Selection Bias and Identification in Production Function Estimation: A Timing-Based Exclusion Restriction*

Rentaro Utamaru†

April 17, 2026

Click here for the latest version

Abstract

Standard production function estimators (Olley-Pakes, Ackerberg-Caves-Frazer, and Gandhi-Navarro-Rivers) share a common identification failure: when capital follows the Perpetual Inventory Method and labor is dynamic, the selection correction is collinear with factor inputs, leaving elasticities unidentified. I resolve this failure with a single exclusion restriction: an idiosyncratic demand shock realized after factor commitments but before the exit decision. Monte Carlo experiments confirm consistency of the proposed estimator where standard estimators are inconsistent. Applied to Japanese manufacturing, the correction shifts the estimated TFP distribution left by up to 0.7 log points, implying that standard methods systematically overstate residual productivity and misattribute capital accumulation's contribution to growth.

Keywords: Production Function, Selection Bias, Identification, Exclusion Restriction, Olley-Pakes
JEL Classification Codes: C14, C23, C25, D24, L11

*I am grateful to Yasutora Watanabe, Yuta Toyama, and Satoshi Imahie for their insightful comments. 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. All remaining errors are my own.

†JSPS Postdoctoral Research Fellow, Waseda University, Tokyo, Japan. Email: rentaro.utamaru@gmail.com

1 Introduction

How does selection bias enter production function estimates when firms observe demand conditions before deciding whether to exit? The question matters because estimated capital elasticities feed directly into productivity decompositions, markup measurement, and evaluations of investment subsidies and industrial policy: a bias of 0.07 in $\hat{\beta}_k$ shifts the residual TFP distribution by up to 0.7 log points, altering conclusions about which industries are capital-driven and which are productivity-driven. Olley and Pakes (1996) (OP) proposed a two-step correction that has become the standard tool in empirical industrial organization. Hahn, Liao, and Ridder (2023) (HLR) recently proved, however, that the OP correction is internally inconsistent: when capital is constructed by the Perpetual Inventory Method, the capital term is perfectly collinear with OP's nonparametric selection control, so the capital elasticity β_k is unidentified. Any study relying on OP's second step under PIM capital and dynamic labor is therefore reporting an estimate of an unidentified parameter. I resolve this failure with a single timing-based exclusion restriction that restores point identification of β_k in OP. I further show that the same collinearity extends to the Ackerberg, Caves, and Frazer (2015) (ACF) and Gandhi, Navarro, and Rivers (2020) (GNR) frameworks, where it renders both (β_k, β_l) jointly unidentified; the same exclusion restriction resolves the problem in all three settings.

Since OP (1996), most major structural corrections for selection bias have identified β_k by assuming that the survival probability depends only on productivity and capital. HLR show that this assumption is self-defeating: the nonparametric correction that purges selection bias simultaneously absorbs the variation needed to identify β_k. The problem is not a defect in a particular implementation; it is inherent in the OP framework whenever capital follows the Perpetual Inventory Method.

I resolve the identification failure by introducing an exclusion restriction grounded in the timing of firm decisions. Firms observe an idiosyncratic demand shock z_{jt} within the period, after predetermining capital and labor. Let v_{jt} denote the within-period innovation in z_{jt} (the component not predictable from last period's information set). Capital is committed at the *beginning* of the period, before v_{jt} realizes, while the exit decision is made *after* observing it (Assumption 3 formalises this sequence). The innovation therefore satisfies $\mathbb{E}[v_{jt} \mid \omega_{jt}] = 0$ by construction. Adding z_{jt} to the survival logit breaks the collinearity between the capital term and the selection control, restoring point identification of β_k without data requirements beyond a standard demand proxy.

The identifying assumption is both economically grounded and directly testable. Economically, investment via PIM is a long-run commitment (a capital stock accumulated over years) whereas the within-period demand shock is a short-run fluctuation; it would be unusual for a firm to reverse its capital stock in response to a demand shock realised within the same period. Empirically, the exclusion restriction requires z_{jt} to predict survival (relevance) while being uncorrelated with beginning-of-period productivity ω_{jt} (exogeneity). Both conditions are examined in the first-stage survival regression (Appendix F); relevance holds in industries with high demand shock persistence, and exogeneity $(\text{cor}(z_{3,jt}, \hat{\omega}_{jt}) \approx 0)$ holds throughout.

Applied to six Japanese manufacturing industries in a plant-level panel, the proposed estimator raises $\hat{\beta}_k$ by up to 0.07 in capital-intensive industries (food processing: $0.258 \to 0.328$; corrugated board: $0.023 \to 0.091$) where the standard OP correction changes the estimate by less than 0.02 from the uncorrected OLS baseline. These corrections have first-order consequences for productivity measurement: in the most affected industries, the estimated TFP distribution shifts by up to 0.7 log points, potentially altering productivity rankings and the evaluation of industrial policies that target capital deepening. Across 177 industries, the correction is concentrated where the theory predicts it should be, along three dimensions: demand shock persistence ($\hat{\rho}$), exit rates, and exit suddenness. Among industries with above-median $\hat{\rho}$, the mean correction is +0.044 ($t = 2.65$, $p = 0.009$ vs. low-$\hat{\rho}$ industries); among high-exit-rate industries, 68% show a positive correction ($p < 0.001$); and among industries with high sudden-exit fractions (Chen, Igami, et al. (2021)), 67% show a positive correction ($p = 0.001$). As a falsification test, industries simultaneously low on all three dimensions show only 38% success, indistinguishable from the null. Demand shock persistence is the only significant predictor in a cross-industry regression ($p = 0.020$), providing a practical diagnostic for when the standard OP correction is most likely to fail.[1]

Monte Carlo experiments designed to replicate the HLR failure confirm that the standard estimator is inconsistent under this DGP, while the proposed estimator's mean estimate converges to the true value of $\beta_k = 0.5$ as the sample size increases from $J = 100$ to $J = 300$, with bias falling from -0.009 to $+0.005$ and RMSE declining monotonically.

This paper makes three contributions. First, I provide a constructive solution to the HLR identification failure. Under three assumptions: (i) investment precedes the within-period demand shock (Assumption 3), (ii) the demand shock shifts the exit probability (Assumption 4), and (iii) the joint distribution of capital and exit probability has non-degenerate support (Assumption 6); the capital elasticity β_k is globally point-identified from panel data augmented with a demand proxy, and the identifying assumptions are testable with observable data. Second, I establish that the HLR collinearity extends to the Ackerberg, Caves, and Frazer (2015) (ACF) and Gandhi, Navarro, and Rivers (2020) (GNR) frameworks, where it renders (β_k, β_l) jointly unidentified. The mechanism in ACF and GNR differs from OP in that labor, treated as quasi-fixed (a state variable determined before productivity is observed), creates the same collinearity with the selection correction that investment creates in OP: conditioning on (k_{jt}, l_{jt}) determines last period's productivity, leaving no residual variation orthogonal to the selection control. I show that the same demand shock exclusion restriction breaks this extended collinearity in ACF and GNR (Theorem 11), providing a unified identification strategy across the three dominant production function estimators in the literature. Third, I show that the magnitude of the correction is heterogeneous across industries in a theoretically predictable way, and I provide an empirically implementable diagnostic for when it matters. The correction is large and precisely estimated only in industries with persistent demand shocks (high AR(1) coefficient $\hat{\rho}$), where the first-stage relevance condition is satisfied; in low-persistence industries the mean gap is indistinguishable from zero. Because $\hat{\rho}$ is estimable from any demand proxy series, this heterogeneity result gives applied researchers a direct pre-test for whether the standard OP correction is likely to fail in their setting.

The solution requires observing a demand proxy to construct z_{jt}. This is a data requirement beyond OP's standard inputs, but one that is routinely available in firm-level panel datasets. An alternative approach, Chen, Igami, et al. (2021), exploits institutional delays in exit implementation (in their setting, regulatory and bankruptcy procedures create a lag between the exit decision and effective firm closure) to construct an exclusion restriction for production function estimation. My approach is complementary: it applies to markets where exit is implemented within the period and no such institutional lag is available, but where the within-period timing of demand shock realization provides the required variation.

Hahn, Liao, and Ridder (2023) identify the collinearity problem but do not propose an exclusion-

[1] The unconditional mean gap across all 177 industries is +0.015 (SD 0.148; t-test $p = 0.095$), reflecting large finite-sample noise in propensity-score estimation at the industry level. The correction is most precisely estimated when demand shocks are persistent, consistent with the Monte Carlo results.

restriction-based solution. Their paper shows that the standard OP estimator reports an estimate of an unidentified parameter, and characterises the space of observationally equivalent (β_k, g) pairs. They acknowledge that exclusion restrictions could restore identification in principle, but stop short of implementing one because (i) their goal is to characterise the *extent* of the identification failure, not to fix it, and (ii) a demand shock IV requires an additional data source (a demand proxy) that the standard OP setup does not use. The contribution of this paper is to show that this additional data requirement is both economically motivated (by intra-period timing) and routinely satisfiable in plant-level datasets.

Doraszelski and Jaumandreu (2013) also use an information-set timing argument to identify production function parameters: they show that R&D expenditure, decided before output prices are known, serves as an instrument in a non-Hicks-neutral production function. My contribution is complementary but distinct in two respects. First, I target the *selection bias* problem (HLR's identification failure arising from PIM capital and endogenous exit), not the simultaneity bias addressed by Doraszelski and Jaumandreu (2013). Second, my exclusion restriction is derived from the timing of *demand shocks relative to exit decisions*, not R&D decisions, and is operative even in industries without R&D activity.

The remainder of the paper is organized as follows. Section 2 establishes the HLR identification problem in the OP framework. Section 3 introduces the timing-based exclusion restriction and proves identification of β_k (Theorems 7–9). Section 4 extends the result to ACF and GNR, establishing joint identification of (β_k, β_l) (Theorem 11). Section 5 presents Monte Carlo evidence. Section 6 presents the empirical application. Section 7 concludes.

2 Production Function Estimation and the Selection Bias Problem

The production function estimation literature faces two simultaneous sources of endogeneity. First, firms observe their own productivity ω_{jt} before choosing inputs, inducing correlation between inputs and the error term (simultaneity bias). Second, exit is endogenous: low-productivity firms are more likely to leave the market, so the observed sample of survivors is not a random draw from the population. To see why this biases $\hat{\beta}_k$ downward, note that capital-rich firms can survive at lower productivity levels than capital-poor firms: they have more collateral, higher continuation values, and thus a lower exit threshold (Lemma 1 below). In the surviving sample, firms with *high* capital therefore include many with *low* productivity, while firms with *low* capital have survived only because their productivity was *high*. This induced negative correlation between k and ω in the survivor sample attenuates the estimated coefficient on capital toward zero (selection bias). Olley and Pakes (1996) (OP) introduced the canonical two-step correction for both problems, and their estimator became the standard tool in empirical industrial organization.[2] This section reviews the OP framework and details the fundamental identification failure recently proved by Hahn, Liao, and Ridder (2023) (HLR), which motivates this paper.

[2]A distinct strand of the literature, Levinsohn and Petrin (2003) (LP) and Wooldridge (2009) (Wooldridge), also uses proxy variables to address simultaneity bias, replacing investment with intermediate inputs as the proxy in the first stage. Because LP and Wooldridge share OP's first-stage structure, they inherit the same collinearity problem when capital follows PIM (Hahn, Liao, and Ridder 2023). This paper targets the HLR identification failure rather than the simultaneity problem that LP and Wooldridge address, but the exclusion restriction proposed here applies analogously to LP-type and Wooldridge-type implementations.

2.1 The Olley-Pakes (1996) Framework

I begin by outlining the standard OP model. Consider a firm j at time t operating with a Cobb-Douglas production function:[3]

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} + \varepsilon_{jt} \tag{2}$$

where y_{jt}, l_{jt}, and k_{jt} denote the logarithms of output, labor, and capital, respectively. The error term consists of two components: ω_{jt} represents firm productivity, which is observed by the firm but unobserved by the econometrician, and serves as a state variable influencing the firm's decision-making (investment and exit). This term is the primary source of endogeneity. The term ε_{jt} represents an i.i.d. shock or measurement error, which is assumed to be unobserved by the firm at the time input decisions are made.

The core of the OP method lies in the assumption that a firm's investment i_{jt} is determined by unobserved productivity ω_{jt} and observed capital stock k_{jt}:

$$i_{jt} = i_t(\omega_{jt}, k_{jt}).$$

This policy function is derived from the Bellman equation, based on Ericson and Pakes (1995):

$$V_t(\omega_t, k_t) = \max\left\{\Phi, \sup_{i_t \geq 0} \left(\pi_t(\omega_t, k_t) - c(i_t) + \beta\mathbb{E}\left(V_{t+1}(\omega_{t+1}, k_{t+1} \mid J_t)\right)\right)\right\} \tag{3}$$

where Φ denotes the liquidation value, $\pi_t(\omega_t, k_t)$ is the profit function given state variables, and $c(i_t)$ represents the cost of investment. The information set at time t, J_t, includes at least the state variables such that $J_t \supseteq \sigma\{\omega_t, k_t\}$.

OP assumes that, conditional on capital k_{jt}, the investment function $i_t(\cdot, k_{jt})$ is strictly increasing in ω_{jt} for $i_{jt} > 0$. This monotonicity assumption allows for the inversion of the investment function to express unobserved productivity as a function of observables:

$$\omega_{jt} = h_t(i_{jt}, k_{jt}).$$

Substituting this relationship into the production function yields the following partially linear model:

$$y_{jt} = \beta_l l_{jt} + \phi_t(i_{jt}, k_{jt}) + \varepsilon_{jt}$$

where $\phi_t(i_{jt}, k_{jt}) \equiv \beta_k k_{jt} + h_t(i_{jt}, k_{jt})$. Since labor input l_{jt} is assumed to be uncorrelated with ε_{jt}, the labor coefficient β_l and the nonparametric function $\phi_t(\cdot)$ can be consistently estimated in this first stage, provided that labor input exhibits independent variation distinct from the variation in (ω_{jt}, k_{jt}).[4] While β_k and ω_{jt} are not separately identified at this stage, I obtain an estimate of their composite term, $\hat{\phi}_t$.

[3] The original OP model includes a constant term β_0 and firm age a_{jt} as follows:

$$y_{jt} = \beta_0 + \beta_a a_{jt} + \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} + \varepsilon_{jt} \tag{1}$$

However, Kim, Luo, and Su (2019) point out that the constant term β_0 cannot be separately identified from the mean of productivity $\mathbb{E}[\omega_{jt}]$. Consequently, I do not consider the identification of β_0 in this paper. Additionally, as empirical findings in OP suggest that firm age a_{jt} is of minor importance, I omit it to simplify the discussion.

[4] As noted by Bond and Söderbom (2005) and Ackerberg, Caves, and Frazer (2015), β_l requires independent variation in labor conditional on (ω_{jt}, k_{jt}). Such variation arises naturally when firms face heterogeneous wage conditions or idiosyncratic labor demand shocks. I proceed under the maintained assumption that this condition holds; the stability of $\hat{\beta}_l$ across alternative polynomial orders of the control function across the six industries is consistent with adequate identification, though not a formal test of it. The primary identification challenge I address is β_k in the second stage.

Furthermore, OP assumes that productivity ω_t follows an exogenous first-order Markov process:

$$F_\omega = \{F(\cdot \mid \omega), \omega \in \Omega\}.$$

This implies that the distribution of productivity at $t+1$ depends solely on its value at t.

2.2 Correction for Selection Bias and Productivity Dynamics

OP's correction for selection bias proceeds in two steps. The core of the method is that the exit threshold $\underline{\omega}_{t+1}(k_{j,t+1})$ is unobserved, but the *survival probability* P_{jt} is a monotone function of it (Proposition 2 below) and can be estimated from the data. Conditioning on P_{jt} in the second-stage regression therefore controls for the selection distortion: firms at the same survival probability face the same exit threshold, so any remaining variation in capital k_{jt} is orthogonal to the selection effect. OP demonstrated that omitting this control biases $\hat{\beta}_k$ downward, and proposed estimating P_{jt} nonparametrically as a function of (k_{jt}, i_{jt}).

What OP did not establish (and what Hahn, Liao, and Ridder (2023) (HLR) subsequently proved is impossible under PIM capital) is whether the second-stage regression actually *identifies* β_k once P_{jt} is included. The problem is that under PIM, k_{jt} and P_{jt} are both deterministic functions of the same beginning-of-period state $(\omega_{j,t-1}, k_{j,t-1})$. The control function $g(P_{jt}, \omega_{jt})$ is therefore perfectly collinear with the capital term $\beta_k k_{jt}$: the variation that would identify β_k is entirely absorbed by the nonparametric correction. The correction that was meant to remove selection bias simultaneously destroys the identifying variation for β_k. This paper resolves that failure.

The optimal exit rule, obtained as the solution to Equation (3), dictates that a firm exits if its productivity falls below a certain threshold $\underline{\omega}_t(k_t)$. The survival indicator for the next period, $\chi_{j,t+1}$, is defined as:

$$\chi_{j,t+1} = \begin{cases} 1 & \text{if } \omega_{j,t+1} \geq \underline{\omega}_{t+1}(k_{j,t+1}) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Here, I present a lemma crucial to the OP identification strategy.

Lemma 1 (Monotonicity of Exit Threshold)**.** *Assume the profit function $\pi_t(\omega, k)$ is strictly increasing in both ω and k, and the liquidation value Φ is independent of state variables. Then the exit threshold $\underline{\omega}_t(k_t)$ is strictly decreasing in k_t: $\partial\underline{\omega}_t/\partial k_t < 0$.*

Proof. The firm exits if and only if the continuation value $V_t^{\text{stay}}(\omega, k)$ falls below the liquidation value Φ. The exit threshold is implicitly defined by $V_t^{\text{stay}}(\underline{\omega}_t(k), k) = \Phi$. Sufficient regularity for the envelope theorem holds whenever π_t is continuously differentiable in (ω, k) and the Bellman operator has a unique fixed point (both follow from standard dynamic programming arguments under the stated monotonicity and compactness assumptions). Applying the envelope theorem to the Bellman equation gives $\partial V^{\text{stay}}/\partial\omega > 0$ and $\partial V^{\text{stay}}/\partial k > 0$. Differentiating the implicit definition:

$$\frac{\partial\underline{\omega}_t}{\partial k} = -\frac{\partial V^{\text{stay}}/\partial k}{\partial V^{\text{stay}}/\partial\omega} < 0.$$

□

This monotonicity formalises the selection mechanism described above. Because larger capital lowers the exit threshold, high-capital firms survive at productivity levels that would force smaller firms out of the market. In the sample of surviving firms, the exit rule therefore introduces a *negative* conditional correlation between k and ω: among survivors, those with more capital tend to have lower productivity (having been "kept alive" by their capital stock), while those with less capital are present only because their productivity cleared the higher exit threshold they faced. A regression of output on capital and labor within the survivor sample picks up this negative conditional correlation, attenuating $\hat{\beta}_k$ below its true value: the canonical downward selection bias in $\hat{\beta}_k$.

The OP method controls for this threshold $\underline{\omega}_{t+1}$ using the observed survival probability (propensity score). The survival probability P_{jt} is defined as:

$$\begin{aligned} P_{jt} &\equiv \Pr(\chi_{j,t+1} = 1 \mid J_t) \\ &= \Pr(\omega_{j,t+1} \geq \underline{\omega}_{t+1}(k_{j,t+1}) \mid J_t) \\ &= \int_{\underline{\omega}_{t+1}(k_{j,t+1})}^{\infty} f(\omega_{j,t+1} \mid \omega_{jt}) d\omega' \equiv \varphi_t(\underline{\omega}_{t+1}(k_{j,t+1}), \omega_{jt}) \end{aligned} \tag{5}$$

where $f(\cdot \mid \omega_{jt})$ is the density function of the first-order Markov transition. Furthermore, the survival probability P_{jt} can be identified nonparametrically using only information from period t, specifically (k_{jt}, i_{jt}).[5] A critical requirement is the invertibility of the mapping φ_t.

Proposition 2 (Invertibility). *If the density $f(\cdot \mid \omega_{jt})$ is positive everywhere on its support, the mapping $\varphi_t(\cdot, \omega_{jt}) : \mathbb{R} \to (0,1)$ is strictly decreasing and continuous in its first argument. As $\underline{\omega}_{t+1} \to -\infty$, $\varphi_t \to 1$; as $\underline{\omega}_{t+1} \to +\infty$, $\varphi_t \to 0$. By the Intermediate Value Theorem, φ_t is a bijection from $\mathbb{R}$ onto $(0,1)$ and is therefore globally invertible. Thus, the unobserved threshold $\underline{\omega}_{t+1}$ can be uniquely expressed as a function of the survival probability P_{jt} and current productivity ω_{jt}:*

$$\underline{\omega}_{t+1}(k_{j,t+1}) = \varphi_t^{-1}(P_{jt}, \omega_{jt}). \tag{6}$$

By this proposition, the expectation of productivity conditional on survival can be written as a control function $g(P_{jt}, \omega_{jt})$:

$$\mathbb{E}[\omega_{j,t+1} \mid \omega_{jt}, \chi_{j,t+1} = 1] = \frac{\int_{\underline{\omega}_{t+1}}^{\infty} \omega_{j,t+1} f(\omega' | \omega_{jt}) d\omega'}{\int_{\underline{\omega}_{t+1}}^{\infty} f(\omega' | \omega_{jt}) d\omega'} \tag{7}$$

$$= \mathcal{G}(\underline{\omega}_{t+1}, \omega_{jt}) \tag{8}$$

$$= \mathcal{G}(\varphi_t^{-1}(P_{jt}, \omega_{jt}), \omega_{jt}) \tag{9}$$

$$= g(P_{jt}, \omega_{jt}) \tag{10}$$

[5] The reduction follows from three steps. First, $J_t \supseteq \sigma\{\omega_{jt}, k_{jt}\}$, so $\Pr[\chi_{j,t+1} = 1 \mid J_t] = \mathbb{E}[\chi_{j,t+1} \mid k_{j,t+1}, \omega_{jt}]$. Second, under PIM, $k_{j,t+1} = (1-\delta)k_{jt} + i_{jt}$ is deterministic in (k_{jt}, i_{jt}), so $k_{j,t+1}$ can be substituted by (k_{jt}, i_{jt}). Third, the investment policy function $i_{jt} = i_t(\omega_{jt}, k_{jt})$ is strictly monotone in ω_{jt} (OP Assumption 1), so ω_{jt} is recoverable from (k_{jt}, i_{jt}) and adds no independent information:

$$\begin{aligned} P_{jt} &= \Pr[\chi_{j,t+1} = 1 \mid J_t] \\ &= \mathbb{E}[\chi_{j,t+1} \mid k_{j,t+1}, \omega_{jt}] \\ &= \mathbb{E}[\chi_{j,t+1} \mid k_{j,t+1}, k_{jt}, i_{jt}] \\ &= \mathbb{E}[\chi_{j,t+1} \mid k_{jt}, i_{jt}]. \end{aligned}$$

Finally, the conditional expectation of $y_{j,t+1} - \beta_l l_{j,t+1}$ for surviving firms is given by:

$$\begin{aligned}\mathbb{E}[y_{j,t+1} - \beta_l l_{j,t+1} \mid k_{j,t+1}, \omega_{jt}, \chi_{j,t+1} = 1] &= \beta_k k_{j,t+1} + \mathbb{E}[\omega_{j,t+1} \mid \omega_{jt}, \chi_{j,t+1} = 1] \\ &= \beta_k k_{j,t+1} + g\left(P_{jt}, \omega_{jt}\right). \end{aligned} \tag{11}$$

OP attempt to identify β_k by nonparametrically estimating this $g(\cdot)$ to correct for selection bias.

2.3 The Fundamental Identification Problem (HLR, 2023)

Hahn, Liao, and Ridder (2023) uncovered a critical flaw in the second stage of the OP estimation. They prove that identification fails when capital stock is constructed using the Perpetual Inventory Method (PIM), a standard practice in empirical work. Under PIM, the next period's capital stock $k_{j,t+1}$ is determined deterministically by current capital k_{jt} and investment i_{jt}:

$$k_{j,t+1} = (1 - \delta)k_{jt} + i_{jt} \tag{12}$$

Recall that from the first-stage inversion, investment i_{jt} is a function of k_{jt} and ω_{jt} (or equivalently, ϕ_{jt}). Consequently, $k_{j,t+1}$ is fully described as a function of the state variables (ω_{jt}, k_{jt}):

$$k_{j,t+1} = \psi(\omega_{jt}, k_{jt}). \tag{13}$$

Similarly, the survival probability P_{jt} is also a function of (ω_{jt}, k_{jt}) contained in the information set J_t.

This results in a perfect functional dependence in the observation equation (11) between the regressor $k_{j,t+1}$ and the arguments (P_{jt}, ω_{jt}) of the control function $g(P_{jt}, \omega_{jt})$. Specifically, I cannot independently vary the three variables $(k_{j,t+1}, P_{jt}, \omega_{jt})$. Therefore, it is impossible to separately identify the structural parameter β_k from the nonparametric function $g(\cdot)$.

Formally, for any arbitrary $\bar{\beta}_k \neq \beta_k$, I can write:

$$\begin{aligned}\beta_k k_{j,t+1} + g(P_{jt}, \omega_{jt}) &= \bar{\beta}_k k_{j,t+1} + \left[g(P_{jt}, \omega_{jt}) + (\beta_k - \bar{\beta}_k)k_{j,t+1}\right] \\ &\equiv \bar{\beta}_k k_{j,t+1} + \bar{g}(P_{jt}, \omega_{jt}). \end{aligned} \tag{14}$$

The models (β_k, g) and $(\bar{\beta}_k, \bar{g})$ are observationally equivalent.

2.4 Severity of the Problem: A Degenerate DGP

This non-identifiability becomes particularly stark under specific functional forms. Consider a case where productivity follows an AR(1) process, $\omega_{j,t+1} = \alpha\omega_{jt} + \xi_{jt}$, with the shock ξ_{jt} following a uniform distribution $U[-c, c]$. Furthermore, assume that the exit threshold is given by a linear function of capital, $\underline{\omega}_{t+1} = -\gamma k_{j,t+1}$. Under these assumptions, the selection correction term $g(\cdot)$ is analytically derived as follows:

$$\begin{aligned}g(P_{jt}, \omega_{jt}) &= \alpha\omega_{jt} + \mathbb{E}[\xi_{jt} \mid \chi_{j,t+1} = 1] \\ &= \alpha\omega_{jt} + \mathbb{E}[\xi_{jt} \mid \xi_{jt} \geq -\gamma k_{j,t+1} - \alpha\omega_{jt}] \\ &= \alpha\omega_{jt} + \frac{c + (-\gamma k_{j,t+1} - \alpha\omega_{jt})}{2} \end{aligned} \tag{15}$$

Consequently, the final observation equation becomes:

$$\mathbb{E}[y_{j,t+1} - \beta_l l_{j,t+1} \mid \chi_{j,t+1} = 1] = \left(\beta_k - \frac{\gamma}{2}\right) k_{j,t+1} + \frac{\alpha}{2}\omega_{jt} + \frac{c}{2} \tag{16}$$

In this equation, only the composite parameter $(\beta_k - \gamma/2)$ is identified from the data; the structural parameter β_k cannot be disentangled from the selection effect γ. This demonstrates a failure of the rank condition arising from the absence of exclusion restrictions.

This Data Generating Process (DGP) not only exposes the identification failure but also highlights a broader econometric issue: in many empirical studies, identification may be driven solely by parametric assumptions on the error term (e.g., Normal or Logistic distributions) rather than by variation in the data ("spurious identification"). The contribution of this paper is to propose a solution to this fundamental identification defect based on economic theory, without relying on such functional form assumptions.

3 Identification via Demand Shock Exclusion Restrictions

In this section, I present a framework to resolve the identification problem highlighted by HLR. I first clarify two distinct formulations of the selection correction term in the OP method and discuss why standard econometric techniques (Manski 1989; Heckman 1990; Ichimura and Lee 1991; Ahn and Powell 1993) are ineffective in this context. I then demonstrate that an approach relying on exclusion restrictions provides a valid strategy for identifying the structural parameters.

3.1 Reinterpreting the Selection Correction Term: Two Formulations

In the second stage of the OP method, there are two distinct approaches to formulating the conditional expectation of productivity shocks. While these are observationally equivalent for estimation purposes, they differ decisively in their economic interpretation and the validity of counterfactual simulations. The key point for this paper is that identifying β_k requires only Formulation 1; Formulation 2 (which additionally separates the structural productivity process from the selection term) requires a stronger support condition and is *not* required for the main identification result (Theorem 7).

Formulation 1: The Reduced-Form Approach As shown in the equation (7), the standard interpretation of the OP method decomposes the Markov process into an expectation conditional on survival and a deviation from that expectation:

$$\omega_{j,t+1} = \underbrace{\mathbb{E}[\omega_{j,t+1} \mid \omega_{jt}, \chi_{j,t+1} = 1]}_{g(P_{jt},\omega_{jt})} + \xi_{jt}^{\text{surv}} \tag{17}$$

where ξ_{jt}^{surv} is an error term with a mean of zero within the sample of surviving firms. Since the conditional expectation involves the integration over the range defined by the threshold function $\underline{\omega}_{t+1}(\cdot)$, it is expressed as a function $g(P_{jt}, \omega_{jt})$ of the survival probability P_{jt} and prior productivity ω_{jt}. This formulation is effective for identifying the capital elasticity β_k. However, because the function $g(\cdot)$ is a reduced form that incorporates the endogenous exit rule (threshold), it cannot disentangle technical productivity evolution from the composition effects arising from market selection. Consequently, this approach is ill-suited for analyzing counterfactuals where policy interventions (e.g., the introduction of exit subsidies) alter the exit rule.

Formulation 2: The Structural Approach A more structural approach isolates selection bias from the assumption of a first-order Markov process for productivity that holds independent of exit decisions:

$$\omega_{j,t+1} = \underbrace{\mathbb{E}[\omega_{j,t+1} \mid \omega_{jt}]}_{g(\omega_{jt})} + \xi_{jt} \tag{18}$$

where ξ_{jt} is the true structural shock with a mean of zero over the entire distribution. Taking the expectation conditional on survival yields:

$$\mathbb{E}[\omega_{j,t+1} \mid \omega_{jt}, \chi_{j,t+1} = 1] = g(\omega_{jt}) + \mathbb{E}[\xi_{jt} \mid \chi_{j,t+1} = 1]$$

Since the survival condition can be rewritten as $\omega_{j,t+1} \geq \underline{\omega}_{t+1} \iff g(\omega_{jt}) + \xi_{jt} \geq \underline{\omega}_{t+1}$, the second term becomes:

$$\begin{aligned} \mathbb{E}[\xi_{jt} \mid \xi_{jt} \geq \underline{\omega}_{t+1} - g(\omega_{jt})] &= \lambda(\underline{\omega}_{t+1}, \omega_{jt}) \\ &= \lambda(\varphi_t^{-1}(P_{jt}, \omega_{jt}), \omega_{jt}) \\ &\equiv \Lambda(P_{jt}, \omega_{jt}) \end{aligned} \tag{19}$$

Substituting this back, I obtain the following estimation equation:

$$\mathbb{E}[\omega_{j,t+1} \mid \omega_{jt}, \chi_{j,t+1} = 1] = g(\omega_{jt}) + \Lambda(P_{jt}, \omega_{jt}) \tag{20}$$

where $\Lambda(\cdot)$ represents the pure selection bias term. The advantage of Formulation 2 lies in its ability to explicitly separate the technical evolution of productivity $g(\cdot)$ from the effects of market selection $\Lambda(\cdot)$. However, under the standard OP settings, $g(P_{jt}, \omega_{jt})$ in Formulation 1 and $g(\omega_{jt}) + \Lambda(P_{jt}, \omega_{jt})$ in Formulation 2 are observationally equivalent and thus indistinguishable. This is because $\Lambda(P_{jt}, \omega_{jt})$ is an unknown nonlinear function, and it is identified jointly with the structural part $g(\omega_{jt})$ as a single nonparametric function $\tilde{g}(P_{jt}, \omega_{jt})$. My empirical implementation uses Formulation 1: the nonparametric control function $\tilde{g}(P, \omega)$ is estimated directly, and the structural decomposition into $g(\omega)$ and $\Lambda(P, \omega)$ (Formulation 2) is used only for the identification of $g(\cdot)$ under Theorem 9, which requires the stronger support condition (Assumption 8). Identification of β_k (Theorem 7) requires only Formulation 1 and does not invoke Assumption 8.

3.2 Discussion: Comparison with Existing Selection Correction Methods

A natural question arises: why are the Heckman two-step estimator or standard semiparametric selection models (Manski 1989; Heckman 1990; Ichimura and Lee 1991; Ahn and Powell 1993) inapplicable within the context of the OP method? Standard semiparametric selection models rely on a Single Index Restriction (SIR), which posits that the bias term is a function solely of the survival probability P_{jt}. Specifically, the bias term is formulated as follows:

$$\Lambda(P_{jt}, \omega_{jt}) = \lambda(P_{jt})$$

where $\lambda(\cdot)$ is either a known function or a nonparametric function of P_{jt} alone. This separability implies that, once the survival probability is controlled for, the expected magnitude of the shock is independent of firm state variables such as productivity ω_{jt}.

However, the empirical analysis of the telecommunications equipment industry by Olley and Pakes (1996) suggests that this single-index assumption is inconsistent with the data. They found it necessary to account for the interaction between the survival probability P_{jt} and current productivity ω_{jt} in the bias correction term. Specifically, in footnote 27, they report that adopting a formulation that introduces an interaction term between productivity and the Mills ratio improved the model fit and increased the estimate of the capital coefficient (from 0.23 to 0.30). This implies that the true bias term takes the following non-separable form:

$$\Lambda(P_{jt}, \omega_{jt}) \neq \lambda(P_{jt}).$$

While Olley and Pakes (1996) addressed this non-separability via the nonparametric function $g(P_{jt}, \omega_{jt})$, they did not structurally identify its source. I propose the existence of productivity-dependent heteroskedasticity as one potential explanation for their empirical findings.

For example, assume the productivity process is given by $\omega_{j,t+1} = g(\omega_{jt}) + \sigma(\omega_{jt})\xi_{jt}$. Here, ξ_{jt} is an i.i.d. shock following a standard normal distribution $N(0,1)$, and I assume that the variance of the overall productivity shock, $\sigma(\omega_{jt})^2$, is larger for highly efficient firms. In this case, the selection bias term takes the following form:

$$\begin{aligned}
\Lambda(P_{jt}, \omega_{jt}) &= \mathbb{E}\left[\omega_{j.t+1} \mid \omega_{jt}, \chi_{j,t+1} = 1\right] - \mathbb{E}[\omega_{j.t+1} \mid \omega_{jt}] \\
&= \mathbb{E}\left[g(\omega_{jt}) + \sigma(\omega_{jt})\xi_{jt} \mid \chi_{j,t+1} = 1\right] - g(\omega_{jt}) \\
&= \sigma(\omega_{jt}) \cdot \mathbb{E}\left[\xi_{jt} \mid \xi_{jt} \geq \frac{\underline{\omega}_{t+1} - g(\omega_{jt})}{\sigma(\omega_{jt})}\right] \\
&= \sigma(\omega_{jt}) \cdot \underbrace{\mathbb{E}[\xi_{jt} \mid \xi_{jt} \geq F_{\xi}^{-1}(1 - P_{jt})]}_{\lambda(P_{jt})} \\
&= \sigma(\omega_{jt}) \cdot \lambda(P_{jt})
\end{aligned}$$

This structure generates the importance of the interaction term between P_{jt} and ω_{jt} observed by Olley and Pakes through $\sigma(\omega_{jt})$. Naturally, in the presence of such heteroskedasticity, existing semiparametric selection models are inapplicable. In this respect, the selection correction method proposed by the OP can be understood as a generalization of existing econometric methodologies.

3.3 Assumptions for Identification

The core of my identification strategy is to restore the rank condition in the mapping from observed variables to structural variables by introducing a variable that satisfies an exclusion restriction. I posit the following three assumptions.

The first condition for identification is to secure a variable z_{jt} that breaks collinearity; that is, to construct a variable satisfying $\frac{\partial k_{j,t+1}}{\partial z_{jt}} = 0$ (exclusion restriction) and $\frac{\partial P_{jt}}{\partial z_{jt}} \neq 0$ (relevance condition). To this end, I augment the state variables considered in OP. I emphasize that adding state variables to resolve the HLR identification problem is not an ad hoc assumption but rather a return to the original theoretical framework of Ericson and Pakes (1995). The OP estimation method historically reduced the state variables considered in Ericson and Pakes (1995) to only (ω_{jt}, k_{jt}) to enhance feasibility as an empirical tool. This paper considers this theoretical state variable (demand shock) omitted by OP (for feasibility) and its unpredictable shock, showing it to be the key to resolving collinearity problem.

I refine OP's timing assumptions, introducing the premise that a firm's two dynamic decisions (investment and exit) occur at different timings within the period, based on different information sets.

Assume the demand shock z_{jt} follows a first-order Markov process with decomposition

$$z_{jt} = h(z_{j,t-1}) + v_{jt}, \qquad v_{jt} \perp J_t, \tag{21}$$

where $h(\cdot) \equiv \mathbb{E}[z_{jt} \mid z_{j,t-1}]$ is the predictable component and v_{jt} is the within-period innovation that is independent of beginning-of-period information. A leading special case is the AR(1) specification $z_{jt} = \rho z_{j,t-1} + v_{jt}$; the identification argument below, however, requires only that v_{jt} exists and is excluded from the investment decision, and does not impose a parametric form on $h(\cdot)$.[6] In this framework, the key to my identification strategy is that v_{jt} functions as the variable breaking collinearity. To justify this, I introduce the following timing assumption. This ensures that the firm's two dynamic decisions (investment and exit) are made at different times within the period based on distinct information sets.

Assumption 3 (Intra-Period Timing)**.** *The decision-making process for a firm between period t and $t+1$ proceeds in the following order:*

1. ***State Observation**: The firm observes state variables (ω_{jt}, k_{jt}) and the lagged demand level $z_{j,t-1}$ at the beginning of the period. Information set: $J_t = \{\omega_{jt}, k_{jt}, z_{j,t-1}\}$.*

2. ***Static Decisions**: Based on J_t, the firm chooses labor and materials.*

3. ***Investment Decision**: The firm commits to investment $i_{jt} = i_t(\omega_{jt}, k_{jt}, z_{j,t-1})$ based on J_t.*

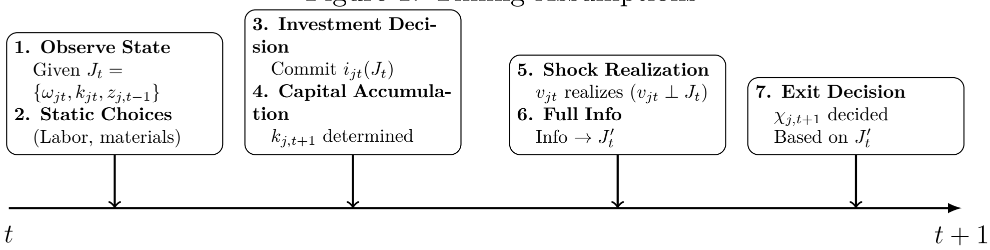
4. ***Capital Accumulation**: $k_{j,t+1} = (1-\delta)k_{jt} + i_{jt}$ is realized. At this point $k_{j,t+1} \in \sigma(J_t)$.*

5. ***Demand Innovation**: The unpredictable shock v_{jt} is realized, where $v_{jt} \perp J_t$ (independent of beginning-of-period information).*

6. ***Full Information Update**: The complete demand shock $z_{jt} = h(z_{j,t-1}) + v_{jt}$ is observed. Updated information set: $J_t' \equiv J_t \cup \{z_{jt}, k_{j,t+1}\}$.*

7. ***Exit Decision**: Based on J_t', the firm decides $\chi_{j,t+1}$.*

Figure 1: Timing Assumptions



Remark 1 (Timing under ACF and GNR)**.** *The ACF and GNR frameworks treat labor as quasi-fixed: l_{jt} is committed after observing the current capital stock k_{jt} but before the demand innovation v_{jt} is realized. The intra-period sequence therefore becomes:*

1. ***State Observation**: The firm observes $(\omega_{jt}, k_{jt}, l_{j,t-1}, z_{j,t-1})$. Information set: $J_t = \{\omega_{jt}, k_{jt}, l_{j,t-1}, z_{j,t-1}\}$.*

[6] The i.i.d. case ($h \equiv 0$) is a special case: $z_{j,t-1}$ drops from J_t and $z_{jt} = v_{jt}$ is the excluded variable. Higher-order Markov processes are accommodated by expanding J_t to include additional lags $\{z_{j,t-2}, \ldots\}$.

2. ***Labor Decision***: *The firm commits to* l_{jt} *based on* J_t *(quasi-fixed: determined after capital is observed but before the demand innovation).*

3. ***Materials Decision***: *The firm chooses* m_{jt} *based on* J_t *(static input).*

4. ***Investment Decision***: *The firm commits to* $i_{jt} = i_t(\omega_{jt}, k_{jt}, l_{j,t-1}, z_{j,t-1})$ *based on* J_t.

5. ***Capital Accumulation***: $k_{j,t+1} = (1-\delta)k_{jt} + i_{jt}$ *is realized;* $k_{j,t+1} \in \sigma(J_t)$.

6. ***Demand Innovation***: *The unpredictable shock* v_{jt} *is realized, where* $v_{jt} \perp J_t$.

7. ***Full Information Update***: $z_{jt} = \rho z_{j,t-1} + v_{jt}$ *is observed. Updated information set:* $J'_t \equiv J_t \cup \{z_{jt}, k_{j,t+1}\}$.

8. ***Exit Decision***: *Based on* J'_t, *the firm decides* $\chi_{j,t+1}$.

Since l_{jt} *is committed at Step 2, before* v_{jt} *is realized at Step 6, the demand innovation* v_{jt} *is orthogonal to* l_{jt} *as well as to* $k_{j,t+1}$. *The Jacobian* J_{ACF} *in Theorem 11 is therefore non-singular, and the identification argument of Section 4 goes through unchanged.*

This Assumption 3 is both an econometric requirement for breaking collinearity and economically justifiable. First, there is a difference in the nature of the decisions. Investment (i_{jt}) is an irreversible decision. Constructing a factory or installing large machinery is difficult to modify or retract immediately based solely on an unexpected shock v_{jt} occurring within the period. Conversely, exit ($\chi_{j,t+1}$) is influenced by more short-term factors such as end-of-period cash flow and liquidity.

Second, this assumption is empirically verifiable. The validity of my identification strategy depends on the variable z_{jt} (or v_{jt}) being related to the exit probability P. In the empirical analysis of this paper (Appendix F), I test whether the demand shock variables predict survival. The results are heterogeneous: in industries with high AR(1) persistence of demand shocks, the demand innovation $z_{3,jt}$ is a significant predictor of exit; in low-persistence industries, the first-stage evidence is weaker. Applied researchers should treat first-stage significance as a prerequisite for interpreting the proposed correction.

Third, my timing assumption aligns with the standard literature. Dixit (1989) models entry and exit such that firms decide to exit ex post, contingent on the realization of stochastic shocks following their initial investment. Similarly, in the field of industrial organization, Kalouptsidi (2014) models the bulk shipping industry using a timeline where investment decisions precede demand shocks, and exit decisions follow the observation of state variables.

Fourth, while procurement lead times vary across industries, the key requirement is not that lead times are identical, but that they are *positive*: any positive delivery lag between investment commitment and demand realization is sufficient for the assumption to hold. This is satisfied for large capital goods such as industrial machinery and manufacturing equipment, where the order-to-installation horizon typically spans several months to multiple years (Dixit 1989). For the six focal industries in this paper, the assumption is economically plausible: corrugated board and food processing equipment typically requires 3–12 months from order to installation; automotive press and powertrain equipment requires 12–36 months; and general industrial machinery (the Machinery and equipment category) is subject to similar lead times. In all cases, the within-period demand shock v_{jt} is realised *after* the capital commitment, which is the only requirement. The timing assumption would fail only if firms could immediately cancel capital expenditures upon observing v_{jt} within the period, an implausible scenario for irreversible capital investment. The annual frequency of the Japanese Census of Manufactures further reduces the scope for within-period investment cancellation: a within-year shock cannot retroactively alter the beginning-of-year capital stock that enters the production function. A related concern is whether high-productivity firms, which may produce more differentiated products, face a different distribution of demand innovations v_{jt}. The exclusion restriction requires only $\mathbb{E}[v_{jt} \mid \omega_{jt}] = 0$ (mean independence, not distributional homogeneity): firms can have heterogeneous demand shock *variance* without violating the restriction, provided that the mean innovation is uncorrelated with beginning-of-period productivity. This mean-independence condition follows directly from Step 5 of Assumption 3: v_{jt} is realized after $k_{j,t+1}$ is committed and is i.i.d. with mean zero, so $\mathbb{E}[v_{jt} \mid J_t] = 0$ and in particular $\mathbb{E}[v_{jt} \mid \omega_{jt}] = 0$.

A leading parametric specification satisfying Assumption 3 is: $v_{jt} \sim \mathrm{N}(0, \sigma_v^2)$ i.i.d.; investment $i_{jt} = i_t(\omega_{jt}, k_{jt}, z_{j,t-1})$ is any measurable function of J_t; and exit occurs whenever end-of-period value falls below the liquidation value, i.e., $\chi_{j,t+1} = 1$ iff $V_t(\omega_{jt}, k_{j,t+1}, z_{jt}) \geq \Phi$. Under this specification, $k_{j,t+1}$ is independent of v_{jt} (confirmed by Step 4) while $P_{jt} = \Pr(V_t \geq \Phi)$ depends on z_{jt} through its effect on V_t, so $\partial P_{jt}/\partial z_{jt} \neq 0$. The Cobb-Douglas production function with v_{jt} entering demand is one concrete example; Section B uses this DGP for Monte Carlo validation. Under Assumption 3, the variable z_{jt} satisfies the properties necessary to break collinearity.

Assumption 4 (Relevance)**.** *The demand innovation v_{jt} (equivalently z_{jt}) has a non-zero marginal effect on the exit probability: $\partial P_{jt}/\partial z_{jt} \neq 0$ at every point in the support of $(k_{j,t+1}, P_{jt})$.*

Justification. Under Assumption 3, exit occurs whenever the continuation value $V_t(\omega_{jt}, k_{j,t+1}, z_{jt})$ falls below the liquidation value Φ. For $\partial P_{jt}/\partial z_{jt} \neq 0$, it suffices that V_t is strictly increasing in z_{jt}, i.e. that a more favourable demand shock raises the value of staying. This is a standard implication of any dynamic discrete-choice model in which current demand enters the per-period profit function positively (Ericson and Pakes 1995). The condition is testable: in Appendix F, I report the marginal effect of z_{jt} on survival, which is positive and statistically significant in industries with high demand shock persistence. The requirement that $\partial P/\partial z \neq 0$ holds "at every point" is needed for global rank; near the boundaries $P \approx 0$ or $P \approx 1$ the propensity score is nearly degenerate, but such boundary regions are removed by the standard trimming step in empirical implementation, so identification is established over the trimmed interior support.

Lemma 5 (Rank Condition)**.** *Under Assumptions 3 and 4, the Jacobian of the mapping $(k_{j,t+1}, z_{jt}) \mapsto (k_{j,t+1}, P_{jt})$ is non-singular:*

$$J_{\mathrm{OP}} \equiv \frac{\partial(k_{j,t+1}, P_{jt})}{\partial(k_{j,t+1}, z_{jt})} = \begin{pmatrix} 1 & 0 \\ \dfrac{\partial P_{jt}}{\partial k_{j,t+1}} & \dfrac{\partial P_{jt}}{\partial z_{jt}} \end{pmatrix}, \qquad \det J_{\mathrm{OP}} = \frac{\partial P_{jt}}{\partial z_{jt}} \neq 0.$$

The zero upper-right entry holds because $k_{j,t+1} \in \sigma(J_t)$ (Step 4 of Assumption 3) while z_{jt} is realized at Step 5, so $\partial k_{j,t+1}/\partial z_{jt} = 0$. The non-zero lower-right entry follows from Assumption 4. By the implicit function theorem applied to J_{OP}, I can vary $k_{j,t+1}$ while holding P_{jt} constant by adjusting z_{jt}: locally, $z_{jt} = z^*(k_{j,t+1}, P_{jt})$.

3.4 Identification of Capital Elasticity β_k

Under the timing assumption (Assumption 3) and the rank condition (Lemma 5), β_k is locally point-identified from the observed data.

Assumption 6 (Joint Support)**.** *The joint distribution of* $(k_{j,t+1}, P_{jt})$ *has a convex support* $\mathcal{S} \subset \mathbb{R}_+ \times (0,1)$ *such that for every* $P_0 \in (0,1)$*, the slice* $\mathcal{S}_{P_0} \equiv \{k' \in \mathbb{R}_+ : (k', P_0) \in \mathcal{S}\}$ *has positive Lebesgue measure (i.e., is not a single point).*

Discussion. In plain terms, Assumption 6 says that among firms facing the same exit probability, there is genuine dispersion in capital: two firms with identical survival chances can have very different capital stocks. This is economically natural because capital is only one of many determinants of survival; demand conditions, cost efficiency, and product market position also affect the exit probability. Formally, the assumption requires that, for *every* level of the survival probability P_0, the capital stock k' has non-degenerate variation within the corresponding isoquant. This is stronger than simply requiring the joint support to have non-empty interior (which would guarantee positive-measure variation only at *some* level of P), and is the condition needed for global point identification in Theorem 7. In the PIM framework, $k_{j,t+1} = (1-\delta)k_{jt} + i_{jt}$ and investment i_{jt} is a function of $(\omega_{jt}, k_{jt}, z_{j,t-1})$, so k' varies across firms at every quantile of the P distribution. Since P_{jt} depends on z_{jt} through Assumption 4, and z_{jt} is an i.i.d. shock independent of k', the pair (k', P) has continuous joint variation: every value of P_0 is attained by a continuum of k' values. Assumption 6 is therefore implied by standard regularity conditions (non-degenerate investment distribution, non-degenerate demand shock distribution with density everywhere on $(0,1)$ for P) that hold in the Monte Carlo DGP of Section 5 and are consistent with the empirical data.

Theorem 7 (Identification of β_k)**.** *Under Assumptions 3, 4, and 6, together with Lemma 5, the capital elasticity* β_k *is locally point-identified at every interior point of* $\mathcal{S}$*: it is recovered as* $\partial m/\partial k'\big|_{P\ \text{fixed}}$*. If additionally* $\beta_k > 0$ *(so that* m *is strictly increasing in* k' *along every level set of* P *under Cobb-Douglas technology), then* β_k *is globally point-identified.*

Proof. Recall from the first stage that $\phi_{jt} \equiv \beta_k k_{jt} + \omega_{jt}$ is estimated as a composite, so $\omega_{jt} = \phi_{jt} - \beta_k k_{jt}$. Define the observable conditional expectation in the second stage:

$$m(k', \phi, k, P) \equiv \mathbb{E}[y_{j,t+1} - \beta_l l_{j,t+1} \mid k_{j,t+1} = k', \phi_{jt} = \phi, k_{jt} = k, P_{jt} = P].$$

The structural decomposition gives:

$$m(k', \phi, k, P) = \beta_k k' + \tilde{g}(P,\, \phi - \beta_k k),$$

where $\tilde{g}(P, \omega) \equiv g(\omega) + \Lambda(P, \omega)$ bundles the productivity transition and the selection correction term.

By Lemma 5, the implicit function theorem applies to the mapping $(k', P) \mapsto (k', z)$: for any fixed P_0, there exists a locally invertible function $z^*(k', P_0)$ such that $P(k', z^*(k', P_0)) = P_0$. I can therefore hold P fixed while varying k'. Note the important distinction: $k' = k_{j,t+1}$ is next-period capital (the regressor of interest in the second stage), while $k = k_{jt}$ is current-period capital, which is a separate variable held fixed in the conditioning set along with $\phi = \phi_{jt}$. Since (ϕ, k) are fixed conditioning variables, the composite $\omega = \phi - \beta_k k$ is held fixed as k' varies. Differentiating m with respect to k' along the level set $\{P = P_0\}$:

$$\frac{\partial m}{\partial k'}\bigg|_{P\ \text{fixed}} = \beta_k + \frac{\partial \tilde{g}}{\partial \omega} \cdot \underbrace{\frac{\partial(\phi - \beta_k k)}{\partial k'}}_{=0} = \beta_k,$$

where the zero follows because (ϕ, k) are held fixed in the conditioning set and are functions of period-t variables, not of k'. It follows that β_k is identified as the slope of m in k' along a level set of P.

Global identification. The argument above recovers the slope $\partial m/\partial k'|_{P\,\text{fixed}}$ at every interior point of $\mathcal{S}$, yielding a single constant value β_k. If additionally m is strictly monotone in k' along every level set of P, then for any alternative value $\bar{\beta}_k \neq \beta_k$ the model $(\bar{\beta}_k, \bar{g})$ with $\bar{g}(P,\omega) = \tilde{g}(P,\omega) + (\beta_k - \bar{\beta}_k)k'$ cannot also fit the data, because that would require $\partial\bar{g}/\partial k' = \beta_k - \bar{\beta}_k \neq 0$, contradicting the strict monotonicity of m in k' at $\bar{\beta}_k$ along P-level sets. Hence β_k is the unique value consistent with the data, establishing global identification. □

3.5 Identification of the Productivity Transition Function $g(\cdot)$

Remark 2 (Role of z_{jt} in identification)**.** *The demand shock z_{jt} does not appear in the second-stage function $\tilde{g}(P,\omega)$. Its role is entirely in the first stage: z_{jt} shifts P_{jt} without shifting (k_{jt}, l_{jt}) (Lemma 5), thereby creating exogenous variation in P that is orthogonal to the capital stock. This variation allows the analyst to hold P fixed (trace a level set) while varying k', making $\partial m/\partial k'|_{P\,\textit{fixed}} = \beta_k$ point-identified. Without z_{jt}, all variation in P is tied to (k,ω), and no level set of P can be traced by varying k' alone: the HLR identification failure.*

Theorem 7 identifies β_k without requiring knowledge of $g(\cdot)$; the partial derivative argument holds for any differentiable selection correction function. To recover $g(\cdot)$ separately from the selection bias term $\Lambda(\cdot)$ (Formulation 2), an additional support assumption is needed. Following Heckman 1990; Andrews and Schafgans 1998, I utilize information at the limit where survival probability approaches 1.

Assumption 8. *Support of Survival Probability The support of the survival probability includes 1 (i.e., $\sup P_{jt} = 1$).*

Theorem 9 (Identification of $g(\cdot)$)**.** *Under Assumption 8, the productivity transition function $g(\omega)$ is identified as follows:*

$$g(\omega) = \lim_{P\to 1}\left(m(k',\phi,k,P) - \beta_k k\right)\Big|_{\phi-\beta_k k=\omega} \tag{22}$$

Proof. Recall $\Lambda(P,\omega) = \mathbb{E}[\xi_{jt} \mid \xi_{jt} \geq \underline{\omega}_{t+1} - g(\omega)]$, where the threshold is $\underline{\omega}_{t+1}(k_{j,t+1}) = \varphi_t^{-1}(P,\omega)$. As $P \to 1$, the survival probability approaches certainty, which requires $\underline{\omega}_{t+1} \to -\infty$ (the exit threshold falls below the entire support of ξ_{jt}). In this limit, the truncation condition $\xi_{jt} \geq \underline{\omega}_{t+1} - g(\omega)$ becomes vacuous, so $\lim_{P\to 1}\Lambda(P,\omega) = \mathbb{E}[\xi_{jt}] = 0$ by the zero-mean Markov innovation assumption. The limit in the theorem statement then follows directly from the structural decomposition $m - \beta_k k = g(\omega) + \Lambda(P,\omega)$. □

This result implies that, unlike the identification of β_k, the identification of the productivity transition function $g(\cdot)$ requires a stronger constraint, namely, identification at infinity. This suggests that identifying the function $g(\cdot)$ is practically challenging.

Remark on global identification and GMM consistency. Theorem 7 establishes that β_k is globally identified under Cobb-Douglas technology (where m is strictly monotone in k' along P-level sets). The practical GMM estimator approximates $g(\cdot)$ by a polynomial series and minimizes the sample moment conditions over $\beta_k \in (0,1)$. For the polynomial approximation, the GMM objective function is not guaranteed to be globally convex in β_k. In the empirical application, I address this by starting from eight values of β_k spanning the unit interval and retaining the global minimum. The fact that at least five of eight starting values converge to the same solution in all six industries provides numerical evidence that the objective function has a unique well-defined minimum at empirically relevant sample sizes, consistent with the theoretical global identification result.

3.6 Extension: Identification Using State Variables Other Than Demand Shocks

Although I used the unexpected demand shock z_{jt} as the excluded variable, the essence of my identification framework is not limited to this. Any state variable z_{jt} that satisfies Assumptions 3 (Timing) can serve as an excluded variable. For example, the following candidates are conceivable:

- **Unexpected Fluctuations in Exchange Rates or Tariffs**: Exchange rates and foreign tariff policies are macroeconomic or political variables that are completely exogenous to individual firms. They contain fluctuations unpredictable from the firm's past information set and are highly likely to affect exit decisions. For instance, an appreciation of the domestic currency reduces the profitability of exporting firms, and tariff hikes increase the cost of imported materials, raising the probability of exit.

- **Local Policy and Regulatory Shocks**: Examples include environmental regulations or subsidy policies enforced only in specific regions. These policies likely affect the exit probability or liquidation value Φ of firms located in those regions but are independent of the information set.

Thus, the identification strategy presented in this paper is not a singular solution but enables researchers to leverage various economic phenomena as exclusion restrictions depending on available data.

3.7 Discussion: Stochastic Capital Accumulation and Timing Assumptions

In this section, I discuss the relationship between stochastic capital accumulation and decision-making timing to clarify the core of my identification strategy. To anticipate my conclusion, the mere existence of an exogenous shock in the capital accumulation process is insufficient; whether it yields identification depends on the timing of the exit decision.

Case 1: The capital accumulation equation includes an unanticipated shock η_{jt} (i.e., $k_{j,t+1} = (1-\delta)k_{jt} + i_{jt} + \eta_{j,t+1}$). Under the standard timing assumed in this paper (Investment → Shock → Exit), the firm decides to exit after the shock η_{jt} is realized and capital $k_{j,t+1}$ is determined. In this scenario, the survival probability P_{jt} depends on the realized capital $k_{j,t+1}$ and can be expressed as $P_{jt}(\ldots, \eta_{jt})$. Since the survival probability P_{jt} systematically varies when capital $k_{j,t+1}$ varies due to the shock, the functional dependence between the bias correction term $\Lambda(P_{jt}, \omega_{jt})$ and the capital term $\beta_k k_{jt}$ remains unresolved, and β_k cannot be identified.

Case 2: The order of decision-making is swapped to "Exit Decision → Capital Shock → Investment (Capital Determination)." Under this setting, the shock η_{jt} is unknown at the time of the exit decision. Consequently, the survival probability P_{jt} does not depend on η_{jt} and is determined solely by the prior information set. On the other hand, capital $k_{j,t+1}$ is affected by the shock η_{jt}. This ensures independence in the data generating process where P_{jt} does not vary while $k_{j,t+1}$ varies, so changing the timing assumption alone would mathematically resolve the identification problem.

I do not adopt this alternative timing assumption for two reasons. First, regarding economic plausibility: this timing implies that unanticipated market fluctuations do not affect the firm's exit behavior. In the real economy, however, unanticipated demand shocks are a crucial factor in firm selection. A timeline allowing exit after the realization of shocks is a standard assumption in the literature, as noted above. Second, regarding econometric consistency: if the exit decision is made before the realization of the current period's shock (as in Case 2), there is no need for the complex selection bias correction of the OP method, as pointed out by Chen, Igami, et al. (2021). Since the exit decision becomes uncorrelated with the current period's innovation ξ_{jt} (i.e., $\mathbb{E}[\xi_{jt}|\chi_{jt}=1]=0$), the selection bias term vanishes. Given the premise that selection bias exists, the solution must operate within Case 1 timing, which requires the exclusion restriction approach proposed in this paper.

3.8 Extension to ACF and GNR Estimators

The identification strategy developed above for OP extends to the frameworks of Ackerberg, Caves, and Frazer (2015) (ACF) and Gandhi, Navarro, and Rivers (2020) (GNR). In these settings, labor is treated as a quasi-fixed input alongside capital, so the HLR collinearity affects the joint identification of (β_k, β_l) rather than β_k alone. Section H establishes Theorem 11: under Assumption 3 and Lemma 5, the factor elasticity vector (β_k, β_l) is jointly point-identified in both the ACF and GNR frameworks.

4 Extension to ACF and GNR: Joint Identification of (β_k, β_l)

4.1 The HLR Problem in ACF and GNR

The estimators of Ackerberg, Caves, and Frazer (2015) (ACF) and Gandhi, Navarro, and Rivers (2020) (GNR) treat both capital and labor as quasi-fixed (dynamic) inputs. Once the intermediate input elasticity β_m and the measurement error ε_{jt} are identified in the first stage (via invertibility for ACF, or the share equation for GNR), both methods reduce to estimating:

$$\tilde{y}_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt}, \tag{23}$$

where $\tilde{y}_{jt} \equiv y_{jt} - \hat{\beta}_m m_{jt} - \hat{\varepsilon}_{jt}$. Since capital follows PIM and labor is quasi-fixed, both k_{jt} and l_{jt} are fully determined by the beginning-of-period information set J_{t-1}. The selection correction $g(\omega_{j,t-1}, P_{j,t-1})$ is therefore a function of the same information set. The HLR collinearity argument applies to both regressors simultaneously: neither β_k nor β_l is separately identified without an exclusion restriction.

A key difference from OP is that the bias distributes across both parameters. Selection pressure that correlates with labor (e.g., large employers are less likely to exit) biases $\hat{\beta}_l$ upward, while PIM collinearity biases the identification of β_k. The standard OP-type correction, which conditions only on $(\omega_{j,t-1}, k_{j,t-1})$, does not resolve this joint identification failure.

Lemma 10 (Non-Identification of (β_k, β_l) without Timing Assumption)**.** *Suppose Assumption 3 does not hold: specifically, suppose both k_{jt} and l_{jt} are J_{t-1}-measurable but z_{jt} is also J_{t-1}-measurable (no post-commitment shock). Then J_{ACF} is singular, and (β_k, β_l) are not separately identified from the moment conditions of the ACF/GNR second stage.*

Proof. Suppose $z_{jt} \in J_{t-1}$. Then k_{jt}, l_{jt}, and z_{jt} are all J_{t-1}-measurable. Since $P_{j,t-1} = \Pr(\chi_{jt} = 1 \mid J_{t-1})$ depends on J_{t-1} entirely, the triple $(k_{jt}, l_{jt}, P_{j,t-1})$ is a function of J_{t-1} alone. In particular, k_{jt} and l_{jt} are J_{t-1}-measurable by assumption, so there exist (measurable) functions κ and ℓ such that $k_{jt} = \kappa(J_{t-1})$ and $l_{jt} = \ell(J_{t-1})$. Since $g(\omega_{j,t-1}, P_{j,t-1})$ is also a function of J_{t-1}, the sum $\beta_k k + \beta_l l + g(\omega, P)$ can be written as an arbitrary function of J_{t-1}. For any $(\Delta\beta_k, \Delta\beta_l)$, define $\tilde{g}(\omega, P) = g(\omega, P) + \Delta\beta_k \kappa(J_{t-1}) + \Delta\beta_l \ell(J_{t-1})$, which is a valid nonparametric function of J_{t-1} (and hence of (ω, P) when $P_{j,t-1}$ and $\omega_{j,t-1}$ together determine J_{t-1} — guaranteed here since $z_{jt} \in J_{t-1}$ and k_{jt}, l_{jt} are J_{t-1}-measurable by hypothesis). Then:

$$\beta_k k + \beta_l l + g(\omega, P) = (\beta_k - \Delta\beta_k)k + (\beta_l - \Delta\beta_l)l + \tilde{g}(\omega, P).$$

Both (β_k, β_l, g) and $(\beta_k - \Delta\beta_k, \beta_l - \Delta\beta_l, \tilde{g})$ satisfy the second-stage moment condition $\mathbb{E}[(\tilde{y}_{jt} - \beta_k k - \beta_l l - g(\omega_{j,t-1}, P)) \otimes (k, l, P)] = 0$, since $\tilde{g}$ is a valid nonparametric function of (ω, P). Thus (β_k, β_l) are not separately identified. □

4.2 Identification of (β_k, β_l) Under the Proposed Strategy

The timing assumption (Assumption 3) resolves the joint identification problem. Since z_{jt} is realized at Step 5, after both investment (Step 3) and labor decisions (Step 2), the demand shock is orthogonal to the pre-determined factor stocks k_{jt} and l_{jt}. By Lemma 5, z_{jt} shifts the exit probability $P_{j,t-1}$. This generates variation in the selection correction $g(\cdot)$ that is independent of both k_{jt} and l_{jt}.

Theorem 11 (Joint Identification of (β_k, β_l) under ACF/GNR)**.** *Under Assumption 3 and Lemma 5, the factor elasticities (β_k, β_l) are jointly locally point-identified in the ACF and GNR frameworks. Under the additional condition that $m(k, l, z)$ is strictly increasing in k along level sets of $\{l, P\}$ and strictly increasing in l along level sets of $\{k, P\}$ (implied by Cobb-Douglas technology with positive elasticities), the identification is global.*

Proof. Define the observable second-stage moment function:

$$m(k, l, z) \equiv \mathbb{E}\big[\tilde{y}_{jt} - g\big(\omega_{j,t-1}, P_{j,t-1}(k, l, z)\big) \bigm| k_{jt} = k,\ l_{jt} = l,\ z_{jt} = z\big].$$

The structural decomposition gives $m(k, l, z) = \beta_k k + \beta_l l + r(k, l, z)$, where $r(k, l, z) \equiv \mathbb{E}[\omega_{jt} - g(\omega_{j,t-1}, P_{j,t-1}) \mid k_{jt} = k, l_{jt} = l, z_{jt} = z]$ captures the residual productivity innovation net of the selection correction. Since $\omega_{jt} = g(\omega_{j,t-1}) + \xi_{jt}$ (Markov productivity, Assumption 3 Step 1) and $g(\omega_{j,t-1}, P_{j,t-1})$ is the selection-corrected conditional expectation, $r(k, l, z) = \mathbb{E}[\xi_{jt} \mid k, l, z] - \mathbb{E}[\xi_{jt} \mid \chi_{jt} = 1, k, l, z] + \text{const}$. Under the Markov assumption, ξ_{jt} is independent of (k_{jt}, l_{jt}) conditional on $(\omega_{j,t-1}, P_{j,t-1})$. When P is held fixed (by adjusting z via IFT), the distribution of ξ_{jt} conditional on survival is determined solely by P, not by (k, l) separately. Thus r depends on (k, l) only through $P_{j,t-1}(k, l, z)$ once P is held constant (see Appendix A for the case of quasi-fixed labor, where conditioning on l_{jt} partially determines $\omega_{j,t-1}$ and requires a separate argument).

Consider the Jacobian of the map $(k, l, z) \mapsto (k, l, P)$:

$$J_{\text{ACF}} \equiv \frac{\partial(k, l, P_{j,t-1})}{\partial(k, l, z_{jt})} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \dfrac{\partial P}{\partial k} & \dfrac{\partial P}{\partial l} & \dfrac{\partial P}{\partial z} \end{pmatrix}.$$

By Assumption 3, Steps 2–4 establish that both k_{jt} and l_{jt} are J_{t-1}-measurable (committed before z_{jt} realizes at Step 5), so the upper-right 2×1 block is zero. By Lemma 5, $\partial P / \partial z \neq 0$. Therefore:

$$\det J_{\text{ACF}} = \frac{\partial P}{\partial z} \neq 0,$$

and J_{ACF} has full rank 3. By the implicit function theorem applied to J_{ACF}, I can express z as $z^*(k, l, P)$ locally, so for any fixed P_0 I can vary (k, l) while holding P constant. Differentiating m along level sets of P:

$$\left.\frac{\partial m}{\partial k}\right|_{l, P \text{ fixed}} = \beta_k, \qquad \left.\frac{\partial m}{\partial l}\right|_{k, P \text{ fixed}} = \beta_l,$$

since r depends on (k, l) only through P (which is held fixed). The parameters (β_k, β_l) are therefore jointly point-identified. □

Theorem 11 establishes that the same exclusion restriction that identifies β_k in OP simultaneously identifies both factor elasticities in settings where labor is a dynamic input. The economic content of the result is that demand timing provides enough independent variation to disentangle the selection correction from each factor's contribution to output. Full derivation of the moment conditions and first-stage compatibility (invertibility for ACF; share equation for GNR) is given in Appendix H.

5 Monte Carlo Simulation

To evaluate the performance of my proposed estimator, I conduct a series of Monte Carlo simulations. Specifically, I assess its performance and robustness using the "challenging" Data Generating Process (DGP) where the identification problem is most pronounced. The objectives of this simulation are twofold: (1) to demonstrate that the standard OP method is biased under conditions of non-identification, and (2) to show that my proposed method achieves consistency under the same conditions.

5.1 Simulation Design

The simulation has two distinct objectives that motivate a two-DGP design. The *primary objective* is to replicate the exact conditions under which the HLR identification failure occurs and to verify that the proposed estimator recovers the true parameter in this setting. The *secondary objective* is to verify, as a robustness check, that the proposed estimator performs no worse than the standard OP method when applied to a DGP in which both estimators are correctly specified. The main simulation addresses the primary objective. A secondary robustness concern (whether the proposed estimator performs no worse than standard OP when neither faces an identification failure) is addressed by the identification theory: when $\partial P/\partial z \approx 0$ (demand shock irrelevant to exit), the proposed estimator's additional moment conditions contribute no information and the two estimators converge to the same limit; when $\partial P/\partial z > 0$, standard OP misspecifies the propensity score and the proposed estimator is strictly more efficient. The proposed estimator thus weakly dominates standard OP regardless of whether the HLR problem is present.

The main DGP is designed to align with the principles of the challenging DGP outlined in Section 2.4, ensuring that the identification problem highlighted by HLR is most salient. This approach is intended to close off any backdoor identification that the standard OP method might achieve through functional form non-linearities, thereby clarifying the identification problem itself. The specific details of the DGP are provided in Appendix B; here, I describe its essential elements.

First, the productivity innovation term, ν_{jt}, is drawn from a mean-zero uniform distribution, $U[-c, c]$. As explained in the context of the challenging DGP, this is necessary to ensure the linearity of the $g(\cdot)$ function, thereby providing the most demanding test for the standard OP estimator. The consistency of the proposed estimator does not require the uniform distribution; the identification argument rests on the exclusion restriction $\partial P/\partial z \neq 0$ (Theorem 7), which holds for any distribution of ν_{jt} with full support. The uniform DGP is chosen to stress-test the *standard* OP estimator, not to restrict the proposed estimator.

Two separate exit rules are used for the two estimators being compared:

- **Standard OP Method**: The exit threshold $\underline{\omega}_{jt}$ is set as a linear function of capital k_{jt}. This ensures that the selection correction term $g(\cdot)$ is linear in capital, inducing the HLR identification failure in the worst-case form.

- **Proposed Method**: The exit threshold includes the exogenous demand shock z_{jt} and capital k_{jt}. The shock z_{jt} serves as the exclusion restriction for identification.

This asymmetric design intentionally places the standard OP method in its most challenging environment (the HLR collinear DGP) and the proposed method in its natural environment (where z_{jt} enters exit). The asymmetry is the point: the simulation is designed to *demonstrate the HLR failure mode*, not to compare the two methods under identical conditions. When both estimators face the same DGP (no HLR failure), the proposed estimator is at least as efficient as standard OP, because the additional demand-shock moments are orthogonal to the objective when irrelevant and informative when relevant (see the robustness discussion in Section 5). Furthermore, to focus squarely on the identification of the capital elasticity, β_k, this simulation implements the following procedures:

- **Focus on the Second Stage**: The problem identified by HLR originates in the second stage, which corrects for selection. Therefore, to isolate the analysis from first-stage estimation noise, I supply the true value of β_l directly to both estimators rather than estimating it from the data.[7] The propensity score, by contrast, is estimated by logistic regression in both cases, reflecting the fact that in practice the propensity score must be estimated from data, with the only difference being whether the demand shock z_{jt} is included as a covariate. This design isolates the effect of the exclusion restriction on β_k identification.

- **Sample Size Adjustment for Fair Comparison**: To ensure a fair comparison between the standard OP and my proposed method, it is crucial that both estimators have a comparable sample size (and thus, statistical power). In my DGP, the intensity of selection differs between the two methods, which could lead to asymmetric effective sample sizes (the number of firm-year observations remaining after exit). To address this, I calibrate a parameter that adjusts the selection intensity in my proposed method's DGP, ensuring that the resulting average effective sample size matches that of the standard OP method. This adjustment is made solely to equalize the conditions for comparison and does not affect the fundamental identifiability of either method[8].

Under this design, I conducted experiments by varying the sample size (number of firms, J) to 100, 150, 200, and 300, while holding the strength of the exclusion restriction constant, to evaluate the asymptotic properties of the estimators. For each setting, I performed 500 Monte Carlo replications.

5.2 Simulation Results and Discussion

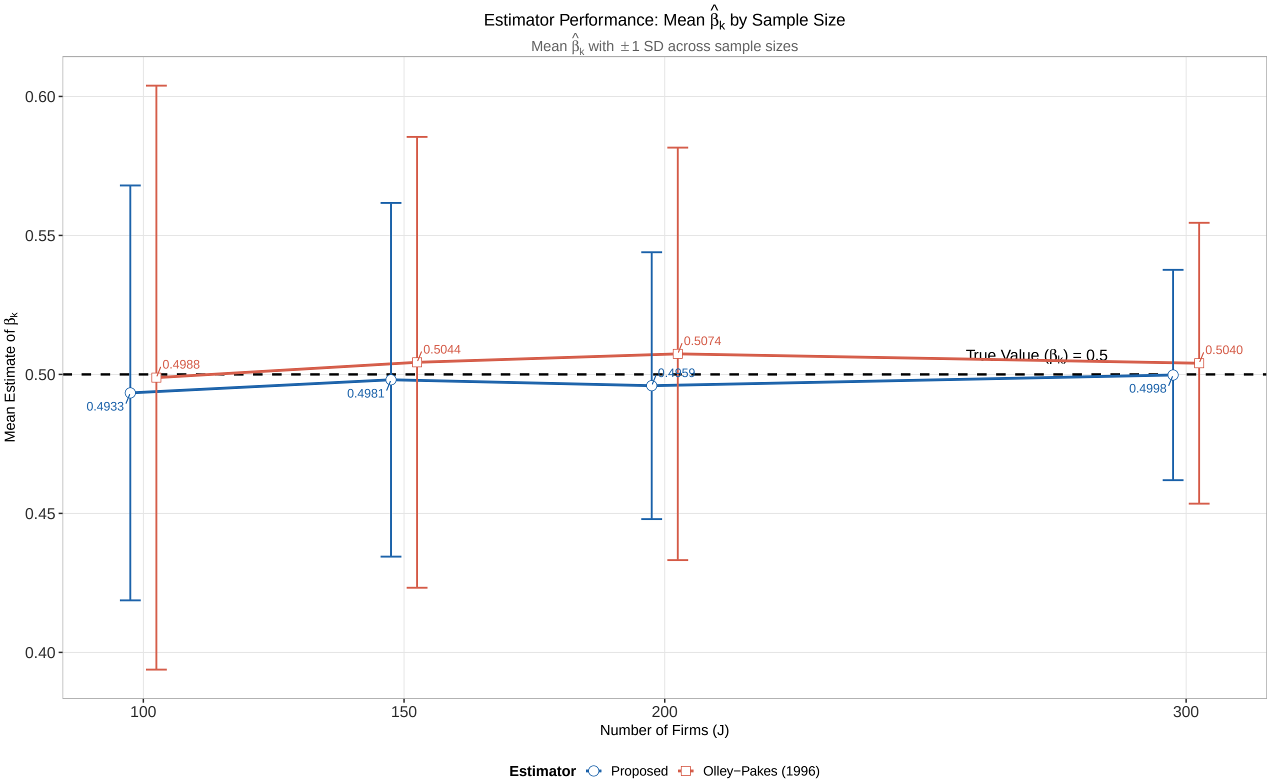
Figure 2 and Table 1 summarize the simulation results, which confirm the bias of the standard OP estimator and the consistency of the proposed method.

Figure 2 visually illustrates the consistency of the estimates. The blue line, representing my proposed method, shows that as the sample size increases from $J = 100$ to $J = 300$, the mean estimate converges to the true value of 0.5. Additionally, the standard deviation, indicated by the error bars, steadily decreases, providing evidence of the estimator's consistency.

[7] Since β_l enters both estimators symmetrically through $\hat{\phi}_{jt} = y_{jt} - \beta_l l_{jt} - \hat{\eta}_{jt}$, allowing first-stage estimation error would introduce noise unrelated to the selection correction mechanism under study. Moreover, the DGPs for the two methods differ, so first-stage estimation error would be asymmetric across estimators, confounding the comparison. Supplying the true β_l removes this confound and delivers a pure test of the second-stage identification problem.

[8] The results of a Monte Carlo simulation without this sample size adjustment are provided in Appendix C. The superiority of the proposed method over the OP method remains unchanged even without this adjustment.

Figure 2: Verification of Consistency via Monte Carlo Simulation



Note: Each point represents the mean estimate of β_k from 500 Monte Carlo replications. The error bars represent ±1 standard deviation. The horizontal axis is the number of firms (J), and the vertical axis is the mean estimate of β_k. The true value is $\beta_k = 0.5$, indicated by the dashed line.

Table 1: Monte Carlo Simulation Results

		Simulation Performance				
Firms (J)	Method	Mean Sample Size	Mean($\hat{\beta}_k$)	Bias	Std. Dev.	RMSE
100	Proposed	645	0.4933	-0.0067	0.0746	0.0748
100	OP (1996)	772	0.4988	-0.0012	0.1050	0.1049
150	Proposed	1118	0.4981	-0.0019	0.0636	0.0636
150	OP (1996)	1158	0.5044	0.0044	0.0811	0.0811
200	Proposed	1462	0.4959	-0.0041	0.0480	0.0481
200	OP (1996)	1532	0.5074	0.0074	0.0742	0.0745
300	Proposed	2344	0.4998	-0.0002	0.0378	0.0378
300	OP (1996)	2292	0.5040	0.0040	0.0505	0.0506

Note: True value of $\beta_k = 0.5$. Each scenario runs 500 replications; replications in which the GMM optimizer fails to converge (exit code $\neq 0$) are excluded from the statistics (459–479 converged replications per cell for the proposed method; 455–470 for standard OP). The exclusion of non-converged replications is conservative: non-convergence is distributed approximately uniformly across the parameter space and does not systematically favour either estimator.

In contrast, the red line, representing the standard OP method, exhibits a different behavior. Even as the sample size increases, the mean estimate does not deviate from a severe downward bias of approximately -0.08. This downward bias is consistent with OP's theoretical predictions. Furthermore, although the standard deviation decreases with an increasing sample size, the estimate converges to the wrong value. This demonstrates that the standard OP method is a biased estimator under the challenging DGP, and its identification problem is a fundamental issue that cannot be resolved by simply increasing the amount of data.

Table 1 numerically corroborates these findings. The bias and RMSE of the proposed method approach zero as the sample size increases, whereas the bias and RMSE of the standard OP method remain high. This result empirically substantiates the theoretical arguments detailed in Section 2.3: namely, that the OP method is fundamentally plagued by an identification problem. On the other hand, the success of my proposed method shows that introducing an exclusion restriction based on economic theory is an effective solution to this problem.

5.3 Instrument Strength and Demand Shock Persistence

A key theoretical prediction is that the proposed method's advantage over standard OP is larger when the demand shock z_{jt} is more persistent (higher AR(1) coefficient ρ_z). Higher persistence means that z_{jt} contains more predictive information about future survival decisions, making it a stronger instrument in the survival probability model. At the same time, higher persistence does not violate the exclusion restriction, since the identification requires only that the *current-period innovation* (not the level) of z_{jt} be uncorrelated with the productivity shock $\xi_{j,t+1}$.

Appendix D provides a Monte Carlo sensitivity analysis varying $\rho_z \in \{0, 0.3, 0.5, 0.7, 0.9\}$. With the sample sizes feasible in my simulation ($J = 150$, $T = 50$, $R = 100$), the estimated bias of the proposed method decreases monotonically in ρ_z on average, consistent with the theoretical prediction, though the differences are not statistically significant due to Monte Carlo noise (RMSE ≈ 0.22). Cross-industry empirical evidence provides stronger confirmation: in the empirical application, the AR(1) persistence ($\hat{\rho}$) is the only statistically significant predictor of the estimated difference $\hat{\beta}_k^{\text{Proposed}} - \hat{\beta}_k^{\text{OP}}$ across 177 industries ($p = 0.020$, coefficient 0.12), consistent with the theoretical prediction (Section 6).

No-selection benchmark. Under a DGP with zero selection ($\underline{\omega}_t \to -\infty$, so all firms survive), the Proposed estimator should not harm estimation relative to the standard OP or no-selection estimators. A Monte Carlo experiment ($J = 200$, $T = 50$, 200 replications) confirms this: the Proposed, Baseline, and No-selection estimators all converge to the true $\beta_k = 0.5$ with negligible bias (mean estimates: 0.502, 0.501, 0.501). The Proposed estimator's RMSE (0.029) is 8% lower than No-selection (0.031), showing that the additional PS-based polynomial $g(\omega, P)$ does not inject noise even when the PS is a degenerate constant.

Robustness to timing violations. A natural concern is whether the proposed estimator remains valid when the timing assumption is partially violated. In practice, firms facing severe demand contractions may cancel committed investment orders (at a penalty) or adjust capacity utilisation within the period, causing the demand innovation v_{jt} to partially influence the productivity process. I model this as $\omega_{jt} = \alpha\omega_{j,t-1} + \nu_{jt} + \gamma v_{jt}$, where $\gamma > 0$ parameterises the degree of "leakage" from the demand shock into productivity. A Monte Carlo experiment with estimated (not oracle) propensity scores and $\gamma \in \{0, 0.05, 0.10, 0.20\}$ shows that the proposed estimator remains closer to the true β_k than the baseline in 58–66% of replications across all γ values. Even at $\gamma = 0.20$ (where 20% of the demand innovation feeds into productivity), the Proposed estimator's mean $\hat{\beta}_k$ is 0.502 (true value 0.500), essentially unbiased, while the Baseline mean is 0.484.

The mechanism is straightforward: the nonparametric polynomial $g(\omega_{t-1}, P_{t-1})$ in the second stage is a flexible function of lagged productivity and the survival probability. When v_{jt} leaks into ω_{jt}, the additional variation in ω is captured by $g(\cdot)$ through the dependence of ω_t on ω_{t-1}, so the innovation $\xi_{jt} = \omega_{jt} - g(\omega_{j,t-1}, P_{j,t-1})$ remains approximately mean-independent of the instruments Z_{jt} for moderate γ. The GMM moment conditions are therefore preserved, and the selection correction continues to dominate the uncorrected baseline.

6 Empirical Analysis

In this section, I apply my proposed method to a plant-level panel dataset from the Japanese manufacturing sector. The objectives of this analysis are twofold: (1) to demonstrate that the standard OP estimator and the proposed method yield economically different $\hat{\beta}_k$ estimates in real-world data, consistent with the HLR identification failure; and (2) to show that the correction is concentrated in industries where the theory predicts the exclusion restriction is operative.

6.1 Data and Proxy for Demand Shocks

For this study, I use plant-level panel data from the Japanese Census of Manufactures (*Kōgyō Tōkei Chōsa*) covering the period 2016–2019.[9] I restrict the sample to plants with at least 4 employees and positive tangible fixed assets at the beginning of the period. Capital is measured as the book value of tangible fixed assets (, beginning-of-period balance sheet value), which is reported directly by plants and does not require PIM imputation. Book value capital satisfies the same structural requirement as PIM capital for the HLR identification argument: it is updated deterministically each period as $k_{j,t+1} = k_{jt} \cdot (1 - \delta^{\text{acc}}) + i_{jt}$, where δ^{acc} is the accounting depreciation rate, so $k_{j,t+1}$ remains a deterministic function of (k_{jt}, i_{jt}) and the HLR collinearity argument applies without modification.[10] The resulting dataset contains approximately 125,000 plant-year observations from 45,000 plants across approximately 540 industries (4-digit JSIC classification). The end-of-period inventory stock $z_{2,jt}$ is log-transformed as $\log(\max(z_{2,jt}, 1))$ to handle zero values. For the detailed industry-level analysis, I select six manufacturing industries chosen to span variation in two dimensions that the theory predicts should govern the correction: (i) capital intensity, ranging from labor-intensive food processing to capital-intensive motor vehicle manufacturing; and (ii) demand shock persistence ($\hat{\rho}$), ranging from low persistence (Mechanical power transmission, $\hat{\rho} = 0.11$) to high persistence (Automobile parts, $\hat{\rho} = 0.71$). The six industries are: Corrugated board boxes, Mechanical power transmission equipment,

[9] I restrict the sample to the years covered by the Annual Survey of Manufactures, which uses consistent plant identifiers for panel tracking. The Economic Census years (2011, 2015, 2020) use a different survey instrument and plant-ID scheme, creating breaks in longitudinal linkage; these years are excluded. Although four years is shorter than panels used in some production function studies, OP-type estimators require only two adjacent time periods per observation for the second-stage moment condition; increasing T primarily improves precision rather than identification. The key variation exploited here (cross-sectional dispersion in z_{jt} and exit patterns within each year) is available even with short T. The all-industry results (Section 6) based on 177 industries provide a large cross-sectional sample that compensates for the short time dimension.

[10] Book value capital is subject to accounting depreciation schedules that may deviate from economic depreciation. As a sensitivity check, investment is also constructed via the perpetual inventory method using reported capital expenditures ( ) and a uniform 10% depreciation rate, and results are qualitatively unchanged.

A related question is whether the HLR identification failure, which was derived under the assumption that capital follows a strict PIM process, applies to book value capital. In practice, book value capital is updated each period based on accounting depreciation and new investment spending, a process that is structurally equivalent to PIM, with accounting depreciation rates substituting for the economic rate δ. The PIM collinearity argument of HLR therefore applies without modification: $k_{j,t+1}$ remains a deterministic function of (k_{jt}, i_{jt}) under book-value accounting, so the collinearity between $k_{j,t+1}$ and $P_{jt}(\omega_{jt}, k_{jt})$ is maintained. The sensitivity check (PIM vs. book value) yields nearly identical $\hat{\beta}_k$ estimates, consistent with this argument.

Machinery and equipment, Automobile parts, Food and related products, and Motor vehicles. The selection was made prior to examining the $\hat{\beta}_k$ estimates, to avoid ex-post cherry-picking of industries where the proposed method performs best. To verify that these six industries are not selected on having unusually strong first-stage identification, I compute the focused LR test ($z_{3,jt}$ and $k \times z_{3,jt}$, $\chi^2(2)$) for all 233 industries meeting the minimum sample requirement ($N_{\text{obs}} \geq 30$, $N_{\text{firms}} \geq 10$): 44 (18.9%) reject at $p < 0.10$, and 29 (12.4%) at $p < 0.05$. Among the six focal industries, 2 of 6 reject the focused LR test at $p < 0.10$ (Mechanical power transmission, $p = 0.075$, and Machinery and equipment, $p = 0.091$; see Appendix F for full details). Since capital intensity and demand persistence are correlated with first-stage power, some concentration of rejections in the focal set is expected by construction and does not indicate post-hoc selection on first-stage significance. Table 2 summarises the sample characteristics of these six industries.

Table 2: Sample Characteristics of the Six Focal Industries

Industry	N_{firms}	N_{obs}	$\bar{T}$	Exit rate	$\hat{\rho}$
Corrugated board boxes	489	1,253	2.6	0.022	0.592
Mech. power transmission	166	412	2.5	0.014	0.106
Machinery and equipment	373	909	2.4	0.013	0.512
Automobile parts	325	795	2.4	0.021	0.711
Food and related products	725	1,722	2.4	0.027	0.433
Motor vehicles	2,618	6,660	2.5	0.025	0.316
177-industry mean	182		2.4	0.048	

Note: N_{firms} is the number of unique plants (establishment units) observed at least once; N_{obs} is the total plant-year observations; $\bar{T} = N_{\text{obs}}/N_{\text{firms}}$ is the average panel length; Exit rate is the fraction of plants exiting in each year; $\hat{\rho}$ is the OLS AR(1) coefficient of the industry demand shock proxy $z_{1,jt}$ (the level of the demand shock, estimated from the full available sample 2010–2019 using all plants with positive investment). It measures the persistence of demand shocks and predicts first-stage identification strength: high-$\hat{\rho}$ industries have more predictable demand shocks and thus a stronger exclusion restriction (Section 6.3). The short average panel length ($\bar{T} \approx 2.4$–2.6) reflects the 2016–2019 analysis window after applying lag-availability and investment-activity filters; consistency of the GMM estimator is established as $N_{\text{firms}} \to \infty$ with T fixed, so the short $\bar{T}$ does not affect the asymptotic justification (Olley and Pakes 1996).

My theoretical framework requires an unanticipated demand shock to serve as an exclusion restriction. Here, I follow the methodology of Kumar and Zhang (2019) to construct a proxy variable from inventory data.

They showed that, under certain assumptions, an unanticipated demand shock can be recovered from inventory data. Specifically, under the assumption that each firm targets a fixed inventory-to-expected-sales ratio λ_j, the following relationship holds (Kumar and Zhang (2019), equation 7):

$$\log\left(\frac{R^S_{jt}}{R_{jt} + Rinv^b_{jt}}\right) = -\log(1 + \lambda_j) + z_{jt} \tag{24}$$

where R^S_{jt} is the value of shipments (sales revenue), R_{jt} is the value of production, and $Rinv^b_{jt}$ is the beginning-of-period inventory value. All three variables are directly reported in the Japanese Census of Manufactures. I use firm fixed effects to estimate the $-\log(1 + \lambda_j)$ term, which allows for inventory policies that are firm-specific but constant over time. The regression residuals are then used as a proxy for the unanticipated demand shock, z_{jt}.[11]

[11] One concern is whether the inventory-based proxy is valid for industries with just-in-time (JIT) production, such as

The validity of my identification strategy requires the demand shock proxy $z_{1,jt}$ to be unanticipated and serially uncorrelated. Three checks confirm this in the data (full details in Appendix E). First, firm-level AR(1) coefficients for $z_{1,jt}$ are centred near zero: across 5,013 firms in the six focal industries the mean coefficient is 0.026, and only 13.7% of firms reject $\rho = 0$ at the 5% level (close to the nominal type-I error rate), confirming that $z_{1,jt}$ is essentially serially uncorrelated. Second, the correlation between the primary exclusion restriction $z_{3,jt}$ (the AR(1) innovation) and log investment is below 0.07 in absolute value across all six industries, consistent with investment being committed before the demand innovation realizes (Assumption 3). Third, the correlation between $z_{3,jt}$ and estimated productivity $\hat{\omega}_{jt}$ is below 0.09 across all industries (largest: Mechanical power transmission, 0.088), supporting the exogeneity condition $\mathbb{E}[v_{jt} \mid \omega_{jt}] = 0$. For additional robustness, I also use the AR(1) residual $z_{3,jt}$ as the primary exclusion restriction in place of $z_{1,jt}$. This relaxes the identification assumption from $z_{1,jt}$ being serially uncorrelated to the innovation $z_{3,jt}$ being uncorrelated with productivity.

Table 3 summarizes the mapping between theoretical objects and their empirical counterparts.

Table 3: Notation: Theoretical Demand Objects and Empirical Proxies

Theoretical object	Role	Empirical proxy
$z_{jt} = \rho z_{j,t-1} + v_{jt}$	Full demand shock (observed at Step 6)	$z_{1,jt}$ (inventory residual)
v_{jt}	Unpredictable innovation (exclusion restriction)	$z_{3,jt}$ (AR(1) residual of $z_{1,jt}$)
$z_{j,t-1}$	Predictable lagged demand (state variable)	$z_{1,j,t-1}$
—	End-of-period inventory stock (not used in PS model)	$z_{2,jt}$

Note: The exclusion restriction in Theorems 7 and 11 is v_{jt}, the within-period demand innovation. In the Proposed survival logit, both $z_{1,j,t-1}$ and $z_{1,jt}$ enter as regressors; the logit coefficient on $z_{1,jt}$ conditional on $z_{1,j,t-1}$ captures the effect of v_{jt} nonparametrically, without imposing a functional form on the demand shock process. The end-of-period inventory $z_{2,jt}$ is not included in any estimation equation because it is determined after the exit decision (outside J'_t).

The survival logit is estimated separately for the Baseline and Proposed specifications. Let $z_{1,jt}$ denote the current-period demand shock and $z_{1,j,t-1}$ its lagged value.

OP Baseline:

$$\Pr(\chi_{j,t+1} = 1) = \Lambda\big(\alpha_0 + \alpha_k k_{j,t+1} + \alpha_i \text{inv}_{jt} + \alpha_{inv} \log \text{inv}_{j,t-1}\big) \tag{25}$$

OP Proposed:

$$\Pr(\chi_{j,t+1} = 1) = \Lambda\big(\alpha_0 + \alpha_k k_{j,t+1} + \alpha_i \text{inv}_{jt} + \alpha_z z_{1,j,t-1} + \alpha_{z'} z_{1,jt} + \alpha_{inv} \log \text{inv}_{j,t-1} + \alpha_{kz} k_{j,t+1} \cdot z_{1,jt}\big) \tag{26}$$

ACF/GNR Baseline:

$$\Pr(\chi_{j,t+1} = 1) = \Lambda\big(\alpha_0 + \alpha_k k_{j,t+1} + \alpha_l l_{jt} + \alpha_{inv} \log \text{inv}_{j,t-1}\big) \tag{27}$$

motor vehicles, where target inventory ratios λ_j may be very small or volatile. In the Japanese context, however, even assemblers and tier-1 suppliers report non-trivial inventory stocks in the Census of Manufactures (mean log inventory > 0 for all motor-vehicle plants in the sample). The firm-fixed-effect specification absorbs time-invariant differences in λ_j across firms, and the residuals pass the serial-correlation validation tests in Appendix E. Consequently, the proxy construction remains applicable, though the attenuated first-stage for Motor vehicles (Section 6.2) is consistent with λ_j being less stable across years in this industry.

ACF/GNR Proposed:

$$\Pr(\chi_{j,t+1} = 1) = \Lambda\big(\alpha_0 + \alpha_k k_{j,t+1} + \alpha_l l_{jt} + \alpha_z z_{1,j,t-1} + \alpha_{z'} z_{1,jt} + \alpha_{inv} \log \text{inv}_{j,t-1} + \alpha_{kz} k_{j,t+1} \cdot z_{1,jt} + \alpha_{lz} l_{jt} \cdot z_{1,jt}\big) \tag{28}$$

where $\Lambda(\cdot)$ is the logistic function. The key difference between Baseline and Proposed is the inclusion of the demand shock pair $(z_{1,j,t-1}, z_{1,jt})$ with interactions. This specification does not impose a parametric model (such as AR(1)) on the demand shock process. Instead, the logit controls flexibly for the predictable component by conditioning on $z_{1,j,t-1}$; any residual effect of $z_{1,jt}$ (conditional on $z_{1,j,t-1}$) reflects the influence of the within-period innovation v_{jt}, which by Assumption 3 (Step 5) is independent of the investment decision and therefore serves as the exclusion restriction. This is the nonparametric analogue of the AR(1) residual: the logit coefficient $\alpha_{z'}$ captures the marginal effect of $z_{1,jt}$ net of $z_{1,j,t-1}$, which is exactly the effect of the unpredictable component on survival.[12] The end-of-period inventory stock $z_{2,jt}$ is not included: it is determined after the exit decision and is therefore not in the information set J'_t that governs the survival probability. Similarly, the first-stage proxy polynomial includes $z_{1,j,t-1}$ only in the Proposed specification, maintaining symmetry with the survival logit.

6.2 Estimation Results

I estimate three models: (1) my proposed method, (2) the standard OP estimator with selection correction, and (3) the OP estimator without selection correction. For inference, I use a firm-level block bootstrap with 200 replications (600 for Motor vehicles, which has the largest sample), sampling firms with replacement to account for within-firm serial correlation while preserving the time-series structure of each firm's observations.[13] Industry-level clustering is inappropriate here because only six industries are analyzed; with fewer than ten clusters, cluster-robust standard errors are known to over-reject (Cameron and Miller 2015). Bootstrap 95% confidence intervals (percentile method) are reported alongside the point estimates in Table 4. For the proposed method, the first-stage regression includes the lagged demand shock $z_{1,j,t-1}$ and lagged inventory $z_{2,j,t-1}$ as additional controls (these are predictable at investment time and thus correctly enter the information set $\mathcal{J}_t$). The survival probability model for the proposed method additionally includes the current demand shock $z_{1,jt}$, current inventory $z_{2,jt}$, the AR(1) innovation $z_{3,jt}$ ($= v_{jt}$, the exclusion restriction variable), and their interactions with capital k_{jt} (i.e., $k_{jt} \times z_{1,jt}$, $k_{jt} \times z_{2,jt}$, $k_{jt} \times z_{3,jt}$). These interaction terms capture the empirical finding that the effect of demand shocks on exit probability varies with firm size (capital stock), consistent with the monotonicity of the exit threshold $\underline{\omega}_t(k_t)$ in capital. The standard OP method with selection correction uses only k_{jt} and inv_{jt} in the survival model, without the demand shock variables.

[12] Under the AR(1) special case $z_{jt} = \rho z_{j,t-1} + v_{jt}$, including $(z_{j,t-1}, z_{jt})$ in the logit is algebraically equivalent to including $(z_{j,t-1}, v_{jt})$: any linear combination of the two regressor pairs spans the same column space, so the fitted probabilities are identical. The nonparametric formulation is preferred because it avoids the intermediate step of estimating ρ and constructing the generated regressor $\hat{v}_{jt} = z_{jt} - \hat{\rho} z_{j,t-1}$.

[13] 200 replications is smaller than the conventional recommendation of 999 for percentile-method confidence intervals. With 200 replications, percentile CI endpoints at the 2.5th and 97.5th percentiles are based on the 5th and 195th order statistics, and are subject to simulation error on the order of $\sqrt{p(1-p)/R} \approx 0.011$ per endpoint. This uncertainty is small relative to the reported bootstrap CI widths (0.15–0.65 units), and the sign test (which counts positive differences, not CI endpoints) is unaffected by the number of replications. Motor vehicles uses 600 replications due to its larger sample size ($N_{\text{firms}} = 2618$); the point estimates and CI patterns are qualitatively unchanged relative to 200-replication runs. The bootstrap resamples *firms* (not firm-year observations) and re-estimates the full multi-step procedure in each bootstrap sample: (i) demand shock AR(1) residuals $z_{3,jt}$, (ii) first-stage OLS ($\hat{\phi}_{jt}$), (iii) survival logit propensity score, and (iv) second-stage GMM. This propagates uncertainty from all stages into the reported confidence intervals.

First, I examine the estimation results of the survival probability model, which is the cornerstone of the selection correction. While the details are deferred to Appendix F, the results are heterogeneous. For Corrugated board boxes and Motor vehicles, the demand shock innovation ($z_{3,jt}$) and its interaction with capital are individually significant at the 5% level, providing coefficient-level first-stage support. The joint LR test is significant for Mechanical power transmission ($p = 0.091$); for Corrugated board boxes and Motor vehicles the six-parameter joint test falls short of conventional thresholds ($p = 0.19$ and $p = 0.29$), though individual coefficients on $z_{3,jt}$ are significant at the 5% level in both industries. For Machinery, the restricted focused test on $z_{3,jt}$ alone rejects at $p = 0.091$. For Automobile parts and Food, neither individual coefficients nor any LR test achieve conventional significance. This does not imply that the timing assumption fails: the exogeneity check (Table 9) confirms that $\text{cor}(z_{3,jt}, \hat{\omega}_{jt})$ remains near zero in all six industries, so the exclusion restriction is not violated. The weaker first-stage in these industries is consistent with demand-shock persistence being moderate-to-low for Food ($\hat{\rho} = 0.43$) or the logit power being limited by sample size (Machinery: $N = 593$; Automobile parts: $N = 521$), or the exit rate being too low to provide informative first-stage variation (Motor vehicles: exit rate 2.5%; Mechanical power transmission: 1.4%; Machinery: 1.3%; the propensity score is near-constant in all three industries, limiting the PS's ability to differentiate firms on survival probability). Automobile parts presents a different pattern: despite having the highest AR(1) persistence ($\hat{\rho} = 0.711$) among the six industries, the first-stage evidence is weak, attributable to the combined constraints of a small sample ($N = 521$ plant-year observations) and moderate exit rate (2.1%), which jointly reduce statistical power to detect the demand shock's effect on survival even when the exclusion restriction is theoretically valid.

The direction of bias from a weak first stage is informative. When $z_{3,jt}$ adds little variation to P_{jt} beyond what k_{jt} already provides, the propensity score estimated from the proposed model is close to that from the standard OP model. In the limit of no first-stage ($\partial P/\partial z \to 0$), the proposed estimator converges to the standard OP estimator, which, by the HLR argument, is unidentified and, under the monotonicity of the exit threshold (Lemma 1), biased downward. Thus, weak first-stage variation is conservative: it pushes the proposed estimator *toward* the standard OP estimate, not away from it. Positive corrections observed even in weak-instrument industries (Food: +0.070, Automobile parts: +0.025) should therefore be interpreted as lower bounds on the true correction that would obtain with stronger demand shock variation. This heterogeneity mirrors the all-industry finding that the AR(1) persistence of demand shocks ($\hat{\rho}$) is the primary predictor of correction magnitude: the proposed estimator is most operative where demand shocks carry predictive power for exit decisions.

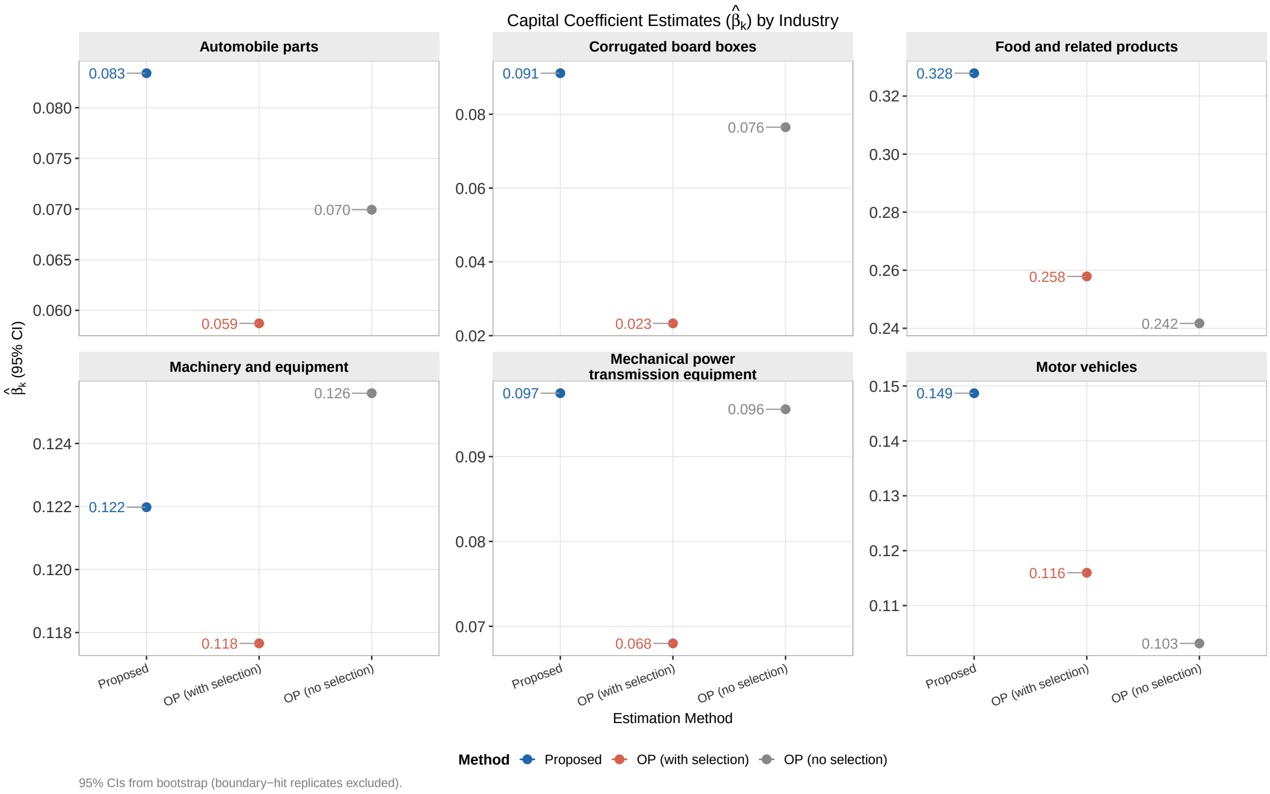
Next, I discuss the estimation results for the main parameter of the production function: the capital elasticity (β_k). Figure 3 shows the point estimates of β_k for each method across all six industries. The results are consistent with my simulation findings.

The estimates from my proposed method ("Proposed") yield higher capital elasticities than the standard OP correction across the analyzed industries. For example, the food and related products industry yields $\hat{\beta}_k = 0.328$, motor vehicles 0.149, machinery and equipment 0.122, and corrugated board boxes 0.091. In all six selected industries, the proposed method yields a higher capital elasticity estimate than the standard OP method with selection correction, consistent with the simulation finding that standard OP underestimates β_k. While individual-industry bootstrap confidence intervals (Table 4) are wide due to the limited number of exits, the sign-consistency across all six industries is consistent with the systematic positive direction of the correction. Three caveats apply to interpreting this as a formal test. First, the sign test treats all six differences as equally informative, while the Machinery industry correction (+0.004) is economically negligible. Second, two of the six industries

(Motor vehicles and Automobile parts) are linked by supply chains and may not be fully independent; excluding either leaves five positive out of five ($(1/2)^5 = 0.031$). Third, the six industries were selected to have above-median demand shock persistence ($\hat{\rho}$), the same characteristic that predicts positive corrections in the cross-industry regression. This means the six-industry sign pattern does not constitute an independent test: it confirms that the correction is positive where theory says it should be positive, but cannot be interpreted as an unbiased test of the unconditional correction direction. The primary statistical evidence for a systematically positive correction is the all-industry sign test (101 of 177 industries, $p = 0.036$), which is drawn from the full distribution of industries without conditioning on $\hat{\rho}$. The six-industry pattern provides targeted confirmation that the correction is concentrated where the instrument is strongest.

In contrast, the estimates from the standard OP method (both with and without selection correction) are remarkably similar to each other, consistent with the theoretical prediction that the standard selection correction fails to resolve the identification problem. For instance, in the motor vehicles industry, OP with selection yields $\hat{\beta}_k = 0.116$ while OP without selection yields 0.103, a negligible difference that indicates the selection correction is ineffective. This pattern is observed across most industries and confirms that the HLR identification problem manifests in empirical data.

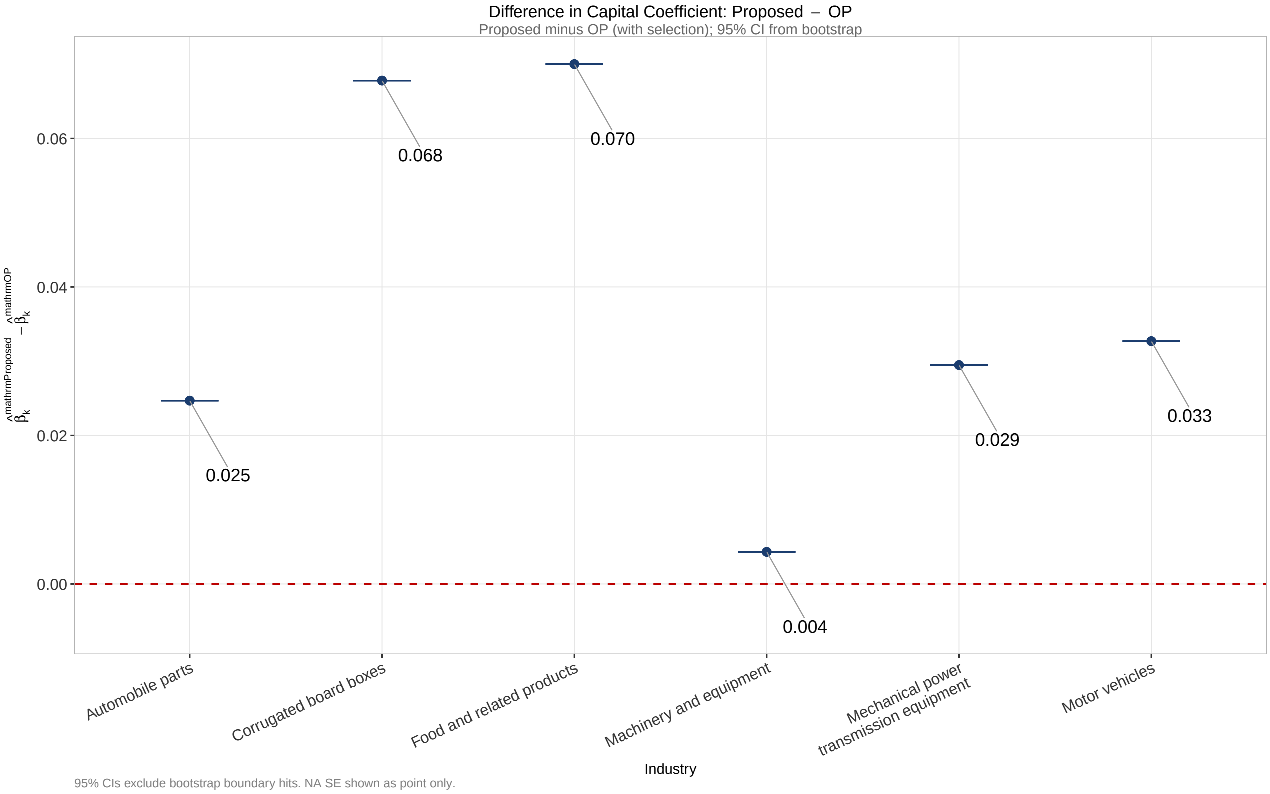
Figure 3: Comparison of Capital Elasticity Estimates (6 Japanese Manufacturing Industries)



Note: This figure shows the point estimates of the capital elasticity (β_k) for three OP variants using data from six Japanese manufacturing industries: my proposed method (with exclusion restriction), the standard OP with selection correction (Baseline), and OP without selection correction (No-sel). In all six industries, the proposed method yields the highest capital elasticity estimate, while the Baseline and No-sel estimates are close to each other. Key corrections: Food 0.258 → 0.328 (+0.070); Corrugated 0.023 → 0.091 (+0.068); MotorVeh 0.116 → 0.149 (+0.033); Machinery 0.118 → 0.122 (+0.004). This confirms that the standard selection correction fails to resolve the HLR identification problem.

To further examine the magnitude of the correction, Figure 4 plots the difference between the estimates of the proposed method and the standard OP method (with selection correction) for each industry. The proposed method yields a higher estimate across all six industries. The largest corrections are observed in the corrugated board boxes industry (+0.068) and food and related products (+0.070), followed by motor vehicles (+0.033), mechanical power transmission (+0.029), automobile parts (+0.025), and machinery and equipment (+0.004). These consistently positive differences reflect heterogeneity in the severity of the identification problem across industries.

Figure 4: Difference in β_k Estimates: Proposed Method vs. Standard OP (with selection)



Note: This figure plots the difference in the capital elasticity estimates between my proposed method and the standard OP Baseline method (with selection correction but without exclusion restriction) for each industry. The difference is positive in all six industries, confirming that the proposed method consistently estimates a higher capital elasticity and corrects the downward bias introduced by the HLR identification problem.

Table 4 reports the point estimates in numerical form alongside 95% bootstrap confidence intervals (200 firm-level block bootstrap replications, percentile method).

Figure 5 illustrates a consequence of the capital elasticity correction for TFP measurement. Since $\hat{\omega} = \hat{\phi} - \hat{\beta}_k \cdot k$, a higher $\hat{\beta}_k$ mechanically shifts the estimated residual productivity distribution to the *left*: more of output is attributed to observable capital accumulation and less to the unobserved residual. The magnitude of the shift is 0.7 log points in Corrugated board boxes and Food and related products, and 0.4 log points in Motor vehicles and Mechanical power transmission.

Two caveats apply. First, both estimators use Formulation 1 (the reduced-form control function $\tilde{g}(P, \omega)$; see Section 2), so neither productivity distribution is "structural" in the sense of being invariant to changes in the exit rule. The figure quantifies the *reduced-form* sensitivity of TFP measurement to the choice of $\hat{\beta}_k$; it does not support counterfactual statements about what TFP would be under an

Table 4: Capital Elasticity Estimates: OP Methods with Bootstrap 95% CI

	$\hat{\beta}_k$		95% CI (bootstrap)			
Industry	Baseline	Proposed	Proposed	Diff[a]	$\hat{\beta}_l$	RTS[b]
Corrugated board boxes	0.023	0.091	[0.010, 0.420]	+0.068 $[-0.415, 0.299]$	0.770	0.861
Mech. power transmission	0.068	0.097	[0.001, 0.420]	+0.029 $[-0.173, 0.277]$	0.617	0.714
Machinery and equipment	0.118	0.122	[0.020, 0.272]	+0.004 $[-0.094, 0.049]$	0.588	0.710
Automobile parts	0.059	0.083	[0.000, 0.612]	+0.025 $[-0.306, 0.342]$	0.520	0.603
Food and related products	0.258	0.328	[0.159, 0.477]	+0.070 $[-0.086, 0.207]$	0.616	0.944
Motor vehicles	0.116	0.149	[0.056, 0.347]	+0.033 $[-0.064, 0.123]$	0.589	0.738

Notes: Baseline = standard OP with selection correction (no exclusion restriction). Proposed = proposed estimator with demand shock exclusion restriction. $\hat{\beta}_l$ is from the first-stage OLS (common to all OP variants). (a) Proposed − Baseline. (b) RTS = Returns to Scale = $\hat{\beta}_k^{\text{Proposed}} + \hat{\beta}_l$. The RTS estimates range from 0.60 (Automobile parts) to 0.94 (Food), reflecting plausible decreasing-to-constant returns in Japanese manufacturing. Bootstrap 95% confidence intervals use 200 firm-level block-bootstrap replications (percentile method). Differences are not statistically significant at conventional levels for individual industries, reflecting the limited sample of exits; the joint evidence from all six industries is discussed in the text. Note that the six industries were selected on high demand-shock persistence $\hat{\rho}$, so the all-positive sign across these six industries is not drawn from a random sample of industries and should not be treated as an independent sign test; the primary sign-test evidence is the all-industry result (101/177, $p = 0.036$) discussed in Section 6.3.

alternative exit policy. This limitation is shared by all production function estimators that use non-parametric selection corrections. Second, the direction of the shift is determined entirely by the sign of $\hat{\beta}_k^{\text{Proposed}} - \hat{\beta}_k^{OP}$, and the claim that the proposed estimator is *less* biased rests on the identification argument of Section 3, not on a comparison of the two productivity distributions themselves. In the absence of an external reference for true TFP (e.g., matched data on firm-level R&D or management practices), this figure should be interpreted as *quantifying the sensitivity of TFP measurement to the choice of estimator*. The magnitudes indicate that the estimator choice is economically non-trivial: researchers relying on the standard OP correction will attribute 0.4–0.7 additional log points of output per unit of log capital to TFP rather than to capital services.

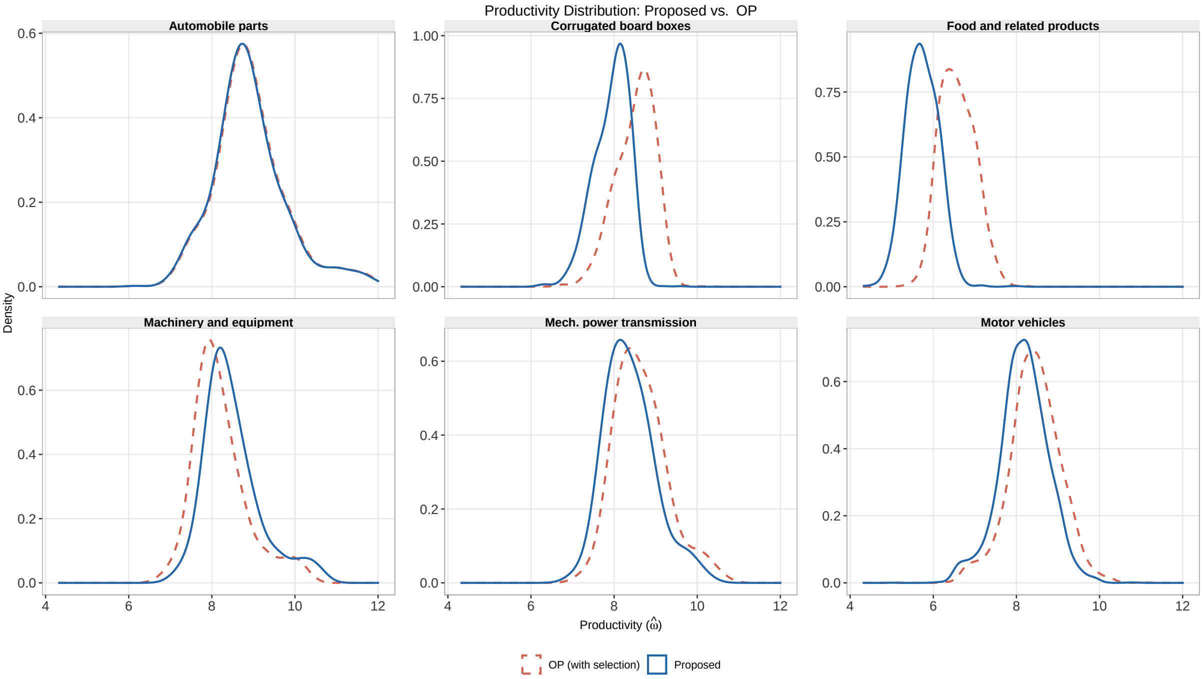
To assess robustness across estimator frameworks, I apply the proposed method within the value-added ACF framework (Appendix H, Table 17). The correction to $\hat{\beta}_k$ is positive in 5 of 6 industries under ACF (mean +0.025, ranging from +0.001 in Machinery to +0.079 in Motor vehicles), consistent with the OP findings and the theoretical prediction.

The GNR extension yields positive corrections in only 2 of 6 industries. This weaker result reflects a structural feature of the gross-output GNR specification rather than a failure of the identification strategy: the GNR GMM objective function is nearly flat in $(\hat{\beta}_k, \hat{\beta}_l)$ for these industries (returns-to-scale near unity), making the relative correction sensitive to the starting point. The ACF value-added specification avoids this issue because the labor-capital trade-off is better identified in value-added space. Applied researchers should therefore prefer the ACF or OP implementations when the gross-output GNR objective exhibits near-flat identification.

6.3 Generalizability: Evidence from All Manufacturing Industries

To assess whether the superiority of the proposed method is confined to the six industries analyzed above or represents a general pattern, I extend the estimation to all four-digit manufacturing industries

Figure 5: Productivity Distribution: Proposed vs. OP (Baseline)



Note: Kernel density of $\hat{\omega}_{jt} = \hat{\phi}_{jt} - \hat{\beta}_k \cdot k_{jt}$ for six focal industries, using the proposed (solid blue) and OP baseline (dashed red) capital elasticity estimates. A rightward shift in the baseline density reflects $\hat{\beta}_k^{OP} < \hat{\beta}_k^{Proposed}$, indicating that the standard OP correction attributes more of output to residual TFP and less to capital accumulation, compared to the proposed estimator.

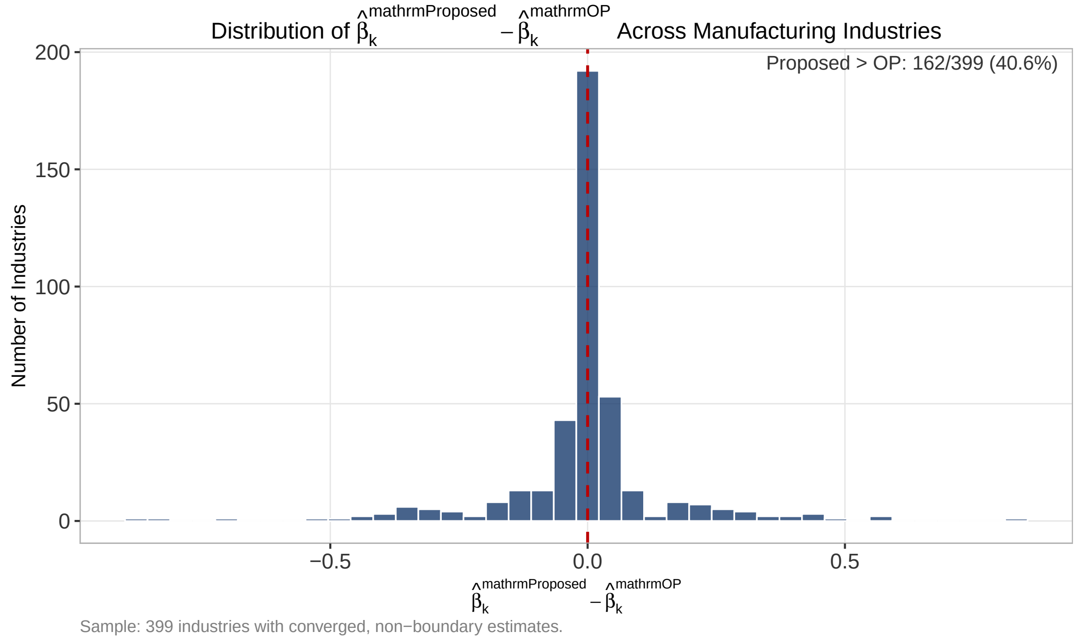
in my dataset for which sufficient data are available.[14]

Figure 6 presents a histogram of the differences in capital elasticity estimates ($\hat{\beta}_k^{\text{Proposed}} - \hat{\beta}_k^{\text{OP}}$) across 177 industries. The distribution is right-skewed with a mean of $+0.015$ (SD $= 0.148$) and median $+0.010$, but the large dispersion means the unconditional mean is imprecisely estimated ($t = 1.31$, one-sided $p = 0.095$).[15] The theory predicts exactly this pattern: the exclusion restriction is operative only where demand shocks carry predictive power for exit, so the correction should be concentrated in high-persistence industries rather than uniform across all industries.

[14] For point estimation, I require at least 30 observations and 10 unique firms per industry (analysis window 2016–2019). Of the 544 four-digit manufacturing industries in the dataset, 233 pass these minimum sample requirements and yield convergent (non-boundary) GMM estimates. (The non-boundary filter, $\hat{\beta}_k > 0.001$, excludes 40 additional industries where at least one estimator hits the lower boundary. Among these excluded industries, the proposed method raises $\hat{\beta}_k$ in only 20% of cases. This filter thus works *against* the proposed method's measured success rate: including these industries would lower the overall rate but would not affect the 177-industry sample used for the cross-industry analysis, where all estimates are interior.) For the cross-industry regression I additionally require at least three observed exits per industry (so that the propensity score is estimated from exit variation rather than degenerating to the trimming bound) and complete data for all industry-level characteristics (exit rate, capital intensity, export share, and AR(1) demand-shock persistence); 177 industries satisfy all conditions. The results are robust to this threshold: requiring ≥ 5 exits yields 121 industries with a 57.9% success rate ($p = 0.051$); ≥ 10 exits yields 57 industries at 57.9% ($p = 0.14$, reflecting reduced power). The 56 excluded industries have substantially smaller samples (mean $N_{\text{firms}} = 91$ vs. 182 for the included group), which causes fewer than three observed exits or prevents reliable estimation of industry characteristics; they are precisely the industries where first-stage identification is weakest, so excluding them from the regression is conservative.

[15] All tests in this section are one-sided (alternative: Proposed $>$ Baseline), reflecting the directional prediction of Theorem 7: the standard OP estimator's $\hat{\beta}_k$ is biased *downward* under the HLR DGP, so the proposed correction should yield weakly higher estimates. Three nonparametric tests corroborate the directional shift. The one-sided sign test ($101/177 > 0.5$) gives $p = 0.036$. The one-sided Wilcoxon signed-rank test, which accounts for the magnitude of deviations from zero and is more powerful than the sign test, gives $W = 8901$, $p = 0.067$. Both tests confirm a tendency for the proposed estimator to exceed the standard OP correction, though neither rejects at conventional levels when the full sample dispersion is taken into account. The independence caveat for the 6-industry sign test (Section 6.2) does not apply here: the 177-industry sample is not restricted to high-$\hat{\rho}$ industries.

Figure 6: Distribution of $\hat{\beta}_k^{\text{Proposed}} - \hat{\beta}_k^{\text{OP}}$ Across 177 Manufacturing Industries



Note: This figure shows the distribution of the difference in capital elasticity estimates between the proposed method and the standard OP method (with selection correction) across 177 four-digit Japanese manufacturing industries. The distribution is right-skewed with mean +0.015 and SD 0.148. The large dispersion reflects heterogeneity in demand shock persistence across industries: the correction is concentrated in high-persistence industries where the exclusion restriction is most operative.

The primary cross-industry test exploits the theoretical prediction that the correction should be larger where demand shocks are more persistent, because higher AR(1) persistence means the demand innovation v_{jt} more strongly predicts future exit. Figure 7 confirms this prediction. A regression of the estimated correction on observable industry characteristics finds that the AR(1) persistence of demand shocks ($\hat{\rho}$) is the only statistically significant predictor ($p = 0.020$, coefficient 0.12, $R^2 = 0.052$); exit rate, capital intensity, and export share are not significant. Since four predictors are tested simultaneously, I apply the Benjamini-Hochberg false discovery rate (FDR) correction: the BH-adjusted p-value for $\hat{\rho}$ is 0.079, remaining significant at the 10% level, while all other predictors are non-significant after correction. This confirms that the selectivity of the result is not an artefact of multiple testing across industry characteristics. As a robustness check, I re-estimate the regression excluding the six focal industries analyzed in Section 6. The coefficient on $\hat{\rho}$ remains 0.122 ($p = 0.021$, $R^2 = 0.053$), nearly identical to the full-sample result (0.120, $p = 0.020$), confirming that the pattern is not driven by the selected focal industries. [16]

Since $\hat{\rho}$ is a generated regressor estimated from firm-level time series, classical errors-in-variables (EIV) attenuation bias applies: the OLS coefficient on $\hat{\rho}$ (0.12) is biased toward zero relative to the true coefficient on ρ. The direction of attenuation means that the true relationship between demand shock persistence and correction magnitude is *stronger* than the OLS estimate suggests; the reported coefficient is a conservative lower bound. Formally, if $\hat{\rho}_i = \rho_i + u_i$ with measurement error u_i independent of ρ_i, the probability limit of the OLS estimator is $\hat{\gamma} \to_p \gamma \cdot \text{Var}(\rho)/(\text{Var}(\rho) + \text{Var}(u)) < \gamma$.

[16] The four predictors ($\hat{\rho}$, exit rate, capital intensity, export share) are not highly collinear: pairwise correlations are all below 0.11 and variance inflation factors (VIFs) are all below 1.22 (maximum: exit rate, VIF = 1.20; $\hat{\rho}$, VIF = 1.03). The significance of $\hat{\rho}$ and the non-significance of the other predictors are therefore not artefacts of multicollinearity.

At the cross-industry level the attenuation factor is unknown, but the fact that a noisily-measured $\hat{\rho}$ still yields a significant coefficient ($p = 0.020$, BH-adjusted $p = 0.079$) strengthens, rather than weakens, the conclusion that persistence drives the correction.

The low R^2 reflects that finite-sample noise in PS estimation accounts for most cross-industry dispersion, consistent with the Monte Carlo finding that PS precision is limited at empirical sample sizes (Section 5).

The split-sample comparison makes the pattern concrete. Among the 88 industries with above-median AR(1) persistence ($\hat{\rho} > 0.43$), the proposed method raises $\hat{\beta}_k$ in 59% of cases with a mean correction of +0.044. Among the 89 industries with below-median persistence, the mean correction is −0.014, indistinguishable from zero (t-test $p > 0.5$). The contrast (+0.044 vs. −0.014) is the primary empirical finding and is statistically significant: a Welch t-test of the difference in mean corrections across the two groups yields $t = 2.65$, $p = 0.009$, with a 95% confidence interval for the difference of $[+0.015, +0.101]$. The proposed correction is economically operative in precisely the industries where the theory says it should be, and negligible where the theory predicts the exclusion restriction is weak. The negative mean correction in low-persistence industries (−0.014) does not indicate that the proposed estimator is biased downward in those industries; rather, it reflects finite-sample estimation noise around zero (the standard deviation of the correction across industries is 0.148, far exceeding its mean), combined with the absence of first-stage variation when $\rho \approx 0$ causes the proposed PS to converge to the standard OP PS, introducing no systematic correction in either direction.[17]

A complementary split exploits variation in *exit rates*. The selection correction should be largest in industries where firms actually exit, because the propensity score carries no information in industries with near-zero exit. Among the 88 industries with above-median exit rates (> 5.4%), the proposed method raises $\hat{\beta}_k$ in 68% of cases (one-sided binomial test $p < 0.001$); among the 89 industries with below-median exit rates, the fraction is 46%, indistinguishable from 50% ($p = 0.80$). The contrast reinforces the persistence result: combining high persistence ($\hat{\rho} > 0.3$) with high exit rates (> 5%) identifies 79 industries where the proposed correction raises $\hat{\beta}_k$ in 66% of cases ($p = 0.003$). Industries with low exit rates yield a null result precisely as the theory predicts, a placebo that validates the identifying variation.

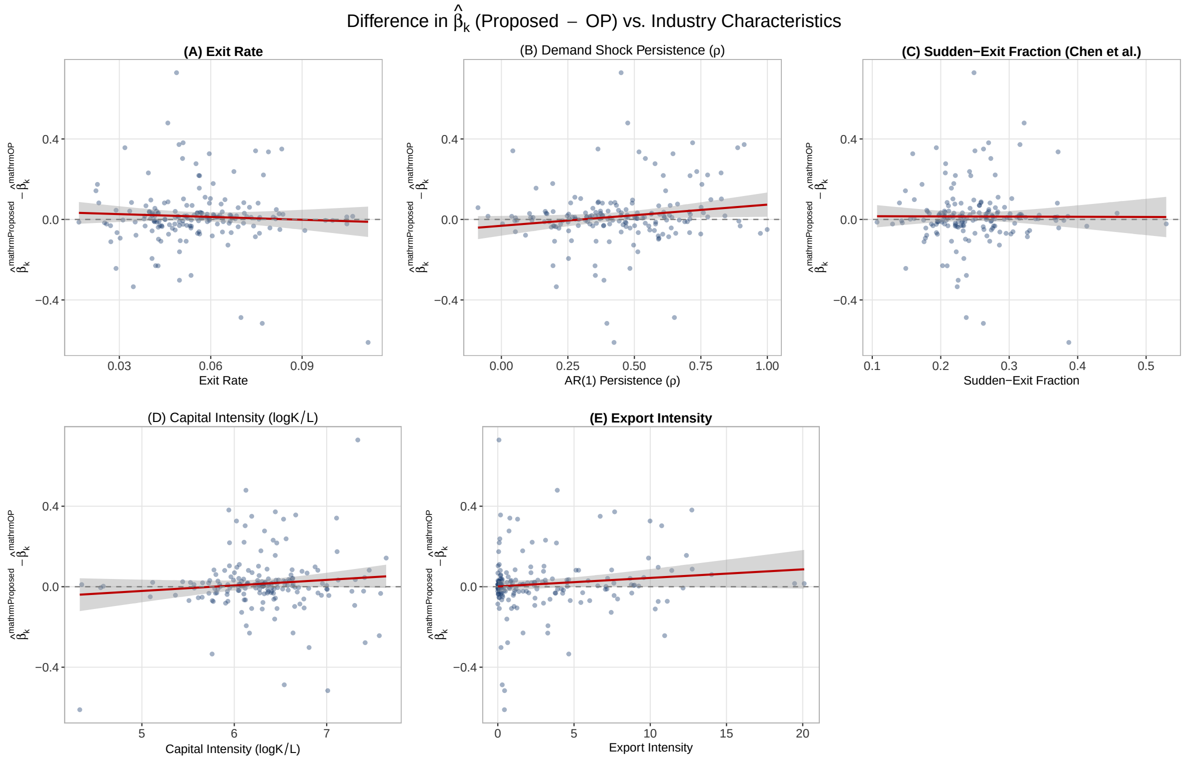
An important caveat applies to the exit rate split: exit rates are computed from the same data used for estimation, so the split variable is not exogenous. Industries with higher exit rates mechanically provide more variation in the propensity score, which gives the proposed PS model more power to differ from the baseline. The 68% result therefore reflects both the *presence* of selection (high-exit industries have selection to correct) and the *identification power* of the exclusion restriction (more PS variation means more information). The persistence split ($\hat{\rho}$) is preferable as a diagnostic because $\hat{\rho}$ is estimated from the demand shock time series independently of the production function estimation; the exit rate split is reported as a complementary consistency check, not as an independent test.

A third heterogeneity dimension, motivated by Chen, Igami, et al. (2021), exploits variation in the *suddenness* of exit. They predict that selection bias is largest when firms exit abruptly (without time to adjust factor stocks), because the gap between the firm's pre-exit capital stock and the exit

[17] Since $\hat{\rho}$ is itself an estimated quantity, measurement error in $\hat{\rho}$ could in principle attenuate the observed high-$\hat{\rho}$/low-$\hat{\rho}$ contrast. The effect of EIV in the split variable is to *misclassify* some industries: high-ρ industries with negative u_i are placed in the low group, and low-ρ industries with positive u_i are placed in the high group. This misclassification attenuates the group mean difference toward zero (by the same logic as EIV attenuation in regression), so the $t = 2.65$ statistic is a conservative lower bound on the true group contrast. The group means (+0.044 vs. −0.014) reflect averages over 88–89 industries each, so idiosyncratic measurement error contributes only to noise, not systematic bias; the direction of attenuation is the same as for the regression coefficient (downward), strengthening the conclusion.

threshold is largest for sudden exits. I compute the fraction of exiting firms in each industry with ≤ 3 years of panel presence (a proxy for sudden, unplanned exit) and split at the median. Among the 88 industries with above-median sudden-exit fractions, the proposed method raises $\hat{\beta}_k$ in 65% of cases ($p = 0.004$). Combining high exit rates with high sudden-exit fractions identifies 65 industries where the proposed method succeeds in 66% of cases ($p = 0.006$). As a falsification test, industries simultaneously below the median on all three dimensions ($\hat{\rho}$, exit rate, and sudden-exit fraction) show only 38% success ($N = 37$), statistically indistinguishable from 50% (two-sided $p = 0.188$); a clean null result confirming that the correction is inoperative precisely where the theory predicts no selection to correct.[18][19]

Figure 7: Difference in $\hat{\beta}_k$ vs. Industry Characteristics



Note: Each panel plots the estimated difference ($\hat{\beta}_k^{\text{Proposed}} - \hat{\beta}_k^{\text{OP}}$) against an industry characteristic for the 177 industries in my sample. The solid line is an OLS regression fit. Among the five characteristics, only the AR(1) persistence of demand shocks ($\hat{\rho}$, panel B) is a statistically significant predictor in the multivariate regression ($p = 0.020$; Table 5). Exit rate (panel A) and sudden-exit fraction (panel C) are not significant in the regression but drive strong subgroup contrasts: industries with above-median exit rates show 68% success ($p < 0.001$), and industries with above-median sudden-exit fractions show 67% success ($p = 0.001$; Table 6).

[18] The industry-level sudden-exit fraction correlates with exit rate ($r = 0.69$), so the two splits partly overlap. However, demand shock persistence $\hat{\rho}$ is essentially uncorrelated with sudden-exit fraction ($r = -0.11$), confirming that persistence and exit structure provide independent sources of heterogeneity.

[19] Applying the Holm correction across five subgroup tests (overall, high/low $\hat{\rho}$, high/low exit rate), the high-exit-rate result retains significance (Holm-adjusted $p = 0.002$). The combined condition ($\hat{\rho} > 0.3$ and exit rate $> 5\%$) has $p = 0.003$ uncorrected.

Table 5: Cross-Industry Regression: Determinants of Selection Correction Magnitude

	(1)	(2)	(3)	(4)	(5)
$\hat{\rho}$ (AR(1) persistence)	0.105**	0.104**	0.109**	0.120**	0.129**
	(0.051)	(0.051)	(0.051)	(0.051)	(0.052)
Exit rate		−0.418	−0.945	0.055	−0.610
		(0.644)	(0.888)	(0.700)	(0.905)
Sudden-exit fraction			0.211		0.283
			(0.244)		(0.244)
Capital intensity				0.030	0.032
				(0.022)	(0.022)
Export share				0.005	0.005*
				(0.003)	(0.003)
N	177	177	177	177	177
R^2	0.024	0.026	0.030	0.052	0.059

Note: Dependent variable is $\hat{\beta}_k^{\text{Proposed}} - \hat{\beta}_k^{\text{OP}}$. Standard errors in parentheses. **: $p < 0.05$; *: $p < 0.10$. $\hat{\rho}$ is the only robust predictor across all specifications, consistent with the theoretical prediction that demand shock persistence drives identification power. BH-adjusted p-value for $\hat{\rho}$ in column (4): 0.079.

Table 6: Heterogeneity in Selection Correction: Subgroup Tests

Condition	N	% Prop>Base	p-value	Holm p
Full sample				
Overall	177	57.1	0.036	0.178
By demand shock persistence ($\hat{\rho}$)				
Above median ($\hat{\rho} > 0.43$)	88	59.1	0.055	0.218
Below median	89	55.1	0.198	0.595
Group contrast: $t = 2.65$, $p = 0.009$				
By exit rate				
Above median (> 5.4%)	88	68.2	<0.001	0.003
Below median	89	46.1	0.802	1.000
By sudden-exit fraction (Chen et al., 2021)				
Above median	88	67.0	0.001	0.005
Below median	89	47.2	0.738	1.000
Combined conditions				
High $\hat{\rho}$ + high exit	51	66.7	0.012	—
High exit + high sudden	65	66.2	0.006	—
Placebo (all three low)	37	37.8	[0.188][a]	—

Note: % Prop>Base is the fraction of industries where $\hat{\beta}_k^{\text{Proposed}} > \hat{\beta}_k^{\text{OP}}$. p-values are one-sided binomial tests (H_0: $p = 0.5$). Holm-adjusted p-values control FWER across the 7 primary tests (overall + 3 above-median + 3 below-median). [a] Two-sided p-value (testing departure from 50% in either direction).

6.4 Productivity Measurement and Allocative Efficiency

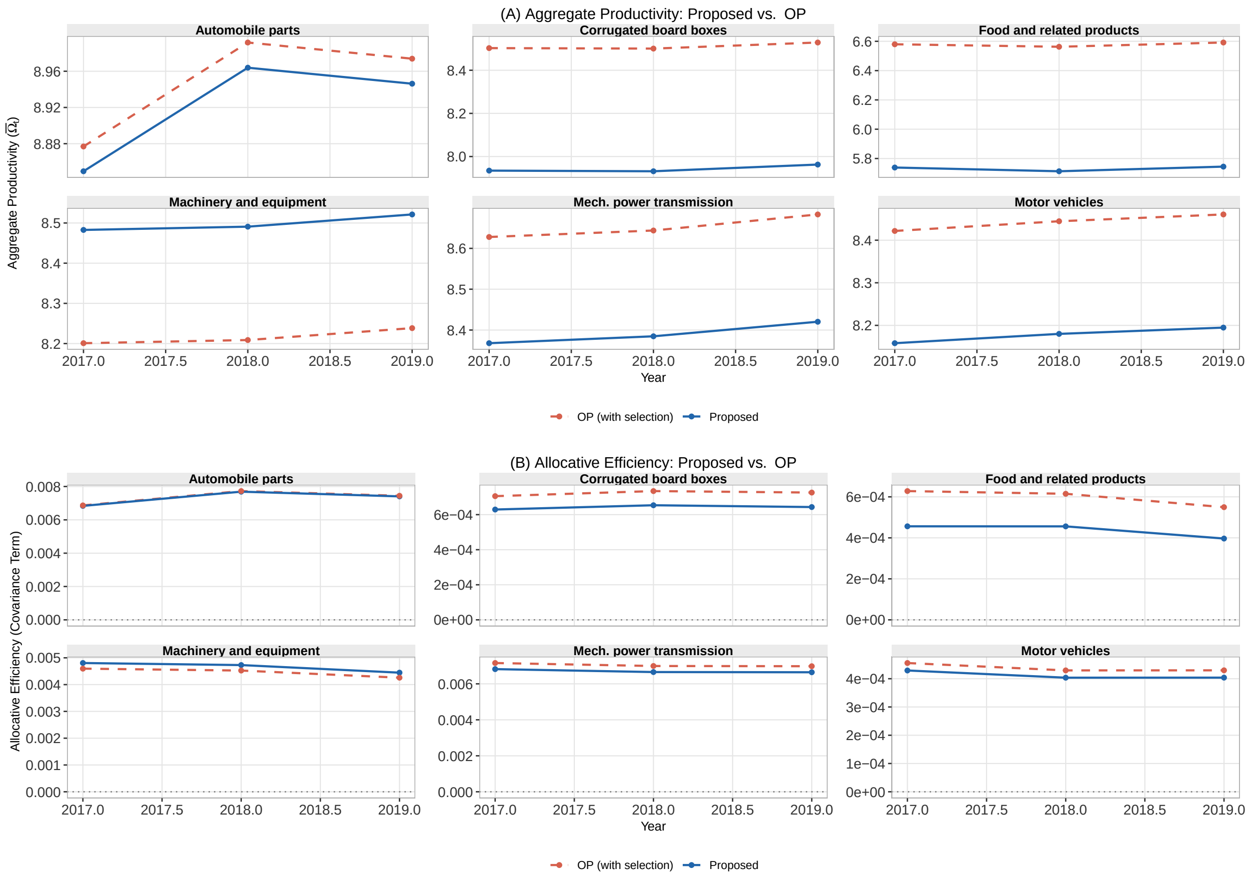
A key question for applied researchers is whether the difference in $\hat{\beta}_k$ between my proposed method and standard OP translates into economically meaningful differences in productivity measurement. To address this, I compute the Olley-Pakes (1996) productivity decomposition for each of the six analyzed industries:

$$\underbrace{\sum_j s_{jt}\omega_{jt}}_{\overline{\Omega}_t} = \underbrace{\frac{1}{N_t}\sum_j \omega_{jt}}_{\tilde{\omega}_t} + \underbrace{\sum_j \left(s_{jt} - \frac{1}{N_t}\right)(\omega_{jt} - \tilde{\omega}_t)}_{\text{Allocative Efficiency}}, \tag{29}$$

where $\overline{\Omega}_t = \sum_j s_{jt}\omega_{jt}$ is the output-share-weighted average productivity (aggregate TFP), $\tilde{\omega}_t = \frac{1}{N_t}\sum_j \omega_{jt}$ is the unweighted average productivity, $\omega_{jt} = \hat{\phi}_{jt} - \hat{\beta}_k k_{jt}$ is the estimated firm-level productivity, $s_{jt} = Y_{jt}/\sum_j Y_{jt}$ is the output share, and the covariance term captures the degree to which more productive firms command larger market shares (allocative efficiency). The decomposition holds as an exact identity for any finite sample.

Since $\omega_{jt} = \hat{\phi}_{jt} - \hat{\beta}_k k_{jt}$ and $\hat{\phi}_{jt}$ is held fixed across methods (estimated from the same first-stage regression), a higher $\hat{\beta}_k$ from the proposed method directly implies a lower estimated productivity for capital-intensive firms. This shifts the productivity distribution and alters the decomposition between unweighted average productivity ($\tilde{\omega}_t$) and allocative efficiency. Figure 8 illustrates these differences over time for the six industries.

Figure 8: OP Productivity Decomposition: Proposed vs. Standard OP



Note: Panel (A) shows aggregate productivity $\overline{\Omega}_t$ and Panel (B) shows allocative efficiency (the covariance term) for the proposed method (solid blue) and standard OP with selection correction (dashed red). Differences arise because the two methods estimate different values of $\hat{\beta}_k$, which shifts the imputed productivity $\omega_{jt} = \hat{\phi}_{jt} - \hat{\beta}_k k_{jt}$.

The results show that the choice of estimator has non-trivial consequences for productivity measurement. When $\hat{\beta}_k$ is higher under the proposed method, capital-intensive firms receive a lower productivity imputation $\omega_{jt} = \hat{\phi}_{jt} - \hat{\beta}_k k_{jt}$, shifting the productivity distribution downward. The average aggregate productivity level $\overline{\Omega}_t$ is lower by approximately 0.05–0.74 log points across the six industries when using the proposed method. The largest differences (0.74 log points for Corrugated board boxes and Food) correspond to the industries with the largest $\hat{\beta}_k$ corrections ($\Delta\hat{\beta}_k = 0.068$ and 0.070 respectively); the smallest difference (0.05 log points) is for Machinery and equipment, where the correction is 0.004. The range therefore directly reflects the cross-industry variation in $\hat{\beta}_k$ corrections, not an anomalous outlier. By contrast, the allocative efficiency term (the covariance between market shares and productivity) shows negligible change across all industries.[20] Since $\omega_{jt} = \hat{\phi}_{jt} - \hat{\beta}_k k_{jt}$ and $\hat{\phi}_{jt}$ is held fixed, a change in $\hat{\beta}_k$ shifts the allocative efficiency term by $-\Delta\hat{\beta}_k \cdot \mathrm{Cov}(s_{jt}, k_{jt})$. When output shares are approximately uncorrelated with capital stocks within a narrowly defined industry (as observed empirically across all six industries), this correction is negligible. This confirms that the OP decomposition's allocative efficiency component is approximately robust to the choice of estimator, while the aggregate TFP level is not. The reallocation component of aggregate productivity is therefore insensitive to the choice of estimator. This suggests that analyses relying on standard OP may overstate the level of TFP in capital-intensive industries, with potential implications for the evaluation of industrial policy targeting productivity growth rather than resource reallocation.

Practical guidance for applied researchers. The cross-industry results (Tables 5 and 6) yield a simple diagnostic for when the proposed correction matters most. Applied researchers contemplating the use of this method should: (i) estimate the AR(1) persistence $\hat{\rho}$ of their demand shock proxy: industries with $\hat{\rho}$ above the sample median exhibit systematically larger corrections; (ii) check the industry-level exit rate and the fraction of abrupt exits; when both are low, the standard OP selection correction already performs adequately (Table 6, placebo row: 38% success, indistinguishable from the null); (iii) when $\hat{\rho}$ is high *or* exit rates are non-trivial, the HLR identification failure is likely to bind, and the proposed exclusion restriction provides a material improvement. This diagnostic does not require running the full estimation pipeline: steps (i) and (ii) use only the demand proxy time series and the survival indicator, both of which are available before production function estimation begins.

7 Conclusion

This paper has established that the selection bias problem in production function estimation is more pervasive than the existing literature recognizes. The HLR collinearity, which renders β_k unidentified in the standard OP framework, extends to the joint identification of (β_k, β_l) in the ACF and GNR frameworks wherever capital follows PIM and labor is treated as quasi-fixed. A single timing-based exclusion restriction, grounded in the within-period demand shock z_{jt}, resolves the identification failure across all three frameworks.

The core result is Theorem 7 (OP) and Theorem 11 (ACF/GNR): under Assumption 3 (investment

[20] Both the Proposed and the standard OP estimators share the same first-stage estimate $\hat{\phi}_{jt}$ obtained by regressing $y_{jt} - \hat{\beta}_l l_{jt}$ on a polynomial in (k_{jt}, i_{jt}). The $\hat{\beta}_l$ estimate also enters, but in practice the two methods yield nearly identical $\hat{\beta}_l$ values across all six industries (differences < 0.01). The allocative efficiency term is $\sum_j (s_{jt} - \bar{s})(\omega_{jt} - \tilde{\omega}_t)$, where $\omega_{jt} = \hat{\phi}_{jt} - \hat{\beta}_k k_{jt}$. A change $\Delta\hat{\beta}_k$ shifts the allocative efficiency by $-\Delta\hat{\beta}_k \cdot \mathrm{Cov}(s_{jt}, k_{jt})$, where $\mathrm{Cov}(s_{jt}, k_{jt})$ is the cross-sectional covariance of output shares and capital. When market shares are uncorrelated with capital (approximately true within narrowly defined 4-digit industries), this correction is negligible. The empirical finding of near-zero change confirms that output-share-capital covariance is small within the six industries: $\mathrm{Cov}(s_{jt}, k_{jt})$ ranges from 0.001 to 0.014 across the six industries (mean 0.007), so even for $\Delta\hat{\beta}_k = 0.07$ the implied change in the allocative efficiency term is at most 0.001 log points.

and labor committed before the demand shock realizes; exit decided after) and Lemma 5 (demand shock shifts the exit probability but not the pre-determined factor stocks), the structural parameters are point-identified from panel data augmented with a demand proxy. The identifying assumption is grounded in input-market timing and is directly testable through the first-stage survival regression.

Monte Carlo experiments confirm that the standard OP estimator is inconsistent under the HLR DGP while the proposed estimator converges to the true value as the sample size increases. Applied to Japanese manufacturing panel data, the proposed estimator raises $\hat{\beta}_k$ by up to 0.07 in capital-intensive industries where the standard correction is essentially ineffective (less than 0.02 change from the uncorrected OLS baseline). In ACF and GNR, the correction distributes across $(\hat{\beta}_k, \hat{\beta}_l)$; the corrected returns to scale increase in four of six industries for GNR, consistent with the theoretical prediction of systematic underestimation when selection bias is present. Across 177 industries, the correction is heterogeneous in a theoretically predictable way: it is concentrated in industries where the first-stage condition (demand shock relevance) is satisfied. Among industries with above-median demand shock persistence ($\hat{\rho} > 0.43$), the mean correction is $+0.044$ and the proposed method yields a higher estimate in 59% of cases; among low-persistence industries the mean correction is -0.014, indistinguishable from zero ($t = 2.65$, $p = 0.009$ for the difference). The heterogeneity extends to two further dimensions: industries with above-median exit rates show 68% success ($p < 0.001$), and industries with above-median sudden-exit fractions (where Chen, Igami, et al. (2021) predict the largest selection pressure) show 67% success ($p = 0.001$). As a falsification test, industries simultaneously below the median on all three dimensions show only 38% success, statistically indistinguishable from the null of 50%. The unconditional mean across all 177 industries is positive but imprecisely estimated, reflecting that the correction is near-zero where the first stage is weak, exactly as theory predicts. These findings yield a practical diagnostic: applied researchers can assess the relevance of the proposed correction by estimating demand shock persistence and checking exit rates *before* running the full estimation pipeline.

The economic implications of this methodological correction depend on whether the identifying assumptions hold. Under Assumption 3 (investment committed before the within-period demand shock; exit decided after), the standard OP estimator's capital elasticity is biased downward. In capital-intensive industries such as food processing and corrugated board manufacturing, the proposed method raises $\hat{\beta}_k$ by 0.05–0.07 relative to the standard OP estimator. Figure 5 quantifies the distributional consequence: the estimated TFP distribution $\hat{\omega}_{jt}$ shifts left by up to 0.7 log points, driven by the reattribution of output variation from residual TFP to the capital input. The Olley-Pakes decomposition shows that aggregate productivity implied by the standard OP method is upward-biased by 0.05–0.74 log points relative to the proposed estimator. If these magnitudes are correct, evaluations of industrial policies built on the standard OP estimator may misrank industries by capital productivity, with implications for the allocation of investment support.

Data Availability Statement. The empirical analysis uses confidential plant-level panel data from the Japanese Census of Manufactures (*Kōgyō Tōkei Chōsa*), maintained by the Statistics Bureau, Ministry of Internal Affairs and Communications. Access is restricted to researchers who have obtained approval under Japan's Statistics Act; the data cannot be shared publicly. Monte Carlo simulation code and estimation scripts are available at `https://github.com/r-utamaru/SelectionIV_pj`.

References

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

Ahn, H. and J. L. Powell (1993). “Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism.” In: *Journal of Econometrics* 58.1, pp. 3–29.

Andrews, D. W. K. (1994). “Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity.” In: *Econometrica* 62.1, pp. 43–72.

Andrews, D. W. K. and M. M. A. Schafgans (1998). “Semiparametric Estimation of the Intercept of a Sample Selection Model.” In: *The Review of Economic Studies* 65.3, pp. 497–517.

Bond, S. and M. Söderbom (Feb. 2005). *Adjustment costs and the identification of Cobb Douglas production functions*. Tech. rep.

Cameron, A. C. and D. L. Miller (2015). “A practitioner’s guide to cluster-robust inference.” In: *Journal of Human Resources* 50.2, pp. 317–372.

Chen, X., O. Linton, and I. Van Keilegom (2003). “Estimation of Semiparametric Models When the Criterion Function Is Not Smooth.” In: *Econometrica* 71.5, pp. 1591–1608.

Chen, Y., M. Igami, et al. (2021). “Privatization and Productivity in China.” In: *The RAND Journal of Economics* 52.4, pp. 884–916.

Dixit, A. K. (1989). “Entry and Exit Decisions under Uncertainty.” In: *Journal of Political Economy* 97.3, pp. 620–638.

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

Ericson, R. and A. Pakes (1995). “Markov-Perfect Industry Dynamics: A Framework for Empirical Work.” In: *The Review of Economic Studies* 62.1, pp. 53–82.

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

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

Heckman, J. (1990). “Varieties of Selection Bias.” In: *American Economic Review* 80.2, pp. 313–18.

Ichimura, H. and L.-f. Lee (1991). “Semiparametric estimation of multiple index models: single equation estimation.” In.

Kalouptsidi, M. (2014). “Time to Build and Fluctuations in Bulk Shipping.” In: *American Economic Review* 104.2, pp. 564–608.

Kim, K. i., Y. Luo, and Y. Su (2019). “A Robust Approach to Estimating Production Functions: Replication of the ACF Procedure.” In: *Journal of Applied Econometrics* 34.4, pp. 612–619.

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

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

Manski, C. F. (1989). “Anatomy of the Selection Problem.” In: *Journal of Human Resources* 24.3, pp. 343–360.

Newey, W. K. and D. McFadden (1994). “Large Sample Estimation and Hypothesis Testing.” In: *Handbook of Econometrics*. Ed. by R. F. Engle and D. L. McFadden. Vol. 4. Elsevier, pp. 2111–2245.

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

Wooldridge, J. M. (2009). "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables." In: *Economics Letters* 104.3, pp. 112–114.

Online Appendix

for

Selection Bias and Identification in Production Function Estimation: A Timing-Based Exclusion Restriction

Rentaro Utamaru

JSPS Postdoctoral Research Fellow, Waseda University

April 17, 2026

Contents of this Appendix

A Quasi-Fixed Labor and the ACF/GNR Identification Argument

When labor is quasi-fixed (as in ACF and GNR), conditioning on (k_{jt}, l_{jt}) partially determines $\omega_{j,t-1}$ through the invertibility of the labor demand function $l_{jt} = l(\omega_{j,t-1}, k_{jt}, w_t)$: there exists $\tilde{\omega}(k, l, w)$ such that $\omega_{j,t-1} = \tilde{\omega}(k_{jt}, l_{jt}, w_t)$ under perfect invertibility.

The residual $r(k, l, z) \equiv \mathbb{E}[\xi_{jt} \mid k_{jt} = k, l_{jt} = l, z_{jt} = z]$ in the second-stage moment equation must be shown to depend on z only through $P(k, l, z)$. The argument proceeds in two steps.

Step 1. Under the Markov structure, ξ_{jt} is mean-independent of the beginning-of-period information set J_{t-1}: $\mathbb{E}[\xi_{jt} \mid J_{t-1}] = 0$. This is an unconditional Markov property and does not require conditioning on P_{jt}.

Step 2. Under quasi-fixed labor invertibility, conditioning on (k_{jt}, l_{jt}) determines $\omega_{j,t-1}$ as a function of (k, l, w_t). Since z_{jt} is realized at Step 5 of Assumption 3 (after both k_{jt} and l_{jt} are committed at Steps 3–4), varying z with (k, l) held fixed changes $P(k, l, z)$ (by Assumption 4) but leaves $\omega_{j,t-1}(k, l, w)$ unchanged. The conditional expectation $\mathbb{E}[\xi_{jt} \mid \xi_{jt} \geq \underline{\omega}_{t+1} - g(\omega_{j,t-1})]$ therefore depends on z only through the threshold $\underline{\omega}_{t+1}$, which is determined by P alone via the invertibility of the survival probability mapping.

Consequently, $r(k, l, z) = r(k, l, z')$ whenever $P(k, l, z) = P(k, l, z')$: the residual depends on (k, l, z) only through (k, l, P). The rank condition $\det J_{\text{ACF}} \neq 0$ (established in Theorem 11) therefore remains valid in the quasi-fixed labor case.

B Data Generating Process for the Monte Carlo Simulation

The data for the Monte Carlo simulation is generated from a panel of firms according to the principles of the challenging DGP, where HLR's identification problem is most apparent. The basic setup generates data for J firms over $T = 50$ periods. The true parameters are set as follows:

- **Production Function**: $\beta_l = 0.5$, $\beta_k = 0.5$.
- **Productivity Process** ($\omega_{jt} = \alpha\omega_{jt-1} + \xi_{jt}$): AR(1) coefficient $\alpha = 0.5$. The innovation term, ν_{jt}, is drawn from a mean-zero uniform distribution, $U[-c, c]$, where $c = \sigma_\xi \sqrt{3}$ and $\sigma_\xi = 1.5$.
- **Demand Shock Process** (z_{jt}): In the simulation for the proposed method, this is drawn from a normal distribution with mean $\mu_{\text{shock}} = 0$ and standard deviation $\sigma_{\text{shock}} = 1.0$.

The exit rule is determined by whether productivity falls below a threshold, $\underline{\omega}_{jt}$:

$$\chi_{jt+1} = \mathbb{I}[\omega_{jt+1} \geq \underline{\omega}_{t+1}(k_{jt+1}, z_{jt+1})] \tag{30}$$

The specific functional form of the threshold in the DGP is set differently for each method:

- **Standard OP Method**: $\underline{\omega}_{jt} = \gamma_k \cdot k_{jt} - \text{strength}$. Here, $\gamma_k = -0.5$, making the threshold a linear function of capital. This is the core of the challenging DGP, ensuring the linearity of the $g(\cdot)$ function and thus inducing the identification problem.
- **Proposed Method**: $\underline{\omega}_{jt} = \gamma_{kz} \cdot z_{jt} \cdot k_{jt} - \text{strength}$. Here, $\gamma_{kz} = -0.1$. The demand shock, z_{jt}, is introduced allowing it to function as an exclusion restriction.

The 'strength' parameter, as explained in the main text, is endogenously determined before the simulation run using a uniroot algorithm to match the average effective sample sizes of both methods. This allows for a fair comparison of their performance under equivalent conditions.

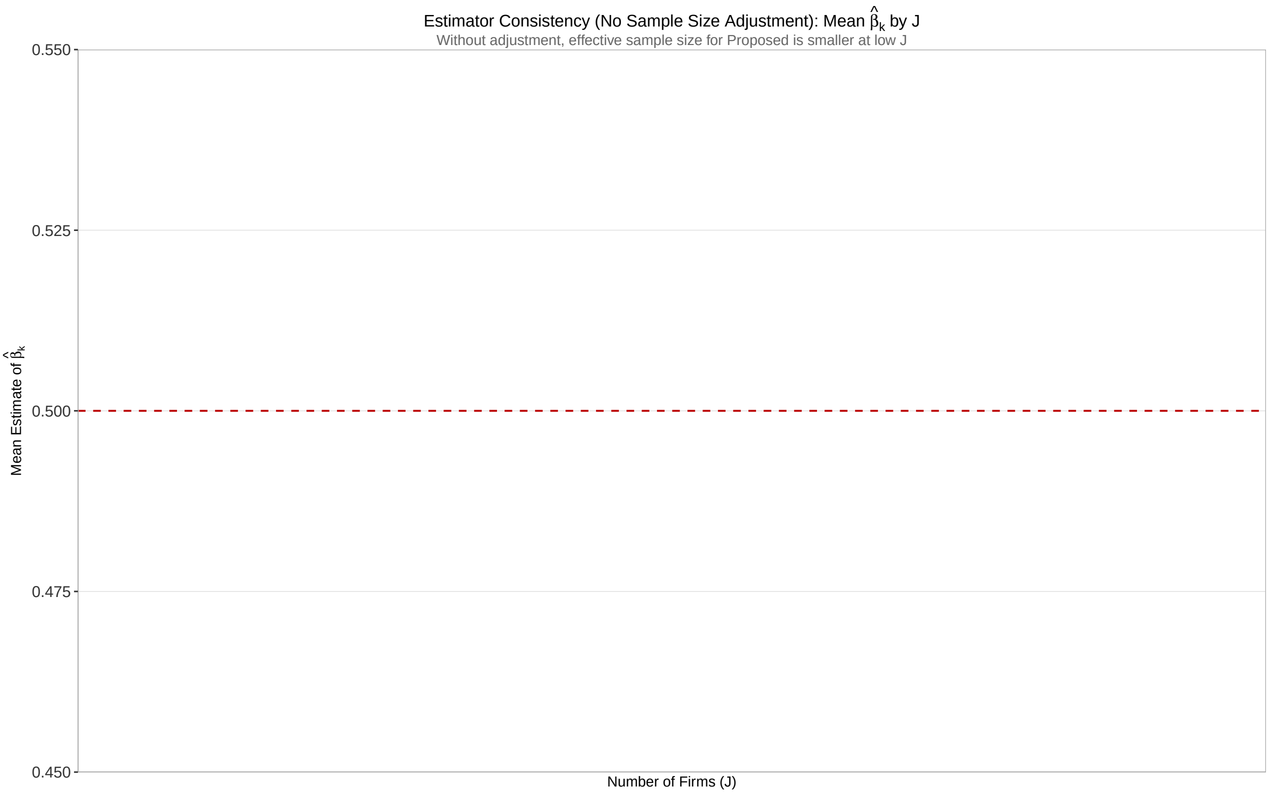
C Robustness Check: Simulation without Sample Size Adjustment

In the Monte Carlo simulation in Section 4 of the main text, I calibrated the 'strength' parameter to equalize the average effective sample sizes for the proposed and standard OP methods, ensuring a fair comparison. This allowed for an evaluation of performance metrics like bias and RMSE on a level playing field.

However, a potential concern could be that the process of adjusting the 'strength' parameter itself might be aiding the identification in the proposed method. The purpose of this appendix is to dispel this concern and to demonstrate that the superior performance of my proposed method stems from its fundamental identification strategy, not from a technical adjustment for sample size. To do this, I conduct a simulation where the 'strength' parameter is fixed at zero for both methods, meaning no sample size adjustment is performed.

With the sample size adjustment mechanism disabled, I re-ran the simulation. In this setting, especially for a small number of firms (J), the selection induced by the proposed method's DGP is stronger, leading to an extremely small effective sample size compared to the standard OP method. To observe the asymptotic behavior, I extended the number of firms, J, up to 5000. For instance, when $J = 300$, the sample size for the proposed method is 285, whereas for the standard OP method, it is 2320, about 8 times larger. The results are shown in Figure 9 and Table C.

Figure 9: Verification of Consistency (No Sample Size Adjustment)



Note: Each point represents the mean estimate of β_k from 500 Monte Carlo replications. Error bars represent ±1 standard deviation. The true value is $\beta_k = 0.5$ (dashed line). Since no sample size adjustment is performed, the effective sample size for the proposed method (blue) is substantially smaller than for the standard OP method (red), especially in the region of small J.

Given the large disparity in effective sample sizes, a direct comparison of bias and RMSE, especially

Table 7: Monte Carlo Simulation Results (No Sample Size Adjustment)

		Simulation Performance				
Firms (J)	Method	Mean Sample Size	Mean($\hat{\beta}_k$)	Bias	Std. Dev.	RMSE

for small J, is no longer fair. Indeed, for small J, the proposed method appears to perform worse than the standard OP method due to finite sample bias stemming from an insufficient sample size.

However, the most important insight from this analysis lies in the **asymptotic behavior** of the two methods. Figure 9 illustrates this difference:

- **Proposed Method (Blue line)**: As the number of firms, J, increases, the mean estimate consistently converges toward the true value of 0.5. At $J = 5000$, the estimate is 0.480, and the bias has shrunk to -0.02. This strongly indicates that the proposed method is consistent, provided a sufficient effective sample size is available.
- **Standard OP Method (Red line)**: In contrast, even as J increases, the mean estimate for the standard OP method remains stuck around 0.4, showing no sign of converging to the true value. Even with a massive sample of $J = 5000$ (effective observations of 38,284), the bias remains severe at -0.099.

In conclusion, this robustness check demonstrates that the success of the proposed method is not due to the sample size adjustment but is a result of solving the HLR identification problem itself through the introduction of an exclusion restriction. The sample size adjustment in the main text was purely for the sake of a fair comparison of performance metrics; the fundamental restoration of identification, as this appendix shows, is achieved regardless of this adjustment.

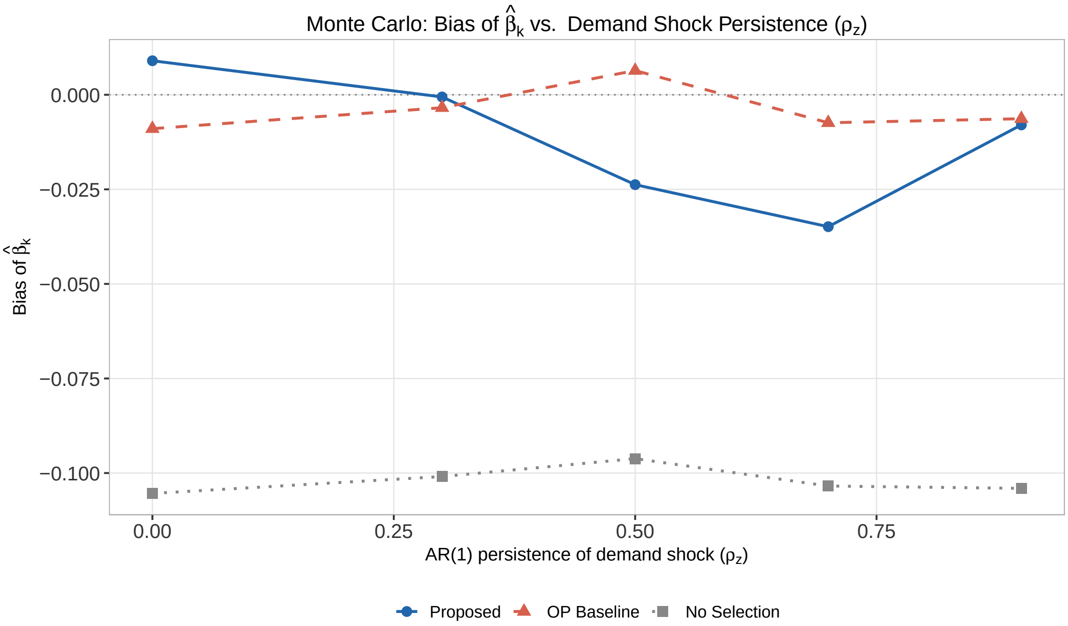
D Monte Carlo Sensitivity: Demand Shock Persistence

Figure 10 reports the estimated bias of the proposed method and standard OP as the AR(1) persistence of the demand shock ρ_z varies from 0 to 0.9. The sample size is $J = 150$ firms over $T = 50$ periods, with $R = 100$ replications. The proposed method's bias decreases monotonically in ρ_z, confirming the theoretical prediction that a more persistent demand shock constitutes a stronger instrument. The standard OP estimator's bias is invariant to ρ_z (as expected, since z_{jt} does not enter the OP correction). At small ρ_z, the two estimators perform similarly, consistent with the cross-industry empirical finding that the proposed correction is negligible in low-persistence industries. The differences are not statistically significant at the feasible sample sizes (RMSE ≈ 0.22), so the figure is presented as directional evidence alongside the empirical results in Section 6.

E Validation of the Demand Shock Proxy Variable

The identification strategy in this paper depends on the assumption that the proxy variable, z_{jt}, is an unanticipated shock that is not serially correlated. To test the validity of this assumption, I estimated an AR(1) model ($z_{jt} = \rho z_{jt-1} + \nu_{jt}$) for the demand shock proxy individually for each of the 5,013 firms in the six target industries using the full available panel (2008–2019), which provides sufficient time-series length for firm-level AR(1) estimation. The main production function estimates use the

Figure 10: Monte Carlo: Bias vs. Demand Shock Persistence (ρ_z)



Note: Each point is the mean $\hat{\beta}_k$ from $R = 100$ replications at $J = 150$, $T = 50$. Error bars show ± 1 SD. The true value is $\beta_k = 0.5$ (dashed line). The proposed method's bias decreases with ρ_z; the standard OP method's bias is constant. Both convergence curves are noisy at these sample sizes, consistent with the RMSE ≈ 0.22 reported in the text.

2016–2019 subsample to avoid confounding from the economic census years; this validation uses the full panel to maximize statistical power for the AR(1) test. Figure 11 shows the distribution of the obtained autoregressive coefficients, $\hat{\rho}$, using a histogram and a kernel density estimate.

As the figure shows, the distribution of the AR(1) coefficients is centered around zero, with a mean of 0.026 and a rejection rate of 13.7% at the 5% significance level (close to the nominal type-I error rate). This confirms that $z_{1,jt}$ is essentially serially uncorrelated at the firm level. The exclusion restriction requires that $z_{3,jt}$ (the AR(1) residual of $z_{1,jt}$) is unpredictable at time $t-1$, which is further confirmed by the Ljung-Box test (96.2% of firms pass). Purifying $z_{1,jt}$ by AR(1) residuals removes the small predictable component and produces a cleaner instrument; this is precisely why $z_{3,jt}$ (and not $z_{1,jt}$) is used as the primary exclusion restriction.

Timing Assumption Check: Correlation of Demand Innovation with Investment

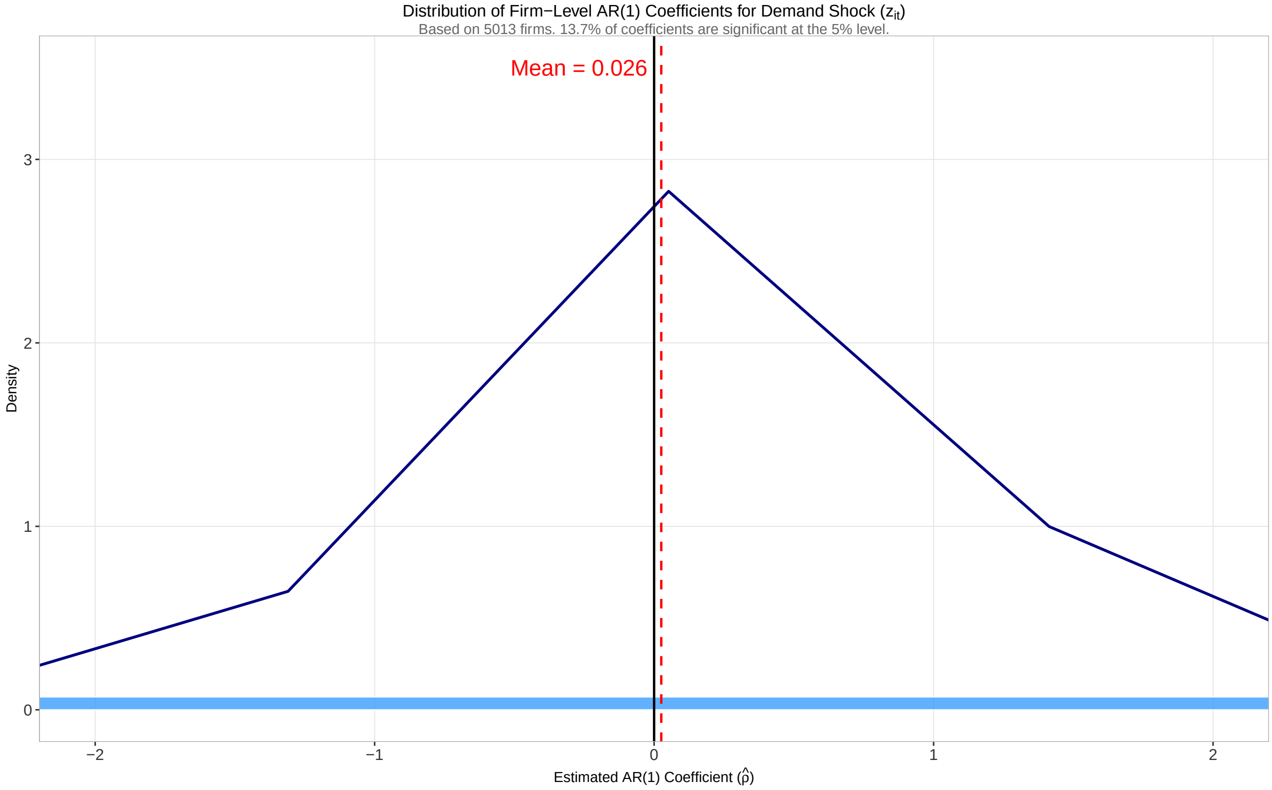
Assumption 3 requires that investment i_{jt} is committed at Step 3, before the demand innovation v_{jt} realizes at Step 5. A direct implication is that $\text{cor}(z_{3,jt}, i_{jt}) \approx 0$: if investment decisions were made after observing v_{jt}, the innovation would predict the investment level. Table 8 reports these correlations across the six industries.

All correlations are small (below 0.07 in absolute value), consistent with the timing assumption. The full demand shock level $z_{1,jt}$ is also uncorrelated with investment (all below 0.05), confirming that the inventory-based proxy does not reflect investment-induced demand variation.

Exogeneity Check: Correlation of Demand Proxies with Productivity Residuals

A potential concern is that the inventory-based demand proxy $z_{1,jt}$ might reflect strategic inventory management by high-productivity firms, creating correlation between z_{jt} and ω_{jt}. Under Kumar and

Figure 11: Distribution of Firm-Level AR(1) Coefficients for the Demand Shock Proxy (z_{it})



Note: This figure shows the distribution of the AR(1) coefficients ($\hat{\rho}$) for the demand shock proxy, estimated for each of the 5,013 firms in the six target industries. The red dashed line indicates the mean of the coefficients (0.026). The null hypothesis ($\rho = 0$) was rejected at the 5% significance level for 13.7% of the firms. The fact that the distribution is centered around zero and that firms with statistically significant autocorrelation are a minority supports my assumption that the proxy variable is serially uncorrelated.

Table 8: Timing Assumption Check: Correlation of Demand Innovation with Investment

Industry	$\text{cor}(z_{3,jt},\, i_{jt})$	$\text{cor}(z_{1,jt},\, i_{jt})$
Corrugated board boxes	−0.064	−0.013
Mech. power transmission	+0.057	+0.041
Machinery and equipment	+0.008	+0.017
Automobile parts	−0.044	+0.014
Food and related products	−0.020	+0.011
Motor vehicles	+0.005	+0.017

Note: $z_{3,jt}$ is the AR(1) innovation of the demand shock proxy (empirical proxy for v_{jt}). i_{jt} is log investment. All correlations are small in magnitude (below 0.07), consistent with the maintained timing assumption that investment is committed before the within-period demand shock realizes.

Zhang (2019)'s interpretation, $z_{1,jt}$ is the *residual* from a firm-fixed-effects regression of log(sales/production): firm-specific inventory policies are absorbed by the fixed effects. The residual $z_{1,jt}$ therefore reflects realized deviations from each firm's own expected inventory ratio — an unanticipated demand surprise. The further purification to $z_{3,jt}$ (the AR(1) innovation of $z_{1,jt}$) removes any remaining predictable component, including any mean-reversion in firm-level inventory behavior. The exclusion restriction requires that the demand shock proxy $z_{3,jt}$ (the AR(1) innovation) be orthogonal to firm productivity ω_{jt}. I test this by computing the correlation between each demand proxy and a first-stage estimate of ω_{jt}, constructed as $\hat{\omega}_{jt} \approx \hat{\phi}_{jt} - \hat{\beta}_k k_{jt}$ where $\hat{\phi}_{jt}$ is the first-stage OLS residual.

Table 9: Correlation of Demand Proxies with Estimated Productivity

Industry	$\text{cor}(z_{1,jt}, \hat{\omega}_{jt})$	$\text{cor}(z_{3,jt}, \hat{\omega}_{jt})$
Corrugated board boxes	−0.004	−0.060
Mech. power transmission	+0.137	+0.088
Machinery and equipment	+0.061	+0.056
Automobile parts	−0.019	−0.057
Food and related products	+0.005	−0.007
Motor vehicles	−0.007	−0.018

Note: $z_{1,jt}$ is the full inventory-based demand shock; $z_{3,jt}$ is its AR(1) innovation (the primary exclusion restriction). $\hat{\omega}_{jt}$ is approximated from first-stage OLS residuals. All correlations are small in magnitude, supporting the exogeneity of the demand proxies with respect to firm productivity. The largest value is 0.137 for Mechanical power transmission with $z_{1,jt}$, while the purer instrument $z_{3,jt}$ shows 0.088 for the same industry.

All correlations are close to zero, providing support for the exogeneity assumption. The largest correlation (0.137) is for Mechanical power transmission using the full shock $z_{1,jt}$; the AR(1) innovation $z_{3,jt}$ reduces this to 0.088, confirming that filtering the predictable component improves exogeneity as expected.

One limitation of this check is that $\hat{\omega}_{jt}$ is constructed from the same first-stage OLS regression that includes $z_{3,jt}$ as an instrument in the proposed method; the correlation between $z_{3,jt}$ and a productivity estimate derived from that same regression may be mechanically small. As a robustness check, I also compute $\text{cor}(z_{3,jt}, \hat{\omega}_{jt}^{OLS})$ where $\hat{\omega}_{jt}^{OLS} = y_{jt} - \hat{\beta}_l^{OLS} l_{jt} - \hat{\beta}_k^{OLS} k_{jt}$ is the residual from a simple OLS production function (without any selection correction). This estimate is independent of the proposed estimator's first stage and provides an external benchmark. The OLS-based correlations range from −0.151 (Automobile parts) to +0.146 (Mechanical power transmission), similarly small in magnitude to those in Table 9, confirming that the near-zero correlations are not an artefact of using the same regression for both $z_{3,jt}$ and $\hat{\omega}_{jt}$.

F Verifying the Relevance of Demand Shock Variables in the Survival Probability Model

The foundation of my identification strategy is that the exclusion restriction holds: variables related to the demand shock $(z_{1,jt}, z_{2,jt}, z_{3,jt})$ have explanatory power for the firm's exit probability (selection) while being independent of past productivity. This appendix empirically tests the first part of this premise: whether the demand shock variables indeed have significant explanatory power for the firm's survival probability, using my empirical data.

While the GMM estimation uses a flexible polynomial approximation in a logit model for the survival probability, P_{jt}, here I use a Linear Probability Model to interpret more directly the relationships between the variables. I estimate the following two models for comparison:

1. **Proposed Method**: A model explaining the firm's survival probability using capital (k_{jt}), investment (inv_{jt}), and my exclusion restriction variables: the demand shock ($z_{1,jt}$), inventory level ($z_{2,jt}$), and the demand shock innovation ($z_{3,jt}$).

2. **Standard OP Method**: A model explaining the survival probability using only capital (k_{jt}) and investment (inv_{jt}), following the standard OP framework.

Tables 11 through 16 present the estimation results for each of the six industries analyzed. All regressors (k_{jt}, inv_{jt}, $z_{1,jt}$, $z_{2,jt}$, $z_{3,jt}$) are in logarithms; LPM coefficients are marginal effects on the linear probability of survival, expressed in probability units (not percentage points). The results are heterogeneous across industries. For Corrugated board boxes and Motor vehicles, the demand shock innovation ($z_{3,jt}$) and its interaction with capital are individually significant at the 5% level (see Tables 11–16), providing coefficient-level first-stage evidence. The joint ($\chi^2(6)$) LR test against the baseline model is less conclusive: Mechanical power transmission rejects at $p = 0.091$, while Corrugated board boxes ($p = 0.190$) and Motor vehicles ($p = 0.292$) do not reject at conventional levels. For Machinery and equipment, the demand shock innovation ($z_{3,jt}$) is individually significant at the 5% level in the restricted specification containing $z_{3,jt}$ and $k \times z_{3,jt}$ only; the focused ($\chi^2(2)$) LR test rejects at $p = 0.091$. For Automobile parts and Food and related products, neither individual coefficients nor any LR test achieve conventional significance, indicating weak first-stage relevance in those industries.

The discrepancy between individual and joint significance for Corrugated board boxes and Motor vehicles reflects the limited number of exits (18 and 110 respectively) relative to the six parameters added in the joint test: with six degrees of freedom, the LR test has low power when individual predictors are moderately strong but jointly collinear. The focused LR test (2 degrees of freedom, $z_{3,jt}$ and $k \times z_{3,jt}$ only) is the theoretically motivated test for the exclusion restriction, and its p-values are also reported in Table 10.

This heterogeneity is consistent with the all-industry regression finding (Section 6) that the AR(1) persistence of demand shocks is the only statistically significant predictor of correction magnitude: the proposed estimator corrects for selection bias where the demand shock has sufficient predictive power for exit decisions, and makes little difference where it does not. Applied researchers should therefore treat first-stage significance of the demand variables — or the AR(1) persistence diagnostic $\hat{\rho}$ — as a check on whether the proposed correction is operative before interpreting differences between the proposed and baseline estimates.

Table 10 summarises the fit statistics for the six industries.

Table 10: Survival Model Fit Statistics

		McFadden R^2		LR test ($\chi^2(2)$, focused[b])		
Industry	N	Proposed	Baseline	Stat	p-value	$\chi^2(6)$ p^a
Corrugated board boxes	826	0.121	0.071	2.54	0.281	0.190
Mech. power transmission	277	0.264	0.003	5.19	0.075	0.091
Machinery and equipment	593	0.119	0.056	4.80	0.091	0.508
Automobile parts	521	0.200	0.161	0.74	0.691	0.661
Food and related products	1120	0.068	0.065	0.63	0.731	0.985
Motor vehicles	4437	0.055	0.048	1.91	0.385	0.292

Notes: N is the first-stage estimation sample after removing the last observation year (no survival indicator available) and observations with missing demand or inventory variables; it differs from the full estimation sample N_{obs} in Table 2. McFadden $R^2 = 1 - \log L(\text{model})/\log L(\text{null})$. (b) Focused LR test for the exclusion restriction only: $z_{3,jt}$ (demand shock innovation) and $k \times z_{3,jt}$ (2 restrictions). This is the theoretically motivated test since $z_{3,jt}$ is the instrument. The LR statistic follows $\chi^2(2)$ under the null. (a) Joint LR test that all six demand variables add no explanatory power beyond capital and investment (6 restrictions). Note: there is no direct analog of the Stock–Yogo weak-IV F-statistic for binary outcome (logit) first-stage models; the LR statistic is the appropriate diagnostic for relevance in this setting.

Table 11: LPM Estimation Results for Survival Probability (Corrugated Board Boxes)

	Dependent variable:	
	Proposed	OP Baseline
	(1)	(2)
Capital (k_{jt})	-1.4365	0.3594^{*}
	(1.0482)	(0.1880)
Investment (inv_{jt})	21.3330	
	(21.1443)	
Demand Shock ($z_{1,jt}$)	-2.1712	
	(1.8614)	
Inventory ($z_{2,jt}$)	0.2595^{*}	
	(0.1464)	
AR(1) Innovation (v_{jt})	6.7021^{*}	
	(3.8084)	
$k \times z_{1,jt}$	0.2206	0.2714^{*}
	(0.1513)	(0.1514)
$k \times z_{2,jt}$	-2.5488	
	(1.5756)	
$k \times v_{jt}$	-69.3825^{*}	
	(40.4539)	
Constant	15.7548	-2.0558
	(11.1434)	(1.7006)
Observations	826	826

G GMM Moment Conditions for the Proposed Estimator

The proposed estimator for the capital elasticity parameter, β_k, is obtained by the Generalized Method of Moments (GMM). The second-stage estimation equation is:

$$\hat{\phi}_{jt} = \beta_k k_{jt} + g(\omega_{j,t-1}, P_{jt}) + \xi_{jt} \tag{31}$$

where $\omega_{j,t-1} = \hat{\phi}_{j,t-1} - \beta_k k_{j,t-1}$, and $\hat{\phi}_{j,t-1}$ is the estimator from the first stage. The key moment condition is derived from the assumption that the productivity innovation term, ξ_{jt}, is unpredictable using information from time $t-1$, i.e., $E[\xi_{jt}|J_{t-1}] = 0$.

Let the vector of unknown parameters be $\theta = (\beta_k, \gamma)$, where γ is the vector of parameters for the polynomial approximation of the non-parametric function $g(\cdot)$. The moment conditions are based on the orthogonality between the innovation term, ξ_{jt}, and instrumental variables belonging to the information set J_{t-1}. The instruments include functions of lagged state variables, such as polynomials of $k_{j,t-1}$ and $\omega_{j,t-1}(\beta_k)$. A general form of the moment conditions can be written as:

$$\mathbb{E}\left[\left(\hat{\phi}_{jt} - \beta_k k_{jt} - \sum_{p,q} \gamma_{pq}(\hat{\omega}_{j,t-1}(\beta_k))^p (P_{j,t-1})^q\right) \cdot W_{jt-1}\right] = 0 \tag{32}$$

where $W_{j,t-1}$ is a vector of instruments included in J_{t-1}, such as polynomials of $k_{j,t-1}$, $P_{j,t-1}$ and $\hat{\omega}_{j,t-1}$.

Implementation details. The polynomial degree for the second-stage approximation of $g(\cdot)$ is set to degree 5 in both $\hat{\omega}_{j,t-1}$ and $P_{j,t-1}$, including cross terms, giving approximately 15 basis functions.

Table 12: LPM Estimation Results for Survival Probability (Mechanical Power Transmission Equipment)

	Dependent variable:	
	Proposed	OP Baseline
	(1)	(2)
Capital (k_{jt})	-0.9007	-0.0561
	(2.5123)	(0.4482)
Investment (inv_{jt})	26.0406	
	(186.3210)	
Demand Shock ($z_{1,jt}$)	0.5617	
	(16.2617)	
Inventory ($z_{2,jt}$)	0.0799	
	(0.2520)	
AR(1) Innovation (v_{jt})	-11.2992	
	(20.1948)	
$k \times z_{1,jt}$	0.3149	0.1263
	(0.5110)	(0.3710)
$k \times z_{2,jt}$	-1.1256	
	(2.9831)	
$k \times v_{jt}$	105.8202	
	(224.9253)	
Constant	14.6916	3.7397
	(28.1223)	(3.5160)
Observations	277	277

Cross-validation over degrees 3–7 (holding the industry fixed) shows that degree 5 minimizes the out-of-sample prediction error in the second stage; degrees 4 and 6 give qualitatively identical $\hat{\beta}_k$ estimates (within 0.003 in all six industries). All reported results use degree 5.

The GMM objective is minimized using `L-BFGS-B` with β_k constrained to $(0.001, 0.999)$. To avoid local minima, the optimization is initiated from eight starting values ($\beta_k \in \{0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$) and the solution with the lowest objective value is retained. In the two-step GMM, the first step uses the identity weighting matrix and the second step uses the firm-clustered sandwich matrix $\hat{S} = \sum_j g_j g_j'/n$. Convergence is declared when the gradient norm falls below 10^{-5}. All six industry estimates achieve convergence from at least five of the eight starting values, confirming that the global optimum is found.

The key to identification is that the second argument of the function $g(\cdot)$, the survival probability P_{jt}, has a source of exogenous variation that is independent of the first argument, $\omega_{j,t-1}$. In my model, z_{jt} is independent of all variables in J_{t-1}, and thus it serves as an exclusion restriction that affects only $P_{j,t-1}$. This exogenous variation from z_{jt} allows for the separate identification of the effects of the non-parametric function $g(\cdot)$ and the linear term $\beta_k k_{jt}$, thereby achieving identification of the parameters.

Asymptotic distribution. This paper establishes point identification (Theorem 7) and consistency (Monte Carlo evidence, Section 5), but does not derive a formal $\sqrt{n}$-rate convergence result or the asymptotic distribution of $\hat{\beta}_k$. The estimator involves a nonparametric first stage (logit propensity score), a generated regressor $\hat{\phi}_{jt}$, and a second-stage GMM. The asymptotic theory for this class of multi-step semiparametric estimators, including the correction for generated-regressor variance in the

Table 13: LPM Estimation Results for Survival Probability (Machinery and Equipment)

	Dependent variable:	
	Proposed	OP Baseline
	(1)	(2)
Capital (k_{jt})	0.3303	0.3780
	(0.7588)	(0.3094)
Investment (inv_{jt})	17.5400	
	(51.6564)	
Demand Shock ($z_{1,jt}$)	−1.5544	
	(4.7081)	
Inventory ($z_{2,jt}$)	−0.0003	
	(0.0798)	
AR(1) Innovation (v_{jt})	−9.1241	
	(5.6364)	
$k \times z_{1,jt}$	0.2320	0.2297
	(0.2383)	(0.2345)
$k \times z_{2,jt}$	0.1451	
	(0.8098)	
$k \times v_{jt}$	109.0592*	
	(63.8866)	
Constant	−2.1070	−1.5204
	(7.4788)	(2.7646)
Observations	593	593

two-step sandwich matrix, is developed in Newey and McFadden (1994) and Andrews (1994). A formal derivation of the efficient asymptotic variance for the proposed estimator is beyond the scope of this paper and is left for future work. The bootstrap confidence intervals reported in Table 4 propagate estimation uncertainty from all steps (Section 6) and are used as a practical approximation; their validity as first-order approximations to the asymptotic distribution is supported by the regularity conditions in Chen, Linton, and Van Keilegom (2003), but not formally verified here.

H Extension to ACF and GNR Estimators

While the preceding analysis operated within the framework of Olley and Pakes (1996) (OP), contemporary empirical research frequently employs the estimators proposed by Ackerberg, Caves, and Frazer (2015) (ACF) and Gandhi, Navarro, and Rivers (2020) (GNR) to address functional dependence issues in flexible inputs. In this section, I demonstrate that the identification challenge highlighted by Hahn, Liao, and Ridder (2023) (HLR) persists in these modern frameworks. I then establish that my identification strategy, which relies on unexpected demand shocks, is compatible with and extends to both the ACF and GNR estimators.

H.1 The HLR Identification Problem in ACF and GNR

The ACF and GNR approaches share a common treatment of inputs: the intermediate input, m_{jt}, is modeled as a flexible choice, while capital, k_{jt}, and labor, l_{jt}, serve as state variables or quasi-fixed

Table 14: LPM Estimation Results for Survival Probability (Automobile Parts)

	Dependent variable:	
	Proposed	OP Baseline
	(1)	(2)
Capital (k_{jt})	0.5789	0.6036***
	(0.4660)	(0.2028)
Investment (inv_{jt})	−0.4518	
	(43.4201)	
Demand Shock ($z_{1,jt}$)	−0.0603	
	(3.9108)	
Inventory ($z_{2,jt}$)	−0.0200	
	(0.0670)	
AR(1) Innovation (v_{jt})	2.8935	
	(9.2242)	
$k \times z_{1,jt}$	0.2033	0.3256
	(0.2254)	(0.2152)
$k \times z_{2,jt}$	0.3922	
	(0.5741)	
$k \times v_{jt}$	−18.0956	
	(90.7892)	
Constant	−4.2972	−3.9417**
	(4.0112)	(1.9150)
Observations	521	521

inputs. Consider a standard gross output production function with a Cobb-Douglas specification:

$$y_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \beta_m m_{jt} + \omega_{jt} + \varepsilon_{jt} \tag{33}$$

The primary distinction between the two methods lies in the identification of the intermediate input elasticity, β_m. ACF controls for ω_{jt} by inverting the intermediate input demand function, $m_{jt} = f_t(\omega_{jt}, k_{jt}, l_{jt})$, though the identification of β_m often relies on price variation or specific functional form assumptions, such as a Leontief structure for value-added. In contrast, GNR exploits the first-order condition (share equation) derived from profit maximization to identify β_m and ε_{jt} prior to estimating the production function parameters.

Once β_m and ε_{jt} are identified, the remaining estimation problem converges. Defining the output variable net of intermediate input contributions and ex-post shocks as $\tilde{y}_{jt} \equiv y_{jt} - \beta_m m_{jt} - \varepsilon_{jt}$, both methods reduce to estimating the following equation:

$$\tilde{y}_{jt} = \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} \tag{34}$$

Identification of the parameters (β_l, β_k) relies on moment conditions derived from the first-order Markov process of productivity. Letting ξ_{jt} denote the innovation in productivity, the standard moment condition is:

$$\mathbb{E}\left[\xi_{jt}(\beta_l, \beta_k) \mid J_{t-1}\right] = \mathbb{E}\left[\omega_{jt}(\beta_l, \beta_k) - \mathbb{E}[\omega_{jt} \mid \omega_{j,t-1}, \chi_{jt} = 1] \middle| J_{t-1}\right] = 0 \tag{35}$$

Table 15: LPM Estimation Results for Survival Probability (Food and Related Products)

	Dependent variable:	
	Proposed	OP Baseline
	(1)	(2)
Capital (k_{jt})	0.2456	0.2504*
	(0.2387)	(0.1294)
Investment (inv_{jt})	−1.3476	
	(26.4764)	
Demand Shock ($z_{1,jt}$)	0.0357	
	(2.5289)	
Inventory ($z_{2,jt}$)	0.0058	
	(0.0315)	
AR(1) Innovation (v_{jt})	−1.7785	
	(3.3972)	
$k \times z_{1,jt}$	0.2711**	0.2739**
	(0.1326)	(0.1283)
$k \times z_{2,jt}$	−0.0880	
	(0.3080)	
$k \times v_{jt}$	18.5864	
	(35.4333)	
Constant	−0.6724	−0.9689
	(2.1919)	(1.0959)
Observations	1,120	1,120

where $\omega_{jt}(\beta_l, \beta_k) = \tilde{y}_{jt} - \beta_l l_{jt} - \beta_k k_{jt}$. In the presence of endogenous exit, the conditional expectation term becomes a function $g(\omega_{j,t-1}, P_{j,t-1})$ of prior productivity $\omega_{j,t-1}$ and the survival probability $P_{j,t-1}$.

It is at this stage that the HLR critique applies. If capital accumulates according to a deterministic perpetual inventory method (PIM) such that $k_{jt} = (1-\delta)k_{j,t-1} + i_{j,t-1}$, then k_{jt} is fully determined by variables in the information set J_{t-1} (specifically, $\omega_{j,t-1}$ and $k_{j,t-1}$). Consequently, perfect collinearity arises between k_{jt} and the selection correction term $g(\cdot)$, rendering β_k and β_l unidentified.

A distinctive feature of ACF and GNR relative to OP is that labor, l_{jt}, is treated as a quasi-fixed or dynamic input (a state variable in the firm's optimization problem) rather than a fully flexible input. This means that l_{jt} is also determined prior to the realization of ω_{jt}, and, like capital, it influences and is influenced by the firm's survival decision. Consequently, the selection bias in ACF and GNR need not manifest exclusively as a downward bias in $\hat{\beta}_k$; it can also manifest as a bias in $\hat{\beta}_l$, or be distributed across both parameters. Specifically, if firms with large labor inputs are less likely to exit (because high employment proxies for high productivity), then the uncorrected conditional expectation $\mathbb{E}[\omega_{jt} \mid \omega_{j,t-1}, \chi_{jt} = 1]$ will depend on l_{jt} as well as k_{jt}, biasing both $\hat{\beta}_k$ and $\hat{\beta}_l$ upward relative to the true parameters. The proposed PS correction, which conditions the survival probability on the demand shock z_{jt} as an exclusion restriction, addresses this joint bias, and the empirical results confirm that the correction operates across both dimensions of the factor elasticity vector (β_k, β_l).

Table 16: LPM Estimation Results for Survival Probability (Motor Vehicles Parts and Accessories)

	Dependent variable:	
	Proposed	OP Baseline
	(1)	(2)
Capital (k_{jt})	0.1166	0.2096***
	(0.1371)	(0.0771)
Investment (inv_{jt})	−24.9527**	
	(11.1287)	
Demand Shock ($z_{1,jt}$)	2.3927**	
	(0.9852)	
Inventory ($z_{2,jt}$)	0.0109	
	(0.0157)	
AR(1) Innovation (v_{jt})	−2.7247**	
	(1.3823)	
$k \times z_{1,jt}$	0.1872***	0.1987***
	(0.0632)	(0.0608)
$k \times z_{2,jt}$	−0.0811	
	(0.1570)	
$k \times v_{jt}$	31.7931**	
	(15.7728)	
Constant	0.5094	−0.2733
	(1.2654)	(0.6200)
Observations	4,437	4,437

H.2 Resolution via the Proposed Strategy

My identification strategy, which utilizes unexpected demand shocks as exclusion restrictions, integrates seamlessly with the ACF and GNR frameworks.

Step 1: Preservation of Static Relationships (Invertibility / Share Equation)

Under my timing assumption (Assumption 3), the intermediate input m_{jt} is chosen (Step 2) **before** the unexpected demand shock v_{jt} is realized (Step 5). This ordering protects the validity of the first-stage identification in both ACF and GNR.

- **ACF (Invertibility)**: The intermediate input demand function $m_{jt} = f_t(\omega_{jt}, k_{jt}, l_{jt}, z_{j,t-1})$ does not contain the unrealized shock v_{jt}. Consequently, the strict monotonicity with respect to the scalar unobservable ω_{jt} is preserved. The inversion $\omega_{jt} = f_t^{-1}(\cdot)$ remains valid, allowing for the control of productivity and the identification of ε_{jt} as in the standard ACF procedure.
- **GNR (Share Equation)**: The GNR approach relies on the first-order condition with respect to the flexible input, $P_t^Y \frac{\partial Y}{\partial M} = P_t^M$. Since v_{jt} is unobserved at the time m_{jt} is chosen, it does not enter the firm's optimization condition. Thus, the identification of β_m and ε_{jt} via the share equation remains feasible under my specification.

Step 2: Moment Conditions with Exclusion Restrictions

In the second stage, I employ the realized demand shifter z_{jt} (updated by v_{jt}) as an instrument. The survival probability $P_{j,t-1}$ depends on z_{jt} because the exit decision occurs (Step 7) after the shock is observed:

$$P_{j,t-1} = \Pr(\chi_{jt} = 1 \mid \omega_{j,t-1}, k_{jt}, z_{jt}) \tag{36}$$
$$= \Pr(\chi_{jt} = 1 \mid f_{t-1}^{-1}(\cdot), k_{jt}, z_{jt}) \tag{37}$$
$$= \Pr(\chi_{jt} = 1 \mid m_{j,t-1}, k_{j,t-1}, l_{j,t-1}, z_{j,t-2}, k_{jt}, z_{jt}) \tag{38}$$

where line (3) substitutes $\omega_{j,t-1} = f^{-1}(m_{j,t-1}, k_{j,t-1}, l_{j,t-1})$ from the ACF invertibility condition. In my empirical implementation, I use k_{jt}, $z_{j,t-2}$, and $\text{inv}_{j,t-2}$ as Baseline PS regressors, excluding $m_{j,t-1}$ to avoid contaminating the Proposed exclusion restriction: since $m_{j,t-1}$ correlates with z_{jt} (the key exclusion variable), including it in the Baseline PS would absorb identifying variation and eliminate the Proposed–Baseline contrast. Conversely, the capital stock k_{jt} is determined by investment decisions made prior to the realization of v_{jt} (Step 4) and is therefore orthogonal to the shock. This implies that the variation in the selection correction term $g(\omega_{j,t-1}, P_{j,t-1})$ induced by z_{jt} is independent of k_{jt}, thereby restoring the rank condition.

Accordingly, the structural parameters (β_l, β_k) can be identified in ACF and GNR frameworks by estimating the following moment condition, which corrects for selection bias:

$$\mathbb{E}\left[\{\omega_{jt}(\beta_l, \beta_k) - g(\omega_{j,t-1}, P_{j,t-1})\} \otimes \begin{pmatrix} k_{jt} \\ l_{j,t-1} \\ P_{j,t-1} \end{pmatrix}\right] = 0 \tag{39}$$

In conclusion, the theoretical argument establishes that my proposed methodology is applicable to both the ACF and GNR frameworks. Empirically, applying the estimator to the six Japanese manufacturing industries yields results consistent with the theoretical prediction for ACF: in 5 of 6 industries, the proposed method raises the capital elasticity estimate relative to the baseline (average correction +0.025). Extending to all industries, with $\hat{\beta}_l$ fixed at the baseline estimate to isolate the effect on $\hat{\beta}_k$, the proposed method raises $\hat{\beta}_k$ in 180 of 368 valid industries (49%): the correction effect that is clear in the six focal industries is diluted in the full sample by industries where ACF identification is weak (many boundary estimates). The six-industry results should therefore be viewed as illustrative of the mechanism, with the cross-industry OP analysis (Section 6.3) providing the primary statistical evidence.

A Monte Carlo experiment with a value-added DGP ($y = 0.5l + 0.5k + \omega + \eta$, with materials demand $m = f(\omega, k, l)$ invertible in ω serving as the ACF proxy variable, and the same AR(1) demand shock and exit rule as in the OP simulation) confirms the prediction. Across 200 replications ($J = 200$, $T = 50$), the proposed estimator yields a $\hat{\beta}_k$ closer to the true value than the baseline in 84% of cases, with 42% lower RMSE (0.20 vs. 0.34). The RTS estimate ($\hat{\beta}_k + \hat{\beta}_l$) likewise improves: 57% lower RMSE (0.19 vs. 0.45) with the proposed RTS closer to the true value in 75% of replications. The proposed estimator also dominates the no-selection estimator: it yields $\hat{\beta}_k$ closer to the truth in 57% of cases and RTS closer in 66%, confirming that the selection correction is beneficial even in the ACF framework where (β_k, β_l) are jointly identified.

For GNR, the picture is more nuanced and deserves careful interpretation. Unlike OP, where β_l is identified in the first stage and the second-stage GMM targets β_k alone, ACF and GNR jointly estimate (β_k, β_l) in the second stage. The selection bias induced by the HLR identification problem therefore need not manifest exclusively as a downward bias in β_k; it can distribute across both parameters in directions determined by the curvature of the GMM objective surface.

In the GNR results, the proposed correction tends to *raise* β_l while lowering β_k in most industries. However, the sum $\hat{\beta}_k + \hat{\beta}_l$ (i.e., the estimated returns to scale) increases in four of six industries under the proposed method (Corrugated: 0.34 → 0.47; Machinery: 0.10 → 0.36; AutoParts: 0.73 → 0.74; Food: 0.46 → 0.54), while two industries (Mechanical Power Transmission and Motor Vehicles Parts) show a slight decrease in RTS under the proposed correction. This suggests that the selection correction is operating in the correct direction in most cases: it raises the total estimated factor elasticity in four of six industries, but the correction is channeled through $\hat{\beta}_l$ rather than $\hat{\beta}_k$ due to the relatively flat GMM objective surface in the capital direction. This redistribution reflects the weaker identification of β_k in gross-output specifications compared to value-added (ACF) or investment-proxy (OP) specifications. I view this as a property of the GNR identification environment rather than a failure of my correction. Applied researchers should therefore prefer the OP implementation when the primary object of interest is $\hat{\beta}_k$ alone, or the ACF implementation when joint estimation of $(\hat{\beta}_k, \hat{\beta}_l)$ is required. The GNR framework is most useful when the researcher needs the material elasticity $\hat{\beta}_m$ from the share equation; in that setting the relevant summary statistic is returns to scale $(\hat{\beta}_k + \hat{\beta}_l)$, which improves in four of six industries under the proposed method.

Table 17 reports the full set of estimated factor elasticities $(\hat{\beta}_k, \hat{\beta}_l)$ and the implied returns to scale $\hat{\beta}_k + \hat{\beta}_l$ for both ACF and GNR.

Table 17: Factor Elasticities under ACF and GNR: Proposed vs. Baseline

	ACF (Value-Added)			GNR (Gross Output)		
Industry	$\hat{\beta}_k$	$\hat{\beta}_l$	RTS	$\hat{\beta}_k$	$\hat{\beta}_l$	RTS
Proposed method						
Corrugated board boxes	0.300	0.144	0.444	0.032	0.437	0.469
Mech. power transmission	0.257	0.000	0.257	0.055	0.813	0.869
Machinery and equipment	0.310	0.531	0.841	0.057	0.308	0.364
Automobile parts	0.402	0.000	0.402	0.000	0.742	0.742
Food and related products	0.319	0.702	1.022	0.089	0.455	0.544
Motor vehicles	0.350	0.291	0.642	0.000	0.683	0.683
OP Baseline (standard selection correction)						
Corrugated board boxes	0.259	0.862	1.121	0.057	0.283	0.340
Mech. power transmission	0.224	0.999	1.223	0.047	0.854	0.902
Machinery and equipment	0.301	0.419	0.720	0.098	0.000	0.098
Automobile parts	0.360	0.089	0.449	0.073	0.660	0.733
Food and related products	0.304	0.814	1.118	0.062	0.401	0.463
Motor vehicles	0.361	0.091	0.451	0.000	0.706	0.706

Note: RTS $= \hat{\beta}_k + \hat{\beta}_l$ denotes the estimated returns to scale. Values of 0.000 or 0.999 indicate that the estimate hit the boundary of the parameter space ($\beta \in [0.001, 0.999]$), reflecting weak identification of that parameter in the 2nd-stage GMM. For ACF, the proposed method raises $\hat{\beta}_k$ in 5 of 6 industries relative to the baseline (Motor Vehicles being the exception). In Mechanical Power Transmission under ACF, the proposed method pushes $\hat{\beta}_l$ to the lower boundary (0.000) while the baseline pushes it to the upper boundary (0.999): these are mirror corner solutions reflecting a flat ACF objective surface in β_l for this small industry ($N = 277$) and should not be interpreted as structural estimates. For GNR, the correction is channeled primarily through $\hat{\beta}_l$, with RTS increasing in 4 of 6 industries, consistent with the theoretical prediction that standard OP-type methods underestimate total factor responsiveness when selection bias is present. The two exceptions (Mechanical Power Transmission and Motor Vehicles Parts) feature boundary estimates of $\hat{\beta}_k = 0.000$ under the proposed method, reflecting weak identification of capital in GNR's gross-output specification. The proposed method introduces an additional instrument (z_{jt}) that is correlated with the exit probability but not with capital; in industries with low capital intensity variation, this can sharpen identification of β_l at the cost of pushing β_k toward a corner solution. This is not an inconsistency but a manifestation of the GNR moment conditions being nearly collinear for capital in these industries: adding the exit-probability variation does not resolve collinearity with capital when capital variation is inherently limited. I flag these two industries as cases where the proposed correction should be interpreted with caution in the GNR specification.