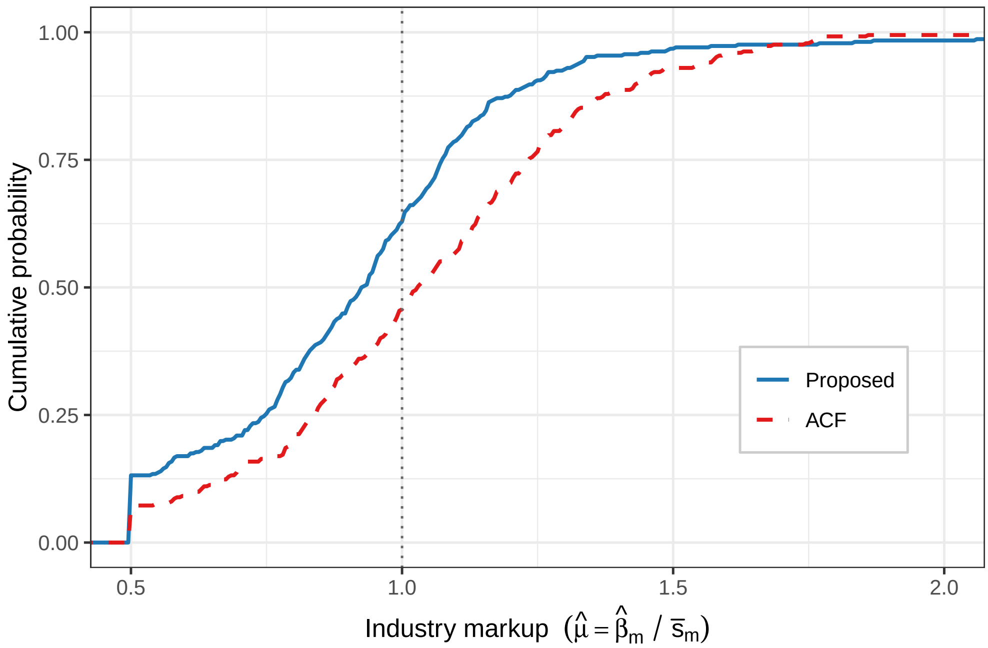


Figure 5: Empirical CDF of Industry Markups: Proposed vs. ACF

Notes: Empirical CDFs of industry-level median markups $\hat{\mu} = \hat{\beta}_m/\bar{s}_m$ under the proposed method (solid, blue) and ACF (dashed, red). Sample restricted to industries with $N_{\text{firms}} \geq 50$ ($N = 372$ industries). Vertical dotted line at $\hat{\mu} = 1$. Summary statistics in Table 4.

Table 4: Markup Distribution: Proposed vs. ACF ($N_{\text{firms}} \geq 50$)

	Proposed	**ACF**
N (industries)	372	372
Mean	0.880	1.026
Std. dev.	0.396	0.360
p10	0.296	0.627
p25	0.748	0.841
Median	0.926	1.027
p75	1.074	1.233
p90	1.240	1.431
Fraction ≥ 1	0.371	0.543
Mean gap (ACF − Proposed)	0.146	—

Notes: Industry-level median markups $\hat{\mu} = \hat{\beta}_m/\bar{s}_m$, where $\bar{s}_m$ is the industry median materials share. Sample: $N_{\text{firms}} \geq 50$. ACF estimates from Ackerberg, Caves, and Frazer (2015); convergence code 0 for 495 of 502 industries.

5.5 Productivity Determinants

The following analysis requires only Blocks A and B. Because the $\Delta(k, l)$ indeterminacy (Theorem 2) varies only through (k_{jt}, l_{jt}), it is absorbed by firm fixed effects. The same argument applies to proportional common shocks (Section 2.3): if an unobserved common component ξ_{jt} loads proportionally on all intermediate input demands, it is absorbed into the recovered productivity $\hat{\omega}_{jt}$, and its (k_{jt}, l_{jt})-dependent component is absorbed by firm fixed effects. The determinants regression therefore identifies the association between covariates and the total latent efficiency measure that drives input allocation decisions, regardless of whether this measure coincides with physical productivity.

As a validation of the recovered productivity measures, I examine their association with observable economic fundamentals. I regress the productivity residual jointly on three firm-level covariates—log investment, exporter status (a binary indicator for export participation, consistent with the learning-by-exporting literature), and log wages—with firm and year fixed effects:

$$\hat{\omega}_{jt} = \alpha_j + \gamma_t + \mathbf{x}'_{jt}\boldsymbol{\beta} + u_{jt},$$

clustering standard errors at the firm level. For the proposed method, I additionally include a cubic polynomial in (k_{jt}, l_{jt}) as nonparametric controls, since the $\Delta(k, l)$ indeterminacy enters through capital and labor. The ACF regression omits these controls, as the ACF residual already subtracts $\hat{\beta}_k k + \hat{\beta}_l l$. Log wages is included as a correlate of productivity; a maintained caveat is that wages may be endogenous, as high-productivity firms can share rents with workers, so the coefficient captures association rather than a causal effect.

Table 5 reports the results. Under the proposed method, log wages are strongly positively associated with estimated productivity ($\hat{\beta} \approx 0.139$, $p < 0.01$), and log investment is positive but small ($\hat{\beta} \approx 0.001$, $p < 0.01$), while exporter status is negligible and imprecisely estimated. The ACF regression yields a smaller wage coefficient ($\hat{\beta} \approx 0.109$), consistent with the theoretical prediction that demand contamination attenuates the signal in the recovered productivity measure. The improvement is quantitatively modest and should be interpreted cautiously; the extent to which it generalizes beyond this application requires further investigation.

5.6 Event Study: 2011 Tohoku Earthquake

Because the proposed estimator recovers productivity from static covariances alone, its estimates are valid under any productivity dynamics—Markov or otherwise (Proposition A.2). Standard proxy variable estimators embed a Markov transition equation; De Loecker (2013) showed that if a policy treatment alters this transition, the equation must be respecified to include the treatment variable, or the production function parameters themselves are inconsistent. Neither problem arises here, since estimation does not employ a transition equation. The proposed estimates are free of this limitation.

As an illustration, I examine the 2011 Tōhoku earthquake using a difference-in-differences design. The treatment group consists of plants in the three core prefectures directly struck by the earthquake and tsunami—Iwate, Miyagi, and Fukushima (seismic intensity $\geq$ 6-strong)—where physical destruction and the nuclear disaster caused severe and sustained disruption to production. The control group consists of plants in Kinki and western prefectures (prefectures 25–47). Supply chain contamination of the control group is mitigated by the industry$\times$year fixed effects, which absorb any industry-level aggregate shocks that propagate nationally. Pre-treatment coefficients are flat ($\max|\hat{\delta}_t| < 0.013$ for the proposed method, < 0.020 for ACF); the full event-study figure is in Appendix G.4.

Table 5: Productivity Determinants: Proposed vs. ACF

	(1) Proposed	(2) ACF
log(Investment)	0.0011***	0.0003**
	(0.0003)	(0.0001)
Exporter Status	0.0131	-0.0010
	(0.0174)	(0.0068)
log(Wage)	0.1388***	0.1085***
	(0.0185)	(0.0127)
Observations	433,425	433,308
R^2	0.86968	0.84551
Firm FE	✓	✓
Time FE	✓	✓

All three covariates enter jointly in a single specification. Firm and year fixed effects included. Standard errors clustered at the firm level in parentheses.
Exporter Status is a binary indicator equal to one if the firm exported in that year, consistent with the learning-by-exporting literature. log(Wage) is included as a correlate of productivity; wage endogeneity is a maintained caveat, as high-productivity firms may pay higher wages through rent-sharing.
Column (1): proposed method with $\text{poly}(k, l, \text{degree} = 3)$ nonparametric controls (coefficients suppressed).
Column (2): ACF residual $\hat{\omega}^{\text{ACF}} = y - \hat{\beta}_k k - \hat{\beta}_l l - \hat{\beta}_m m - \hat{\beta}_e e - \hat{\beta}_w w$.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6 reports the difference-in-differences estimates under both methods. For the proposed method, cubic polynomial controls in (k, l) are included to absorb the $\Delta(k, l)$ indeterminacy in the residual $\hat{\omega}$; the ACF method requires no such controls, as $\hat{\omega}^{\text{ACF}}$ already subtracts $\hat{\beta}_k k + \hat{\beta}_l l$. Both methods detect a negative and statistically significant post-treatment effect on productivity. Under the proposed method, the DiD estimate is -1.28 percent (s.e. 0.52, $p < 0.05$); under ACF, the corresponding estimate is -1.68 percent (s.e. 0.41, $p < 0.01$). The gap between the two estimates is approximately 0.40 percentage points.[24] Proposition A.2 guarantees that the proposed estimates recover $\mathbb{E}[\omega_{jt} \mid D_{jt}]$ under Conditions (i)–(ii) of that proposition. Condition (ii) is satisfied by construction: the earthquake is a natural disaster whose occurrence is orthogonal to firm-level input demand shocks (τ, ν, η). Condition (i) is empirically supported: difference-in-differences estimates of the post-treatment change in intermediate input shares show no significant shift in the materials share ($t = 0.87$) or water share ($t = 1.14$).[25] The ACF estimator does not carry this guarantee. Its residual subtracts $\hat{\beta}_k k + \hat{\beta}_l l$, so

$$\mathbb{E}[\hat{\omega}_{jt}^{\text{ACF}} \mid D_{jt}] = \mathbb{E}[\omega_{jt} \mid D_{jt}] + (\beta_k^{\text{true}} - \hat{\beta}_k^{\text{ACF}})\,\mathbb{E}[k_{jt} \mid D_{jt}] + (\beta_l^{\text{true}} - \hat{\beta}_l^{\text{ACF}})\,\mathbb{E}[l_{jt} \mid D_{jt}]. \tag{38}$$

[24] As a back-of-envelope illustration of the economic magnitude, Japan's manufacturing value added averaged approximately ¥100 trillion per year over 2003–2020 (National Accounts, Cabinet Office of Japan). A 0.40 percentage-point discrepancy in estimated productivity translates to roughly ¥400 billion per year in mismeasured value-added productivity—an amount comparable in scale to major industrial policy programs and sufficient to materially affect the allocation of post-disaster recovery subsidies across industries. This calculation assumes the measured ATT gap reflects a uniform proportional bias in the productivity level, and should be interpreted as indicative rather than structural.

[25] The electricity share shows a small post-treatment increase ($t = 8.91$, $\Delta s_e \approx 0.002$). This is plausibly a compositional effect: the simultaneous contraction of materials usage ($t = -4.53$) raises the electricity share mechanically without altering the structural electricity demand function g_e.

The bias terms vanish only if $\hat{\beta}^{\text{ACF}} = \beta^{\text{true}}$ (exact identification) or if treatment is orthogonal to (k, l). Monte Carlo evidence (Section 4) shows that ACF incurs positive bias in $\hat{\beta}_m$ under Markov misspecification; the condition of orthogonality also fails here (DiD(l) $= -0.029$, $t = -7.86$). The proposed estimates, resting on the theoretical guarantee of Proposition A.2, provide a theoretically justified point of comparison.

Table 6: 2011 Tōhoku Earthquake: Difference-in-Differences

	(1) Proposed	(2) ACF
Treated × Post	-0.0128**	-0.0168***
	(0.0052)	(0.0041)
Observations	219,573	219,573
R^2	0.98729	0.94606
poly(k, ℓ) control	✓	
Firm FE	✓	✓
Ind.×Year FE	✓	✓

Treatment: Iwate, Miyagi, Fukushima (seismic intensity ≥ 6-strong). Control: West Japan (prefectures 25–47).
Firm and industry×year fixed effects included. Heteroskedasticity-robust standard errors in parentheses.
Pre-treatment coefficients are flat ($\max|\hat{\delta}_t| < 0.013$ for proposed, < 0.020 for ACF); year-by-year coefficient estimates in Table 18 (Appendix G.4).
Column (1): proposed method with poly$(k, \ell, \text{degree} = 3)$ nonparametric controls for $\Delta(k, \ell)$ (coefficients suppressed).
Column (2): ACF residual already subtracts $\hat{\beta}_k k + \hat{\beta}_l l$; no polynomial control.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.7 Capital and Labor Inputs

Identifying (β_k, β_l) requires closing the $\Delta(k, l)$ indeterminacy documented in Theorem 2. The paper provides two independent routes: the exclusion restriction (Corollary 1) and the homothetic regularity condition (Theorem 3).

The exclusion-based OLS recovery (Proposition A.1) is applied to all 502 manufacturing industries and produces mutually consistent estimates of β_k across the three proxy equations (materials, electricity, water), while β_l estimates diverge systematically—confirming the diagnostic pattern in Figure 4 that the exclusion restriction holds for capital but not labor.

To identify (β_k, β_l) jointly, I apply the Block C homothetic CES approach (Section 2.4.5). Block C is the primary identification route for both β_k and β_l; the exclusion restriction provides an independent check on β_k only, since the restriction fails for labor (Figure 4, Panel b). The two strategies yield mutually consistent estimates of β_k for the 302 industries where the exclusion restriction is validated for capital (Table 16, Appendix G).

Table 7 summarizes the cross-industry distributions of $\hat{\beta}_k$ and $\hat{\beta}_l$ across three approaches: exclusion restriction (broken out by proxy input), Block C (homothetic CES), and ACF.

Estimates of $\hat{\beta}_k$ are broadly consistent across all three approaches (median ≈ 0.01–0.04), corroborating the identification cross-check in Table 16 (Appendix G). Labor elasticity estimates diverge more substantially: Block C yields a median $\hat{\beta}_l = 0.33$, while ACF produces a higher median of 0.50.

Table 7: Cross-Industry Distribution of $\hat{\beta}_k$ and $\hat{\beta}_l$: Exclusion Restriction, Block C, and ACF

	Excl. (m)	Excl. (e)	Excl. (w)	Block C	ACF
$\hat{\beta}_k$ Median	0.0086	0.0220	0.0106	0.0350	0.0297
$\hat{\beta}_k$ Mean	0.0121	0.0010	-0.0077	0.0481	0.0528
$\hat{\beta}_k$ SD	0.2018	0.2724	0.1935	0.0544	0.0921
$\hat{\beta}_l$ Median	0.2106	0.2600	0.2010	0.3316	0.2753
$\hat{\beta}_l$ Mean	-0.1625	-0.2815	-0.2317	0.3357	0.3217
$\hat{\beta}_l$ SD	3.8991	5.1635	3.7362	0.2239	0.2657
N	389	389	389	502	499

Notes: Industries with $|\hat{\beta}| > 2$ excluded. Excl. ($m/e/w$): exclusion restriction OLS using each proxy (Proposition A.1). Block C: homothetic CES approach (Theorem 3). ACF: Ackerberg, Caves, and Frazer (2015).

The Monte Carlo simulations (Table 13) show that under DGP 3, ACF $\hat{\beta}_l$ collapses to near zero (bias ≈ -0.30), while the proposed method recovers the true value accurately (bias $\approx +0.01$). The empirical ACF estimate lies above the proposed estimate, which is the opposite direction from the MC collapse. Both patterns reflect the same fragility: ACF labor elasticity identification breaks down when the Markov assumption is violated, with the direction of the deviation depending on the specific dynamics of the data-generating process. Figure 6 shows the full cross-industry density distributions for all three methods. A four-group comparison across identification strategies is reported in Table 16 (Appendix G).

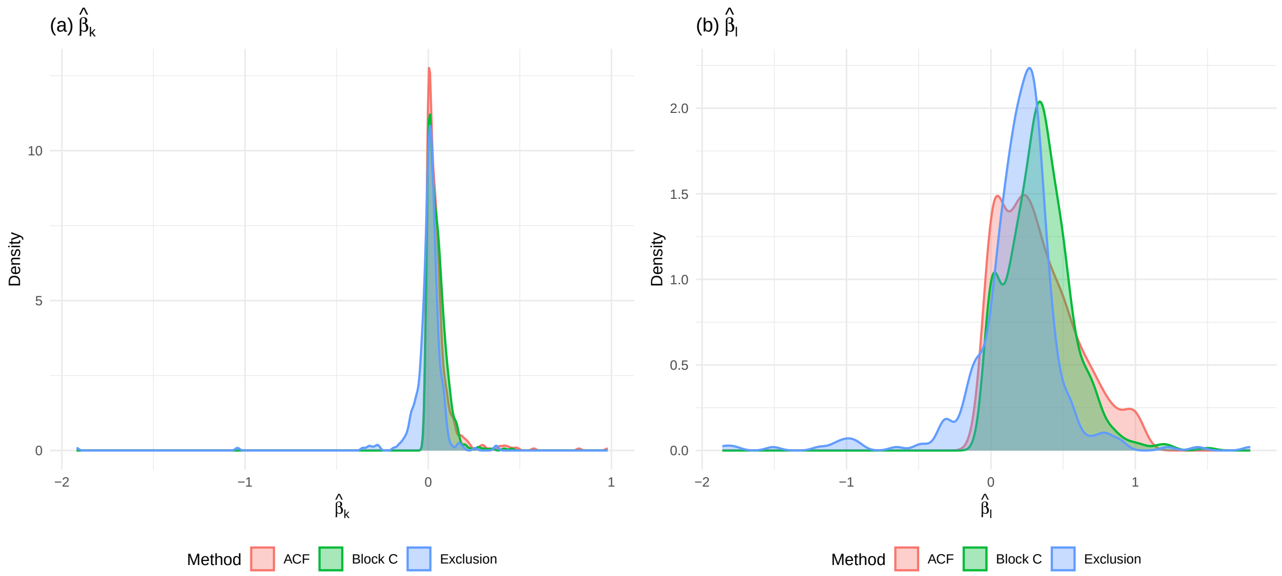


Figure 6: Cross-Industry Distribution of $\hat{\beta}_k$ and $\hat{\beta}_l$: Three Methods

Notes: Panel (a) shows the density of $\hat{\beta}_k$ from Exclusion, Homothetic (Block C), and ACF. Panel (b) shows $\hat{\beta}_l$ for all three methods; Exclusion estimates use the materials proxy (Proposition A.1). Industries with $|\hat{\beta}| > 2$ are excluded. Summary statistics in Table 7; four-group identification cross-check in Table 16 (Appendix G.2).

6 Conclusion

The Markov assumption is not a feature of the data; it is an assumption that existing methods impose to achieve identification. This paper shows that the cross-sectional covariance structure among three flexible intermediate inputs fully replaces this time-series restriction: the production function and the distribution of productivity are nonparametrically identified from a single cross-section. The empirical evidence confirms that the substitution matters. The convergence diagnostic reveals that the exclusion restriction holds for capital but fails for labor; the two frameworks yield systematically different markup distributions across 502 industries; and the recovered productivity measures show stronger associations with economic fundamentals, consistent with the higher signal-to-noise ratio from separating input-specific demand shocks.

The residual indeterminacy is characterized completely (Theorem 2) and closed by two routes: an exclusion restriction with a testable implication (Corollary 1) and a homothetic regularity condition (Theorem 3). Neither requires the Markov assumption. The two routes yield mutually consistent estimates in the subset of industries where the exclusion restriction is validated, providing cross-strategy confirmation of the recovered (β_k, β_l).

These differences imply materially different conclusions for market power measurement, for the magnitude of gains from trade liberalization, and for productivity analysis. The productivity determinant analysis further illustrates the stakes: the proposed method yields associations with economic fundamentals that differ substantially from those of the standard method (Table 5)—the log-wage coefficient is roughly 25% larger—consistent with the prediction that separating input-specific demand shocks raises the signal-to-noise ratio in the recovered productivity measure. Beyond estimation, the static identification strategy opens a methodological gap with the existing literature. Standard proxy variable estimators require a Markov transition equation for ω_{jt}; when a policy treatment alters the transition path, the assumed Markov structure is violated during estimation, making the recovered productivity measures uninformative about counterfactual outcomes. The proposed method identifies the production function from static covariances alone, so its productivity estimates are invariant to how the treatment operates on productivity dynamics (Section 5.6, Proposition A.2).

Three limitations bound the applicability of the results. First, the identification strategy requires at least three intermediate inputs with separately observable quantity data, though this requirement is met in several settings beyond the Japanese Census of Manufactures, including the U.S. EIA Form 923 (Fabrizio, Rose, and Wolfram 2007; Cicala 2015) and emissions data in environmental economics.[26] When labor adjustment is rapid, labor itself serves as an additional productivity signal, reducing the required intermediate inputs from three to two (footnote 6). Second, no targeted test of the conditional independence assumption alone exists; the convergence diagnostic of Remark 1 provides a necessary condition for the exclusion restriction. Extending the moment system to achieve overidentification—for instance via Block C structural constraints or cross-equation demand restrictions under Cobb–Douglas—would enable formal specification testing and is left for future research. Third, Block C identification of (β_k, β_l) requires non-negligible curvature in $h(v)$; when the capital-labor ratio varies little, the exclusion restriction route becomes preferable.

These limitations point to three directions for future work: extending the framework to settings with fewer intermediate inputs by exploiting labor as a productivity signal (footnote 6); extending the moment system to achieve overidentification—for instance via Block C structural constraints or cross-equation demand restrictions—which would enable formal specification testing of the conditional independence assumption; and combining the static identification of flexible input elasticities with

[26] Additional datasets satisfying this requirement include India's Annual Survey of Industries (ASI), which reports firm-level electricity and fuel consumption alongside materials; Canada's Annual Survey of Manufacturing and Logging (ASML), which covers electricity and water use at the establishment level; and the World Bank Enterprise Survey (WBES), which collects firm-level electricity expenditure and water source data across over 100 countries. These datasets enable direct application of the proposed estimator in diverse institutional settings.

semiparametric methods for the capital-labor component when the homothetic regularity condition is not supported by the data.

References

Ackerberg, D., J. Hahn, and Q. Pan (2022). “Nonparametric Identification Using Timing and Information Set Assumptions with an Application to Non-Hicks Neutral Productivity Shocks.” en. In.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). “Identification Properties of Recent Production Function Estimators.” en. In: *Econometrica* 83.6, pp. 2411–2451. ISSN: 1468-0262.

Ackerberg, D. A. and J. De Loecker (2024). “Production Function Identification Under Imperfect Competition.” en. In: *Working Paper*.

Arellano, M. and S. Bond (Apr. 1991). “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” In: *The Review of Economic Studies* 58.2, pp. 277–297. ISSN: 0034-6527.

Avignon, R. and E. Guigue (Mar. 2025). “Markups and Markdowns in the French Dairy Market.” en. In: *Working Papers*.

Blundell, R. and S. Bond (Nov. 1998). “Initial conditions and moment restrictions in dynamic panel data models.” en. In: *Journal of Econometrics* 87.1, pp. 115–143. ISSN: 0304-4076.

Bond, S. and M. Söderbom (Feb. 2005). “Adjustment costs and the identification of Cobb Douglas production functions.” en. In: *IFS Working Papers*.

Brand, J. (2020). “Estimating Productivity and Markups Under Imperfect Competition.” en. In: *Job Market Paper*.

Chen, Z., M. Liao, and K. Schurter (2024). “Identifying Treatment Effects on Productivity: Theory with An Application to Production Digitalization.” en. In.

Cicala, S. (2015). “When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation.” In: *American Economic Review* 105.1, pp. 411–444.

De Loecker, J. (Sept. 2007). “Do exports generate higher productivity? Evidence from Slovenia.” In: *Journal of International Economics* 73.1, pp. 69–98. ISSN: 0022-1996.

— (Aug. 2013). “Detecting Learning by Exporting.” en. In: *American Economic Journal: Microeconomics* 5.3, pp. 1–21. ISSN: 1945-7669.

De Loecker, J. and F. Warzynski (2012). “Markups and Firm-Level Export Status.” In: *American Economic Review* 102.6, pp. 2437–2471.

Demirer, M. (2022). “Production Function Estimation with Factor-Augmenting Technology: An Application to Markups.” en. In.

Doraszelski, U. and J. Jaumandreu (Oct. 2013). “R&D and Productivity: Estimating Endogenous Productivity.” In: *The Review of Economic Studies* 80.4, pp. 1338–1383. ISSN: 0034-6527.

— (June 2018). “Measuring the Bias of Technological Change.” In: *Journal of Political Economy* 126.3, pp. 1027–1084. ISSN: 0022-3808.

Doraszelski, U. and L. Li (June 2025). *Production Function Estimation without Invertibility: Imperfectly Competitive Environments and Demand Shocks*.

Doty, J. (2022). “A Dynamic Framework for Identification and Estimation of Nonseparable Production Functions.” en. In: *Job Market Paper*.

Dunford, N. and J. T. Schwartz (1971). *Linear Operators, Part II: Spectral Theory*. New York: Wiley-Interscience.

Fabrizio, K. R., N. L. Rose, and C. D. Wolfram (2007). “Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency.” In: *American Economic Review* 97.4, pp. 1250–1277.

Gandhi, A., S. Navarro, and D. Rivers (2017). “How Heterogeneous is Productivity? A Comparison of Gross Output and Value Added.” en. In.

Gandhi, A., S. Navarro, and D. A. Rivers (Aug. 2020). “On the Identification of Gross Output Production Functions.” In: *Journal of Political Economy* 128.8, pp. 2973–3016. ISSN: 0022-3808.

Hahn, J., Z. Liao, and G. Ridder (Oct. 2023). “Identification and the Influence Function of Olley and Pakes’ (1996) Production Function Estimator.” en. In: *Econometric Theory* 39.5, pp. 1044–1066. ISSN: 0266-4666, 1469-4360.

Hsieh, C.-T. and P. J. Klenow (Nov. 2009). “Misallocation and Manufacturing TFP in China and India.” In: *The Quarterly Journal of Economics* 124.4, pp. 1403–1448. ISSN: 0033-5533.

Hu, Y., G. Huang, and Y. Sasaki (Apr. 2020). “Estimating production functions with robustness against errors in the proxy variables.” In: *Journal of Econometrics* 215.2, pp. 375–398. ISSN: 0304-4076.

Hu, Y. and S. M. Schennach (2008). “Instrumental Variable Treatment of Nonclassical Measurement Error Models.” en. In: *Econometrica* 76.1, pp. 195–216. ISSN: 1468-0262.

Imbens, G. W. and W. K. Newey (2009). “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity.” en. In: *Econometrica* 77.5, pp. 1481–1512. ISSN: 1468-0262.

Jaumandreu, J. (June 2025). “Robust Production Function Estimation when there is Market Power.” en. In: *Working Paper*.

Jaumandreu, J. and U. Doraszelski (Apr. 2021). “Reexamining the De Loecker & Warzynski (2012) method for estimating markups.” en. In: *CEPR Discussion Papers*.

Kasahara, H., P. Schrimpf, and M. Suzuki (May 2023). *Identification and Estimation of Production Function with Unobserved Heterogeneity*.

Kasahara, H. and Y. Sugita (Oct. 2020). *Nonparametric Identification of Production Function, Total Factor Productivity, and Markup from Revenue Data*.

Kumar, P. and H. Zhang (2019). “Productivity or Unexpected Demand Shocks: What Determines Firms’ Investment and Exit Decisions?” en. In: *International Economic Review* 60.1, pp. 303–327. ISSN: 1468-2354.

Levinsohn, J. and A. Petrin (Apr. 2003). “Estimating Production Functions Using Inputs to Control for Unobservables.” In: *The Review of Economic Studies* 70.2, pp. 317–341. ISSN: 0034-6527.

Li, T. and Y. Sasaki (2024). “Identification of heterogeneous elasticities in gross-output production functions.” In: *Journal of Econometrics* 238.2, p. 105637.

Matzkin, R. L. (2003). “Nonparametric Estimation of Nonadditive Random Functions.” en. In: *Econometrica* 71.5, pp. 1339–1375. ISSN: 1468-0262.

Navarro, S. and D. A. Rivers (2018). “Nonparametric Identification of Productivity in Nonseparable Production Functions.” en. In.

Newey, W. K. and D. McFadden (Jan. 1994). “Chapter 36 Large sample estimation and hypothesis testing.” en. In: *Handbook of Econometrics*. Vol. 4. Elsevier, pp. 2111–2245.

Olley, G. S. and A. Pakes (1996). “The Dynamics of Productivity in the Telecommunications Equipment Industry.” en. In: *Econometrica* 64.6, pp. 1263–1297.

Pan, Q. (2022). “Identification of Gross Output Production Functions with a Nonseparable Productivity Shock.” en. In: *Job Market Paper*.

Raval, D. (2023). “Testing the Production Approach to Markup Estimation.” In: *Review of Economic Studies* 90.5, pp. 2592–2611.

Raval, D. R. (2019). “The micro elasticity of substitution and non-neutral technology.” en. In: *The RAND Journal of Economics* 50.1, pp. 147–167. ISSN: 1756-2171.

Sun, L. and S. Abraham (2021). "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." In: *Journal of Econometrics* 225.2, pp. 175–199.

Zeng, J. (July 2023). *Identification and estimation of a nonseparable production function*. en. SSRN Scholarly Paper. Rochester, NY.

Appendix

Nonparametric Identification and Estimation of Production Functions
Invariant to Productivity Dynamics

Rentaro Utamaru

A Identification Details

This appendix collects the regularity conditions (Assumptions A.1–A.3), the density identification theorem (Theorem 1), and selected propositions that supplement the main identification results in Section 2.

A.1 Regularity Conditions

Assumption A.1 (Injectivity)**.** *The integral operator $L_{e_{jt}|\omega_{jt},x_{jt}}$ with kernel $f_{e_{jt}|\omega_{jt},x_{jt}}$ and the integral operator $L_{\omega_{jt}|w_{jt},x_{jt}}$ with kernel $f_{\omega_{jt}|w_{jt},x_{jt}}$ are both injective.*

Role and economic content. Injectivity requires that distinct productivity levels generate distinct conditional distributions of e_{jt} and w_{jt}: a firm cannot be more productive without systematically altering its input demand. This is weaker than the strict monotonicity plus scalar unobservability required by Ackerberg, Caves, and Frazer (2015): strict monotonicity of the input demand function in productivity is one sufficient condition for injectivity, but injectivity holds more broadly in the presence of idiosyncratic shocks (τ, ν, η) that would violate scalar unobservability. The condition could fail in settings where input allocation is determined by administrative rules rather than optimization; for example, publicly operated utilities where electricity consumption follows fixed schedules regardless of productivity. In competitive manufacturing, the condition is generically satisfied.

Weak identification and eigenvalue decay. Injectivity is a point-identification condition, not a strength condition. Even when the operators $L_{e|\omega,x}$ and $L_{\omega|w,x}$ are injective, identification can be weak in finite samples if the eigenvalues of the associated integral operators decay rapidly to zero. Rapid eigenvalue decay arises when the demand shocks $(\tau_{jt}, \nu_{jt}, \eta_{jt})$ have large variance relative to the productivity signal: a high noise-to-signal ratio compresses the spectrum of $L_{e|\omega,x}$, making the inversion ill-conditioned. Formally, let $\{\lambda_s\}_{s=1}^{\infty}$ denote the eigenvalues of $L_{e|\omega,x}$ in descending order. Rapid decay $\lambda_s/\lambda_1 \to 0$ as $s \to \infty$ implies that the conditional density of e_{jt} given ω_{jt} concentrates its informational content in a low-dimensional subspace, reducing effective identification to that subspace. In the current application, the use of three independent intermediate inputs mitigates this concern: the mutual independence of (τ, ν, η) (Assumption 2) ensures that the information in (e_{jt}, w_{jt}) about ω_{jt} is not colinear, and the empirical diagnostics in Section F.5 provide indirect evidence on identification strength through the significance of $\hat{\rho}_2$ and $\hat{\rho}_3$. Industries where these parameters are insignificant should be interpreted with caution as potentially subject to weak identification of the Block C moments.

Assumption A.2 (Distinct Eigenvalues)**.** *For any $\omega_{jt} \neq \bar{\omega}_{jt}$, the conditional densities $f_{w_{jt}|\omega_{jt},x_{jt}}$ and $f_{w_{jt}|\bar{\omega}_{jt},x_{jt}}$ are not identical as functions of w_{jt}.*

Role and economic content. This condition ensures that the eigenvalues in the spectral decomposition are distinct, which is required for unique identification of the eigenfunctions. It requires that no two distinct productivity levels produce identical distributions of water demand, conditional on (k, l, z). The condition fails if water demand is degenerate or if the conditioning set x_{jt} contains information that renders w_{jt} uninformative about ω_{jt}. As with Assumption A.1, this could be violated in regulated industries where water allocation is rationed (e.g., public irrigation systems), but is generically satisfied in manufacturing where water consumption responds to the scale of production.

Assumption A.3 (Productivity Labeling)**.** *For each fixed (k_{jt}, l_{jt}), the location (labeling) of ω_{jt} is fixed by a normalization corresponding to Assumption 5 in HS08. Specifically, there exists a location* *functional M (e.g., the conditional mean) such that*

$$M\big[f_{m_{jt}|\omega_{jt},x_{jt}}(\cdot \mid \omega_{jt})\big] = \omega_{jt} \quad \textit{for all } \omega_{jt} \in \Omega$$

holds for each (k_{jt}, l_{jt}) separately.

Role and economic content. Assumption A.3 fixes the scale and location of the latent productivity index within each capital-labor cell. The HS08 spectral decomposition identifies the densities up to a relabeling of the latent variable; Assumption A.3 resolves this by requiring that the conditional mean of material demand (as a function of ω) equals ω itself, normalizing productivity to the metric of material demand. The choice of M as the conditional mean is conventional; other location functionals that are equivariant under location shifts yield equivalent identification results, and the linear GMM in Section 3 is invariant to this choice.

Critically, the normalization is applied independently at each (k_{jt}, l_{jt}): for each capital-labor cell, the HS08 decomposition solves a separate measurement error model instance with its own latent variable. The consistency of ω levels across different values of (k, l) is not guaranteed by this assumption alone; see Section 2.4.2.

The analysis further requires that the conditional densities $f_{m|\omega,k,l}$, $f_{e|\omega,k,l}$, and $f_{w|\omega,k,l}$ depend continuously on (k, l) in the L^1 norm for each fixed ω. This regularity condition is satisfied whenever the demand functions g_m, g_e, g_w are continuous in (k, l) and the demand shocks have smooth densities.

A.2 Proof of Theorem 1

Proof of Theorem 1. The proof applies HS08 to the present setting. Consider the joint density $f_{m_{jt},e_{jt}|x_{jt},w_{jt}}$, which is directly identifiable from the data. Using the law of total probability, I introduce unobserved productivity ω_{jt} as a latent variable of integration:

$$f_{m_{jt},e_{jt}|x_{jt},w_{jt}} = \int f_{m_{jt},e_{jt},\omega_{jt}|x_{jt},w_{jt}}\, d\omega_{jt}.$$

Since w_{jt} is a function of $(\omega_{jt}, x_{jt}, \eta_{jt})$ and η_{jt} is independent of (τ_{jt}, ν_{jt}) conditional on (ω_{jt}, x_{jt}) (Assumption 2), conditioning on w_{jt} does not alter the conditional distribution of m_{jt} or e_{jt} given (ω_{jt}, x_{jt}). Applying the chain rule and the conditional independence assumption, I decompose the integrand into the product of three unknown conditional densities:

$$f_{m_{jt},e_{jt}|x_{jt},w_{jt}} = \int f_{e_{jt}|\omega_{jt},x_{jt}} \cdot f_{m_{jt}|\omega_{jt},x_{jt}} \cdot f_{\omega_{jt}|x_{jt},w_{jt}}\, d\omega_{jt}. \tag{39}$$

This equation has the structure of Equation (5) in HS08, whose Theorem 1 establishes uniqueness of the three unknown densities under the maintained assumptions.

The conditional distributions of input demand and productivity are therefore nonparametrically identified from static data alone. □

The roles of m, e, and w in Theorem 1 are interchangeable: any permutation of the three inputs yields the same identification result. The asymmetric instrument strategy in Section 3.1 breaks this symmetry at the estimation stage for efficiency, but the underlying identification is symmetric.

As a consequence of Theorem 1, the density $f_{w_{jt}|\omega_{jt},x_{jt}}$ is also identified. This follows from Bayes' rule:

$$f_{w_{jt}|\omega_{jt},x_{jt}}(w \mid \omega, x) = \frac{f_{\omega_{jt}|x_{jt},w_{jt}}(\omega \mid x, w) \cdot f_{w_{jt}|x_{jt}}(w \mid x)}{f_{\omega_{jt}|x_{jt}}(\omega \mid x)}, \tag{40}$$

where the numerator's first factor is identified by Theorem 1, $f_{w_{jt}|x_{jt}}$ is directly computable from the data, and the denominator is obtained by marginalizing over w_{jt}.

By applying Bayes' rule, the full posterior density of productivity given all observable inputs is identified:

$$f_{\omega_{jt}|x_{jt},m_{jt},e_{jt},w_{jt}}(\omega \mid \cdot) = \frac{f_{m_{jt}|\omega_{jt},x_{jt}}(m \mid \omega, \cdot)\; f_{e_{jt}|\omega_{jt},x_{jt}}(e \mid \omega, \cdot)\; f_{\omega_{jt}|x_{jt},w_{jt}}(\omega \mid \cdot)}{f_{m_{jt},e_{jt}|x_{jt},w_{jt}}(m, e \mid \cdot)}, \tag{41}$$

where all densities in the numerator are identified by Theorem 1 and the denominator is directly computable from the data. This conditional density is used in Section 2.4 for production function identification.

A.3 Identification up to $\mathbb{E}[\omega \mid k, l]$

Theorem A.1 (Identification up to $\mathbb{E}[\omega \mid k, l]$)**.** *Under Assumptions 1–2 and Assumptions A.1–A.3, the production function f_t is identified up to the specification of $\mathbb{E}[\omega \mid k, l]$. Specifically:*

(a) If an additional restriction specifying the functional form of $\mathbb{E}[\omega \mid k, l]$ is introduced, f_t is point-identified.

(b) Without any restriction on $\mathbb{E}[\omega \mid k, l]$, f_t is not point-identified.

Proof. (a) By Theorem 2, observationally equivalent structures are indexed by $\Delta(k, l)$. Since $\mathbb{E}[\tilde{\omega} \mid k, l] = \mathbb{E}[\omega \mid k, l] + \Delta(k, l)$, fixing $\mathbb{E}[\omega \mid k, l]$ uniquely pins down $\Delta(k, l) = 0$, and hence f_t is point-identified.
(b) Without restrictions, $\Delta(k, l)$ can be any continuous function, and Theorem 2 implies that point identification cannot be achieved. □

A.4 Identification via Exclusion Restriction (Proposition)

Proposition A.1 (Identification via Exclusion Restriction)**.** *Under Assumptions 1–2, Assumptions A.1–A.3, and the linear demand specification, let $\tilde{y}_{jt} \equiv y_{jt} - \hat{\beta}_m m_{jt} - \hat{\beta}_e e_{jt} - \hat{\beta}_w w_{jt}$ denote the partially identified output using Block A+B estimates.*

***Case 1** (Joint exclusion). If $a_k^h = a_l^h = 0$ for some input h, construct the productivity proxy*

$$\hat{\omega}_{jt}^h = \frac{h_{jt} - \hat{a}_z^{h\prime} z_{jt}}{\hat{a}_\omega^h}. \tag{42}$$

Then OLS regression of $(\tilde{y}_{jt} - \hat{\omega}_{jt}^h)$ on $(1, k_{jt}, l_{jt})$ consistently estimates $(\beta_0, \beta_k, \beta_l)$.

***Case 2** (Marginal exclusion). If $a_k^{h_1} = 0$ for input h_1 and $a_l^{h_2} = 0$ for input h_2 ($h_1 \neq h_2$), construct the proxies*

$$\hat{\omega}_{jt}^{h_1} = \frac{h_{1,jt} - \hat{a}_l^{h_1 *} l_{jt} - \hat{a}_z^{h_1\prime} z_{jt}}{\hat{a}_\omega^{h_1}}, \tag{43}$$

$$\hat{\omega}_{jt}^{h_2} = \frac{h_{2,jt} - \hat{a}_k^{h_2 *} k_{jt} - \hat{a}_z^{h_2\prime} z_{jt}}{\hat{a}_\omega^{h_2}}, \tag{44}$$

where $\hat{a}_l^{h_1 *}$ and $\hat{a}_k^{h_2 *}$ denote the Block A+B estimates (which include the $\Delta(k, l)$ indeterminacy). Then:

(i) the coefficient on k in the OLS regression of $(\tilde{y}_{jt} - \hat{\omega}_{jt}^{h_1})$ on $(1, k_{jt}, l_{jt})$ consistently estimates β_k;

(ii) the coefficient on l in the OLS regression of $(\tilde{y}_{jt} - \hat{\omega}^{h_2}_{jt})$ on $(1, k_{jt}, l_{jt})$ consistently estimates β_l.

This procedure requires neither the Markov assumption nor the homothetic regularity condition (Assumption 3).

Proof sketch (Case 1; full proof in Appendix H.2). Under the linear specification, the observational equivalence of Theorem 2 implies that Block A+B estimates satisfy $\hat{a}^{h*}_k \xrightarrow{p} a^h_k + a^h_\omega c_k$ for constants (c_k, c_l). The proxy (42) uses only the invariant estimates $\hat{a}^h_z$ and $\hat{a}^h_\omega$, yielding $\hat{\omega}^h \xrightarrow{p} \omega + \eta^h / a^h_\omega$. Subtracting from $\tilde{y}$ leaves a regression with error $\varepsilon - \eta^h / a^h_\omega$, which is mean-independent of (k, l) by iterated expectations and the exclusion restriction $a^h_k = a^h_l = 0$. Case 2 follows by a symmetric argument using separate proxies for β_k and β_l. □

Replacing the linear subtraction of $\hat{\omega}^h$ in Proposition A.1 with a polynomial regression is not consistent in general; see Appendix H.3 for details.

A.5 Validity of Recovered Productivity for Policy Evaluation

Proposition A.2 (Validity of Recovered Productivity for Policy Evaluation)**.** *Under Assumptions 1–2 and Assumptions A.1–A.3, suppose additionally that:*

(i) D does not alter the functional form of the intermediate input demand functions g_m, g_e, g_w; that is, D may affect the level of ω, but the mapping from ω to (m, e, w) is structurally invariant to D.

(ii) The determination of D may depend on ω and the state variables (k, l, z), but does not depend on the input-specific demand shocks (τ, ν, η).

Then $\mathbb{E}[\hat{\omega}_{jt} \mid D_{jt}] = \mathbb{E}[\omega_{jt} \mid D_{jt}]$.

The proof is given in Appendix H.4. The idea is that under conditions (i) and (ii), the intermediate inputs serve as sufficient statistics for ω: once (x, m, e, w) are observed, D provides no additional information about ω, so the law of iterated expectations yields the result.[27]

Implication for ATT identification in event studies. Proposition A.2 provides the measurement guarantee needed to identify the average treatment effect on the treated (ATT) using $\hat{\omega}_{jt}$ as the outcome variable. Suppose that the standard parallel trends assumption holds for the latent productivity ω_{jt}: for all $t \geq t_0$,

$$\mathbb{E}[\omega_{jt}(0) \mid D_j = 1] - \mathbb{E}[\omega_{jt_0-1}(0) \mid D_j = 1] = \mathbb{E}[\omega_{jt}(0) \mid D_j = 0] - \mathbb{E}[\omega_{jt_0-1}(0) \mid D_j = 0], \tag{45}$$

where $\omega_{jt}(0)$ denotes the potential outcome under no treatment. Under this assumption, the ATT at period t,

$$\text{ATT}_t = \mathbb{E}[\omega_{jt}(1) - \omega_{jt}(0) \mid D_j = 1],$$

is identified by the difference-in-differences estimand

$$\begin{aligned}\hat{\text{ATT}}_t = &\big(\mathbb{E}[\hat{\omega}_{jt} \mid D_j = 1] - \mathbb{E}[\hat{\omega}_{jt_0-1} \mid D_j = 1]\big) \\ &- \big(\mathbb{E}[\hat{\omega}_{jt} \mid D_j = 0] - \mathbb{E}[\hat{\omega}_{jt_0-1} \mid D_j = 0]\big).\end{aligned} \tag{46}$$

[27] Condition (i) may fail if the policy fundamentally changes how firms use intermediate inputs. In such cases, Theorem 1 should be applied separately to subpopulations defined by $D = 0$ and $D = 1$. Condition (ii) may fail if, for example, a specific input demand shock influences the firm's decision to participate in the policy.

To see that (46) converges to ATT_t, apply Proposition A.2 to replace each $\mathbb{E}[\hat{\omega}_{jt} \mid D_j]$ with $\mathbb{E}[\omega_{jt} \mid D_j]$, and then invoke parallel trends (45). No assumption on the time-series dynamics of ω is required beyond (45) itself.

Standard proxy variable estimators do not share this property. Their residual satisfies

$$\hat{\omega}_{jt}^{\text{proxy}} = \omega_{jt} + (\beta_k^{\text{true}} - \hat{\beta}_k^{\text{proxy}})\, k_{jt} + (\beta_l^{\text{true}} - \hat{\beta}_l^{\text{proxy}})\, l_{jt} + o_p(1),$$

so the DiD estimand (46) applied to $\hat{\omega}^{\text{proxy}}$ converges to

$$\text{ATT}_t + (\beta_k^{\text{true}} - \hat{\beta}_k^{\text{proxy}})\, \Delta\mathbb{E}[k_{jt} \mid D_j] + (\beta_l^{\text{true}} - \hat{\beta}_l^{\text{proxy}})\, \Delta\mathbb{E}[l_{jt} \mid D_j],$$

where $\Delta\mathbb{E}[k_{jt} \mid D_j]$ denotes the DiD in capital between treatment and control groups. The bias vanishes only if $\hat{\beta}^{\text{proxy}} = \beta^{\text{true}}$ or if the treatment is orthogonal to (k, l). In settings where the treatment induces capital or labor adjustment—as in the earthquake application of Section 5.6—neither condition is guaranteed.

A.6 GMM Moment Conditions

Assumption A.4 (Moment Conditions for GMM)**.** *Let $\Xi_{jt} = (\tau_{jt}, \nu_{jt}, \eta_{jt}, \varepsilon_{jt})$ denote the vector of all structural shocks, and define $W_{jt} = (1, \omega_{jt}, k_{jt}, l_{jt}, z_{jt})$ as the vector of relevant state variables and functions thereof. The following conditions are maintained:*

1. *Zero Mean Shocks: $\mathbb{E}[\Xi_{jt}^n] = 0$ for all n.*

2. *State Exogeneity: All shocks are uncorrelated with productivity ω_{jt} and primary inputs k_{jt}, l_{jt} (and functions thereof):*
$$\mathbb{E}[\Xi_{jt}^n \cdot W_{jt}^p] = 0 \quad \textit{for all } n, p.$$

3. *Mutual Exogeneity of Shocks: Different structural shocks are mutually uncorrelated:*
$$\mathbb{E}[\Xi_{jt}^n \cdot \Xi_{jt}^p] = 0 \quad \textit{for all } n \neq p.$$

Assumption A.4 is implied by the zero conditional mean condition $\mathbb{E}[\Xi_{jt}^n \mid \omega_{jt}, x_{jt}] = 0$ together with Assumption 2 (conditional independence).[28]

B Microfoundations of Intermediate Input Factor Demand

In this appendix, I provide microfoundations for the factor demand functions of the three intermediate inputs introduced in Section 2 (Equations (2)–(4)). Specifically, I demonstrate that the unobserved shock terms $\tau_{jt}, \nu_{jt}, \eta_{jt}$ included in each demand function are structurally derived from the firm's optimization behavior and market frictions. I thereby offer a theoretical rationale for the independence conditions required for identification.

[28] Strictly speaking, Assumption A.4 is weaker than imposing zero conditional mean on top of Assumption 2. The zero conditional mean condition implies uncorrelatedness with *any* measurable function of (ω, x), whereas Assumption A.4(2) requires uncorrelatedness only with the specific functions in W_{jt}.

B.1 Primitives

The production technology of firm j at time t is described by the following general nonparametric production function:

$$Y_{jt} = F_t(K_{jt}, L_{jt}, M_{jt}, E_{jt}, W_{jt}, \Omega_{jt}) \exp(\varepsilon_{jt}) \tag{47}$$

where Ω_{jt} is productivity observed by the firm but unobserved by the econometrician, and ε_{jt} is an ex-post production shock realized after input decisions are made. The firm faces an inverse demand function $P_{jt}(Y_{jt}, A_{jt})$ in the product market, where A_{jt} denotes a demand shock. Additionally, the firm faces an inverse supply function $P_{h,jt}(h_{jt})$ in the market for each intermediate input $h \in \{M_{jt}, E_{jt}, W_{jt}\}$.

I allow for deviations from perfect competition. I define the markup μ_{jt} in the product market and the markdown $\psi_{h,jt}$ for input h as follows:

$$\mu_{jt} \equiv \frac{P_{jt}}{MC_{jt}}, \quad \psi_{h,jt} \equiv \frac{ME_{h,jt}}{P_{h,jt}} \tag{48}$$

where MC_{jt} denotes marginal cost and $ME_{h,jt}$ denotes marginal expenditure. By definition, $\mu_{jt} \geq 1$ and $\psi_{h,jt} \geq 1$ hold.

Following Hsieh and Klenow (2009), I define a wedge $\Upsilon_{h,jt}$ representing exogenous distortions specific to each input h (e.g., taxes, adjustment costs, or optimization errors).

B.2 Expected Cost Minimization

The firm minimizes the cost of achieving a target expected output $\bar{Y}_{jt}$:

$$\min_{M_{jt}, E_{jt}, W_{jt}} \quad \sum_{h \in \{M,E,W\}} P_{h,jt}(h_{jt}) \, \Upsilon_{h,jt} \, h \tag{49}$$

$$\text{s.t.} \quad F_t(K_{jt}, L_{jt}, M_{jt}, E_{jt}, W_{jt}, \Omega_{jt}) \geq \bar{Y}_{jt} \tag{50}$$

The Lagrange multiplier λ_{jt} is interpreted as the marginal cost (MC_{jt}). The first-order condition with respect to input h is:

$$ME_{h,jt} \, \Upsilon_{h,jt} = MC_{jt} \frac{\partial F_t}{\partial h_{jt}} \tag{51}$$

Rewriting using (48), I obtain $\psi_{h,jt} P_{h,jt} \Upsilon_{h,jt} = \frac{P_{jt}}{\mu_{jt}} (\partial F_t / \partial h_{jt})$. Taking logarithms:

$$\ln\left(\frac{\partial F_t}{\partial h}\right) = \underbrace{\ln(\psi_{h,jt} \, P_{h,jt} \, \Upsilon_{h,jt})}_{\text{Idiosyncratic Input Cost}} - \underbrace{\ln \frac{P_{jt}}{\mu_{jt}}}_{\text{Common Market Factor}} \tag{52}$$

The second term, $\ln(P_{jt}/\mu_{jt})$, is a "Common Market Factor" that affects the demand for all variable inputs symmetrically. This term aggregates the effects of product market demand shocks and markups. The markdown $\psi_{h,jt}$ is input-specific: different intermediate inputs may face different degrees of buyer power. This heterogeneity in markdowns across inputs generates different productivity loading coefficients $(\gamma_\omega, \delta_\omega, \zeta_\omega)$ in the reduced-form demand functions, even when the underlying production technology is common.

B.3 Correspondence with Factor Demand Functions

From equation (52), the optimal input M_{jt} is a function of state variables, productivity, the input price and wedge, and the common factor $\ln(P_{jt}/\mu_{jt})$. Assuming strict concavity of the production function and applying the Implicit Function Theorem yields:

$$m_{jt} = g_m\left(k_{jt}, l_{jt}, \omega_{jt}, \ln(\psi_{m,jt}\, P_{m,jt}\, \Upsilon_{m,jt}) - \ln\frac{P_{jt}}{\mu_{jt}}\right) \tag{53}$$

The challenge for identification is that the Common Market Factor $\ln(P_{jt}/\mu_{jt})$ may induce correlation among the error terms across inputs $(\tau_{jt}, \nu_{jt}, \eta_{jt})$, threatening the conditional independence assumption (Assumption 2). To address this, I include additional control variables z_{jt} in the observable state variables x_{jt}.

The econometric error term τ_{jt} is defined as the "orthogonal residual" obtained by projecting the combined term onto $(\omega_{jt}, k_{jt}, l_{jt}, z_{jt})$:

$$\tau_{jt} \equiv \left[\ln(\psi_{m,jt}\, P_{m,jt}\, \Upsilon_{m,jt}) - \ln\frac{P_{jt}}{\mu_{jt}}\right] - \mathbb{E}\left[\ln(\psi_{m,jt}\, P_{m,jt}\, \Upsilon_{m,jt}) - \ln\frac{P_{jt}}{\mu_{jt}} \;\middle|\; \omega_{jt}, k_{jt}, l_{jt}, z_{jt}\right] \tag{54}$$

Since the common factor $\ln(P_{jt}/\mu_{jt})$ is spanned by z_{jt}, it is removed from the residual τ_{jt}. The residual reflects only idiosyncratic input costs: input-specific price fluctuations, the outcomes of negotiations with individual suppliers, procurement frictions, or optimization errors. It is economically reasonable to assume that specific supply shocks in the markets for raw materials, industrial water, and electricity are mutually independent. This definition provides the microfoundations for $\tau_{jt} \perp \nu_{jt} \perp \eta_{jt} \mid (\omega_{jt}, x_{jt})$.

Applying this definition to equation (53) yields the form of equation (2) in the main text:

$$m_{jt} = g_m(x_{jt}, \omega_{jt}, \tau_{jt}).$$

C Production Function Recovery

This appendix shows that, for each fixed (k_0, l_0), the production function $f_t(k_0, l_0, m, e, w, \omega)$ is identified as a function of (m, e, w, ω). I treat three cases of increasing generality.

Theorem C.1 (Hicks-Neutral Case)**.** *Under Assumptions 2–A.3, in the additively separable model $y = g_t(k, l, m, e, w) + \omega + \varepsilon$, for each fixed (k_0, l_0), $g_t(k_0, l_0, m, e, w)$ is nonparametrically identified as a function of (m, e, w).*

Proof. By Assumption 1,

$$\mathbb{E}[y \mid x, m, e, w] = g_t(k_0, l_0, m, e, w) + \mathbb{E}[\omega \mid x, m, e, w],$$

where $\mathbb{E}[\varepsilon \mid x, m, e, w] = 0$ follows from the law of iterated expectations. Therefore

$$g_t(k_0, l_0, m, e, w) = \mathbb{E}[y \mid x, m, e, w] - \mathbb{E}[\omega \mid x, m, e, w].$$

The first term is identified from the data, and the second is computable from $f_{\omega|x,m,e,w}$ (equation (41)). □

Theorem C.2 (Non-Hicks-Neutral Case: No Ex-Post Shock)**.** *Under Assumptions 2–A.2 and A.3, in the model $y = f_t(k, l, m, e, w, \omega)$ with no ex-post shock, suppose $f_t(k_0, l_0, m, e, w, \omega)$ is strictly* *monotone in ω for each fixed (m, e, w). Then, for each fixed (k_0, l_0), f_t is nonparametrically identified as a function of (m, e, w, ω).*

Proof. Fix (k_0, l_0) and (m_0, e_0, w_0). With $\varepsilon = 0$, $y = f_t(k_0, l_0, m_0, e_0, w_0, \omega)$, and strict monotonicity in ω yields

$$F_{y|x,m_0,e_0,w_0}(y \mid \cdot) = F_{\omega|x,m_0,e_0,w_0}\big(f_t^{-1}(\cdot, y) \,\big|\, \cdot\big).$$

Both sides are identified (the left from data, the right from Theorem 1 and equation (41)). By quantile matching:

$$f_t(k_0, l_0, m_0, e_0, w_0, \omega) = F_{y|\cdot}^{-1}\big(F_{\omega|\cdot}(\omega) \,\big|\, \cdot\big).$$

□

Theorem C.3 (Non-Hicks-Neutral Case: Known f_ε)**.** *Under Assumptions 2–A.3, suppose additionally that ε is independent of (ω, x, m, e, w), f_ε is known, its characteristic function has no zeros on $\mathbb{R}$, and $f_t(k_0, l_0, m, e, w, \omega)$ is strictly monotone in ω for each fixed (m, e, w). Then, for each fixed (k_0, l_0), f_t is nonparametrically identified.*

Proof. Fix (k_0, l_0) and (m_0, e_0, w_0). Setting $s = f_t(k_0, l_0, m_0, e_0, w_0, \omega)$, the observed conditional density becomes a convolution: $f_{y|\cdot}(y) = (f_\varepsilon * \tilde{K})(y)$. Taking characteristic functions and using $\varphi_\varepsilon(t) \neq 0$, one can recover $\tilde{K}$ by deconvolution. With $K \equiv f_{\omega|x,m_0,e_0,w_0}$ identified, quantile matching recovers f_t. □

Remark C.1. *Theorem C.3 assumes f_ε is fully known. When f_ε belongs to a parametric family (e.g., $N(0, \sigma^2)$) with unknown parameters, these can typically be recovered from the decay rate of the characteristic function.*

D Proof of Theorem 2

Proof. Sufficiency. Take any continuous function $\Delta(k, l)$ and define $\tilde{\omega}$ and $\tilde{f}_t$ by (8). Then $\tilde{f}_t(k, l, m, e, w, \tilde{\omega}) + \varepsilon = f_t(k, l, m, e, w, \omega) + \varepsilon = y$, so the distribution of output is unchanged. For intermediate input demands, $f_{m|\tilde{\omega},k,l}(m \mid \tilde{\omega}) = f_{m|\omega,k,l}(m \mid \tilde{\omega} - \Delta(k, l))$, so the observable joint distribution $f_{y,m,e,w|k,l}$ is invariant for each fixed (k, l).

Necessity. Suppose $(\tilde{f}_t, \tilde{\omega})$ is observationally equivalent to (f_t, ω). Fix (k_0, l_0) and apply the identification procedure of Theorem 1 (HS08, Theorem 1).

The proof of HS08's Theorem 1 proceeds in four stages: (1) uniqueness of the spectral decomposition (Dunford and Schwartz (1971), Theorem XV.4.5); (2) fixing the scale of eigenfunctions by the density integration condition; (3) resolving degenerate eigenvalues via Assumption A.2; and (4) fixing the indexing via the location normalization (Assumption A.3).

Stages (1)–(3) determine the family of conditional densities $\{f_{m|\omega,k_0,l_0}(\cdot \mid \omega) : \omega \in \Omega\}$ uniquely *as an unordered set.* Since $(\tilde{f}_t, \tilde{\omega})$ is observationally equivalent, the spectral decomposition based on $\tilde{\omega}$ must produce the same unordered set. Therefore, there exists a bijection $R_{k_0,l_0} : \Omega \to \Omega$ such that

$$f_{m|\tilde{\omega},k_0,l_0}(\cdot \mid \tilde{\omega}) = f_{m|\omega,k_0,l_0}(\cdot \mid R_{k_0,l_0}^{-1}(\tilde{\omega})) \quad \text{for all } \tilde{\omega}. \tag{55}$$

Applying Stage (4): under the ω-normalization $M[f_{m|\omega}(\cdot \mid \omega)] = \omega$, one obtains

$$M\big[f_{m|\tilde{\omega}}(\cdot \mid \tilde{\omega})\big] = R^{-1}(\tilde{\omega}).$$

If the $\tilde{\omega}$-normalization is imposed, then $R^{-1}(\tilde{\omega}) = \tilde{\omega}$ and $R = \text{id}$.

However, the normalization is defined independently for each (k_0, l_0). In the true structure, the normalization functional $M[f_{m|\omega^{\text{true}},k,l}(\cdot \mid \omega^{\text{true}})]$ generally depends on (k, l), so the normalizations at different (k, l) fix ω at reference points differing by

$$c(k_0, l_0) \equiv M\big[f_{m|\omega^{\text{true}},k_0,l_0}(\cdot \mid \omega^{\text{true}})\big] - \omega^{\text{true}}.$$

It follows that $\tilde{\omega} = \omega + \Delta(k, l)$ where $\Delta(k, l) = c_{\tilde{\omega}}(k, l) - c_{\omega}(k, l)$ is continuous (by the assumed continuous dependence of $f_{m|\omega,k,l}$ on (k, l)), and $\tilde{f}_t(\cdot, \tilde{\omega}) = f_t(\cdot, \tilde{\omega} - \Delta(k, l))$. □

E Data Generating Process for Monte Carlo Simulation

This appendix describes the detailed parameter settings for the Data Generating Process (DGP) of the Monte Carlo simulation outlined in Section 4.

E.1 Common Parameter Settings

The following parameters are common to all DGPs.

Production Function (Equation (36)):

- Intercept $\beta_0 = 0.1$
- Capital $\beta_k = 0.2$
- Labor $\beta_l = 0.3$
- Intermediate Input m: $\beta_m = 0.3$
- Intermediate Input e: $\beta_e = 0.15$
- Intermediate Input w: $\beta_w = 0.1$
- Measurement Error $\varepsilon_{jt} \sim N(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon = 0.05$

Intermediate Input Demand Functions. The demand function coefficients are derived from the first-order conditions of cost minimization under input-specific markdowns (Appendix B). The theoretical benchmark under perfect competition yields $\gamma_\omega^{\text{bench}} = 1/(1 - S_m) \approx 2.22$, where $S_m = \beta_m + \beta_e + \beta_w = 0.55$. Input-specific markdowns ψ_h generate heterogeneity in the productivity loading coefficients across inputs.

- $m_{jt} = \gamma_k k_{jt} + \gamma_l l_{jt} + \gamma_\omega \omega_{jt} + \tau_{jt}$
  - $(\gamma_k, \gamma_l, \gamma_\omega) = (0.45, 0.65, 2.2)$
  - $\tau_{jt} = \rho_\tau \tau_{j,t-1} + e_{\tau,jt}$, $\rho_\tau = 0.5$, $\text{Var}(\tau_{jt}) = \sigma_\tau^2 = 0.15^2$
- $e_{jt} = \delta_k k_{jt} + \delta_l l_{jt} + \delta_\omega \omega_{jt} + \nu_{jt}$
  - $(\delta_k, \delta_l, \delta_\omega) = (0.40, 0.60, 2.0)$
  - $\nu_{jt} = \rho_\nu \nu_{j,t-1} + e_{\nu,jt}$, $\rho_\nu = 0.5$, $\text{Var}(\nu_{jt}) = \sigma_\nu^2 = 0.15^2$

- $w_{jt} = \zeta_k k_{jt} + \zeta_l l_{jt} + \zeta_\omega \omega_{jt} + \eta_{jt}$
  - $(\zeta_k, \zeta_l, \zeta_\omega) = (0.50, 0.70, 1.8)$
  - $\eta_{jt} = \rho_\eta \eta_{j,t-1} + e_{\eta,jt}$, $\rho_\eta = 0.5$, $\mathrm{Var}(\eta_{jt}) = \sigma_\eta^2 = 0.15^2$

The demand shock innovations $e_{\tau,jt}, e_{\nu,jt}, e_{\eta,jt}$ follow mutually independent normal distributions. The AR(1) structure preserves the conditional independence assumption (Assumption 2) at each time point, since each shock is generated from its own independent chain.

Dynamic Decisions and Firms' Beliefs:

- **Firms' Beliefs** (Assumed AR(1) in all DGPs): $\omega_{jt} = \rho_{\text{belief}}\, \omega_{j,t-1} + \xi_{\text{belief},jt}$, $\rho_{\text{belief}} = 0.8$, stationary variance $\sigma^2_{\omega,\text{belief}} = 0.2^2/(1-0.8^2) \approx 0.111$.
- **Capital Accumulation:** $k_{j,t+1} = \log((1-\delta_{\text{capital}})\exp(k_{jt}) + i_{jt})$, $\delta_{\text{capital}} = 0.2$.
- **Investment Function:** i_{jt} is determined to maximize expected returns under the AR(1) belief, following the mechanism in the Monte Carlo simulation of Ackerberg, Caves, and Frazer (2015).
- **Labor Decision:** $l_{j,t+1}$ is determined based on the productivity forecast under the AR(1) belief, the predetermined $k_{j,t+1}$, and the exogenous wage $\ln w_{j,t+1}$.
- **Wage Process (AR(1)):** $\ln w_{jt} = \rho_{\ln w} \ln w_{j,t-1} + \xi_{\ln w,jt}$, $\rho_{\ln w} = 0.3$, $\sigma_{\ln w} = 0.1$.
- **Others:** Discount factor 0.95, investment cost heterogeneity $\sigma_b = 0.6$.

E.2 DGP-Specific Productivity Process Settings

DGP1: AR(1) Process (Baseline)

- $\omega_{jt} = \rho_{\text{dgp1}}\, \omega_{j,t-1} + \xi_{\text{dgp1},jt}$
- $\rho_{\text{dgp1}} = 0.8$
- Innovation standard deviation $\sigma_{\xi,\text{dgp1}} = 0.2$ (consistent with firms' beliefs)

DGP2: AR(2) Process

- $\omega_{jt} = \rho_{1,\text{dgp2}}\, \omega_{j,t-1} + \rho_{2,\text{dgp2}}\, \omega_{j,t-2} + \xi_{\text{dgp2},jt}$
- $\rho_{1,\text{dgp2}} = 0.6$, $\rho_{2,\text{dgp2}} = 0.3$
- Innovation standard deviation $\sigma_{\xi,\text{dgp2}} = 0.15$

DGP3: Potential Outcome Model

- **Potential Process (Untreated ω^0):** $\omega^0_{jt} = \rho_0\, \omega^0_{j,t-1} + \xi_{0,jt}$, $\rho_0 = 0.8$, $\sigma_{\xi 0} = 0.2$
- **Potential Process (Treated ω^1):**
  - From treated state ($D_{j,t-1} = 1$): $\omega^1_{jt} = \rho_1\, \omega^1_{j,t-1} + \Delta + \xi_{1,jt}$
  - From untreated state ($D_{j,t-1} = 0$): $\omega^1_{jt} = \rho_1\, \omega^0_{j,t-1} + \Delta + \xi_{1,jt}$
  - $\rho_1 = 0.5$, $\Delta = 0.15$, $\sigma_{\xi 1} = 0.25$

- **Reversible Selection:** $D_{jt} = \mathbb{I}(\omega_{jt}^0 > 0)$. Treatment is not an absorbing state: firms enter and exit treatment as ω_{jt}^0 crosses the threshold. Both potential processes ω^0, ω^1 evolve independently regardless of the current treatment state (Diagonal Markov; Chen, Liao, and Schurter (2024), Assumption 2.1).

- **Realized Productivity:** $\omega_{jt} = (1 - D_{jt})\,\omega_{jt}^0 + D_{jt}\,\omega_{jt}^1$

E.3 Simulation Execution Settings

- Burn-in period $T_{\text{burnin}} = 30$
- **Part 1** (Block A+B): $R = 100$, $N \in \{50, 200, 500\}$, $T_{\text{obs}} \in \{10, 20, 50\}$
- **Part 2** (Block A+B+C): $R = 100$, $(N, T) = (200, 50)$

F Additional Results for Monte Carlo Simulation

This appendix presents detailed simulation results. Tables report bias, standard deviation, and RMSE for each parameter, estimation method, and DGP at $(N, T) = (500, 50)$.

F.1 Part 1: Flexible Input Parameters (Block A+B)

F.2 Part 2: All Parameters (Block A+B+C)

F.3 Additional Monte Carlo Figures

F.4 DGP4: Conditional Independence Violation

F.5 Block C Identification Diagnostics

The significance of $\hat{\rho}_2$ and $\hat{\rho}_3$ provides a practical diagnostic for Block C identification strength. When these coefficients are statistically significant, the nonlinear component $h(v)$ is identified beyond the linear term, enabling separate estimation of β_k and β_l. When both are statistically insignificant, the capital-labor aggregate $v(k, l)$ is empirically indistinguishable from a Cobb–Douglas form, and the rank condition in Theorem 3 is not met in the data: (β_k, β_l) cannot be separately identified through Block C for such industries. This is a testable, data-driven indicator of the proximity to the identification boundary described in Section 2.4.5. In the empirical application (Section 5), I report t-statistics for $\hat{\rho}_2$ and $\hat{\rho}_3$ alongside the J-test to assess Block C reliability for each industry.

F.6 Block A+B Only vs. Block A+B+C Comparison

A robustness check compares estimates obtained using Block A+B moments alone against estimates using the full Block A+B+C system. Parameters identified by Block A+B (namely $(\beta_m, \beta_e, \beta_w)$ and

Table 8: Part 1: DGP 1: AR(1) (N=500, T=50)

		ACF			ACF-Mod			GNR			Proposed		
Parameter	True	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
β_m	0.30	0.0004	0.0009	0.0010	0.0004	0.0010	0.0011	0.5886	0.0011	0.5886	0.0007	0.0035	0.0035
β_e	0.15	0.0006	0.0015	0.0016	0.0005	0.0012	0.0013	0.5480	0.0019	0.5480	-0.0008	0.0034	0.0034
β_w	0.10	0.0001	0.0007	0.0007	0.0001	0.0008	0.0008	1.0366	0.0027	1.0366	0.0001	0.0037	0.0037

Table 9: Part 1: DGP 2: AR(2) (N=500, T=50)

		ACF			**ACF-Mod**			**GNR**			**Proposed**		
Parameter	True	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
β_m	0.30	0.0157	0.0190	0.0245	0.0181	0.0200	0.0268	0.5877	0.0013	0.5877	0.0004	0.0037	0.0036
β_e	0.15	0.0178	0.0219	0.0280	0.0206	0.0229	0.0306	0.5476	0.0029	0.5476	-0.0005	0.0032	0.0033
β_w	0.10	0.0021	0.0032	0.0038	0.0014	0.0030	0.0033	1.0347	0.0033	1.0347	0.0004	0.0041	0.0041

Table 10: Part 1: DGP 3: Potential (N=500, T=50)

		ACF			**ACF-Mod**			**GNR**			**Proposed**		
Parameter	True	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
β_m	0.30	0.1947	0.0144	0.1952	0.2325	0.0285	0.2342	0.5888	0.0014	0.5888	0.0012	0.0039	0.0041
β_e	0.15	0.1868	0.0090	0.1871	0.1707	0.0204	0.1719	0.5408	0.0029	0.5408	-0.0014	0.0037	0.0039
β_w	0.10	0.0886	0.0117	0.0893	0.0602	0.0122	0.0614	1.0299	0.0027	1.0300	0.0001	0.0031	0.0031

Table 11: Part 2: DGP 1: AR(1) (N=200, T=50)

		ACF			**Proposed**		
Parameter	True	Bias	SD	RMSE	Bias	SD	RMSE
β_k	0.20	0.0005	0.0064	0.0063	0.0016	0.0119	0.0119
β_l	0.30	-0.0017	0.0047	0.0050	0.0019	0.0226	0.0225
β_m	0.30	0.0005	0.0021	0.0021	0.0002	0.0126	0.0125
β_e	0.15	0.0009	0.0037	0.0038	0.0023	0.0089	0.0091
β_w	0.10	-0.0000	0.0009	0.0009	-0.0012	0.0115	0.0115

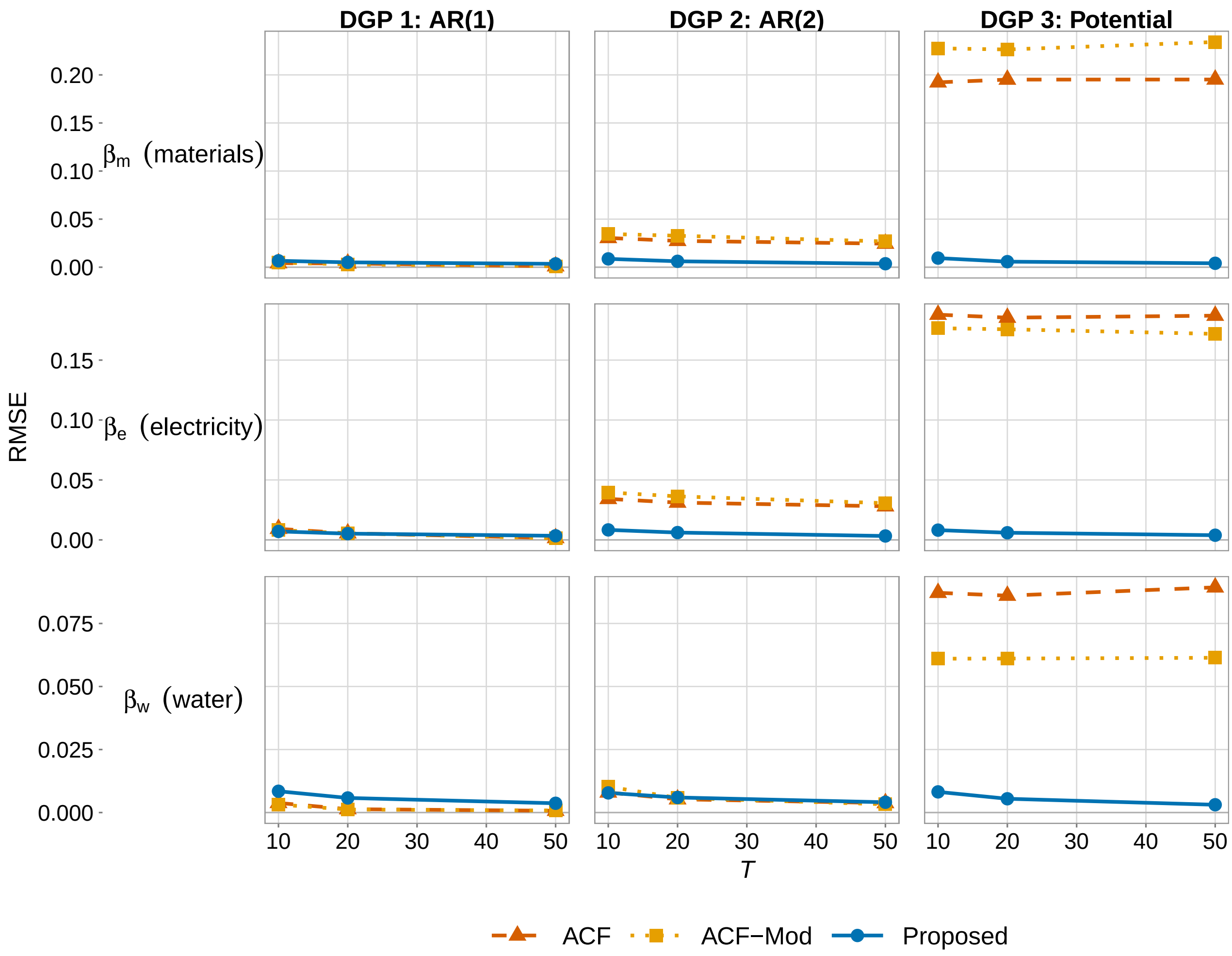
Table 12: Part 2: DGP 2: AR(2) (N=200, T=50)

		ACF			**Proposed**		
Parameter	True	Bias	SD	RMSE	Bias	SD	RMSE
β_k	0.20	-0.0211	0.0269	0.0340	0.0108	0.0133	0.0170
β_l	0.30	-0.0394	0.0540	0.0665	-0.0021	0.0192	0.0191
β_m	0.30	0.0187	0.0239	0.0301	-0.0019	0.0128	0.0128
β_e	0.15	0.0228	0.0303	0.0376	0.0047	0.0102	0.0111
β_w	0.10	0.0022	0.0045	0.0050	-0.0002	0.0099	0.0098

Table 13: Part 2: DGP 3: Potential (N=200, T=50)

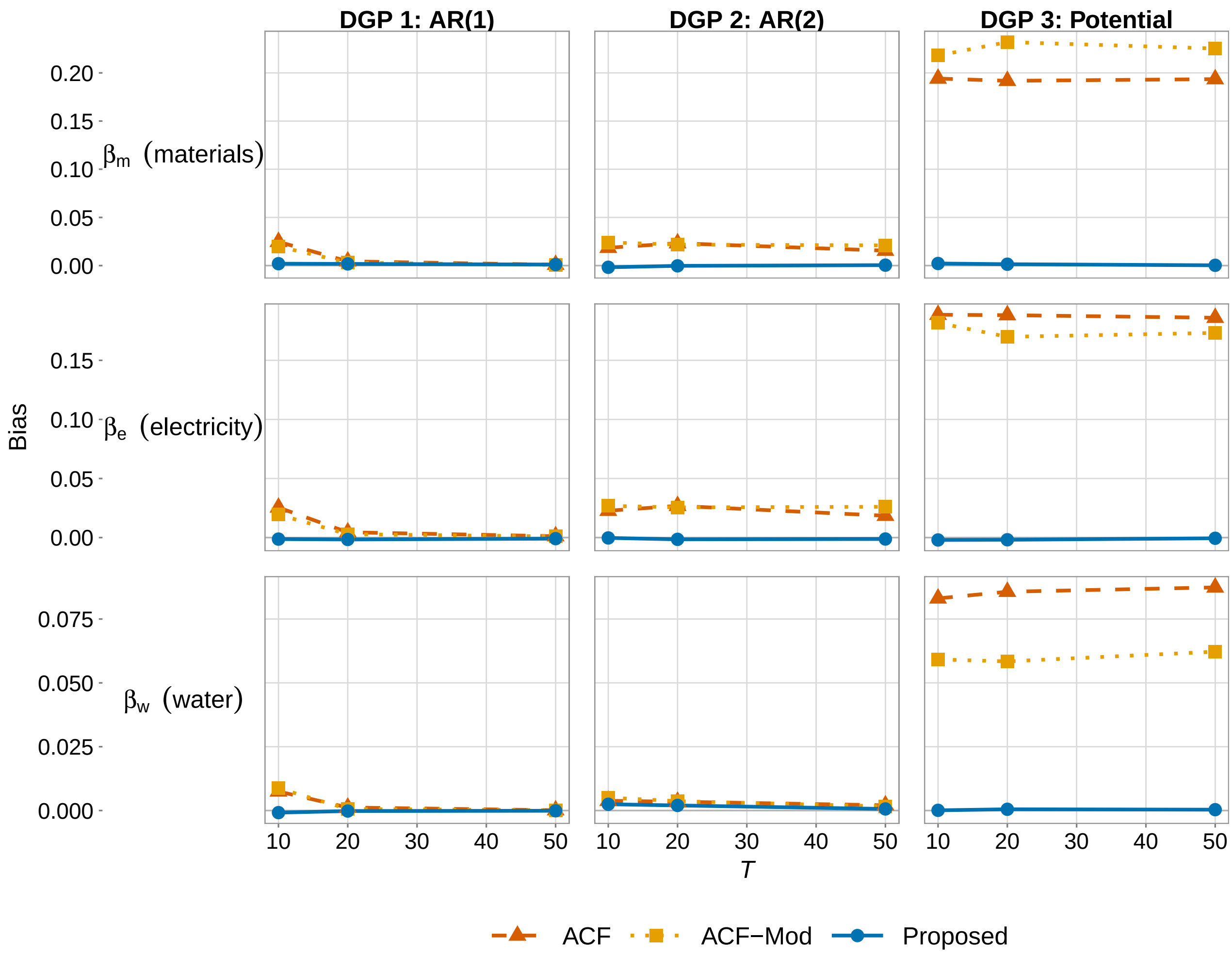
		ACF			**Proposed**		
Parameter	True	Bias	SD	RMSE	Bias	SD	RMSE
β_k	0.20	-0.1972	0.0190	0.1981	0.0123	0.0104	0.0161
β_l	0.30	-0.2954	0.0287	0.2967	0.0116	0.0139	0.0180
β_m	0.30	0.1902	0.0254	0.1918	-0.0057	0.0110	0.0122
β_e	0.15	0.1860	0.0218	0.1872	-0.0009	0.0087	0.0087
β_w	0.10	0.0878	0.0144	0.0890	-0.0033	0.0083	0.0088

Figure 7: Part 1: Mean RMSE Convergence ($N = 500$)



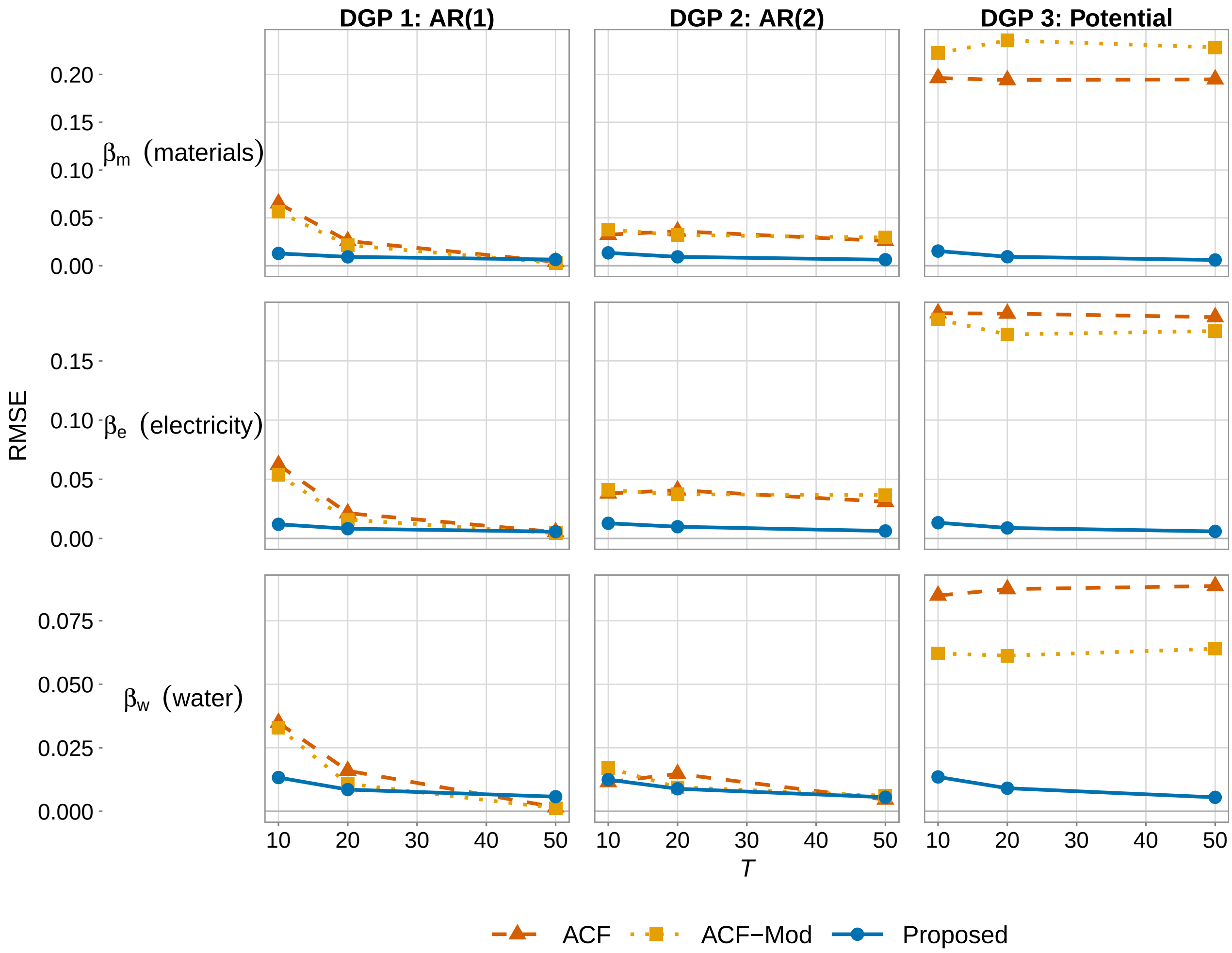
Notes: Mean RMSE of $(\hat{\beta}_m, \hat{\beta}_e, \hat{\beta}_w)$ *as a function of* T $(N = 500,\ R = 100)$*. Under DGP 2 and DGP 3, the ACF and ACF-Mod RMSEs do not vanish with* T*, reflecting asymptotic bias. The proposed method's RMSE declines monotonically.*

Figure 8: Part 1: Mean Bias Convergence ($N = 200$)



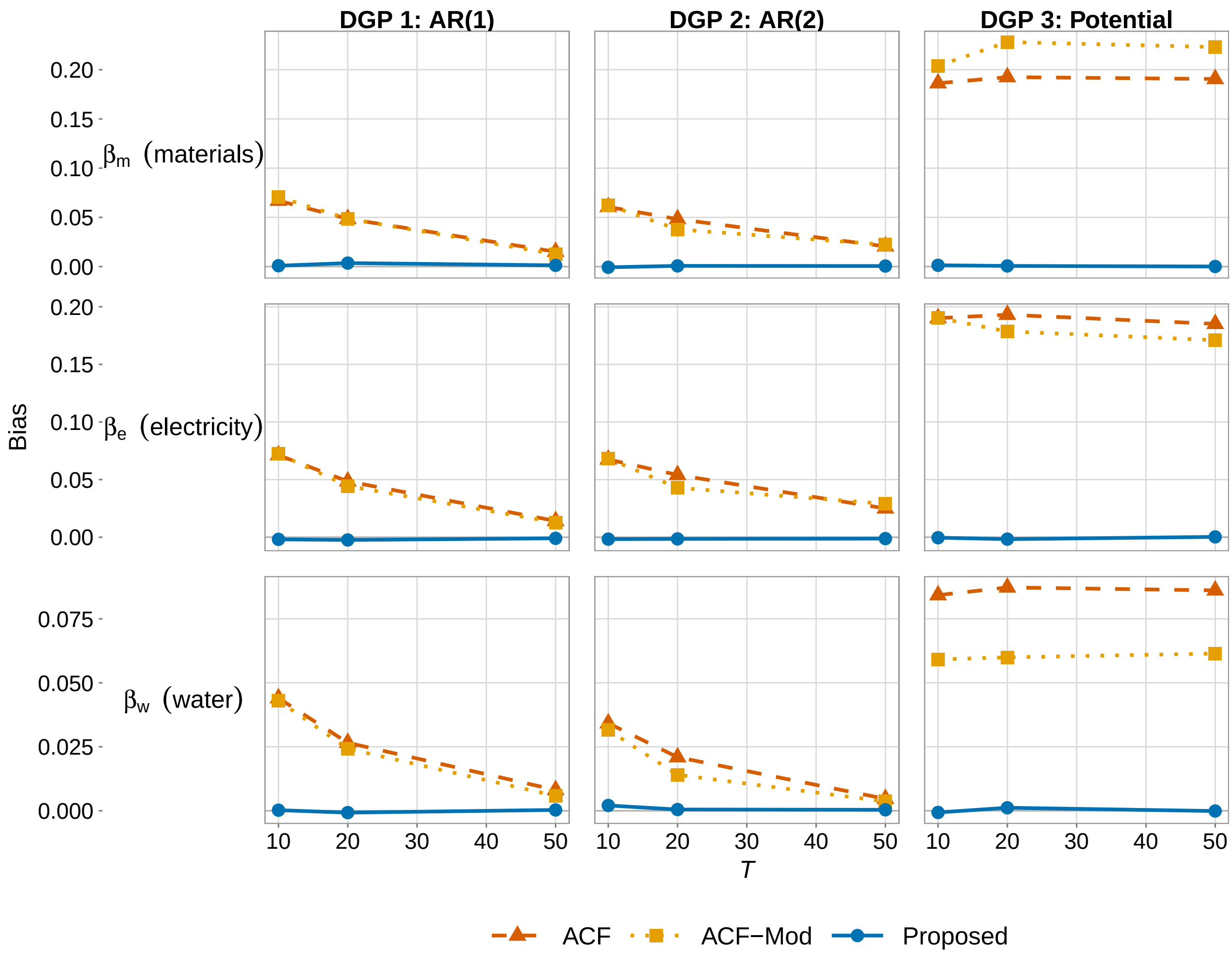
Notes: Same as Figure 1 but for $N = 200$. The qualitative patterns are preserved; the larger variance reflects the smaller cross-section.

Figure 9: Part 1: Mean RMSE Convergence ($N = 200$)



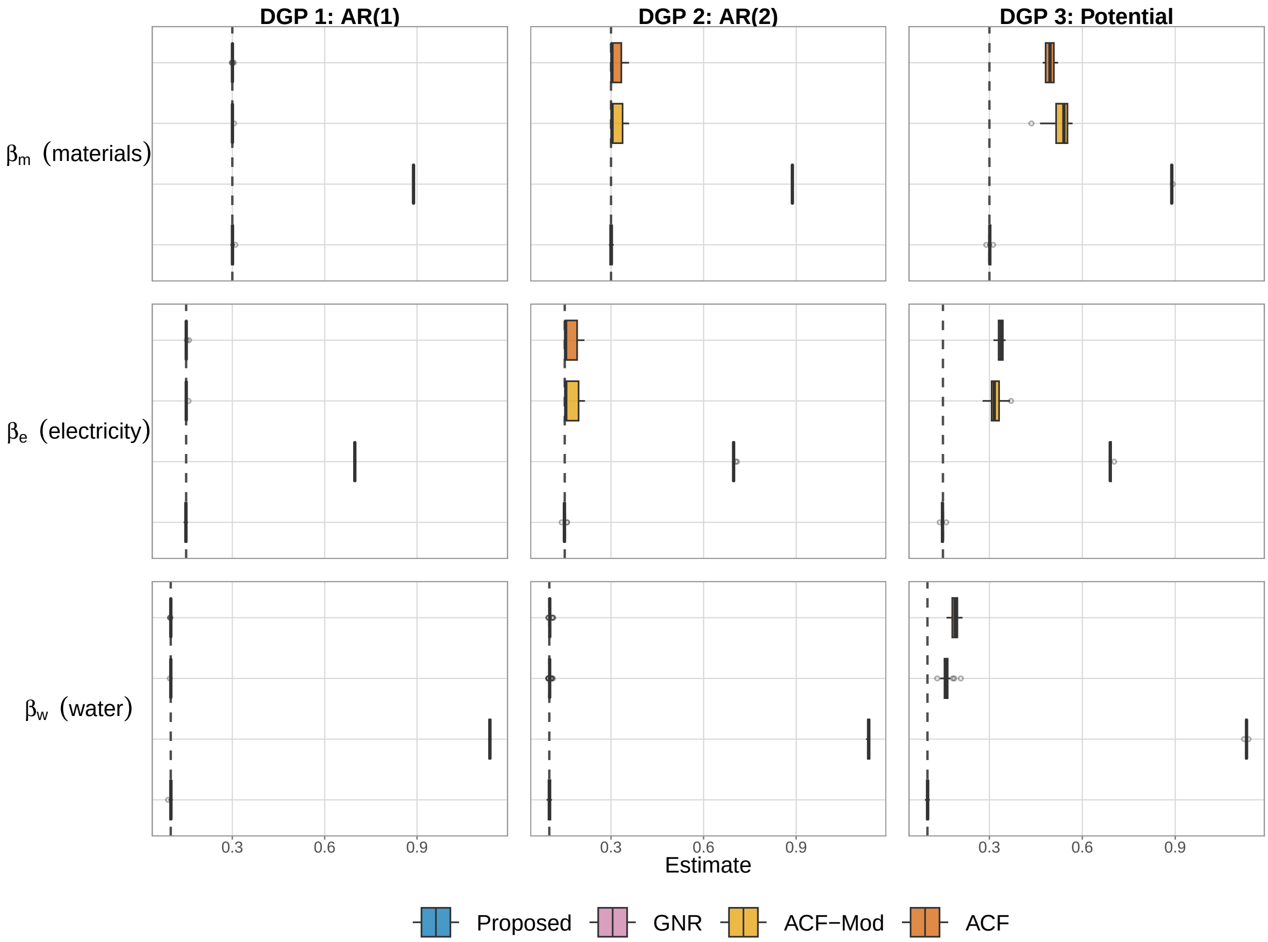
Notes: Mean RMSE of $(\hat{\beta}_m, \hat{\beta}_e, \hat{\beta}_w)$ *as a function of* T $(N = 200,\ R = 100)$.

Figure 10: Part 1: Mean Bias Convergence ($N = 50$)



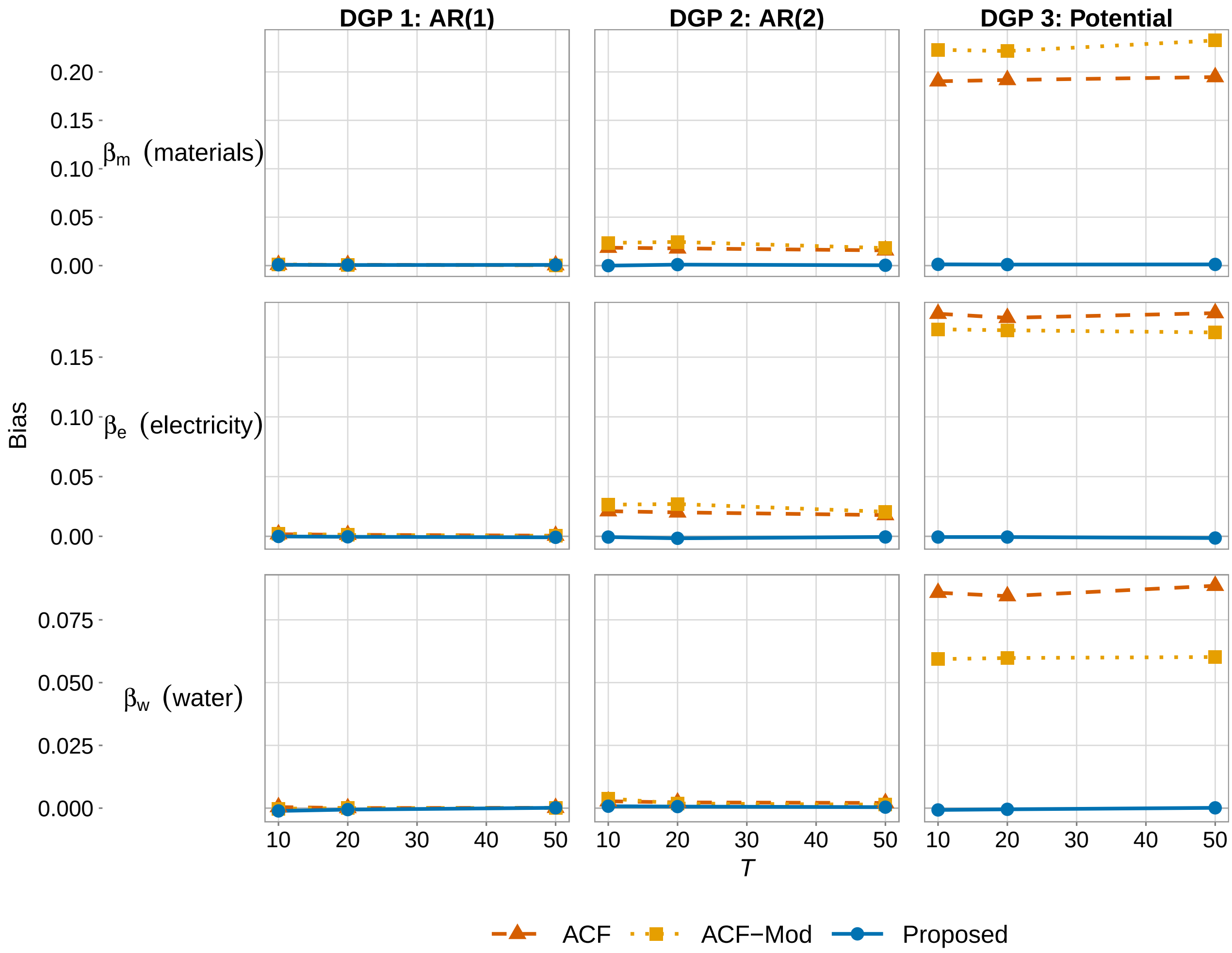
Notes: Same as Figure 1 but for $N = 50$. The qualitative patterns are preserved; the larger variance reflects the smaller cross-section.

Figure 11: Part 1: Four-Method Comparison ($N = 500$, $T = 50$)



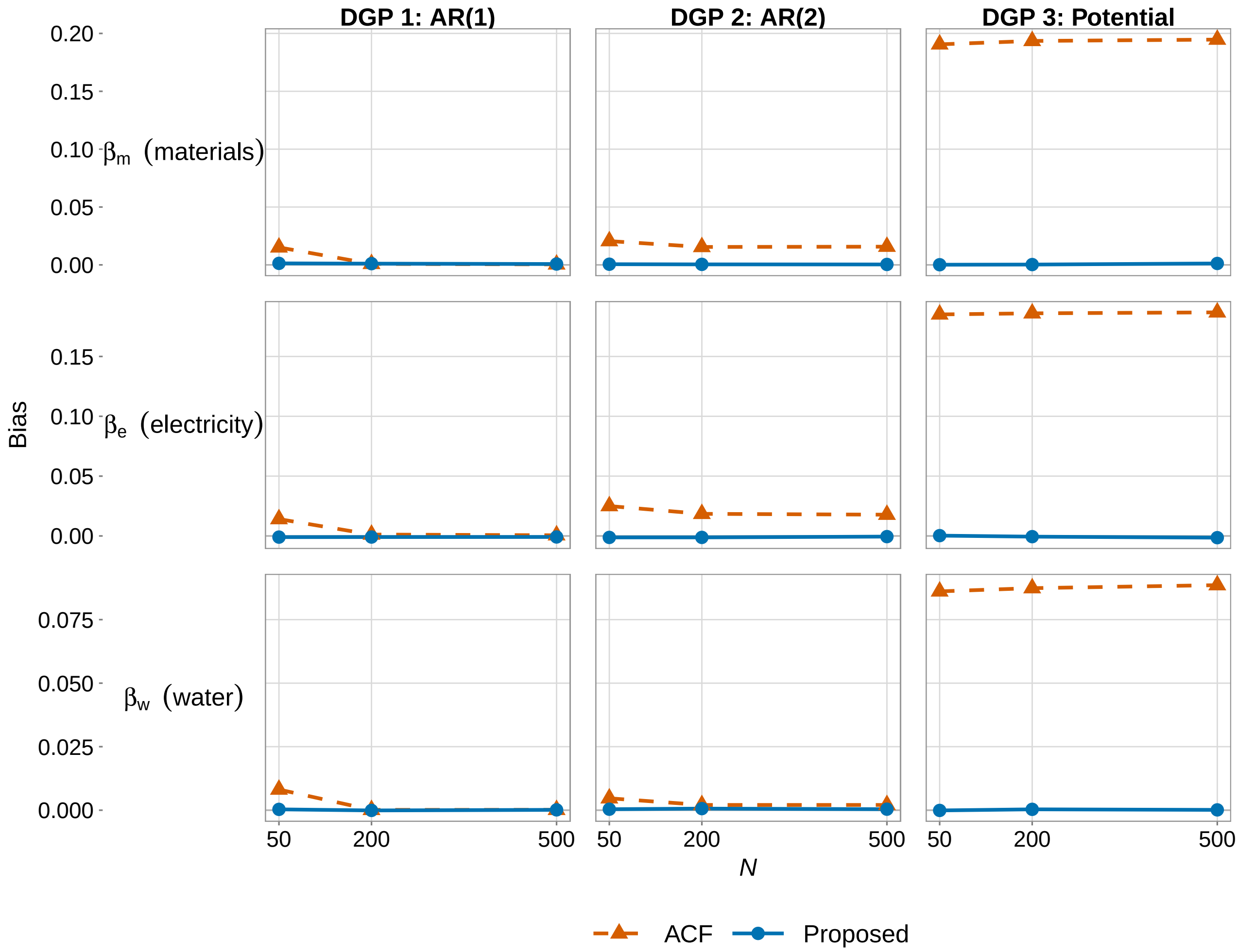
Notes: Distribution of $(\hat{\beta}_m, \hat{\beta}_e, \hat{\beta}_w)$ *including the GNR estimator* ($N = 500$, $T = 50$)*. The GNR estimates are severely biased (Bias*$(\hat{\beta}_m) \approx 0.59$*) due to persistent demand shocks* ($\rho_\tau = 0.5$) *violating the scalar unobservability assumption. The main text figures (Figures 1–2) exclude GNR to preserve visual clarity for the ACF–Proposed comparison.*

Figure 12: Part 1: Three-Method Bias Convergence ($N = 500$)



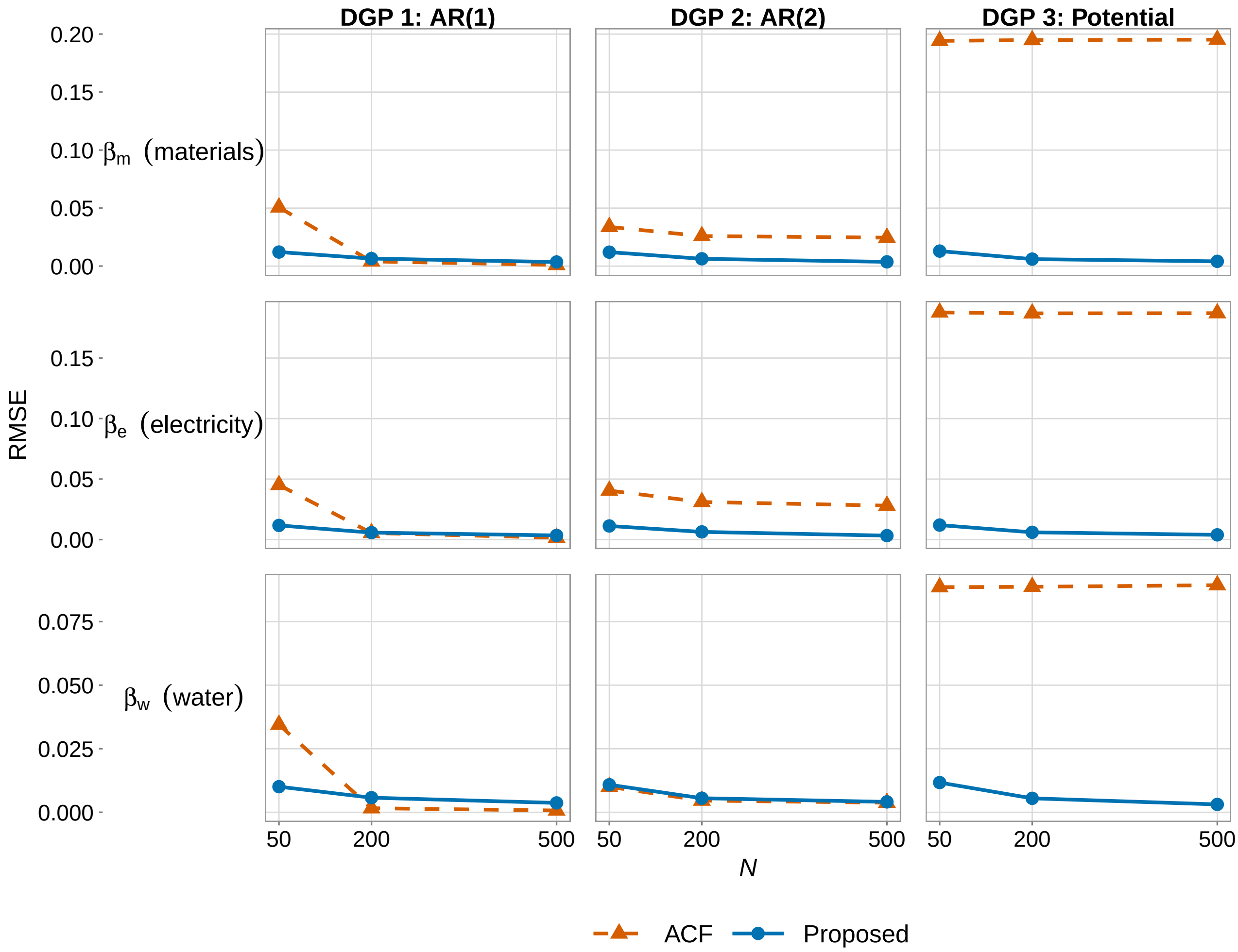
Notes: Same as Figure 1 but including the ACF-Mod (oracle) estimator that observes the true demand shock τ_{jt}. Under DGP 2 and 3, ACF-Mod bias is comparable to or larger than standard ACF.

Figure 13: Part 1: Mean Bias as a Function of N ($T = 50$)



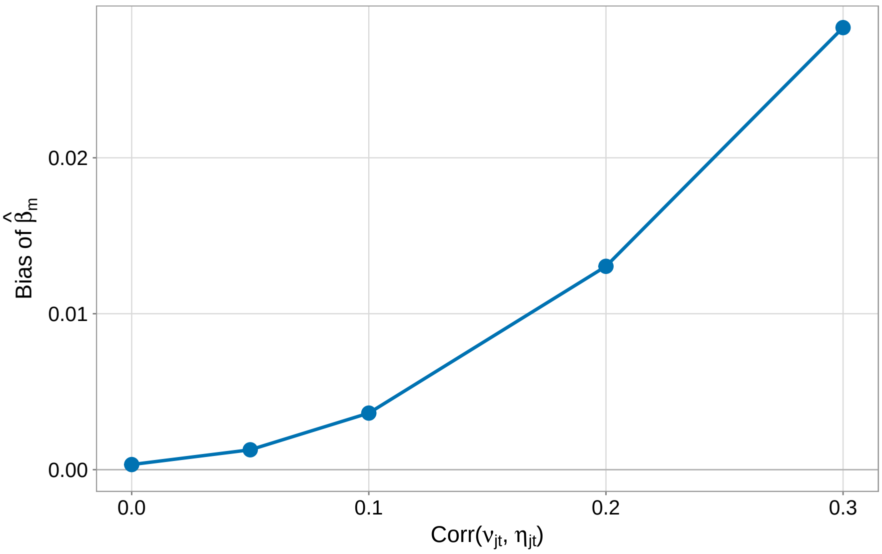
Notes: Mean bias of $(\hat{\beta}_m, \hat{\beta}_e, \hat{\beta}_w)$ as a function of N for $T = 50$ ($R = 100$). Increasing N reduces variance for all estimators. ACF bias under DGP 2 and 3 does not vanish with N, confirming asymptotic bias.

Figure 14: Part 1: RMSE as a Function of N ($T = 50$)



Notes: Root mean squared error of $(\hat{\beta}_m, \hat{\beta}_e, \hat{\beta}_w)$ *as a function of* N *for* $T = 50$ $(R = 100)$*. Under DGP 1 (correct Markov specification), the RMSE of both estimators converges to zero at similar rates. Under DGP 2 and 3, ACF RMSE is bounded away from zero because bias dominates, whereas the proposed estimator's RMSE continues to decrease with* N*.*

Figure 15: DGP 4: Bias of $\hat{\beta}_m$ as a Function of $\mathrm{Corr}(\nu, \eta)$



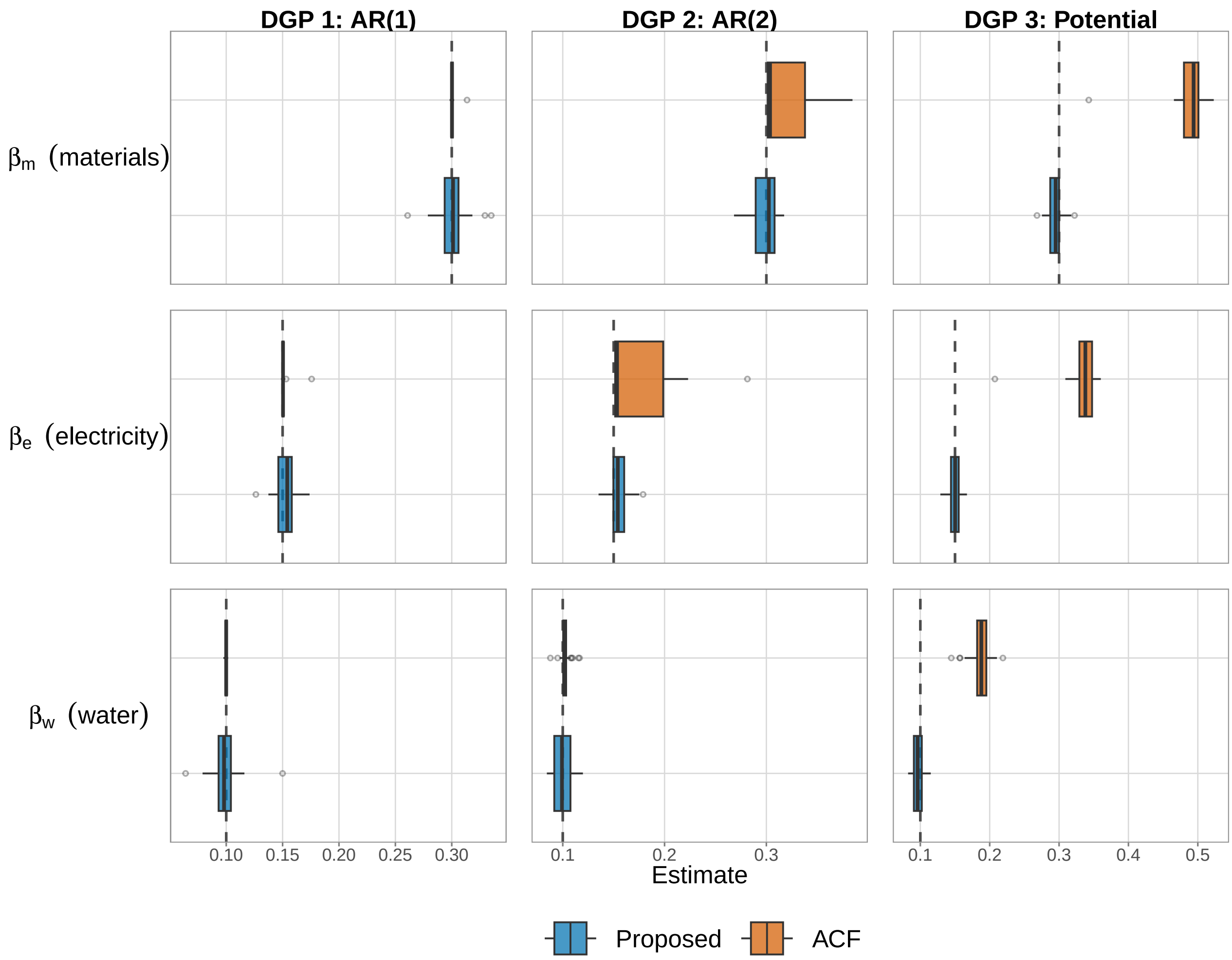
Notes: Mean bias of $\hat{\beta}_m$ *as a function of* $\mathrm{Corr}(\nu_{jt}, \eta_{jt})$ *for the proposed method (*$N = 200$*,* $T = 50$*,* $R = 20$*). When* $\mathrm{Corr}(\nu, \eta) = 0$*, the estimator is approximately unbiased at the true value* $\beta_m = 0.30$*. As the electricity–water correlation increases,* $\hat{\zeta}_\omega$ *is overestimated, causing upward bias in* $\hat{\beta}_m$ *(Appendix J). The bias direction is the same as the Markov misspecification bias in ACF, so the empirical gap (ACF > Proposed) cannot be attributed to CI violation.*

Table 14: DGP 4 (CI Violation): $\hat{\beta}_m$ Bias, SD, and RMSE by $\mathrm{Corr}(\nu, \eta)$. $N = 200$, $T = 50$, $R = 20$.

$\mathrm{Corr}(\nu, \eta)$	Bias	SD	RMSE
0.0000	0.0003	0.0056	0.0054
0.0500	0.0013	0.0052	0.0052
0.1000	0.0036	0.0052	0.0062
0.2000	0.0130	0.0048	0.0139
0.3000	0.0283	0.0044	0.0287

derived quantities such as the markup $\mu = \beta_m / s_m$) should remain stable across the two specifications. Instability would indicate misspecification of the Block C moments or weak identification of the CES aggregator structure. See Table 16 and Section 5 for the empirical comparison.

Figure 16: Part 2: Flexible Input Parameters Under Block A+B+C



Notes: Distribution of $(\hat{\beta}_m, \hat{\beta}_e, \hat{\beta}_w)$ *from Part 2 (Block A+B+C). These parameters are identified by Blocks A and B alone; comparison with Part 1 confirms that the addition of Block C moments does not contaminate the flexible input estimates.*

G Additional Results for Empirical Analysis

G.1 Cross-Industry Distribution of Parameter Estimates

Table 15 reports the cross-industry distribution (median, mean, SD) of the intermediate input elasticity estimates ($\hat{\beta}_m$, $\hat{\beta}_e$, $\hat{\beta}_w$) for both the proposed method (Panel A) and ACF (Panel B). Capital and labor elasticities ($\hat{\beta}_k$, $\hat{\beta}_l$) across all three methods are reported in Table 7.

Table 15: Cross-Industry Distribution of Intermediate Input Elasticity Estimates

Parameter	N	Median	Mean	SD
Panel A: Proposed Method ($N = 502$)				
$\hat{\beta}_m$ (Material)	502	0.491	0.422	0.271
$\hat{\beta}_e$ (Electricity)	502	0.001	0.079	0.191
$\hat{\beta}_w$ (Water)	502	0.006	0.142	0.300
Panel B: ACF ($N = 500$)				
$\hat{\beta}_m$ (Material)	500	0.565	0.535	0.191
$\hat{\beta}_e$ (Electricity)	500	0.072	0.115	0.143
$\hat{\beta}_w$ (Water)	500	0.017	0.055	0.095

Note:
Outliers $|\hat{\beta}| > 2$ excluded from ACF panel.

G.2 Identification Cross-Check: Exclusion Restriction vs. Block C

Table 16 presents the four-group comparison of $\hat{\beta}_k$ and $\hat{\beta}_l$. Groups (i) and (i-a) form the identification cross-check: the exclusion restriction and Block C applied to the same 302 exclusion-consistent industries yield closely aligned estimates of β_k (median 0.010 vs. 0.039) and β_l (median 0.219 vs. 0.341), indicating that the two conceptually distinct identification strategies converge. Groups (ii) and (iii) provide diagnostic contrast.

Group	Method	N	$\hat{\beta}_k$ Median	$\hat{\beta}_k$ Mean	$\hat{\beta}_k$ SD	$\hat{\beta}_l$ Median	$\hat{\beta}_l$ Mean	$\hat{\beta}_l$ SD
(i) Excl. (consistent, $N = 302$)	Excl. restriction	302	0.010	0.010	0.038	0.219	0.204	0.208
(i-a) Block C (consistent, $N = 302$)	Block C	302	0.039	0.048	0.046	0.341	0.339	0.192
(ii) Excl. (inconsistent, $N = 200$)	Excl. restriction	200	-0.013	-0.073	0.555	0.184	0.093	0.573
(iii) Block C (all, $N = 502$)	Block C	502	0.035	0.048	0.054	0.332	0.336	0.224

Notes: Outliers $|\hat{\beta}| > 2$ excluded. Groups (i) and (i-a) are applied to the identical set of industries, forming the identification cross-check.

Table 16: Four-Group Comparison of $\hat{\beta}_k$ and $\hat{\beta}_l$: Exclusion Restriction vs. Block C

G.3 Demand Function Parameter Estimates

Table 17 reports the cross-industry distribution (median, mean, SD) of the input demand function parameter estimates. Panel A covers the productivity loading parameters ($\hat{\gamma}_\omega$, $\hat{\delta}_\omega$, $\hat{\zeta}_\omega$) identified by Block B; Panel B covers the demand slopes on capital and labor.

Table 17: Cross-Industry Distribution of Input Demand Function Parameter Estimates (Block A+B)

Parameter	N	Median	Mean	SD
Panel A: Productivity Loading (Block B)				
$\hat{\gamma}_\omega$ (Material)	502	1.547	2.498	3.239
$\hat{\delta}_\omega$ (Water)	502	1.334	4.537	5.819
$\hat{\zeta}_\omega$ (Electricity)	502	1.325	3.210	4.676
Panel B: Demand Slopes on (k, l)				
$\hat{\gamma}_k$ (Capital)	502	0.017	0.002	0.252
$\hat{\gamma}_l$ (Labor)	502	0.010	-0.253	1.541
$\hat{\delta}_k$ (Capital)	502	-0.050	-0.045	0.484
$\hat{\delta}_l$ (Labor)	502	-0.978	-1.280	2.444
$\hat{\zeta}_k$ (Capital)	502	-0.028	-0.040	0.466
$\hat{\zeta}_l$ (Labor)	502	-0.246	-0.604	2.203

Note:
502 manufacturing industries (Block A+B convergence). Outliers $|\hat{\theta}| > 10$ excluded from summary statistics.

G.4 Event Study Results

See Section 5.6 for the main DiD analysis (Table 6). Table 18 below reports the full Sun and Abraham (2021) year-by-year coefficient estimates, confirming flat pre-trends.

Table 18: 2011 Tohoku Earthquake: Sun-Abraham (2021) Event Study

	Proposed (1)	ACF (2)
time = -8	0.0022	0.0124
	(0.0129)	(0.0104)
time = -7	0.0024	0.0167*
	(0.0126)	(0.0099)
time = -6	0.0086	0.0192**
	(0.0123)	(0.0095)
time = -5	0.0015	0.0047
	(0.0121)	(0.0098)
time = -4	0.0116	0.0099
	(0.0121)	(0.0098)
time = -3	0.0131	-0.0037
	(0.0097)	(0.0080)
time = -2	0.0014	0.0042
	(0.0096)	(0.0080)
time = 0	-0.0288**	-0.0253**
	(0.0138)	(0.0122)
time = 1	-0.0010	-0.0029
	(0.0097)	(0.0085)
time = 2	-0.0024	-0.0042
	(0.0096)	(0.0081)
time = 3	-0.0031	-0.0052
	(0.0096)	(0.0081)
time = 4	-0.0120	-0.0295***
	(0.0130)	(0.0114)
time = 5	-0.0061	-0.0117
	(0.0102)	(0.0082)
time = 6	-0.0054	-0.0062
	(0.0102)	(0.0081)
time = 7	0.0005	-0.0097
	(0.0100)	(0.0081)
time = 8	-0.0004	-0.0126
	(0.0103)	(0.0082)
time = 9	-0.0380**	-0.0261**
	(0.0154)	(0.0114)
Observations	219,573	219,573
R^2	0.98729	0.94606
poly(k, ℓ) control	✓	
Firm fixed effects	✓	✓
Ind.×Year fixed effects	✓	✓

Sun and Abraham (2021) estimator. Single cohort 2011; SA estimator coincides with plain event study numerically. Treatment: Iwate, Miyagi, Fukushima (seismic intensity $\geq$ 6-strong). Control: West Japan (prefectures 25–47). Fixed effects: firm + industry×year. Proposed: nonparametric poly(k, ℓ) degree 3 control for $\Delta(k, \ell)$. ACF: no polynomial control (k, ℓ already subtracted in $\hat{\omega}^{\text{ACF}}$). Heteroskedasticity-robust standard errors. Reference period: $t = -1$.

G.5 Time-Varying Parameter Estimates

Because the proposed estimator identifies the production function from static covariances alone, it can be applied to each cross-section separately, tracking parameter evolution over time without imposing structural stability. Table 19 reports annual Block A+B estimates of $(\beta_m, \beta_e, \beta_w)$ for four representative industries. Production elasticities exhibit variation across years; where confidence intervals do not overlap, the variation is statistically significant and inconsistent with time-invariant parameters. Annual cross-sections are smaller than the pooled sample, yielding wider confidence intervals in some periods.

Table 19: Annual Production Function Parameter Estimates (Proposed Method)

Parameter	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Bread																		
β_m (Material)	0.679	0.633	0.681	0.000	0.681	0.638	0.649	0.000	0.486	0.000	0.679	0.000	0.000	0.266	0.692	0.547	0.553	0.000
	(0.037)	(0.160)	(0.030)	(0.394)	(0.041)	(0.076)	(0.054)	(1.142)	(0.452)	(0.236)	(0.036)	(0.245)	(1.042)	(0.224)	(0.046)	(0.069)	(0.049)	(0.625)
β_e (Electricity)	0.000	0.001	0.005	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.057	0.001	0.000	0.000	0.000
β_w (Water)	0.013	0.017	0.023	0.019	0.040	0.026	0.039	0.000	0.013	0.000	0.049	0.027	0.000	0.014	0.001	0.000	0.000	0.000
Corrugated board boxes																		
β_m (Material)	-	-	-	-	-	0.663	0.611	0.655	0.060	0.547	0.539	0.682	0.556	0.591	0.536	0.000	0.636	0.587
						(0.045)	(0.032)	(0.051)	(1.147)	(0.074)	(0.070)	(0.055)	(0.068)	(0.030)	(0.026)	(0.322)	(0.035)	(0.057)
β_e (Electricity)	-	-	-	-	-	0.000	0.000	0.000	0.113	0.000	0.000	0.010	0.008	0.000	0.000	0.000	0.000	0.000
β_w (Water)	-	-	-	-	-	0.000	0.000	0.002	0.000	0.007	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.007
Plastic film																		
β_m (Material)	-	-	-	-	-	0.502	0.421	0.503	0.000	0.552	0.492	0.104	0.009	0.426	0.425	0.455	0.103	0.573
						(0.052)	(0.042)	(0.040)	(0.189)	(0.049)	(0.552)	(0.275)	(0.712)	(0.046)	(0.050)	(0.040)	(0.194)	(0.082)
β_e (Electricity)	-	-	-	-	-	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.001	0.000	0.000	0.000
β_w (Water)	-	-	-	-	-	0.000	0.001	0.001	0.000	0.008	0.002	0.016	0.000	0.028	0.021	0.020	0.000	0.024
Robots																		
β_m (Material)	0.425	0.395	0.413	0.384	0.406	1.000	0.715	0.588	-	0.000	0.000	0.585	-	0.439	0.304	0.445	0.000	-
	(0.040)	(0.037)	(0.060)	(0.039)	(0.041)	(0.167)	(0.182)	(0.381)		(0.563)	(0.486)	(0.070)		(0.584)	(0.331)	(1.013)	(0.306)	
β_e (Electricity)	0.034	0.135	0.000	0.007	0.000	0.067	0.241	0.009	-	0.000	0.000	0.002	-	0.403	0.000	0.079	0.000	-
β_w (Water)	0.008	0.000	0.004	0.002	0.000	0.130	0.094	0.950	-	0.000	0.124	0.094	-	0.324	0.000	0.000	0.000	-

Note: Estimates from annual cross-sectional GMM (Proposed Method). Analytical standard errors in parentheses; SEs < 0.001 reported as such.

G.6 Block C: Homothetic Recovery of (β_k, β_l)

Table 20 reports *internal* Block C diagnostics: the cross-industry distribution of the estimated CES substitution parameter $\hat{\rho}_v$ and capital share $\hat{\alpha}$ (left panel), and the significance rates of the higher-order curvature terms $\hat{\rho}_2, \hat{\rho}_3$ together with the stability of $\hat{\beta}_m$ when Block C moments are added (right panel). These diagnostics assess whether the CES curvature assumption holds within each industry. The subsequent cross-tabulation (Table 16) is an *external* cross-check: it asks whether industries that pass the internal exclusion diagnostic also tend to pass the Block C J-test, providing evidence that the two identification strategies are consistent with each other.

Table 20: Block C Estimation Diagnostics ($N = 502$ industries)

CES parameters			*Specification diagnostics*	
Parameter	Median	Mean	Statistic	Value
$\hat{\rho}_v$	-1.000	-0.149	$\hat{\rho}_2$ significant ($\lvert t \rvert > 1.96$)	153/502
$\hat{\alpha}$	0.500	0.533	$\hat{\rho}_3$ significant ($\lvert t \rvert > 1.96$)	178/502
			$\Delta\hat{\beta}_m$ median (A+B→A+B+C)	0.0003

Notes: $\hat{\rho}_v$: CES substitution parameter; $\hat{\alpha}$: capital share in CES aggregator. $\hat{\rho}_2, \hat{\rho}_3$: higher-order CES terms (Block C curvature instruments). $\Delta\hat{\beta}_m$: change in materials elasticity when Block C moments are added.

Cross-validation of two identification strategies. The exclusion restriction (Section 2.4.4) and the homothetic regularity condition (Section 2.4.5) provide two independent routes to identifying

(β_k, β_l). Their joint behavior across industries provides a powerful indirect validity check that does not rely on either restriction alone.

I classify each industry along two dimensions: (i) *exclusion consistency*, defined as the maximum absolute gap in $\hat{\beta}_k$ (or $\hat{\beta}_l$) across the three proxy-specific OLS estimates being below 0.2 (chosen as roughly one within-group standard deviation of $\hat{\beta}_k$ across all industries; results are qualitatively robust to thresholds of 0.1 and 0.3); and (ii) *Block C specification*, defined as non-rejection of the Block C J-test at the 5% level (the full Block A+B+C system is overidentified).

Among the 502 industries, 39.2% of exclusion-consistent industries also pass the Block C J-test, compared to 12.0% among exclusion-inconsistent industries (302 consistent, 200 inconsistent). Among the 38 industries where both criteria are satisfied, the cross-method correlation of $\hat{\beta}_k$ reaches 0.75, indicating that the two conceptually distinct identification strategies converge to similar estimates when their respective maintained assumptions are empirically supported.

This pattern is informative about the source of Block C J-test failures. A logistic regression of Block C J-test passage on industry characteristics finds that $|d_k|$—the exclusion restriction diagnostic (Remark 1)—is the only statistically significant predictor ($p = 0.004$); sample size, $\hat{\rho}_v$, and $\hat{\alpha}$ are all insignificant. Industries with large $|d_k|$ (median 1.16 among those failing both criteria) exhibit demand functions that respond strongly to capital and labor conditional on productivity, violating the exclusion restriction. In such industries, the Block A+B estimates of the g-function slopes carry substantial contamination from the (k, l)-direction, which propagates into Block C through the constructed productivity index.

These findings support the interpretation that the Block C J-test rejection primarily reflects misspecification transmitted from the demand-side moment conditions, rather than failure of the homothetic production function assumption per se. The 38 industries satisfying both criteria serve as an internally validated subsample in which the full three-block GMM system is well-specified.

H Identification Proofs and Technical Remarks

This appendix collects proofs and technical remarks that supplement the identification results in Section 2.

H.1 Testability of the Exclusion Restriction

This subsection provides the detailed derivation supporting Remark 1.

The Wald test of $d_k = d_l = 0$ described in the main text has 2 degrees of freedom, corresponding to the two parametric restrictions.

With three inputs, two independent pairwise differences for β_k and two for β_l provide four testable implications; the formal test has two degrees of freedom, corresponding to the parametric constraints $d_k = d_l = 0$. The null hypothesis is slightly weaker than the full exclusion restriction: it requires a_k^h / a_ω^h to be common across inputs, a condition satisfied by the exclusion restriction but also by a knife-edge proportional response with no structural basis when the three inputs involve distinct production technologies.

H.2 Proof of Proposition A.1

Proof. **Preliminary: the observational equivalence and demand function parameters.** Under the linear specification, the observational equivalence of Theorem 2 implies that Block A+B estimates satisfy, for each input h,

$$\hat{a}_k^{h*} \xrightarrow{p} a_k^h + a_\omega^h\, c_k, \qquad \hat{a}_l^{h*} \xrightarrow{p} a_l^h + a_\omega^h\, c_l, \tag{56}$$

for some constants (c_k, c_l) characterizing the equivalence class, while $\hat{a}_z^h \xrightarrow{p} a_z^h$ and $\hat{a}_\omega^h \xrightarrow{p} a_\omega^h$ are unaffected by the indeterminacy (since z and ω are orthogonal to the (k,l)-direction of the shift).

Case 1. Under $a_k^h = a_l^h = 0$, equation (56) gives $a_k^{h*} = a_\omega^h\, c_k$ and $a_l^{h*} = a_\omega^h\, c_l$. Since $a_\omega^h \neq 0$, the exclusion restriction forces $c_k = c_l = 0$, resolving the indeterminacy completely. Note, however, that the OLS consistency established below does not require this global identification: it follows directly from the proxy construction (42), which uses only the invariant estimates $\hat{a}_z^h$ and $\hat{a}_\omega^h$ and does not involve (k,l).

The proxy (42) satisfies

$$\hat{\omega}_{jt}^h \xrightarrow{p} \frac{a_z^{h\prime}\, z + a_\omega^h\, \omega + \eta^h - a_z^{h\prime}\, z}{a_\omega^h} = \omega + \frac{\eta^h}{a_\omega^h},$$

and the regression equation becomes

$$\tilde{y}_{jt} - \hat{\omega}_{jt}^h = \beta_0 + \beta_k\, k_{jt} + \beta_l\, l_{jt} + \varepsilon_{jt} - \frac{\eta_{jt}^h}{a_\omega^h}.$$

OLS consistency requires $\mathbb{E}[\varepsilon - \eta^h / a_\omega^h \mid k, l] = 0$. The first term vanishes by Assumption 1. For the second, the law of iterated expectations yields

$$\mathbb{E}[\eta^h \mid k, l] = \mathbb{E}\big[\underbrace{\mathbb{E}[\eta^h \mid k, l, \omega, z]}_{=0} \mid k, l\,\big] = 0,$$

where the inner expectation vanishes because $a_k^h = a_l^h = 0$ ensures $\eta^h = h - a_z^{h\prime} z - a_\omega^h\, \omega$ has no residual dependence on (k,l). Both β_k and β_l are identified.

Case 2. Under $a_k^{h_1} = 0$ (with $a_l^{h_1}$ possibly nonzero), the Block A+B estimate satisfies $\hat{a}_l^{h_1 *} \to a_l^{h_1} + a_\omega^{h_1}\, c_l$. The proxy (43) satisfies

$$\begin{aligned}\hat{\omega}^{h_1} &\xrightarrow{p} \frac{a_l^{h_1}\, l + a_z^{h_1\prime}\, z + a_\omega^{h_1}\, \omega + \eta^{h_1} - (a_l^{h_1} + a_\omega^{h_1}\, c_l)\, l - a_z^{h_1\prime}\, z}{a_\omega^{h_1}} \\ &= \omega - c_l\, l + \frac{\eta^{h_1}}{a_\omega^{h_1}},\end{aligned}$$

The regression equation becomes

$$\tilde{y}_{jt} - \hat{\omega}_{jt}^{h_1} = \beta_0 + \beta_k\, k_{jt} + (\beta_l + c_l)\, l_{jt} + \varepsilon_{jt} - \frac{\eta_{jt}^{h_1}}{a_\omega^{h_1}}.$$

By the same iterated expectations argument, $\mathbb{E}[\varepsilon - \eta^{h_1} / a_\omega^{h_1} \mid k, l] = 0$, so OLS consistently estimates the coefficient on k as β_k and the coefficient on l as $\beta_l + c_l$. The *capital* elasticity β_k is identified; the labor coefficient carries the indeterminacy c_l.

By a symmetric argument using h_2 (with $a_l^{h_2} = 0$), the proxy (44) yields a regression where the coefficient on l equals β_l (identified) and the coefficient on k equals $\beta_k + c_k$ (biased). Combining the two regressions identifies both β_k and β_l. □

H.3 Nonlinear Specifications

Remark H.1 (On nonlinear specifications)**.** *One might consider replacing the subtraction of $\hat{\omega}^h$ in Proposition A.1 with a polynomial regression of $\tilde{y}$ on $(k, l, \hat{\omega}^h, (\hat{\omega}^h)^2, (\hat{\omega}^h)^3)$. This is* not *consistent for β_k and β_l in general. Since $\hat{\omega}^h = \omega + \eta^h/a^h_\omega$ is a noisy proxy, the conditional expectation $\mathbb{E}[\omega \mid k, l, \hat{\omega}^h]$ depends on (k, l) through signal extraction, and a polynomial in $\hat{\omega}^h$ alone cannot absorb this component. By contrast, fixing the coefficient on $\hat{\omega}^h$ to unity (as in Proposition A.1) eliminates ω algebraically, avoiding this problem.*

H.4 Proof of Proposition A.2

Proof. Under conditions (i) and (ii), the intermediate inputs serve as sufficient statistics for ω given the state variables: once (x, m, e, w) are observed, knowing D provides no additional information about ω. Formally,

$$\mathbb{E}[\omega_{jt} \mid D_{jt}, x_{jt}, m_{jt}, e_{jt}, w_{jt}] = \mathbb{E}[\omega_{jt} \mid x_{jt}, m_{jt}, e_{jt}, w_{jt}]. \tag{57}$$

To see this, note that condition (i) implies that the demand functions $m = g_m(x, \omega, \tau)$, $e = g_e(x, \omega, \nu)$, and $w = g_w(x, \omega, \eta)$ have the same structure regardless of D. Hence, for any given (x, ω), the conditional distribution of (m, e, w) is unaffected by D. Condition (ii) implies that $D = d(\omega, x)$ for some function d that does not depend on (τ, ν, η). It follows that conditional on (x, m, e, w), the posterior distribution of ω already incorporates all the information that D could provide about ω. By the law of iterated expectations:

$$\mathbb{E}[\hat{\omega}_{jt} \mid D_{jt}] = \mathbb{E}\big[\mathbb{E}[\omega_{jt} \mid x, m, e, w] \,\big|\, D_{jt}\big] = \mathbb{E}[\omega_{jt} \mid D_{jt}]. \tag{58}$$

□

H.5 Assumption 3: Necessity and Testability Details

Each condition in Assumption 3 is necessary for Theorem 3. The failure mode differs by condition.

Necessity of condition (A). If $h' \equiv \gamma$ is constant, then $\bar{\omega}(k, l) = \gamma v(k, l) + c_0$ is linear in v. For any (c_k, c_l), define $\tilde{v}(k, l) = v(k, l) - (c_k k + c_l l)/\gamma$; then $\tilde{\bar{\omega}}(k, l) = \gamma\tilde{v}(k, l) + c_0$ preserves the structure. The shift is fully absorbed, and (β_k, β_l) remain unidentified.

Necessity of condition (B). Under translation homogeneity, the alternative index $\tilde{v}(k, l) = v(k, l) - (c_k k + c_l l)/\gamma$ (from the necessity argument for (A)) is also translation homogeneous only if $c_k k + c_l l$ is translation homogeneous, which requires $c_k + c_l = 1$; for general (c_k, c_l) this need not hold. The requirement that $\tilde{\bar{\omega}} = \tilde{h}(\tilde{v})$ for some $\tilde{h}$ and translation homogeneous $\tilde{v}$ constrains (c_k, c_l) through the nonlinearity of h (condition (A)) and the non-constancy of the MRS (condition (C)). Without translation homogeneity, $\tilde{v}$ can absorb the shift through higher-order terms without contradicting the structural form.

Necessity of condition (C). If v_k/v_l is constant on (k, l), then under translation homogeneity $v(k, l) = \alpha k + (1-\alpha)l$ (the Cobb–Douglas form in logs). The proof of Theorem 3 requires $c_l v_k - c_k v_l = 0$ everywhere, which is automatically satisfied when $v_k/v_l = \alpha/(1-\alpha)$ for any (c_k, c_l) satisfying $c_l/c_k =$

$\alpha/(1-\alpha)$. The one-dimensional manifold of solutions $\{(c_k, c_l) : c_l/c_k = \alpha/(1-\alpha)\}$ represents a residual indeterminacy that cannot be eliminated.

Economic interpretation. Condition (A) requires that the cross-sectional relationship between the capital-labor index and expected productivity is nonlinear: identical absolute increases in the log index at different levels have different effects on expected productivity. Condition (B) corresponds to constant returns to scale in the level variables (translation homogeneity on the log scale is equivalent to degree-one homogeneity in levels); if relaxed, the aggregator can absorb shifts that would otherwise identify the parameters. Condition (C) requires a finite and non-unit elasticity of substitution between capital and labor: firms with different capital-labor ratios face different marginal rates of technical substitution, and this variation provides the cross-sectional nonlinearity needed for identification.

Testability. Conditions (A)–(C) concern $\bar{\omega}(k,l)$, which is a function of the structural parameters estimated in Blocks A and B. Using Block A+B estimates, one can recover $\hat{\bar{\omega}}(k,l)$ up to the $\Delta(k,l)$ shift. Condition (C) can be assessed by testing whether the estimated $\hat{\rho}_v$ differs significantly from zero. In the CES specification, $\hat{\rho}_v$ is directly estimated by grid search in Block C (Section 3.1); a likelihood ratio or information criterion comparison between the CES and Cobb–Douglas nested models tests (C) directly. Condition (A) can be assessed by examining whether the relationship between the estimated index and production residuals exhibits significant nonlinearity, using a RESET-type test on the Block C residuals. Section 3.1.7 describes the implementation.

I Estimation Details

This appendix collects technical details of the GMM estimation procedure that supplement Section 3.

I.1 Block A: Instrument Assignment and Invariance

Instrument assignment logic. Since $u_{1,jt}$ consists only of τ_{jt} and ν_{jt}, it is uncorrelated with η_{jt} (Assumption A.4(3)) and hence with w_{jt} (which contains η_{jt} as the sole unobserved component uncorrelated with Z_{base}). Thus w_{jt} serves as an additional instrument for $u_{1,jt}$. The reasoning for $u_{2,jt}$ and $u_{3,jt}$ is analogous: $u_{2,jt}$ contains τ_{jt} and η_{jt} but not ν_{jt}, so e_{jt} is a valid instrument; $u_{3,jt}$ contains ε_{jt} and τ_{jt} but not ν_{jt} or η_{jt}, so both e_{jt} and w_{jt} are valid instruments.

Invariance to $\Delta(k,l)$. Block A is invariant to the observationally equivalent transformation $\tilde{\omega} = \omega + c_k k + c_l l$ (Theorem 2). Under this relabeling, $\tilde{\beta}_k = \beta_k + c_k$ and $\tilde{\gamma}_k = \gamma_k + \gamma_\omega c_k$ (and analogously for l and for e, w). All residuals $\tilde{m}_{jt}$, $\tilde{e}_{jt}$, $\tilde{w}_{jt}$, and $\tilde{y}_{jt}$ are individually invariant to this transformation: for instance, $\tilde{y}_{jt} = \tilde{\beta}_k k + \tilde{\beta}_l l + \tilde{\omega} + \varepsilon = \beta_k k + \beta_l l + \omega + \varepsilon$. Hence $u_{i,jt}$ $(i = 1,2,3)$ are invariant, and Block A cannot separately identify β_k and the demand slopes on (k,l).

I.2 Block B: Covariance Derivation

Using the mutual exogeneity of shocks (Assumption A.4(3)), the cross-covariances among residuals satisfy:

$$\mathrm{Cov}(\tilde{m}, \tilde{e}) = \gamma_\omega \, \delta_\omega \, \mathrm{Var}(\omega), \tag{59}$$

$$\mathrm{Cov}(\tilde{m}, \tilde{w}) = \gamma_\omega \, \zeta_\omega \, \mathrm{Var}(\omega), \tag{60}$$

$$\mathrm{Cov}(\tilde{e}, \tilde{w}) = \delta_\omega \, \zeta_\omega \, \mathrm{Var}(\omega), \tag{61}$$

$$\mathrm{Cov}(\tilde{y}, \tilde{m}) = \gamma_\omega \, \mathrm{Var}(\omega), \tag{62}$$

$$\mathrm{Cov}(\tilde{y}, \tilde{e}) = \delta_\omega \, \mathrm{Var}(\omega), \tag{63}$$

$$\mathrm{Cov}(\tilde{y}, \tilde{w}) = \zeta_\omega \, \mathrm{Var}(\omega). \tag{64}$$

Eliminating $\mathrm{Var}(\omega)$ across pairs yields six moment conditions. For instance, dividing (59) by (62) gives $\mathrm{Cov}(\tilde{m}, \tilde{e}) = \delta_\omega \mathrm{Cov}(\tilde{y}, \tilde{m})$. Expressing all six relations in expectation form (using de-meaned data):

$$\mathbb{E}\big[\tilde{m}\,\tilde{e} - \delta_\omega\,\tilde{y}\,\tilde{m}\big] = 0, \tag{65}$$

$$\mathbb{E}\big[\tilde{m}\,\tilde{e} - \gamma_\omega\,\tilde{y}\,\tilde{e}\big] = 0, \tag{66}$$

$$\mathbb{E}\big[\tilde{m}\,\tilde{w} - \zeta_\omega\,\tilde{y}\,\tilde{m}\big] = 0, \tag{67}$$

$$\mathbb{E}\big[\tilde{m}\,\tilde{w} - \gamma_\omega\,\tilde{y}\,\tilde{w}\big] = 0, \tag{68}$$

$$\mathbb{E}\big[\tilde{e}\,\tilde{w} - \zeta_\omega\,\tilde{y}\,\tilde{e}\big] = 0, \tag{69}$$

$$\mathbb{E}\big[\tilde{e}\,\tilde{w} - \delta_\omega\,\tilde{y}\,\tilde{w}\big] = 0. \tag{70}$$

Redundancy with Block A. Of the six moment conditions (65)–(70), four are algebraically implied by the Block A instrumental variable moments. Specifically, any four of the six conditions that involve cross-products of the demand residuals $\tilde{e}$ or $\tilde{w}$ with $\tilde{m}$ (or equivalently with $\tilde{y}$ via u_3) are already encoded in the Block A moment conditions through the instruments $Z_3 = (k, l, \tilde{e}, \tilde{w})$ (equation (27)); which four are labeled "redundant" depends on the chosen basis, but the rank reduction by four is basis-independent. The two independent contributions (in any basis) correspond to cross-covariance ratios not captured by Block A instruments. Consequently, the combined Block A+B system has $\mathrm{rank}(\Omega) = 12$, matching the 12 free parameters, and is just-identified. The concentrated covariance-ratio formulas remain useful for obtaining closed-form scale parameter estimates, improving computational efficiency.

Invariance to $\Delta(k, l)$. As with Block A, Block B is invariant to the $\Delta(k, l)$ transformation, since each residual $\tilde{m}_{jt}$, $\tilde{e}_{jt}$, $\tilde{w}_{jt}$, and $\tilde{y}_{jt}$ is individually invariant to the relabeling $\tilde{\omega} = \omega + c_k k + c_l l$ (Appendix I.1).

I.3 De-Meaning, Intercept Recovery, and Two-Step Procedure

De-meaning. Non-zero demand intercepts $(\gamma_0, \delta_0, \zeta_0)$ cause the raw-level moment conditions to be misspecified. To see this, note that $\mathbb{E}[u_{1,jt}] = \delta_\omega \gamma_0 - \gamma_\omega \delta_0$, which is generally nonzero. All variables are therefore de-meaned prior to estimation and the constant is excluded from all instrument vectors.

Post-estimation intercepts. After obtaining the GMM estimates $\hat{\Theta}$, the production function intercept is recovered as:

$$\hat{\beta}_0^{\text{new}} = \bar{y} - \hat{\beta}_k \bar{k} - \hat{\beta}_l \bar{l} - \hat{\beta}_m \bar{m} - \hat{\beta}_e \bar{e} - \hat{\beta}_w \bar{w}, \tag{71}$$

where overbars denote sample means of the original (non-de-meaned) data. This formula follows from the normalization $\mathbb{E}[\omega] = 0$, which implies $\mathbb{E}[h(v)] = \mathbb{E}[\mathbb{E}[\omega \mid k, l]] = \mathbb{E}[\omega] = 0$ by the law of iterated expectations. Hence the function h does not appear in the intercept. The intercept absorbs the demand function intercepts and the constant ρ_0.

Stacked GMM objective. Define the integrated moment vector by stacking all three blocks:

$$g_{jt}(\Theta) = \left[g_{jt,A}(\Theta)',\ g_{jt,B}(\Theta)',\ g_{jt,C}(\Theta)'\right]',$$

where $g_{jt,A}$ collects the Block A moments, $g_{jt,B}$ collects the Block B covariance moments, and $g_{jt,C}$ collects the Block C structural moments. All parameters $\Theta = (\theta_1, \theta_2)$ are estimated simultaneously by minimizing:

$$\hat{\Theta} = \arg\min_{\Theta}\ g_N(\Theta)' \hat{W} g_N(\Theta), \qquad g_N(\Theta) = \frac{1}{N}\sum_{j=1}^{N} \bar{g}_j(\Theta), \tag{72}$$

where $\bar{g}_j(\Theta) = T^{-1}\sum_{t=1}^{T} g_{jt}(\Theta)$ is the time-averaged moment for firm j.

Two-step procedure.

1. **Step 1:** Minimize (72) with an initial weighting matrix $\hat{W}^{(0)}$ (e.g., a block-diagonal matrix that normalizes the scale of each block) to obtain $\hat{\Theta}^{(1)}$.
2. **Optimal weight:** Estimate the long-run covariance matrix $\hat{\Sigma} = \frac{1}{N}\sum_{j=1}^{N} \bar{g}_j(\hat{\Theta}^{(1)})\, \bar{g}_j(\hat{\Theta}^{(1)})'$ and set $\hat{W}_{\text{opt}} = \hat{\Sigma}^{-1}$.
3. **Step 2:** Re-minimize (72) with $\hat{W}_{\text{opt}}$ to obtain the efficient estimator $\hat{\Theta}^{(2)}$.
4. **Intercepts:** Recover $\hat{\beta}_0^{\text{new}}$ via (71).

I.4 Computational Details

Computational cost. The two-step GMM requires numerical optimization over Θ ($\dim \Theta = 18 + 3\,d_z$ when control variables with a polynomial basis of dimension d_z are included; $d_z = 0$ in the baseline specification). Block C introduces nonlinearity through the CES index (16), and (ρ_v, α) are optimized by profile GMM over a discrete grid. Each grid point requires a standard GMM optimization over the remaining parameters, making the total cost approximately $|\text{grid}| \times$ cost of a single GMM evaluation. In the empirical application with $N \approx 5{,}000$ and $T \approx 20$, a single industry estimation completes in under one minute on a 12-core workstation. Bootstrap standard errors (200 replications) require proportionally more time.

Iterative profile estimation of scale parameters. The scale parameters $(\gamma_\omega, \delta_\omega, \zeta_\omega)$ and the slope parameters $(\beta_m, \beta_e, \beta_w, \theta_g)$ enter the moment conditions multiplicatively, creating a ridge in the GMM objective surface. I employ an iterative profile strategy: given current scale values, the slope parameters are estimated by minimizing the GMM objective; the scale parameters are then updated via closed-form covariance ratios derived from the Block B conditions (28); and the procedure iterates until convergence. A final joint optimization step refines all parameters simultaneously, using the iterative profile solution as starting values.

I.5 Regularity Conditions and Asymptotic Proof

Assumption I.1 (Standard Conditions for Asymptotics). *1. The sample $\{(y_{jt}, k_{jt}, l_{jt}, m_{jt}, e_{jt}, w_{jt})_{t=1}^{T}\}_{j=1}^{N}$ consists of N independent draws (independence across firms), with T fixed.*

2. *The weighting matrix $\hat{W}$ converges in probability to a positive definite matrix W ($\hat{W} \xrightarrow{p} W$).*

3. *The true parameter vector Θ_0 lies in the interior of a compact parameter space.*

4. *Identification Condition: $\mathbb{E}[\bar{g}_j(\Theta)] = 0$ if and only if $\Theta = \Theta_0$.*

5. *The variables necessary to compute the moment function $g_{jt}(\Theta)$ have finite moments of sufficiently high order.*

6. *$g_{jt}(\Theta)$ is continuously differentiable in Θ in a neighborhood of Θ_0, and the expected Jacobian matrix $G \equiv \mathbb{E}[\nabla_\Theta \bar{g}_j(\Theta_0)]$ has full column rank.*

Proof of Theorem 4. The result follows from Theorems 2.6 and 3.4 in Newey and McFadden (1994), with Assumption I.1(1) ensuring the applicability of cross-sectional LLN and CLT. The time-averaged moment $\bar{g}_j(\Theta) = T^{-1} \sum_{t=1}^{T} g_{jt}(\Theta)$ treats each firm's T-period panel as a single observation, so the asymptotic framework is cross-sectional ($N \to \infty$, T fixed). The optimal weighting matrix $W = \Sigma^{-1}$ yields the efficient two-step GMM estimator with asymptotic variance $(G'\Sigma^{-1}G)^{-1}$.

Standard errors are computed from a consistent estimate $\hat{V}$ of the asymptotic variance (35), using the second-step estimates $\hat{\Theta}^{(2)}$ to evaluate the sample Jacobian $\hat{G}$ and the moment covariance $\hat{\Sigma}$. The standard error for the post-estimation intercept $\hat{\beta}_0^{\text{new}}$ is computed via the delta method. In practice, standard errors are clustered at the firm level. Although the time-averaged moment $\bar{g}_j$ aggregates across periods, within-firm serial dependence can still inflate the variance of $\bar{g}_j$ relative to the i.i.d. case. Clustering at the firm level provides a heteroskedasticity- and autocorrelation-consistent estimate of Σ that accommodates arbitrary within-firm temporal dependence, analogous to cluster-robust variance estimation in panel regressions. □

J Direction of Bias under Conditional Independence Violation

This appendix derives the direction of bias in $\hat{\beta}_m$ when the conditional independence assumption (Assumption 2) is violated through a positive covariance between the electricity and water demand shocks.

J.1 Setup

After the Frisch–Waugh–Lovell projection, the residual structure takes the form:

$$\tilde{y}_{jt} = \omega_{jt} + \varepsilon_{jt}, \tag{73}$$

$$\tilde{m}_{jt} = \gamma_\omega \, \omega_{jt} + \tau_{jt}, \tag{74}$$

$$\tilde{e}_{jt} = \delta_\omega \, \omega_{jt} + \nu_{jt}, \tag{75}$$

$$\tilde{w}_{jt} = \zeta_\omega \, \omega_{jt} + \eta_{jt}, \tag{76}$$

where $(\tau_{jt}, \nu_{jt}, \eta_{jt})$ are the input-specific demand shocks. Under Assumption 2, all pairwise covariances among these shocks are zero.

J.2 CI Violation: Electricity–Water Common Utility Shock

Suppose:

$$\sigma_{\nu\eta} \equiv \mathrm{Cov}(\nu_{jt}, \eta_{jt}) > 0, \tag{77}$$

while $\mathrm{Cov}(\tau, \nu) = \mathrm{Cov}(\tau, \eta) = 0$. This arises naturally when a common energy price shock or seasonal supply constraint raises both electricity and water costs simultaneously—the most economically salient threat to conditional independence, since electricity and water are both utility services subject to common regulatory and infrastructure conditions. The materials demand shock τ_{jt}, which reflects raw material procurement through distinct supply chains, remains independent.

J.3 Bias in Scale Parameters

The concentrated scale estimator for ζ_ω uses the cross-covariance between the electricity and water residuals. From (75) and (76):

$$\mathbb{E}[\tilde{e} \cdot \tilde{w}] = \delta_\omega \zeta_\omega \,\mathrm{Var}(\omega) + \sigma_{\nu\eta}. \tag{78}$$

The normalizing moment $\mathbb{E}[\tilde{y} \cdot \tilde{e}] = \delta_\omega \,\mathrm{Var}(\omega)$ is unaffected by $\sigma_{\nu\eta}$. The ratio gives:

$$\hat{\zeta}_\omega \xrightarrow{p} \zeta_\omega + \frac{\sigma_{\nu\eta}}{\delta_\omega \,\mathrm{Var}(\omega)}. \tag{79}$$

The bias is positive when $\sigma_{\nu\eta} > 0$: $\hat{\zeta}_\omega$ overestimates the true scale parameter. Since $\mathrm{Cov}(\tau, \nu) = \mathrm{Cov}(\tau, \eta) = 0$, the other two scale parameters $\hat{\gamma}_\omega$ and $\hat{\delta}_\omega$ remain consistently estimated.

J.4 Propagation to $\hat{\beta}_m$: Upward Bias

The overestimation of ζ_ω propagates to $\hat{\beta}_m$ through the Block A moment conditions. The Block A error $u_{2,jt} = \zeta_\omega \tilde{m}_{jt} - \gamma_\omega \tilde{w}_{jt}$ eliminates ω_{jt} when the scale parameters are correctly specified. When $\hat{\zeta}_\omega > \zeta_\omega$, a positive fraction of ω_{jt} leaks into $\hat{u}_{2,jt}$:

$$\hat{u}_{2,jt} = (\zeta_\omega + b)\,\tilde{m}_{jt} - \gamma_\omega\,\tilde{w}_{jt} = u_{2,jt} + b\,\tilde{m}_{jt},$$

where $b > 0$ is the bias in (79) and $\tilde{m}_{jt} = \gamma_\omega \omega_{jt} + \tau_{jt}$ is positively correlated with productivity. This contamination biases the moment conditions for the production function coefficients. In particular, the GMM estimator compensates for the positive ω leakage in the u_2-based moments by *increasing* $\hat{\beta}_m$, producing an **upward bias**.

Monte Carlo simulations (Section 4, Table 14) confirm this direction: $\hat{\beta}_m$ increases monotonically with $\mathrm{Corr}(\nu, \eta)$.

J.5 Implications

1. The bias is *upward*: if CI is violated through a common electricity–water utility shock, the proposed estimator *overestimates* β_m and implied markups.

2. This bias direction is the *same* as the Markov misspecification bias in ACF-type estimators (which also overestimates β_m under DGPs 2 and 3). Therefore, the empirical finding that the

proposed estimator yields *lower* $\hat{\beta}_m$ than ACF cannot be attributed to CI violation; it must reflect Markov misspecification bias in ACF.

3. Including additional control variables in z_{jt} (e.g., regional energy price indices, seasonal indicators) reduces $\sigma_{\nu\eta}$ by absorbing common sources of utility cost variation, providing a partial remedy.

K Parametric Implementation under Flexible Functional Forms

This appendix extends the parametric GMM implementation of Section 3 to flexible functional forms. The identification source throughout is the conditional independence of demand shocks (Assumption 2)—the parametric counterpart of the HS08 spectral decomposition (Theorems 1–2). Under Cobb–Douglas, conditional independence yields the linear covariance structure of Blocks A and B. Under translog, the same condition yields nonlinear moment conditions derived from the structure of input demand functions.

K.1 Translog Production Function

Consider the translog production function:

$$\begin{aligned} y_{jt} = \; & \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \beta_e e_{jt} + \beta_w w_{jt} \\ & + \beta_{kk} k_{jt}^2 + \beta_{ll} l_{jt}^2 + \beta_{mm} m_{jt}^2 + \beta_{ee} e_{jt}^2 + \beta_{ww} w_{jt}^2 \\ & + \beta_{kl} k_{jt} l_{jt} + \beta_{km} k_{jt} m_{jt} + \beta_{ke} k_{jt} e_{jt} + \beta_{kw} k_{jt} w_{jt} \\ & + \beta_{lm} l_{jt} m_{jt} + \beta_{le} l_{jt} e_{jt} + \beta_{lw} l_{jt} w_{jt} \\ & + \beta_{me} m_{jt} e_{jt} + \beta_{mw} m_{jt} w_{jt} + \beta_{ew} e_{jt} w_{jt} + \omega_{jt} + \varepsilon_{jt}. \end{aligned} \tag{80}$$

The log marginal products are:

$$\frac{\partial f}{\partial m} = \beta_m + 2\beta_{mm} m + \beta_{km} k + \beta_{lm} l + \beta_{me} e + \beta_{mw} w, \tag{81}$$

$$\frac{\partial f}{\partial e} = \beta_e + 2\beta_{ee} e + \beta_{ke} k + \beta_{le} l + \beta_{me} m + \beta_{ew} w, \tag{82}$$

$$\frac{\partial f}{\partial w} = \beta_w + 2\beta_{ww} w + \beta_{kw} k + \beta_{lw} l + \beta_{mw} m + \beta_{ew} e. \tag{83}$$

K.2 Demand Structure under Translog

From the first-order condition for cost minimization (Appendix B), each intermediate input $h \in \{m, e, w\}$ satisfies:

$$f + \omega - h + \ln\left(\frac{\partial f}{\partial h}\right) = \phi_h(z_{jt}) + \tau_h, \tag{84}$$

where $\phi_h(z)$ captures price and markdown terms absorbed by control variables, and τ_h is the input-specific demand shock. Under translog, the log marginal product $\ln(\partial f/\partial h)$ depends on the levels of all inputs, so input demands are implicitly defined and nonlinear in productivity.

K.3 Moment Conditions from Conditional Independence

The key observation is that subtracting the first-order conditions for two inputs eliminates both f and ω. For inputs m and e:

$$\ln\left(\frac{\partial f/\partial m}{\partial f/\partial e}\right) - (m-e) = (\phi_m - \phi_e) + (\tau_m - \nu_e). \tag{85}$$

The production function f and productivity ω cancel exactly. De-meaning removes the price terms $\phi_m - \phi_e$, yielding:

$$\ln\widetilde{\left(\frac{\partial f/\partial m}{\partial f/\partial e}\right)} - (\tilde{m} - \tilde{e}) = \tilde{\tau}_m - \tilde{\nu}_e, \tag{86}$$

where tildes denote de-meaned variables.

By Assumption 2, τ_m, ν_e, and η_w are mutually independent conditional on (ω, k, l, z). Therefore, η_w is uncorrelated with $\tau_m - \nu_e$, and w serves as a valid instrument. Defining:

$$u_{me}^{TL} \equiv \ln\left(\frac{\beta_m + 2\beta_{mm}m + \beta_{km}k + \beta_{lm}l + \beta_{me}e + \beta_{mw}w}{\beta_e + 2\beta_{ee}e + \beta_{ke}k + \beta_{le}l + \beta_{me}m + \beta_{ew}w}\right) - (m-e) - \text{mean}, \tag{87}$$

the moment condition is:

$$\mathbb{E}\big[u_{me}^{TL} \cdot (k,\, l,\, w,\, z)\big] = 0. \tag{88}$$

Analogous conditions hold for the pairs (m, w) and (e, w):

$$\mathbb{E}\big[u_{mw}^{TL} \cdot (k,\, l,\, e,\, z)\big] = 0, \tag{89}$$

$$\mathbb{E}\big[u_{ew}^{TL} \cdot (k,\, l,\, m,\, z)\big] = 0. \tag{90}$$

These three sets of nonlinear moment conditions identify the intermediate input parameters $(\beta_m, \beta_e, \beta_w, \beta_{mm}, \beta_{ee}, \beta$

K.4 Identification of Primary Input Parameters

The $\Delta(k, l)$ indeterminacy of Theorem 2 persists under translog: the moment conditions (88)–(90) do not identify parameters involving (k, l), namely $(\beta_k, \beta_l, \beta_{kk}, \beta_{ll}, \beta_{kl}, \beta_{km}, \beta_{ke}, \beta_{kw}, \beta_{lm}, \beta_{le}, \beta_{lw})$.

Corollary 1 (exclusion restrictions) extends directly to translog. Condition (i)—that some input demand is independent of (k, l) conditional on productivity—implies:

$$\beta_{kw} = \beta_{lw} = 0. \tag{91}$$

Under this restriction, the log marginal product of w (equation 83) does not depend on (k, l):

$$\frac{\partial f}{\partial w} = \beta_w + 2\beta_{ww}w + \beta_{mw}m + \beta_{ew}e. \tag{92}$$

The productivity proxy constructed from input w is then independent of (k, l), and the (k, l) parameters can be recovered by the following procedure.

Define the partially residualized output using the intermediate input estimates from Section K.3:

$$\begin{aligned}\tilde{y}_{jt}^{TL} \equiv\ & y_{jt} - \hat{\beta}_m m - \hat{\beta}_e e - \hat{\beta}_w w \\ & - \hat{\beta}_{mm}m^2 - \hat{\beta}_{ee}e^2 - \hat{\beta}_{ww}w^2 - \hat{\beta}_{me}me - \hat{\beta}_{mw}mw - \hat{\beta}_{ew}ew.\end{aligned} \tag{93}$$

This equals:

$$\tilde{y}_{jt}^{TL} = \beta_k k + \beta_l l + \beta_{kk} k^2 + \beta_{ll} l^2 + \beta_{kl} kl + \beta_{km} km + \beta_{ke} ke + \beta_{lm} lm + \beta_{le} le + \omega + \varepsilon. \tag{94}$$

Construct the productivity proxy from input w:

$$\hat{\omega}_{jt}^{w} \equiv w_{jt} - f_{jt} - \ln\left(\frac{\partial f}{\partial w}\right)_{jt} - \hat{\phi}_w(z_{jt}), \tag{95}$$

where all terms on the right-hand side are evaluated at the estimated parameters. Under the exclusion restriction (91), $\hat{\omega}^w$ does not depend on (k, l).

The moment condition for the (k, l) parameters is:

$$\mathbb{E}\Big[\big(\tilde{y}^{TL} - \hat{\omega}^w - g_{kl}(k, l, m, e; \beta_{kl})\big) \cdot Z_{kl}\Big] = 0, \tag{96}$$

where g_{kl} collects all terms involving (k, l):

$$g_{kl} \equiv \beta_k k + \beta_l l + \beta_{kk} k^2 + \beta_{ll} l^2 + \beta_{kl} kl + \beta_{km} km + \beta_{ke} ke + \beta_{lm} lm + \beta_{le} le, \tag{97}$$

and $Z_{kl} = (k, l, k^2, l^2, kl, km, ke, lm, le)$. The error term $\varepsilon - \eta_w/\zeta_\omega$ is orthogonal to Z_{kl} under the exclusion restriction, identifying all (k, l) parameters.

K.5 Reduction to Cobb–Douglas

Setting all second-order coefficients to zero ($\beta_{hh'} = 0$ for all h, h'), the translog reduces to Cobb–Douglas. The log marginal product ratio in equation (87) becomes:

$$\ln\left(\frac{\partial f/\partial m}{\partial f/\partial e}\right) = \ln\frac{\beta_m}{\beta_e}, \tag{98}$$

a constant that vanishes under de-meaning. The residual u_{me}^{TL} reduces to $-(\tilde{m} - \tilde{e})$, and the nonlinear moment conditions collapse to linear orthogonality conditions.

Similarly, the (k, l) identification (equation 96) reduces to:

$$\tilde{y} - \hat{\omega}^w = \beta_k k + \beta_l l + \text{error}, \tag{99}$$

which is the OLS regression of Remark 1. The Cobb–Douglas implementation of Section 3 is thus a computationally tractable special case of the general framework.

L Empirical Estimation Flowchart

Figure 17 provides an overview of the full empirical estimation and inference pipeline.

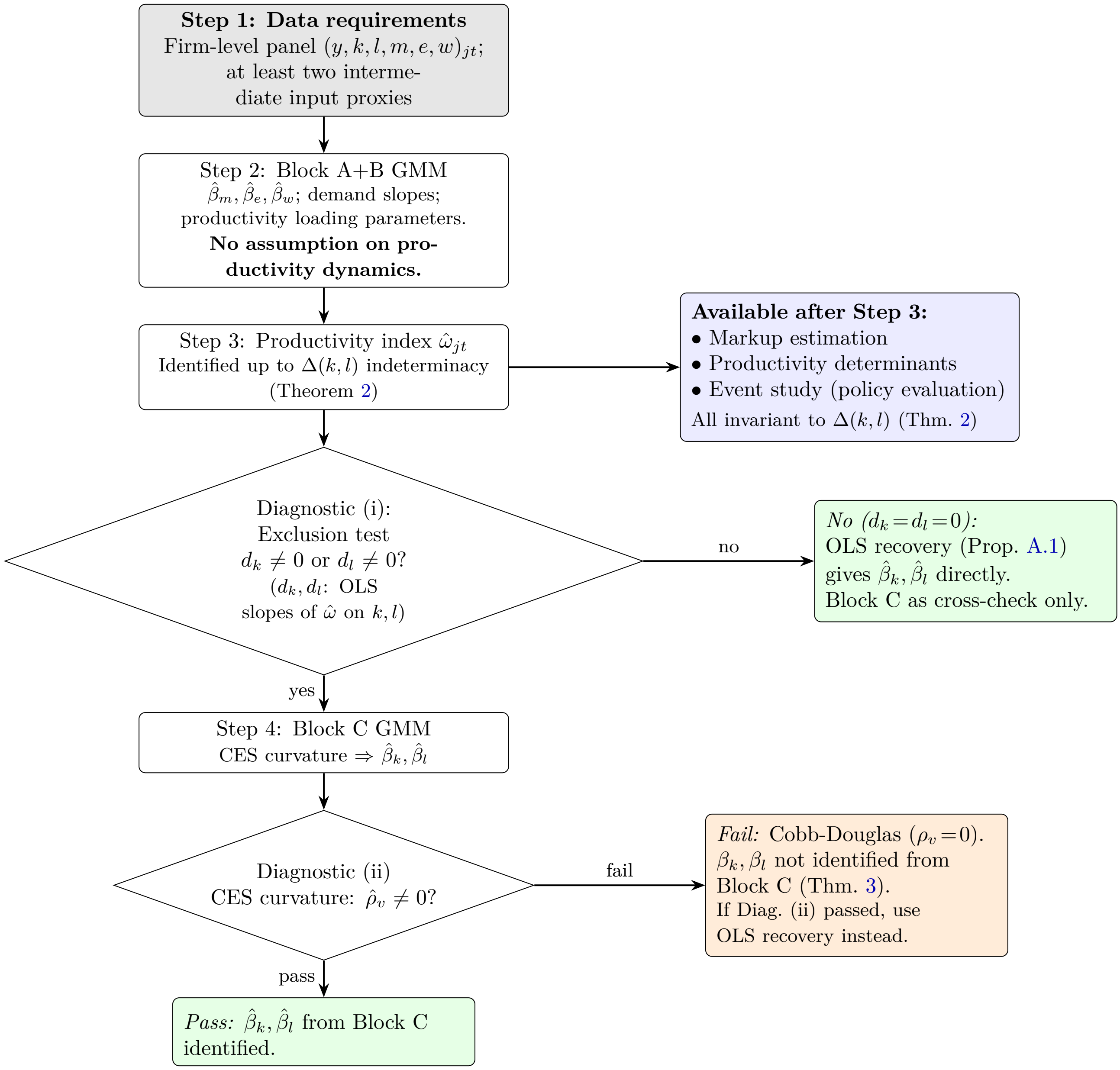


Figure 17: Implementation Guide: Proposed Estimation Pipeline

Notes: Step-by-step guide for applying the proposed estimator to firm-level panel data. The key message is that ***Step 3 alone*** *(Block A+B GMM) suffices for the most common downstream applications—markup estimation, productivity analysis, event studies, and Olley–Pakes decomposition—because these rely only on* $\hat{\beta}_m$ *and* $\hat{\omega}_{jt}$*, which are invariant to the* $\Delta(k,l)$ *indeterminacy (Theorem 2). Step 4 (Block C GMM) is needed only when capital and labor elasticities* (β_k, β_l) *are themselves of interest. Diagnostics (i)–(iii) check the maintained assumptions at each stage.*