ONLINE SUPPLEMENTARY MATERIALS FOR "ESTIMATING CORPORATE INVESTMENT EFFICIENCY WITH BIAS CORRECTION: A SEMIPARAMETRIC PANEL MODEL APPROACH"

TAINING WANG

International School of Economics and Management
Capital University of Economics and Business
Beijing, 100070, China.
taining.wang@cueb.edu.cn

ZHAO WANG

International School of Economics and Management
Capital University of Economics and Business
Beijing, 100070, China.
zhaowang@cueb.edu.cn

FENG YAO

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
feng.yao@mail.wvu.edu

SUBAL C. KUMBHAKAR

Department of Economics
Binghamton University
Binghamton, NY 13902, USA
email: kkar@binghamton.edu
kkar@binghamton.edu

Abstract This appendix provides supporting materials for "Estimating Corporate Investment Efficiency with Bias Correction: A Semiparametric Panel Model Approach." *Appendix 1* presents the multi-step estimator for a general form of SEM-SP in *Appendix 1.1*, discusses asymptotic properties of the estimator in *Appendix 1.2*, and illustrates the finite-sample performance of the estimator through simulation studies in *Appendix 1.3*. A robust version of SEM-SP is proposed in *Appendix 1.4*. *Appendix 2* presents the implementation procedures of three test statistics for correct specifications on efficiency functions and composite error of SEM-SP. *Appendix 2.1* provides a test statistic for the null where a parametric efficiency function is correctly specified; *Appendix 2.2* provides a test statistic for the null where only under-investment exists due to financial constraints; and *Appendix 2.3* provides a test statistic for the null where the three-regime composite error under SEM-SP is correctly specified. The numeric properties of all the three tests are illustrated through Monte-Carlo studies. *Appendix 3* provides descriptive data summary in application section. *Appendix 4* presents the complete empirical results comparing SEM-SP, SFM-SP, and SFM-NP, as discussed in Section 4.3 of the paper. *Appendix 5* provide instructions for the data and code used in this study for replications.

Keywords: Investment efficiency, semiparametric panel model, three-regime composite error, additive model, fixed effects.

Appendix 1: Construction and Properties of SEM-SP Estimator

In this section, we consider a general form of SEM-SP. We detail the estimation procedure in Appendix 1.1, assumptions for asymptotic properties in Appendix 1.2, and finite sample performance in Appendix 1.3.

Appendix 1.1: Estimation procedure of SEM-SP

Our SEM-SP can be expressed in a general fashion as

$$I_{it} = m_0 + \alpha_i + \eta_t + \sum_{j=1}^{d} m_j(X_{j,it}) + \epsilon_{it}, \quad i = 1, ..., n, \ t = 1, ..., T, \tag{A.1}$$

where the investment I_{it} of firm i at time t is influenced by a global constant m_0, firm fixed effect α_i, time fixed effect η_t, and a total of d efficiency variables $X_{j,it}$ with its individual unknown efficiency function $m_j(\cdot)$, for $j = 1, \ldots, d$ (i.e., in our application, $X_{1,it} = PQ_{it}$, $X_{2,it} = CS_{it}$, and $X_{3,it} = PS_{it}$, so $d = 3$). To identify $m_j(\cdot)$, we impose the condition $m_j(X_{j,it} = 0) = 0$ (Li, 2000), and to identify m_0, we require $\sum_{i=1}^{n} \alpha_i = \sum_{t=1}^{T} \eta_t = 0$ (Baltagi, 2013). The composite error ϵ_{it} captures investment inefficiency under imperfect capital markets due to different financial frictions. We specify ϵ_{it} with the following structure

$$\epsilon_{it} = v_{it} - u_{it}\mathbb{1}(u_{it} > 0, e_{it} = 0) + e_{it}\mathbb{1}(u_{it} = 0, e_{it} > 0), \tag{A.2}$$

where v_{it} is the conventional unobserved randomness that captures deviation from efficient investment under perfect market; $u_{it} \geq 0$ represents the degree of under-investment due to financial constraints (i.e., the extent to which firm's investment falls below efficient level); and $e_{it} \geq 0$ reflects the degree of over-investment due to agency problem (i.e., the extent to which firm's investment rises above efficient level) [1] Notably, the composite error in (A.2) contains three regimes, from which firms may under-invest with probability $p_u(W_{it}^u; \gamma_0^u)$ ($\epsilon_{it} = v_{it} - u_{it}$); over-invest with probability $p_e(W_{it}^e; \gamma_0^e)$ ($\epsilon_{it} = v_{it} + e_{it}$); and efficiently invest with probability $1 - p_u(W_{it}^u; \gamma_0^u) - p_e(W_{it}^e; \gamma_0^e)$, where variables $W_{it}^u \in \Re^{q_u}$ and $W_{it}^e \in \Re^{q_e}$ are determinants of under-investment and over-investment (multi-logit) probability function $p_u(\cdot)$ and $p_e(\cdot)$, respectively, as defined in Section 2.2. Thus, our specification in (A.2) rules out the case where firms under-invest and over-invest concurrently by imposing the condition that $u_{it} > 0$ and $e_{it} > 0$ appear simultaneously with probability zero.

[1] We emphasize that a firm's efficient investment is defined as I_{it} under *perfect market*, which is determined only by X_j plus unobserved randomness $(\alpha_i, \eta_t, v_{it})$.

Given our assumptions in Section 2, let $W_{it} = (W_{it}^{u\top}, W_{it}^{e\top})^\top$ and recall that

$$E(\epsilon_{it}|\alpha_i, \eta_t, X_{it}, W_{it}) = -\left[\mu_0^u p_u(W_{it}^u; \gamma_0^u) - \mu_0^e p_e(W_{it}^e; \gamma_0^e)\right] \equiv -\pi(W_{it}; \mu_0, \gamma_0) \neq 0, \tag{A.3}$$

where $\mu_0 = (\mu_0^u, \mu_0^e)$ are positive constants and $\gamma_0 = (\gamma_0^{u\top}, \gamma_0^{e\top})^\top$ are unknown coefficients in inefficiency probabilities. With $\omega_0 = (\mu_0, \gamma_0^\top) \in \Re^{q_u+q_e+2}$ and $\tilde{\epsilon}_{it} = \epsilon_{it} + \pi(W_{it}; \omega_0)$, we have $E(\tilde{\epsilon}_{it}|\alpha_i, \eta_t, G_{it}) = 0$ by construction. For this reason, (A.1) can be transformed into a proper regression form as

$$I_{it} = m_0 + \alpha_i + \eta_t + \sum_{j=1}^{d} m_j(X_{j,it}) - \pi(W_{it}; \omega_0) + \tilde{\epsilon}_{it}. \tag{A.4}$$

Define $\eta_t = D_t^\top \eta$, where D_t is a $(T \times 1)$ zero vector except for the t^{th} position being one, and $\eta \equiv \eta_0 = [\eta_{0,1}, ..., \eta_{0,T}]^\top$. Since $\eta_{0,1} = -\sum_{s=2}^{T} \eta_{0,s}$ by identification condition, we replace η_0 with $\eta_{0,-1} = [\eta_{0,2}, ..., \eta_{0,T}]^\top$ which removes the first element $\eta_{0,1}$, and replace D_t with $D_{t,-1}$ defined similarly. Then, we eliminate firm fixed effect α_i through first-differencing as

$$\Delta I_{it} = \Delta D_{t,-1}^\top \eta_{0,-1} + \sum_{j=1}^{d} \Delta m_j(X_{j,it}) - \Delta\pi(W_{it}; \omega_0) + \Delta\tilde{\epsilon}_{it}, \tag{A.5}$$

where $\Delta\zeta_{it} = \zeta_{it} - \zeta_{it-1}$ for any random variable/function ζ_{it}.

We follow Wang et al. (2024) to implement a two-step estimator for our interested parameters ω_0, efficiency functions $\{m_j(\cdot)\}$, and derivative functions $\{m_j^{(1)}\}_{j=1}^3$. In the first step, we estimate ω_0 by a profile nonlinear least square estimator (PNLS) $\widehat{\omega}$, which involves series approximation on time fixed effects η_t and unknown functions $m_j(\cdot)$. Specifically, let $\mathcal{X}_j = [a_j, b_j] \subset \Re$ be a compact support of X_j for some finite constants a_j and b_j. Without loss of generosity, we consider $\mathcal{X}_j = \mathcal{X}$ for all j with $a_j = a = -1$ and $b_j = b = 1$. For any point $x_j \in \mathcal{X}$, we define the basis function $\phi^\kappa(x_j) = [\phi_1(x_j), \ldots, \phi_\kappa(x_j)]^\top$ and series coefficients $\theta_j \in \Re^\kappa$. Then, we approximate each efficiency function through $\Delta m_j(X_{j,it}) = m_j(X_{j,it}) - m_j(X_{j,it-1}) \approx (\phi^\kappa(X_{j,it}) - \phi^\kappa(X_{j,it-1}))^\top \theta_j \equiv \Delta\phi^\kappa(X_{j,it})^\top \theta_j$. Let $X_{it} = [X_{1,it}, \ldots, X_{d,it}]^\top$ and $\tilde{X}_{it} = [D_{t,-1}^\top, X_{it}^\top]^\top \in \Re^{T-1+d}$. Also, define $\Delta\Phi(\tilde{X}_{it}) = [\Delta D_{t,-1}^\top, \Delta\phi^\kappa(X_{1,it})^\top, \ldots, \Delta\phi^\kappa(X_{d,it})^\top]^\top$ be a $S(\kappa) \times 1$ vector, with $S(\kappa) = T - 1 + d\kappa$. Then, we approximate all efficiency functions, plus time effects, through $\Delta\eta_t + \sum_{j=1}^{d} \Delta m_j(X_j) \approx \Delta\Phi(\tilde{X}_{it})^\top \theta$, where $\theta = (\eta_{0,-1}^\top, \theta_1^\top, \ldots, \theta_d^\top)^\top$. Notably, our basis vector $\Delta\Phi(\tilde{X}_{it})$ shares a similar structure considered by Horowitz and Mammen (2004), who employ series estimator for one intercept in regression model. In our case, we employ the series estimator for $T-1$ intercepts to control for unobserved heterogeneities in time dimension, with T being fixed.

Let $N = n(T-1)$, we define $N \times 1$ vectors $\Delta I = \{\Delta I_{it}\}_{i=1,t=2}^{n,T}$ and $\Delta\pi(\omega) = \{\Delta\pi(W_{it}; \omega)\}_{i=1,t=2}^{n,T}$. Also, we define a $N \times S(\kappa)$ basis matrix $\Delta\Phi = \{\Delta\Phi(\tilde{X}_{it})^\top\}_{i=1,t=2}^{n,T}$. Minimizing the sum of squared

residuals, we estimate θ and ω_0 with

$$(\hat{\theta}, \hat{\omega}) = \underset{\{\theta,\omega\}}{argmin}\ [\Delta I - \Delta\Phi\theta + \Delta\pi(\omega)]^\top [\Delta I - \Delta\Phi\theta + \Delta\pi(\omega)],$$

where parameters θ and ω are additively separable. Therefore, for given ω, we obtain

$$\hat{\theta}(\omega) = [\Delta\Phi^\top \Delta\Phi]^{-1} \Delta\Phi^\top (\Delta I + \Delta\pi(\omega)). \tag{A.6}$$

Denote $M_{\Delta\Phi} = I_N - P_{\Delta\Phi} = I_N - \Delta\Phi(\Delta\Phi^\top\Delta\Phi)^{-1}\Delta\Phi^\top$ for I_N being an $N \times N$ identity matrix, we estimate ω with PNLS as

$$\hat{\omega} = \underset{\{\omega\in\Omega\}}{argmin}\ [\Delta I + \Delta\pi(\omega)]^\top M_{\Delta\Phi} [\Delta I + \Delta\pi(\omega)], \tag{A.7}$$

where $\Omega \subset \Re^{q_u+q_e+2}$ is a compact support of ω. From (A.6) and (A.7), we obtain profile series-based estimators $\hat{\theta} \equiv \hat{\theta}(\hat{\omega})$, providing estimates for time fixed effects $\hat{\eta}_t$ and estimates for functions $\hat{m}_j(x_j) = \phi^\kappa(x_j)^\top \hat{\theta}_j$ for $j = 1, \ldots, d$.

In the second step, we propose a one-step backfitting for $m_j(\cdot)$ and $m_j^{(1)}(\cdot)$ using kernel estimation to improve estimation efficiency and facilitate the inference. Based on (A.5), we observe the following regression model for function $m_j(\cdot)$ as

$$I_{it,-j} = m_j(X_{j,it}) + \tilde{\epsilon}_{it},$$

where $I_{it,-j} = \Delta I_{it} + m_j(X_{j,it-1}) - \sum_{l=1,l\neq j}^{d} \Delta m_l(X_{l,it}) + \Delta\pi(W_{it};\omega)$ is treated as a new dependent variable which collects all the differenced $\pi(W_{it};\omega_0)$ and efficiency functions except for $m_j(X_{it}^j)$. Using our estimates from the first step, we construct a feasible version of $I_{it,-j}$ as

$$\hat{I}_{it,-j} = \Delta I_{it} + \hat{m}_j(X_{j,it-1}) - \sum_{l=1,l\neq j}^{d} \Delta\hat{m}_l(X_{l,it}) + \Delta\pi(W_{it};\hat{\omega}).$$

For any point x_j from its compact support, we obtain the backfitting estimator for $m_j(x_j)$ with a local linear regression of $\hat{I}_{it,-j}$ on $X_{j,it}$ as $\tilde{m}_j(x_j) \equiv \tilde{a}$, and for $m_j^{(1)}(x_j)$ as $\tilde{m}_j^{(1)}(x_j) \equiv \tilde{b}$, by minimizing the following kernel weighted sum of squared residuals

$$(\tilde{a}, \tilde{b}) = \underset{\{a,b\}}{argmin} \sum_{i=1}^{n}\sum_{t=2}^{T} \left[\hat{I}_{it,-j} - a - (X_{j,it} - x_j)b\right]^2 K\left(\frac{X_{j,it} - x_j}{h_j}\right), \tag{A.8}$$

where $K(\cdot)$ is a univariate kernel function and h_j is the bandwidth for X_j. Finally, we can estimate the global intercept m_0 in (A.1) using using parameter estimates in (A.7) and backfitting estimates

in (A.8) by

$$\tilde{m}_0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(I_{it} - \sum_{j=1}^{d}\tilde{m}_j(X_{j,it}) + \pi(W_{it};\hat{\omega})\right). \tag{A.9}$$

Appendix 1.2: Assumptions for asymptomatic properties of SEM-SP estimator

Our SEM-SP is constructed as a modified version of a recently developed semiparametric additive stochastic frontier model (SFM-AM) by Wang et al. (2024). Thus, the SEM-SP is different from the SFM-AM in three aspects. First, our SEM-SP is applied to capital investment model, assuming that two distinct financial frictions may arise to cause inefficient investment. Correspondingly, we specify a three-regime composite error ϵ_{it} in (A.2), covering the case of under-investment ($\epsilon_{it} = v_{it} - u_{it}$), over-investment ($\epsilon_{it} = v_{it} + e_{it}$), or efficient investment ($\epsilon_{it} = v_{it}$) with corresponding probabilities. In contrast, the SFM-AM is applied to production model, assuming that firms produce inefficiently with certainty (i.e., output produced lower than its maximum given the same amount of inputs). As a result, SFM-AM specifies a one-regime composite error $\epsilon_{it} = v_{it} - u_{it}$, which occurs with probability one so that all firms are deemed inefficient. Second, our SEM-SP focuses on revealing individual effects of efficiency variables through functions $\{m_j(\cdot)\}_{j=1}^{d}$, but not on the interactive effects among the efficiency variables (i.e., Tobin's q and sales ratio) as they are not clearly documented from existing literature. SFM-AM specifies both individual and interactive functions of production inputs to accommodate complex structure of the underlying production function beyond a simple linear fashion. Thus, the frontier functions under SFM-AM reduces to the efficiency functions under SEM-SP when interactive effects among $\{X_j\}_{j=1}^{d}$ are not considered. Finally, our SEM-SP controls for latent heterogeneities in both firm and time dimension through additive fixed effects (α_i, η_t), whereas SFM-AM incorporates only firm fixed effects (α_i).

Below, we list out assumptions for the asymptotic characterizion of our proposed estimator based on assumptions **A**-**C** in Wang et al. (2024).

A1* (1) $\{(Y_{it}, X_{it}, W_{it}^u, W_{it}^e) : i = 1,\ldots,n,\ t = 1,\ldots,T\}$ is i.i.d. (identically and independently distributed) across $i = 1,\ldots,n$, and T is finite. The support of vector X_{it} is a compact subset of $\Re^d$, assumed to be $[-1,1]^d$. (2) The fixed effects are α_i and η_t with $E(\alpha_i) = E(\eta_t) = 0$. (3) v_{it} is i.i.d. across both i and t independent of other variables, with mean 0 and finite variance σ_v^2. (4) u_{it} and e_{it} are i.i.d. across $i = 1,\ldots,n$, satisfying $E(u_{it}|\alpha_i,\eta_t,X_{it},W_{it}) = \mu_0^u > 0$, $E(e_{it}|\alpha_i,\eta_t,X_{it},W_{it}) = \mu_0^e > 0$, $E(u_{it}^2|\alpha_i,\eta_t,X_{it},W_{it}) < C$, and $E(e_{it}^2|\alpha_i,\eta_t,X_{it},W_{it}) < C$ for all i and t, with $W_{it} = (W_{it}^u, W_{it}^e) \in \Re^{q_u+q_e}$. (5) For $\tilde{\epsilon}_{it} = \epsilon_{it} - \pi(W_{it};\omega_0)$, $E(\tilde{\epsilon}_{it} - \tilde{\epsilon}_{it-1}|X_{it}, X_{it-1}, W_{it}, W_{it-1}) = 0$.

A2* (1) For Ω, a compact subset of $\Re^{\mathbf{q}}$ with $\mathbf{q} = q_u + q_e + 2$, ω_0 is contained in the interior of Ω. (2) For $\omega \in \Omega$ and $\tilde{X}_{it} = (D_{t,-1}^\top, X_{it})$, define $Q(\omega) = \frac{1}{T}\sum_{t=1}^{T} E[\Delta\pi(W_{it};\omega) - h(\tilde{X}_{it}, \tilde{X}_{it-1};\omega) - (\Delta\pi(W_{it};\omega_0) - h(\tilde{X}_{it}, \tilde{X}_{it-1};\omega_0))]^2$, where $h(\tilde{X}_{it}, \tilde{X}_{it-1};\omega) = E_G(E(\Delta\pi(W_{it};\omega)|\tilde{X}_{it}, \tilde{X}_{it-1}))$.

$Q(\omega) = 0$ only when $\omega = \omega_0$. (3) $\pi(W_{it};\omega)$ is continuous in $\omega \in \Omega$ uniformly such that $|\pi(W_{it};\omega) - \pi(W_{it};\omega')| \leq B_\pi(W_{it})||\omega - \omega'||$, where $B_\pi(\cdot)$ is continuous and $EB_\pi^2(W_{it}) < \infty$, $\forall\ \omega, \omega' \in \Omega$. (4) $(\Delta\pi(W_{it};\omega) - \Delta\pi(W_{it};\omega_0))^2$, $\Delta\epsilon_{it}\Delta\pi(W_{it};\omega)$, $\Delta_1\pi_{it}(\omega)$, $(\Delta_1\pi_{it}(\omega))^2$, and $\Delta_1\pi_{it}(\omega)\Delta_1\pi_{it}(\omega_0)$ satisfy the Cramer's condition, where $\Delta_1\pi_{it}(\omega) = \Delta\pi(W_{it};\omega) - h(\tilde{X}_{it}, \tilde{X}_{it-1};\omega)$. (5) $E\sup_{\omega\in\Omega}|\pi(W_{it};\omega)|^2 < \infty$.

A3* (1) Define $S(\kappa) = T - 1 + d\kappa$, $\exists$ some $\delta > 0$ and a $S(\kappa)\times 1$ real vector B_h such that $\sup_{\omega\in\Omega}\sup_{X_{it}\in[-1,1]^d, X_{it-1}\in[-1,1]^d}|h(\tilde{X}_{it}, \tilde{X}_{it-1};\omega) - \Delta\Phi(\tilde{X}_{it})^\top B_h| = O(d\kappa^{-\delta})$. (2) $\exists$ some $\delta > 0$ and a $\kappa\times 1$ real vector B_{m_j} such that $\sup\limits_{x_j\in[-1,1]}|m_j(x_j) - \phi^\kappa(x_j)^\top B_{m_j}| = O(\kappa^{-\delta})$. Also, $\exists$ a $S(\kappa)\times 1$ real vector B_m such that $\sup_{X_{it}\in[-1,1]^d}|\eta_t + \sum_{j=1}^d m_j(X_{j,it}) - \Phi(\tilde{X}_{it})^\top B_m| = O(d\kappa^{-\delta})$ for all $t = 1, ..., T$. (3) $\kappa\to\infty$, and $\sqrt{n}(d\kappa^{-\delta}) \to 0$ as $n\to\infty$. (4) For some nonstochastic sequences ξ_κ and $\xi_{S(\kappa)} = 2d\xi_\kappa$, $\sup\limits_{x_j\in[-1,1]}||\phi^\kappa(x_j)|| \leq \xi_\kappa$ and $\sup_{X_{it}\in[-1,1]^d, X_{it-1}\in[-1,1]^d}||\Delta\Phi(\tilde{X}_{it})|| \leq \xi_{S(\kappa)}$. Furthermore, as $n\to\infty$, $\xi_{S(\kappa)}^2 S(\kappa)/n \to 0$, and $\xi_{S(\kappa)}^2(\ln(n))^{1/2}/n \to 0$. (5) The smallest eigenvalue of $E(\Delta\Phi(\tilde{X}_{it})\Delta\Phi(\tilde{X}_{it})^\top)$ is bounded away from zero for all t.

B1* For $\partial_j\pi_{it}(\omega) = \frac{\partial\pi(W_{it};\omega)}{\partial\omega_j}$, $\partial_{jl}\pi_{it}(\omega) = \frac{\partial^2\pi(W_{it};\omega)}{\partial\omega_j\partial\omega_l}$, $|\partial_j\pi_{it}(\omega) - \partial_j\pi_{it}(\omega')| \leq B_{\partial_j\pi}(W_{it})||\omega - \omega'||$, and $|\partial_{jl}\pi_{it}(\omega) - \partial_{jl}\pi_{it}(\omega')| \leq B_{\partial_{jl}\pi}(W_{it})||\omega - \omega'||$, for all ω, $\omega' \in \Omega$, $t = 1, \ldots, T$, j or $l = 1, \ldots, \mathbf{q}$ with $EB_{\partial_j\pi}^2(W_{it}) < \infty$ and $EB_{\partial_{jl}\pi}^2(W_{it}) < \infty$. The condition holds almost surely with respect to the probability measure associated with W_{it}.

B2* For $\partial_j h(X_{it}, X_{it-1};\omega) = E_\mathcal{G}(E(\partial_j\Delta\pi_{it}(\omega)|X_{it}, X_{it-1}))$, $\partial_{jl}h(X_{it}, X_{it-1};\omega) = E_\mathcal{G}(E(\partial_{jl}\Delta\pi_{it}(\omega)|X_{it}, X_{it-1}))$, where $\partial_j\Delta\pi_{it}(\omega) = \frac{\partial\Delta\pi(W_{it};\omega)}{\partial\omega_j}$, $\partial_{jl}\Delta\pi_{it}(\omega) = \frac{\partial^2\Delta\pi(W_{it};\omega)}{\partial\omega_j\partial\omega_l}$, $\exists$ some $\delta, \delta_1 > 0$, $B_{\partial_j h}$ and $B_{\partial_{jl}h}$, both $S(\kappa)\times 1$ real vectors, such that $\sup_{\omega\in\Omega}\sup_{X_{it}\in[-1,1]^d, X_{it-1}\in[-1,1]^d}|\partial_j h(X_{it}, X_{it-1};\omega) - \Delta\Phi(X_{it})' B_{\partial_j h}| = O(d\kappa^{-\delta})$ and $\sup_{\omega\in\Omega}\sup_{X_{it}\in[-1,1]^d, X_{it-1}\in[-1,1]^d}|\partial_{jl}h(X_{it}, X_{it-1};\omega) - \Delta\Phi(X_{it})' B_{\partial_{jl}h}| = O(d\kappa^{-\delta})$.

B3* Define $\partial_j\Delta_1\pi_{it}(\omega) = \partial_j\Delta\pi_{it}(\omega) - \partial_j h(X_{it}, X_{it-1};\omega)$ and $\partial_{jl}\Delta_1\pi_{it}(\omega) = \partial_{jl}\Delta\pi_{it}(\omega) - \partial_{jl}h(X_{it}, X_{it-1};\omega)$. For all $t = 2, \ldots, T$, $j, l \in \{1, \ldots, p\}$, $E(\partial_j\Delta_1\pi_{it}(\omega)\partial_l\Delta_1\pi_{it}(\omega))$, $E(\partial_{jl}\Delta_1\pi_{it}(\omega))^2$ and $E(\Delta_1\pi_{it}(\omega) - \Delta_1\pi_{it}(\omega_0))\partial_{jl}\Delta\pi_{it}(\omega)$ are finite $\forall\ \omega \in \Omega$ and continuous at $\omega_0 \in \Omega$.

B4* $\Sigma_T(\omega_0)$ is a $\mathbf{q}\times\mathbf{q}$ nonstochastic invertible matrix with its $(j,l)^{th}$ element as

$$\Sigma_{T,jl}(\omega_0) = \frac{2}{T-1}\sum_{t=2}^{T} E(\partial_j\Delta_1\pi_{it}(\omega_0)\partial_l\Delta_1\pi_{it}(\omega_0)).$$

B5* $E\sup_{\omega\in\Omega}|\partial_j\pi_{it}(\omega)|^2 < C$, $E\sup_{\omega\in\Omega}|\partial_{jl}\pi_{it}(\omega)|^2 < C$, $E[\sup_{\omega\in\Omega}|\partial_j\theta(X_{it}, X_{it-1};\omega)|]^2 < C$, and $E[\sup_{\omega\in\Omega}|\partial_{jl}\theta(X_{it}, X_{it-1};\omega)|]^2 < C$, where $\partial_j\theta(X_{it}, X_{it-1};\omega) = E_\mathcal{G}(E(\partial_j\pi_{it}(\omega)|X_{it}, X_{it-1}))$

and
$\partial_{jl}\theta(X_{it}, X_{it-1}; \omega) = E_{\mathcal{G}}(E(\partial_{jl}\pi_{it}(\omega)|X_{it}, X_{it-1}))$.

B6* For all $j, l \in \{1, \ldots, \mathbf{q}\}$, $t, t' \in \{2, \ldots, T\}$, $\partial_j \pi_{it}(\omega)\partial_l \pi_{it'}(\omega)$, $\partial_j \Delta_1 \pi_{it}(\omega)$, $\partial_j \Delta_1 \pi_{it}(\omega)\partial_j \Delta_1 \pi_{it'}(\omega)$, $\partial_j \theta(X_{it}, X_{it-1}; \omega)\partial_l \theta(X_{it}, X_{it-1}; \omega)$, $\partial_j h(X_{it}, X_{it-1}; \omega)\partial_l h(X_{it}, X_{it-1}; \omega)$, $(\partial_{jl}\pi_{it}(\omega))^2$, $(\partial_{jl}\theta(X_{it}, X_{it-1}; \omega))^2$, $(\partial_{jl}h(X_{it}, X_{it-1}; \omega))^2$, $\partial_{jl}\Delta_1 \pi_{it}(\omega)$, $\partial_{jl}\Delta_1 \pi_{it}(\omega)\partial_{jl}\Delta_1 \pi_{it'}(\omega)$, $\Delta\epsilon_{it}\partial_{jl}\Delta\pi_{it}(\omega)$, $\Delta_1 \pi_{it}(\omega)\partial_{jl}\pi_{it}(\omega)$ and $\Delta_1 \pi_{it}(\omega_0)\partial_{jl}\pi_{it}(\omega)$ satisfy the Cramér's condition.

C1* The kernel function $K(\psi) : \Re \to \Re$ is symmetric and satisfies (1) $\left|K(\psi)\psi^j\right| \leq C$ for all $\psi \in \Re$ with $j = 0, 1, \ldots, 3$; (2) $\int |\psi^j K(\psi)| d\psi \leq C$ for $j = 0, 1, \ldots, 3$; (3) $\int K(\psi) d\psi = 1$, $\int \psi K(\psi) d\psi = 0$, and $\mu_{k,s} = \int \psi^s K(\psi) d\psi$; (4) $K(\psi)$ is continuously differentiable on $\Re$ with $|\psi^j \frac{d}{d\psi} K(\psi)| \leq C$ for all $\psi \in \Re$ and $j = 0, 1, \ldots, 3$.

C2* (1) $\forall\ j, l \in \{1, \ldots, d\}$, $j < l$, $m_j(\cdot) \in C^2$. (2) For some $\rho > 0$, $E|v_{it}|^{2+\rho} < C$, $E|u_{it}|^{2+\rho} < C$, and $E(|\Delta\epsilon_{it}|^{2+\rho}|X_{it}^j, X_{it}^l) < C$. (3) $E(\Delta\epsilon_{it}^2|X_{it}^j = x^j) = \sigma^2_{\Delta\epsilon_{it}}(x^j)$ is continuous at x^j.

C3* (1) The conditional density of $X_{j,it}, X_{l,it}$ given $\tilde{u}_{it}, \tilde{u}_{it-1}$ is $f_{X_{j,t},X_{l,t}|\tilde{u}_t,\tilde{u}_{t-1}}(X_{j,it}, X_{l,it}) < C$. (2) For G, a compact subset of $[-1, 1]$, $\inf_{X_{j,it}\in G, X_{l,it}\in G} f_{X_{j,t},X_{l,t}}(X_{j,it}, X_{l,it}) > 0$, where $f_{X_{j,t},X_{l,t}}(X_{j,it}, X_{l,it})$ is the joint density of $X_{j,it}, X_{l,it}$, $\tau = t$ or $t-1$, $\forall\ 1 \leq j < l \leq d$. (3) The joint density of X_{it} by is $f_{X_t}(X_{it}) \in C^2$. (5) The joint density of W_{it} is $f_{W_t}(W_{it}) \in C^2$. The joint density of X_{it}^j and W_{it-1} is $f_{X_{j,t},W_{t-1}}(X_{j,it}, W_{it-1}) \in C^2$.

C4 (1) $\phi_k(x^j) \in C^2$. (2) $h = O(n^{-1/5})$, $\xi^2_{S(\kappa)} = O(S(\kappa))$, $\kappa = o(n^{3/10})$.

Assumption **A1*** revises original **A1** by allowing both fixed effects α_i and η_t to be arbitrary correlated with covariates for all t. The zero conditional mean of α_i and η_t ensures the identification of global constant m_0 in (A.1). The incidental parameter problem does not arise in our SEM-SP because **A1*** requires fixed T for time effects to be estimated consistently given that $n \to \infty$. The constant means of u_{it} and e_{it} follow conventional specifications in SFM with zero inefficiency (Yao et al., 2018; Wang et al., 2020). Thus, we allow inefficiency variables W_{it}^u and W_{it}^e to influence investment inefficiencies in a more stochastic fashion through altering probability functions $p_u(\cdot)$ and $p_e(\cdot)$, respectively. Assumption **A2*** revises the original **A2** by imposing standard conditions of uniform continuity and boundedness on the inefficiency function $\pi(W; \omega_0)$. The additive function $h(.; \omega)$ is obtained by projecting $\Delta\pi(\cdot)$ onto an additive space spanned by $\tilde{X}_{it}$ and $\tilde{X}_{it-1}$ (both including continuous variables X and $T-1$ dummies). This is similar to the generalized additive regression model by Horowitz and Mammen (2004), where only one constant is considered (i.e., $\eta_t = \eta$) and the link function is identify. Assumption **A3*** revises the original **A3** mainly by imposing new conditions **A3***(1)-**A3***(2), which ensure that the series approximation error of $h(\cdot)$ and $m_j(\cdot)$ vanishes sufficiently fast in the presence of $T-1$ constants for time fixed effects. Similar conditions can be found from Horowitz and Mammen (2004) where one constant is estimated

besides additive functions. As in Wang et al. (2024), the rate at which the approximation errors decay toward zero is governed by **A3***(3)-**A3***(4) in order to achieve the $1/\sqrt{n}$ convergence rate of $\hat{\omega}$ in the first step. The rest assumptions **B*** and **C*** are revised to accommodate the structure of our SEM-SP in which efficient investment function (EIF) contains only univariate functions $\{m_j(x_j)\}_{j=1}^d$.

We first characterize the asymptotics of parameter estimators $\hat{\omega}$. For $\partial_j \Delta_1 \pi_{it}(\omega) = \partial_j \Delta \pi_{it}(\omega) - \partial_j h(\tilde{X}_{it}, \tilde{X}_{it-1}; \omega)$, we denote $\Omega_T(\omega_0)$ as a $\mathbf{q} \times \mathbf{q}$ matrix whose $(j,l)^{th}$ element is given by

$$\Omega_{T,jl}(\omega_0) = \frac{4}{(T-1)^2} \sum_{t=2}^{T} \sum_{\tau=2}^{T} E(\Delta\tilde{\epsilon}_{it} \Delta\tilde{\epsilon}_{i\tau} \partial_j \Delta_1 \pi_{it}(\omega_0) \partial_l \Delta_1 \pi_{i\tau}(\omega_0)).$$

Also, define the $(j,l)^{th}$ element of $\Sigma_T(\omega_0)$ defined in **B4*** as

$$\Sigma_{T,jl}(\omega_0) = \frac{2}{T-1} \sum_{t=2}^{T} E(\partial_j \Delta_1 \pi_{it}(\omega_0) \partial_l \Delta_1 \pi_{it}(\omega_0)).$$

Theorem 1. *As $n \to \infty$ and T is fixed,*

*a) With Assumption **A****, $\hat{\omega} \xrightarrow{p} \omega_0$.

*b) With Assumptions **A*** and **B***, $\sqrt{n}(\hat{\omega} - \omega_0) \xrightarrow{d} N(0, \Sigma_T(\omega_0)^{-1} \Omega_T(\omega_0) \Sigma_T(\omega_0)^{-1})$.*

The consistent estimator for $\Sigma_T(\omega_0)$ and $\Omega_T(\omega_0)$ can be obtained following similar arguments in Appendix 4 of Wang et al. (2024).

We next present asymptotics of function and derivative estimators $\{\tilde{m}_j(x_j), \tilde{m}_j^{(1)}(x_j)\}_{j=1}^d$. Define the conditional variance function of $\tilde{m}_j(x_j)$ as

$$\sigma^2_{m_j}(x_j) = \left[\frac{1}{T-1} \sum_{t=2}^{T} f_{X_{j,t}}(x_j)\right]^{-2} \frac{1}{(T-1)^2} \sum_{t=2}^{T} \sigma^2_{\Delta\tilde{\epsilon}_{it}}(x_j) f_{X_{j,t}}(x_j) \int K^2(\psi) d\psi.$$

Similar, the variance function of $\tilde{m}_j^{(1)}(x_j)$ is defined as

$$\sigma^{2(1)}_{m_j}(x_j) = \left[\frac{\mu_{k,2}}{T-1} \sum_{t=2}^{T} f_{X_{j,t}}(x_j)\right]^{-2} \frac{1}{(T-1)^2} \sum_{t=2}^{T} \sigma^2_{\Delta\tilde{\epsilon}_{it}}(x_j) f_{X_{j,t}}(x_j) \int \psi^2 K^2(\psi) d\psi.$$

Theorem 2. *With Assumption **A***-**C***, as $n \to \infty$ with T fixed,*

(a) $\sqrt{nh}(\tilde{m}_j(x_j) - m_j(x_j) - \frac{h^2}{2} B_j(x_j) + o_p(h^2)) \xrightarrow{d} N(0, \sigma^2_{m_j}(x_j))$.

(b) $\sqrt{nh}(\tilde{m}_j(x_j) - m_j(x_j) - \frac{h^2}{2} + o_p(h^2))$ *and* $\sqrt{nh}(\tilde{m}_l(x_l) - m_l(x_l) - \frac{h^2}{2} B_j(x_j) + o_p(h^2))$ *are asymptotically independently normally distributed for $l \neq j = 1, ..., d$.*

(c) $\sqrt{nh^3}(\tilde{m}_j^{(1)}(x_j) - m_j^{(1)}(x_j) - \frac{h^2}{2}B_j^{(1)}(x_j) + o_p(h^2)) \xrightarrow{d} N(0, \sigma_{m_j}^{2(1)}(x_j))$.

(d) $\sqrt{nh^3}(\tilde{m}_j^{(1)}(x_j) - m_j^{(1)}(x_j) - \frac{h^2}{2}B_j^{(1)}(x_j) + o_p(h^2))$ *and* $\sqrt{nh^3}(\tilde{m}_l^{(1)}(x_l) - m_l^{(1)}(x_l) - \frac{h^2}{2}B_l^{(1)}(x_l) + o_p(h^2))$ *are asymptotically independently normally distributed for* $l \neq j = 1, ..., d$.

Here, the leading bias term $B_j(x_j)$ in (a)-(b) is given by $B_j(x_j) = m_j^{(2)}(x_j)\mu_{k,2}$ for $\tilde{m}_j(x_j)$, and $B_j^{(1)}(x_j)$ in (c)-(d) takes the form

$$B_j^{(1)}(x_j) = \left[\frac{\mu_{k,2}}{T-1}\sum_{t=2}^{T} f_{X_{j,t}}(x_j)\right]^{-1} \frac{\mu_{k,4}}{T-1}\sum_{t=2}^{T} m_j^{(2)}(x_j) f_{X_{j,t}}^{(1)}(x_j),$$

where $f_{X_{j,t}}^{(1)}(x_j) = \partial_j f_{X_{j,t}}^{(1)}(x_j)/\partial_j x_j$ and $m_j^{(2)}(x_j) = \partial^2 m_j(x_j)/\partial^2 x_j$. The proof of Theorems 1-2 follows closely to the arguments in the appendix of Wang et al. (2024) and thus are omitted here for brevity.

Appendix 1.3: Numerical properties of SEM-SP estimator

This section investigates the finite-sample performance of our propose parameter estimator $\hat{\omega}$ in (A.7) and functions and derivatives estimator $\{\tilde{m}_j(x_j), \tilde{m}_j^{(1)}(x_j)\}_{j=1}^d$ in (A.8). To mimic the structure of our empirical dataset, we simulate our data below by setting $T = 15$ and, for reducing computational burden, we choose $n = (100, 200, 400)$. As in our application, we set $d = 3$ and highlight that our estimator's performance is not affected by the dimension d due to the combined use of series-kernel estimation (Wang et al., 2024). Accordingly, we generate data according to the following data generating process (DGP):

$$I_{it} = m_0 + \alpha_i + \eta_t + m_1(X_{1,it}) + m_2(X_{2,it}) + m_3(X_{3,it}) + \epsilon_{it}, \tag{A.10}$$

where $\epsilon_{it} = v_{it} - u_{it}$ with probability $p_u(W_{it}^u; \gamma_0^u)$, $\epsilon_{it} = v_{it} + e_{it}$ with probability $p_e(W_{it}^e; \gamma_0^e)$, and $\epsilon_{it} = v_{it}$ with probability $1 - p_u(W_{it}^u; \gamma_0^u) - p_e(W_{it}^e; \gamma_0^e)$. As above, we mimic our empirical dataset by considering $W^u = [W_1^u, W_2^u]^\top$ (i.e., $q_u = 2$), and $W^e = [W_1^e, W_2^e, W_3^e, W_4^e, W_5^e, W_6^e]^\top$ with $W_1^e = W_1^u$ and $W_2^e = W_2^u$ (i.e., $q_e = 6$). We specify two DGPs with different functional forms $m_j(\cdot)$ and parameters ω_0. In DGP_1, we set $m_0 = 3$, $m_1(v) = v + \frac{1}{2}v^2$, $m_2(v) = v^2 - \frac{1}{2}v^3$, $m_3(v) = 1.5v^4$, $(\mu_0^u, \mu_0^e) = (1, 2)$, $[\gamma_{0,1}^u, \gamma_{0,2}^u] = [0.5, 1]^\top$, and $[\gamma_{0,1}^e, \gamma_{0,2}^e, \gamma_{0,3}^e, \gamma_{0,4}^e, \gamma_{0,5}^e, \gamma_{0,6}^e] = [1.5, 2, 0.5, -0.5, 2.5, -1]^\top$. In DGP_2, we have $m_0 = -2$, $m_1(v) = \exp(v) - 1$, $m_2(v) = -v^2$, $m_3(v) = \sin(v)$, $(\mu_0^u, \mu_0^e) = (3, 0.5)$, $[\gamma_{0,1}^u, \gamma_{0,2}^u] = [-1, 0.5]^\top$, and $[\gamma_{0,1}^e, \gamma_{0,2}^e, \gamma_{0,3}^e, \gamma_{0,4}^e, \gamma_{0,5}^e, \gamma_{0,6}^e] = [2.5, -2, 0.5, -0.5, 1, -1]^\top$. Efficiency functions exhibit quadratic structures under DGP_1, and exhibit exponential, quadratic, or periodic structures under DGP_2. The derivative functions $m_j^{(1)}(\cdot)$ under each DGP are apparent with both linear and nonlinear forms, and thus omitted for brevity. The parameters $\gamma_0 = (\gamma_0^{u\top}, \gamma_0^{e\top})^\top$ are carefully chosen so that the mean of $(p_u(\cdot), p_e(\cdot))$ are $(0.7, 0.08)$ in DGP_1 and $(0.26, 0.40)$ in DGP_2.

Table A1.1: SEM-SP Estimation Results for Inefficiency Parameters

$T = 15$		DGP_1			DGP_2		
		$n = 100$	200	400	$n = 100$	200	400
$\hat{\mu}^u$	RMSE	0.0468	0.0357	0.0215	0.0336	0.0216	0.0151
	BIAS	0.0374	0.0280	0.0180	0.0258	0.0171	0.0119
	SD	0.0278	0.0212	0.0130	0.0215	0.0132	0.0093
$\hat{\mu}^e$	RMSE	0.0509	0.0425	0.0270	0.0410	0.0264	0.0183
	BIAS	0.0406	0.0332	0.0224	0.0324	0.0210	0.0145
	SD	0.0319	0.0278	0.0160	0.0252	0.0160	0.0111
$\hat{\gamma}_1^u$	RMSE	0.0989	0.0752	0.0552	0.0735	0.0510	0.0334
	BIAS	0.0749	0.0582	0.0435	0.0574	0.0410	0.0266
	SD	0.0614	0.0462	0.0333	0.0459	0.0305	0.0202
$\hat{\gamma}_2^u$	RMSE	0.0598	0.0410	0.0269	0.0418	0.0282	0.0188
	BIAS	0.0454	0.0322	0.0215	0.0332	0.0226	0.0149
	SD	0.0373	0.0235	0.0157	0.0255	0.0169	0.0115
$\hat{\gamma}_1^e$	RMSE	0.0784	0.0572	0.0379	0.0650	0.0431	0.0294
	BIAS	0.0583	0.0481	0.0308	0.0516	0.0337	0.0236
	SD	0.0491	0.0336	0.0215	0.0396	0.0268	0.0176
$\hat{\gamma}_2^e$	RMSE	0.0658	0.0474	0.0306	0.0509	0.0341	0.0232
	BIAS	0.0516	0.0374	0.0249	0.0400	0.0270	0.0187
	SD	0.0395	0.0267	0.0181	0.0315	0.0209	0.0138
$\hat{\gamma}_3^e$	RMSE	0.0535	0.0345	0.0246	0.0367	0.0238	0.0165
	BIAS	0.0413	0.0273	0.0195	0.0291	0.0190	0.0132
	SD	0.0323	0.0207	0.0149	0.0224	0.0144	0.0099
$\hat{\gamma}_4^e$	RMSE	0.0399	0.0265	0.0193	0.0350	0.0234	0.0153
	BIAS	0.0319	0.0216	0.0157	0.0278	0.0186	0.0122
	SD	0.0232	0.0166	0.0118	0.0212	0.0142	0.0092
$\hat{\gamma}_5^e$	RMSE	0.0706	0.0450	0.0315	0.0528	0.0341	0.0238
	BIAS	0.0544	0.0377	0.0254	0.0423	0.0274	0.0189
	SD	0.0442	0.0242	0.0188	0.0317	0.0203	0.0146
$\hat{\gamma}_6^e$	RMSE	0.0673	0.0402	0.0304	0.0482	0.0317	0.0223
	BIAS	0.0538	0.0323	0.0250	0.0385	0.0256	0.0175
	SD	0.0439	0.0233	0.0185	0.0291	0.0188	0.0138

Thus, a relatively higher likelihood to under-invest (over-invest) is reflected in DGP_1 (DGP_2) as in non-SOEs (SOEs and whole sample) considered in our study.

To allow correlations among variables, we first generate auxiliary variables $\{X_{j,it}^0\}_{j=1}^3$, $\{W_{j,it}^{u,0}\}_{j=1}^2$, and $\{W_{j,it}^{e,0}\}_{j=3}^6$ from a multivariate normal distribution $\mathcal{N}(\mu_0, \Sigma_0)$, where μ_0 is a 9×1 vector of 0, and Σ_0 is the covariance matrix with its $(j,l)^{th}$ element as $\rho^{|j-l|}$ for $j, l = 1, ..., 9$. We set $\rho = 0.5$ to allow regressors to be reasonably correlated. We next introduce dependence across time by generating $X_{j,it} = X_{j,it}^0 + \zeta_{j,it}^x$ (with $j = 1, 2, 3$), $W_{s,it}^u = W_{s,it}^{u,0} + \zeta_{s,it}^u$ (with $s = 1, 2$), $W_{l,it}^e = W_{l,it}^{e,0} + \zeta_{l,it}^e$ (with $l = 3, 4, 5, 6$), where $\zeta_{j,it}^x = 0.25\zeta_{j,it-1}^x + \xi_{it}$, $\zeta_{s,it}^u = 0.5\zeta_{s,it-1}^u + \xi_{it}$, and $\zeta_{s,it}^e = 0.75\zeta_{l,it-1}^u + \xi_{it}$ follow an AR(1) process with $\xi_{it} \sim \mathcal{N}(0, 0.25^2)$. Here, we generate $X_{j,it}$ with values falling into $[-2, 2]$ to be in line with our assumption of the compact support for $X_{j,it}$.

Table A1.2: SEM-SP Estimation Results for Efficiency Functions

DGP_1	$T = 15$	$n = 100$		200		400	
		SEM-SP	SEM-SP-Oracle	SEM-SP	SEM-SP-Oracle	SEM-SP	SEM-SP-Oracle
$\tilde{m}_1$	RAMSE	0.0956	0.0927	0.0684	0.0674	0.0513	0.0507
	ABIAS	0.0660	0.0644	0.0489	0.0482	0.0374	0.0369
	ASD	0.0787	0.0760	0.0588	0.0578	0.0449	0.0443
$\tilde{m}_2$	RAMSE	0.1775	0.1747	0.1378	0.1373	0.1068	0.1061
	ABIAS	0.1431	0.1416	0.1117	0.1112	0.0872	0.0869
	ASD	0.0851	0.0808	0.0614	0.0611	0.0455	0.0446
$\tilde{m}_3$	RAMSE	0.3598	0.3587	0.2321	0.2314	0.1819	0.1815
	ABIAS	0.2963	0.2950	0.1782	0.1773	0.1351	0.1346
	ASD	0.0842	0.0839	0.0515	0.0512	0.0302	0.0301
$\tilde{m}_1^{(1)}$	RAMSE	0.1534	0.1517	0.1281	0.1267	0.1051	0.1024
	ABIAS	0.1109	0.1101	0.0918	0.0916	0.0724	0.0712
	ASD	0.1129	0.1127	0.0916	0.0912	0.0767	0.0742
$\tilde{m}_2^{(1)}$	RAMSE	0.4017	0.4007	0.3402	0.3393	0.2905	0.2902
	ABIAS	0.2534	0.2521	0.2089	0.2086	0.1713	0.1705
	ASD	0.1018	0.1013	0.0759	0.0758	0.0624	0.0616
$\tilde{m}_3^{(1)}$	RAMSE	0.4494	0.4490	0.3993	0.3990	0.3431	0.3430
	ABIAS	0.2718	0.2711	0.2364	0.2363	0.1972	0.1972
	ASD	0.1767	0.1751	0.1181	0.1176	0.0866	0.0863
DGP_2	$T = 15$	$n = 100$		200		400	
		SEM-SP	SEM-SP-Oracle	SEM-SP	SEM-SP-Oracle	SEM-SP	SEM-SP-Oracle
$\tilde{m}_1$	RAMSE	0.1144	0.1135	0.0855	0.0851	0.0679	0.0677
	ABIAS	0.0821	0.0817	0.0645	0.0641	0.0502	0.0500
	ASD	0.0758	0.0730	0.0570	0.0566	0.0454	0.0452
$\tilde{m}_2$	RAMSE	0.1052	0.1029	0.0797	0.0793	0.0622	0.0617
	ABIAS	0.0711	0.0698	0.0539	0.0538	0.0433	0.0428
	ASD	0.0739	0.0722	0.0578	0.0575	0.0463	0.0461
$\tilde{m}_3$	RAMSE	0.0832	0.0801	0.0677	0.0671	0.0534	0.0524
	ABIAS	0.0617	0.0597	0.0499	0.0495	0.0396	0.0389
	ASD	0.0705	0.0675	0.0560	0.0555	0.0465	0.0459
$\tilde{m}_1^{(1)}$	RAMSE	0.2178	0.2143	0.1820	0.1811	0.1507	0.1499
	ABIAS	0.1209	0.1191	0.0989	0.0983	0.0791	0.0788
	ASD	0.0713	0.0702	0.0574	0.0566	0.0471	0.0469
$\tilde{m}_2^{(1)}$	RAMSE	0.2866	0.2829	0.2372	0.2372	0.1970	0.1966
	ABIAS	0.2001	0.1972	0.1595	0.1591	0.1322	0.1319
	ASD	0.1410	0.1375	0.1142	0.1130	0.0978	0.0977
$\tilde{m}_3^{(1)}$	RAMSE	0.1870	0.1865	0.1670	0.1650	0.1298	0.1286
	ABIAS	0.1346	0.1336	0.1192	0.1191	0.0935	0.0926
	ASD	0.1346	0.1330	0.1156	0.1145	0.0971	0.0965

We construct the firm fixed effect $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} c_0(\sum_{j=1}^{d} X_{j,it} + \sum_{s=1}^{2} W^u_{s,it} + \sum_{l=3}^{6} W^e_{l,it}) + \xi^\alpha_i$, where $\xi^\alpha_i \sim \mathcal{N}(0,1)$ and $c_0 \neq 0$ reflects correlation between α_i and covariates. Likewise, we generate time fixed effect $\eta_t = \frac{1}{n}\sum_{i=1}^{n} c_0(\sum_{j=1}^{d} X_{j,it} + \sum_{s=1}^{2} W^u_{s,it} + \sum_{l=3}^{6} W^e_{l,it}) + \xi^\eta_t$, where $\xi^\eta_t \sim \mathcal{N}(0,1)$ and $c_0 \neq 0$ reflects correlation between η_t and covariates. We consider a fixed effect model by setting $c_0 = 1$, and empirically center (α_i, η_t) to be consistent with our assumption that $E(\alpha_i) = E(\eta_t) = 0$. Finally, we employ a cubic B-spline estimator in (A.7) with each knot evenly spaced on the support of X_j, and a 2nd-order Gaussian kernel function in (A.8). We impose the identification condition for $m_j(\cdot)$ by empirically centering basis functions around zero.

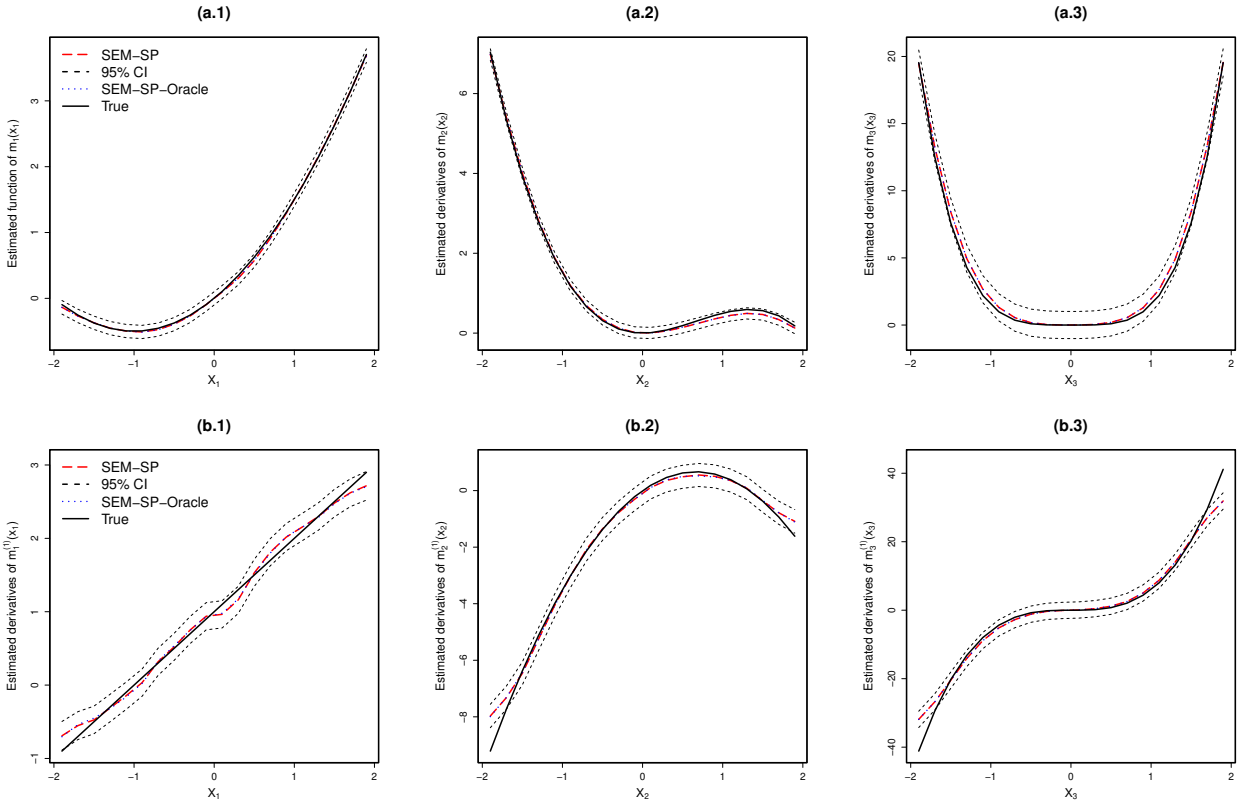
Our two-step estimation necessitates choosing the number of basis functions κ in the series estimator (A.7) and the bandwidth h in the kernel backfitting estimator (A.8). To ensure that the series estimation in the first step do not affect the convergence rate of $\hat{\omega}$ asymptotically, we adjust the order of κ through a constant ϱ to implement $\hat{\omega}$ by setting $\kappa = [n^{\frac{1}{5}+\varrho}]$, where $[v]$ refers to the integer part of a real number v. For the backfitting estimator, we adopt a rule-of-thumb (ROT) bandwidth $h_j = C\hat{\sigma}_j n^{-\frac{1}{5}}$, with $\hat{\sigma}_j$ the empirical standard deviation of $\{X_{j,it}\}_{i=1,t=1}^{n,T}$. Thus, we need to choose constants ϱ and C. Below, we first set $\varrho = 0.1$ and $C = 1$ as benchmarks in our simulation study, and we choose (ϱ, C) from different pairs to investigate their impact on simulation results.

We perform 1000 repetitions throughout the experiment. To evaluate the performance of estimates $\hat{\omega}$, we report the root mean squared error (RMSE), the absolute bias (BIAS), and the standard deviation (SD). To evaluate the performance of function and derivatives estimates $(\tilde{m}_j(\cdot), \tilde{m}^{(1)}_j(\cdot))$, we report the root averaged mean squared error (RAMSE), the averaged BIAS (ABIAS), and averaged standard deviation (ASD) across 50 grid points in $[-2, 2]$.

Table A1.1 reports the performance of $\hat{\omega}$ for DGP_1 on the left panel and DGP_2 on the right panel, with $(\varrho, C) = (0.1, 1)$. Across both $DGPs$, the magnitude of each measure for $\hat{\omega}$ is fairly small. The performance of two constant estimates $(\hat{\mu}^u, \hat{\mu}^e)$ are relatively superior than most of other parameters, except for $\hat{\gamma}^e_4$. Moreover, we observe a clear improvement in the performance of each parameter in $\hat{\omega}$ as n doubles, indicated by the reduction in RMSE, BIAS, and SD. This observation holds across two different $DGPs$. Thus, the results above suggest consistency properties of our $\hat{\omega}$ in (A.7).

Table A1.2 summarizes the performance of kernel-backfitting estimates under DGP_1 (upper panel) and DGP_2 (lower panel), from each the function estimates $\tilde{m}_j(\cdot)$ are reported in the first three rows and the derivative estimates $\tilde{m}^{(1)}_j(\cdot)$ in the last three rows. For comparison purpose, in each DGP we compare the performance of our estimator (SEM-SP) with the oracle estimator (SEM-SP-Oracle), which assumes that ω_0 and all the other efficient functions (except the one being estimated) were known. The results are qualitatively similar to our parameter estimators in Table A1.1. Across both DGPs and as n gets large, both function and derivative estimators clearly become superior in terms of decreasing RAMSE, ABIAS, and ASD, illustrating their consistency properties. The performance of $\tilde{m}^{(1)}_j(\cdot)$ is uniformly worsen than $\tilde{m}_j(\cdot)$, which is consistent with

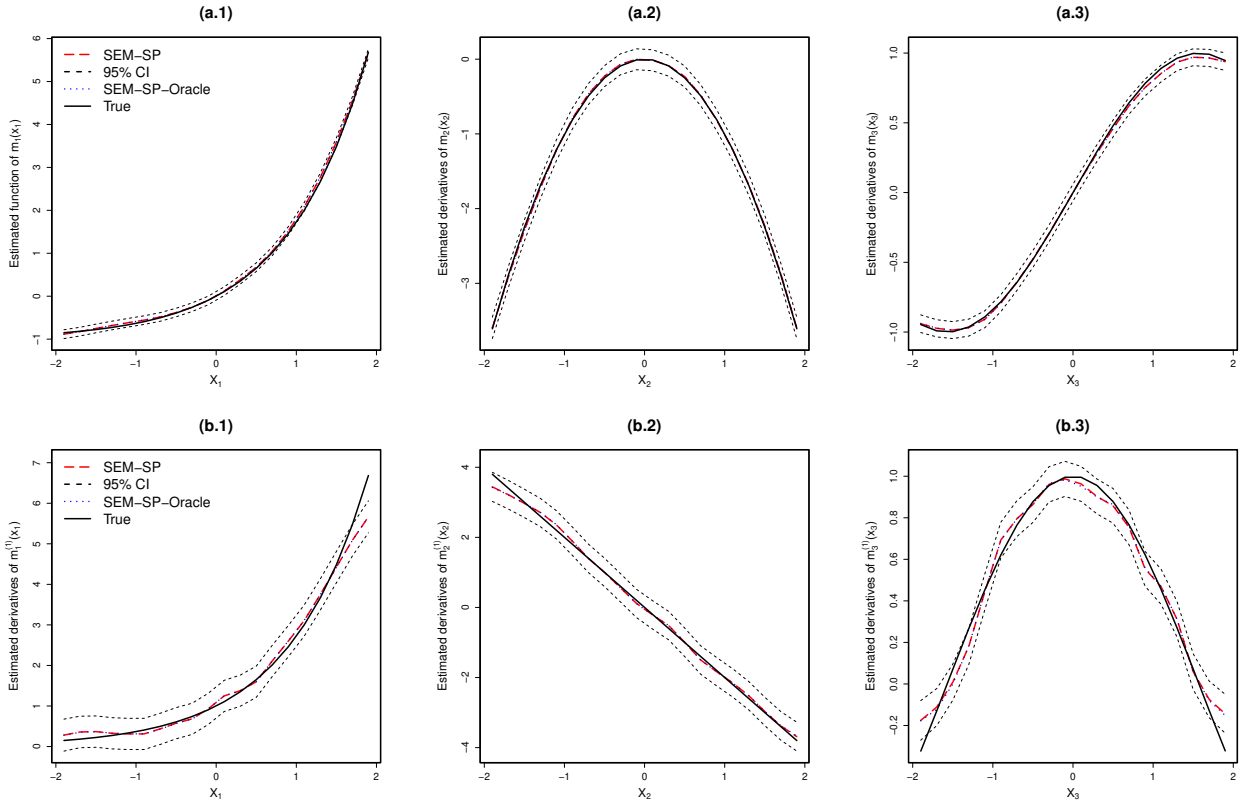
Figure A.1: SEM-SP Estimation from DGP_1 for Efficiency Function (First row) and Derivatives (Second row)



Theorem 3 of Wang et al. (2024) on the slower convergence rate of derivative estimator. Given each sample size of n, SEM-SP reveals a quantitatively similar estimates compared to SEM-SP-Oracle, which is also expected due to the fast convergence speed of parametric parameter $\hat{\omega}$ given in Theorem 1 of Wang et al. (2024).

To provide a vivid picture on the performance of our estimators, we plot estimates for three functions (derivatives) correspondingly in panels (a.1)-(a.3) ((b.1)-(b.3)) of Figure A.1 under DGP_1 and Figure A.2 under DGP_2. In each panel, we plot the true function $m_j(\cdot)$ or $m_j^{(1)}(\cdot)$ (black solid line), together with SEM-SP estimates $\tilde{m}_j(\cdot)$ or $\tilde{m}_j^{(1)}(\cdot)$ (red dash line), 95% asymptotic confidence interval constructed based on Theorem 3 and Corollary 2 of Wang et al. (2024) (black dot line), and the SEM-SP-Oracle estimates (blue dot line). The estimation is performed with the largest sample size considered in our simulation study, i.e., $(n, T) = (400, 15)$. We observe that our function estimator closely captures the nonlinearity of $m_j(\cdot)$ across both DGPs. Similar observations are made for our derivative estimator, which clearly reveals the linear or nonlinear functional form of $m_j^{(1)}(\cdot)$ evaluated at most observations, except for x_j around boundaries in those derivative functions with a high degree of curvature. Overall, our SEM-SP demonstrates appealing numerical

Figure A.2: SEM-SP Estimation from DGP_2 for Efficiency Function (First row) and Derivatives (Second row)



performance in estimating unknown parameters, functions, and derivatives of our interests. We expect that its application to our empirical study provides a reliable empirical findings given our large number of firms.

Finally, we investigate the sensitivity of our simulation results to the choice of scaling parameters (ϱ, C). Maintaining the same experimental setup, we explore additional values by selecting ϱ from $(0.05, 0.15)$ and C from $(0.5, 1.5)$, resulting in four combinations as $(0.05, 0.5)$, $(0.05, 1.5)$, $(0.15, 0.5)$, and $(0.15, 1.5)$. For each constant, the larger value is three times the smaller one, allowing for a broad exploration of these constants. Here, a larger ϱ indicates higher degree of undersmoothing, implying more basis functions in the series estimation. Conversely, a larger C corresponds to a larger bandwidth and a higher degree of oversmoothing in the kernel backfitting estimation.

Table A1.3 summarizes the performance of selected parameter estimates $(\hat{\mu}^u, \hat{\gamma}_1^u, \hat{\gamma}_1^e)$ and function estimates $(\tilde{m}_1, \tilde{m}_1^{(1)})$ for DGP_1 (top) and DGP_2 (bottom). Results for different (ϱ, C) combinations are reported under each sample size n. We find no clear pattern in how variations in ϱ or C influence the numerical properties, as the differences in all statistical measures remain minimal and there does not seem to be a clear pattern on the changes. For instance, increasing ϱ from 0.05 to

Table A1.3: SEM-SP Estimation Results with Different Scaling Parameters (ρ, C) for $T = 15$

DGP_1		$n = 100$				$n = 200$				$n = 400$			
	(ϱ, C)	(0.05,0.5)	(0.05,1.5)	(0.15,0.5)	(0.15,1.5)	(0.05,0.5)	(0.05,1.5)	(0.15,0.5)	(0.15,1.5)	(0.05,0.5)	(0.05,1.5)	(0.15,0.5)	(0.15,1.5)
$\hat{\mu}^u$	RMSE	0.0495	0.0531	0.0614	0.0551	0.0363	0.0384	0.0372	0.0473	0.0263	0.0270	0.0259	0.0292
	ABIAS	0.0393	0.0432	0.0492	0.0431	0.0309	0.0299	0.0308	0.0365	0.0206	0.0214	0.0203	0.0232
	ASD	0.0303	0.0310	0.0370	0.0345	0.0192	0.0242	0.0210	0.0302	0.0165	0.0166	0.0161	0.0179
$\hat{\gamma}_1^u$	RMSE	0.0387	0.0385	0.0369	0.0324	0.0266	0.0251	0.0263	0.0266	0.0203	0.0176	0.0185	0.0171
	ABIAS	0.0301	0.0323	0.0295	0.0255	0.0224	0.0202	0.0215	0.0216	0.0153	0.0143	0.0154	0.0132
	ASD	0.0245	0.0210	0.0223	0.0201	0.0145	0.0150	0.0153	0.0157	0.0135	0.0104	0.0103	0.0109
$\hat{\gamma}_1^e$	RMSE	0.0875	0.0863	0.0947	0.0866	0.0639	0.0551	0.0594	0.0552	0.0457	0.0437	0.0410	0.0410
	ABIAS	0.0666	0.0664	0.0756	0.0663	0.0533	0.0435	0.0453	0.0447	0.0367	0.0345	0.0327	0.0324
	ASD	0.0570	0.0554	0.0573	0.0560	0.0354	0.0341	0.0386	0.0326	0.0274	0.0270	0.0249	0.0253
$\tilde{m}_1$	RAMSE	0.1046	0.1046	0.1074	0.1072	0.0798	0.0804	0.0828	0.0834	0.0598	0.0630	0.0612	0.0630
	ABIAS	0.0766	0.0784	0.0780	0.0784	0.0590	0.0604	0.0624	0.0628	0.0456	0.0466	0.0464	0.0470
	ASD	0.0926	0.0921	0.0958	0.0938	0.0718	0.0714	0.0754	0.0738	0.0556	0.0540	0.0556	0.0551
$\tilde{m}_1^{(1)}$	RAMSE	0.1592	0.1476	0.1512	0.1479	0.1217	0.1197	0.1253	0.1287	0.1060	0.1105	0.1056	0.1106
	ABIAS	0.1125	0.1161	0.1096	0.1098	0.0896	0.0898	0.0912	0.0917	0.0767	0.0770	0.0755	0.0784
	ASD	0.1182	0.1089	0.1120	0.1118	0.0958	0.0922	0.0969	0.0942	0.0845	0.0830	0.0839	0.0824
DGP_2		$n = 100$				$n = 200$				$n = 400$			
	(ϱ, C)	(0.05,0.5)	(0.05,1.5)	(0.15,0.5)	(0.15,1.5)	(0.05,0.5)	(0.05,1.5)	(0.15,0.5)	(0.15,1.5)	(0.05,0.5)	(0.05,1.5)	(0.15,0.5)	(0.15,1.5)
$\hat{\mu}^u$	RMSE	0.0761	0.0835	0.0693	0.0814	0.0605	0.0545	0.0608	0.0606	0.0423	0.0404	0.0383	0.0365
	ABIAS	0.0591	0.0698	0.0539	0.0658	0.0487	0.0417	0.0513	0.0479	0.0332	0.0330	0.0303	0.0285
	ASD	0.0482	0.0462	0.0438	0.0481	0.0360	0.0353	0.0328	0.0373	0.0263	0.0234	0.0235	0.0230
$\hat{\gamma}_1^u$	RMSE	0.0250	0.0292	0.0239	0.0279	0.0203	0.0188	0.0193	0.0210	0.0131	0.0145	0.0137	0.0135
	ABIAS	0.0191	0.0245	0.0186	0.0228	0.0169	0.0150	0.0156	0.0170	0.0106	0.0119	0.0109	0.0109
	ASD	0.0162	0.0160	0.0151	0.0162	0.0114	0.0115	0.0115	0.0123	0.0078	0.0084	0.0084	0.0079
$\hat{\gamma}_1^e$	RMSE	0.0354	0.0346	0.0420	0.0398	0.0266	0.0284	0.0267	0.0276	0.0172	0.0228	0.0202	0.0186
	ABIAS	0.0290	0.0281	0.0348	0.0317	0.0208	0.0230	0.0225	0.0224	0.0133	0.0185	0.0167	0.0142
	ASD	0.0204	0.0202	0.0236	0.0241	0.0167	0.0168	0.0146	0.0162	0.0109	0.0134	0.0114	0.0120
$\tilde{m}_1$	RAMSE	0.1454	0.1466	0.1434	0.1378	0.1120	0.1134	0.1106	0.1122	0.0866	0.0820	0.0890	0.0898
	ABIAS	0.1090	0.1098	0.1072	0.1084	0.0842	0.0847	0.0844	0.0856	0.0620	0.0648	0.0674	0.0686
	ASD	0.0966	0.0938	0.0944	0.0922	0.0754	0.0698	0.0710	0.0676	0.0542	0.0500	0.0604	0.0554
$\tilde{m}_1^{(1)}$	RAMSE	0.1881	0.1812	0.1829	0.1787	0.1562	0.1524	0.1532	0.1562	0.1274	0.1265	0.1333	0.1285
	ABIAS	0.0955	0.0963	0.0923	0.0950	0.0772	0.0793	0.0775	0.0781	0.0626	0.0636	0.0659	0.0660
	ASD	0.0569	0.0561	0.0579	0.0564	0.0481	0.0469	0.0469	0.0454	0.0403	0.0391	0.0428	0.0415

Note: The performance of selected parameter estimates $(\hat{\mu}^u, \hat{\gamma}_1^u, \hat{\gamma}_1^e)$ and function estimates $(\tilde{m}_1, \tilde{m}_1^{(1)})$ are evaluated using different scaling constants ϱ and $C = (C_\kappa, C_h)$, both are used to choose tuning parameters $\kappa = [C_\kappa n^{\frac{1}{5}-\varrho}]$ in series estimator and $h_j = C_{h_j} \hat{\sigma}_j n^{-\frac{1}{5}}$ in kernel backfitting estimator. Here, $[v]$ refers to the integer part of a real number v.

0.15 while holding $C = 0.5$ and $n = 100$ raises the RMSE of $\hat{\mu}^u$ from 0.0495 to 0.0614 under DGP_1 but lowers it from 0.0761 to 0.0693 under DGP_2. More generally, RMSE for all parameter estimates remain relatively stable across different choices of ϱ. Even with a larger sample size of $n = 400$, no specific pattern appears regarding these scaling parameters and estimation performance. For the function estimates, when $C = 0.5$ or $C = 1.5$, we observe that the RAMSE of $\tilde{m}_1$ under DGP_1 decreases with smaller ϱ, but this difference diminishes as n increases and is not observed under DGP_2. Overall, changes in RAMSE for both $\tilde{m}_1$ and $\tilde{m}_1^{(1)}$ remain small, particularly for larger n. Additionally, for a fixed ϱ, increasing C (i.e., increasing the bandwidth h_j) reduces variance while increasing bias in kernel estimates, which is expected from the well-known bias-variance trade-off in nonparametric smoothing.

Overall, the performance of our estimator for parameters and functions is fairly robust as the values of (ϱ, C) vary. The effect of changing tuning parameters diminishes as n rises. Given the large sample size in our empirical application, these results provide support for our empirical findings, which are expected to be relatively robust to varying the number of basis functions or bandwidths.

Appendix 1.4: Extension of SEM-SP

In Appendix 1.1, we propose SEM-SP and demonstrate a two-step estimation for parameters ω_0 by PNLS (A.6) and unknown functions $\{m_j(\cdot), m_j^{(1)}(\cdot)\}_{j=1}^3$ by kernel backfitting (A.8). Specifically, recall that the determinants of under- (over-) investment probability are $W_{it}^u \in \Re^{q_u}$ ($W_{it}^e \in \Re^{q_e}$), and $W_{it} = (W_{it}^u, W_{it}^e) \in \Re^q$, with $q = q_u + q_e$. With our notations $\omega_0 = (\mu_0, \gamma_0)$ with $\mu_0 = (\mu_0^u, \mu_0^e)$ and $\gamma_0 = (\gamma_0^u, \gamma_0^e)$, the PNLS estimates ω_0 from a nonlinear parametric function $\pi(W; \omega_0)$ in (A.3) as the difference between (scaled) under- and over-investment probabilities:

$$\pi(W_{it}; \omega_0) = \mu_0^u p_u(W_{it}; \gamma_0) - \mu_0^e p_e(W_{it}; \gamma_0), \tag{A.11}$$

where the conditional under- and over-investment probabilities take the following parametric multi-logit structure

$$p_u(W_{it}; \gamma_0) = \frac{\exp(W_{it}^{u\top}\gamma_0^u)}{1 + \exp(W_{it}^{u\top}\gamma_0^u) + \exp(W_{it}^{e\top}\gamma_0^e)}, \quad p_e(W_{it}; \gamma_0) = \frac{\exp(W_{it}^{e\top}\gamma_0^e)}{1 + \exp(W_{it}^{u\top}\gamma_0^u) + \exp(W_{it}^{e\top}\gamma_0^e)}. \tag{A.12}$$

In this setup, the conditional efficient investment probability is $1 - p_u(W_{it}; \gamma_0) - p_e(W_{it}; \gamma_0)$. The estimates $\hat{\omega}$ is then used in the second step to perform kernel smoothing on efficient investment function and their derivatives.

The consistency of the two-step estimation critically depends on correct functional form assumptions for 1) efficient investment functions (EIFs) and 2) multi-logit probability functions. Violating either assumption leads to model misspecification in SEM-SP and therefore inconsistent estimates. On one hand, the current SEM-SP excludes interactive effects among efficient investment variables (EIVs) X_l and X_s, for $1 \leq l \neq s \leq p$. While such interactions (e.g., between Tobin's q

and sales) lack theoretical support in finance, omitting them in other applications can oversimplify joint effects and bias the model estimation. On the other hands, the structure of $p_u(\cdot)$ and $p_e(\cdot)$ may deviate from the assumed multi-logit form, such as exhibiting non-monotonic structure, which yields inconsistent estimates of $\hat{\omega}$ as well as second-step estimates.

To mitigate misspecification issues arising from the two sources, we propose a robust version of SEM-SP, denoted as SEM-SP-R, with two main modifications. First, the SEM-SP-R follows Wang et al. (2024) to improve model flexibility of EIFs by adding interactions as

$$I_{it} = m_0 + \alpha_i + \eta_t + \sum_{j=1}^{d} m_j(X_{j,it}) + \sum_{1\le j<l\le d}\sum H_{jl}(X_{j,it}, X_{l,it}) + \epsilon_{it}, \tag{A.13}$$

where $H_{jl}(\cdot)$ is a bivariate unknown function that captures pair-wise interactive effects among X beyond a simple multiplicative form. Second, the SEM-SP-R replaces the parametric conditional probabilities in (A.12) with a general nonparametric form as $p_u(W_{it}^u) \in \Re^{q_u}$ and $p_e(W_{it}^e) \in \Re^{q_e}$. As a result, $\pi(W_{it};\gamma_0)$ in (A.11) is generalized as an additive nonparametric function $\pi(W_{it}) = \mu_0^u p_u(W_{it}) - \mu_0^e p_e(W_{it})$, where $W_{it} = (W_{it}^u, W_{it}^e)$. Nonetheless, the structure of $\pi(W)$ is delicate because μ_0^u (μ_0^e) cannot be identified from the nonparametric $p_u(W^u)$, and $p_u(\cdot)$ and $p_e(\cdot)$ must contain strictly separable W for their individual identification.

To facilitate our discussion on SEM-SP-R while preserving the model's applicability, below we focus on a simplified scenario in which firms may under-invest or over-invest, but never efficiently invest. This case is simplified only from a modeling perspective because the previous three-regime structure (allowing under-, over-, optimally-invest) is reduced to a two-regime structure (allowing under- and over-investment), restricting the probability of making efficient investment (i.e., $1 - p_u(W) - p_e(W)$) to be zero. However, this feature is very likely to hold in practice because firms are almost always invest inefficiently due to frictions under imperfect financial markets (Jensen and Meckling, 1976).

Under the two-regime structure, firms may under-invest with a conditional probability $E(\mathbb{1}(u_{it} > 0, e_{it} = 0)|W_{it}) = p_u(W_{it}) \ge 0$ and over-investment with $p_e(W_{it}) = 1 - p_u(W_{it})$. Thus, $\epsilon_{it} = v_{it} - u_{it}$ with probability $p_u(W_{it})$ and $\epsilon_{it} = v_{it} + u_{it}$ with probability $1 - p_u(W_{it})$ in (A.13). Equivalently,

$$\epsilon_{it} = v_{it} - u_{it}\mathbb{1}(u_{it} > 0, e_{it} = 0) + e_{it}[1 - \mathbb{1}(u_{it} > 0, e_{it} = 0)]. \tag{A.14}$$

Recall that $E(u_{it}|u_{it} > 0, e_{it} = 0, W_{it}) \equiv \mu_0^u > 0$ and $E(e_{it}|e_{it} > 0, u_{it} = 0, W_{it}) \equiv \mu_0^e > 0$ by our assumption. Define $\tilde{\mu}_0^u = \mu_0^u + \mu_0^e > 0$, we have

$$E(\epsilon_{it}|\alpha_i, \eta_t, W_{it}) = -\left[\mu_0^u p_u(W_{it}) - \mu_0^e(1 - p_u(W_{it}))\right] = -[\tilde{\mu}_0^u p_u(W_{it}) - \mu_0^e] \equiv -\pi(W_{it}),$$

where $\pi(\cdot)$ a linear transformed function of $p_u(W_{it})$ with a location shifter $-\mu_0^e$ and scaling factor

$\tilde{\mu}_0^u$. The regression form of the SEM-SP-R is

$$I_{it} = m_0 + \alpha_i + \eta_t + \sum_{j=1}^{d} m_j(X_{j,it}) + \sum\sum_{1\le j<l\le d} H_{jl}(X_{j,it}, X_{l,it}) - \pi(W_{it}) + \tilde{\epsilon}_{it}, \tag{A.15}$$

where $\tilde{\epsilon}_{it} = \epsilon_{it} + \pi(W_{it})$. Compared to the regression form of SEM-SP in (A.4), SEM-SP-R replaces parametric functions $\pi(W_{it};\omega_0)$ with its nonparametric version $\pi(W_{it})$, and further includes a total of $d(d-1)/2$ interaction functions $H_{jl}(\cdot)$. Define $p_u^*(W_{it}) = \tilde{\mu}_0^u p_u(W_{it})$ as a scaled probability function. Since $\Delta\pi(W_{it}) = [p_u^*(W_{it}) - p_u^*(W_{it-1})] = \Delta p_u^*(W_{it})$, we eliminate the fixed effects by applying the first-difference on (A.15) to have

$$\Delta I_{it} = \Delta D_{t,-1}^{\top}\eta_{0,-1} + \sum_{j=1}^{d} \Delta m_j(X_{j,it}) + \sum\sum_{1\le j<l\le d} \Delta H_{jl}(X_{j,it}, X_{l,it}) - \Delta p_u^*(W_{it}) + \Delta\tilde{\epsilon}_{it}. \tag{A.16}$$

Our estimation targets are $\{m_j(\cdot), H_{jl}(\cdot)\}_{j=1,j<l}^{d-1,d}$ and $p_u^*(\cdot)$. The main challenge in (A.16) is that $p_u^*(W) = \tilde{\mu}_0^u p_u(W) \ge 0$ is non-negative, although it is not upper bounded by one as $\tilde{\mu}_0^u > 0$. To impose its non-negativity, we assume $p_u^*(W)$ take a flexible structure as follows

$$p_u^*(W) = \tilde{\mu}_0^u p_u(W) = \exp\left(\ln(\tilde{\mu}_0^u) + g(W)\right) \ge 0, \tag{A.17}$$

where $g(W)$ is a unbounded and auxiliary nonparametric function of W. Since the sign of $p_u^*(W)$'s derivative only depends on the sign of $g(W)$'s derivatives, our use of $g(W)$ can be useful to capture arbitrary nonlinear, particularly non-monotonic functional form of the unknown probability function.

Estimating the model in (A.16) using assumption in (A.17) calls for proper identification conditions. First, as in SEM-SP, we normalize fixed effects as $\sum_{i=1}^n \alpha_i = \sum_{t=1}^T \eta_t = 0$, implying $\eta_{0,1} = -\sum_{s=2}^T \eta_{0,s}$. Second, we "anchor" the EIFs in (A.16) by assuming that each function passes through origin points as

$$m_j(X_j = 0) = 0, \forall j = 1, \ldots, d,\ H_{jl}(X_j = 0, X_l) = H_{jl}(X_j, X_l = 0) = 0, \forall 1 \le j < l \le d. \tag{A.18}$$

Last, the identification of $p_u^*(W)$ is subtle. Notice first that $\ln(\tilde{\mu}_0^u)$ cannot be separately identified from $g(W)$ because $\ln(\tilde{\mu}_0^u) + g(W)$ is additive invariant, i.e., $[\tilde{\mu}_0^u + c] + [g(W) - c] \equiv \tilde{\mu}_0^{u*} + g^*(W)$ for any constant c. To identify $g(W)$ from $p_u^*(W)$, we adopt the following identification condition

$$g(W_1 = 0, W_2, \ldots, W_d) = g(W_1, W_2 = 0, \ldots, W_d) = \cdots = g(W_1, W_2, \ldots, W_q = 0) = 0. \tag{A.19}$$

(A.19) implies that $g(\cdot)$ passes through origin point or has a zero intercept.

It should be clear that one cannot identify $p_u(W)$ by $g(W)$, but only identify $p_u(W)$ up to scale

by $p_u^*(W)$ through $\exp\left(\ln(\tilde{\mu}_0^u) + g(W)\right)$. Identifying $p_u^*(W)$ is important to access under-investment probability $p_u(W)$. For instance, if we impose distributional assumption on u and e, say each follows a standard half-normal distribution, we can identify $p_u(W) = p_u^*(W)/\tilde{\mu}_0^u$, where $\tilde{\mu}_0^u = 2\sqrt{2/\pi}$. In general without distributional assumptions, $p_u^*(W)$ can be identified up to a constant because it enters the model additively along with other additive EIFs. This does not create an issue if we are interested in the percentage change of $p_u(W_{it})$, which is invariant to arbitrary constant and scale parameter $\tilde{\mu}_0^u$ and can be easily obtained through $(p_u^*(W_{it}) - p_u^*(W_{it-1}))/p_u^*(W_{it-1})$. Instead, if we are interested in identifying $p_u^*(W)$, we propose a similar zero-intercept identification condition as in (A.19):

$$p_u^*(W_1 = 0, W_2, \ldots, W_d) = p_u^*(W_1, W_2 = 0, \ldots, W_d) = \cdots = p_u^*(W_1, W_2, \ldots, W_q = 0) = 0. \quad \text{(A.20)}$$

Because $p_u^*(W)$ is non-negative, this above condition implies that $p_u^*(W)$ contains a zero intercept. For instance, if $q = 1$, the assumed structure of $p_u^*(W)$ in (A.17) and the identification conditions (A.19)-(A.20) implies that $p_u^*(W) = \exp\left(\ln(\tilde{\mu}_0^u) + g(W)\right) - \exp\left(\ln(\tilde{\mu}_0^u) + g(0)\right) = \tilde{\mu}_0^u[\exp(g(W)) - 1]$, satisfying $p_u^*(0) = 0$.

The improved flexibility of SEM-SP-R significantly alleviates model misspecification by allowing for unknown interaction and probability functions. However, it comes with cost of the curse-of-dimentionality, i.e., the local smoothing needs to be performed for two-dimensional $H_{jl}(\cdot)$ and q-dimensional $p_u^*(\cdot)$. Given moderately large sample size, Wang et al. (2024) show that the bivariate smoothing works fairly well in a semiparametric SFM with only individual and interactive functions. Different from them, our SEM-SP-R needs to account for an additional $p_u^*(W)$, but a large q would significantly downgrade the precision of estimation for $p_u^*(\cdot)$. For this reason, we restrict the dimension of W to be small in practice to alleviate the curse of dimensionality.

Below, we propose our nonparametric estimator for the SEM-SP-R in Appendix 1.4.1 and demonstrate their finite-sample performance in Appendix 1.4.2.

Appendix 1.4.1

Focusing on the differenced regression of SEM-SP-R in (A.16), we estimate unknown functions $m_j(\cdot)$, $H_{jl}(\cdot)$, and $p_u^*(\cdot)$ in two steps. In the first step, we approximate all unknown functions through series estimator. Recall that $\mathcal{X}_j = [a_j, b_j] \subset \Re$ is a compact support of X_j for some finite constants a_j and b_j. Without loss of generosity, we consider $\mathcal{X}_j = \mathcal{X}$ for all j with $a_j = a = -1$ and $b_j = b = 1$. Similarly, define $\mathcal{W}$ being a common compact support of W_j, for $j = 1, \ldots, q$. To approximate each individual function $m_j(x_j)$, we adopt the basis function $\phi^\kappa(x_j) = [\phi_1(x_j), \ldots, \phi_\kappa(x_j)]^\top$ and series coefficients $\theta_j \in \Re^\kappa$ to obtain $m_j(x_j) \approx \phi^\kappa(x_j)^\top \theta_j$. So $\Delta m_j(X_{j,it}) = m_j(X_{j,it}) - m_j(X_{j,it-1}) \approx (\phi^\kappa(X_{j,it}) - \phi^\kappa(X_{j,it-1}))^\top \theta_j \equiv \Delta\phi_\kappa(X_{j,it})^\top \theta_j$, where $\Delta\phi^\kappa(X_{j,it}) = [\Delta\phi_1(X_{j,it}), \ldots, \Delta\phi_\kappa(X_{l,it})]^\top$. Thus, $\sum_{j=1}^d \Delta m_j(X_{j,it}) \approx \Delta\phi_m(X_{it})^\top \theta_m$, where $\Delta\phi_m(X_{it}) = [\Delta\phi^\kappa(X_{1,it})^\top, \ldots, \Delta\phi^\kappa(X_{d,it})^\top]^\top$ and

$\theta_m = [\theta_1^\top, \ldots, \theta_d^\top]^\top \in \Re^{d\kappa}$. To approximate interaction functions, we have $\Delta H_{jl}(X_{j,it}, X_{l,it}) \approx \Delta\phi^{\kappa_1}(X_{j,it}, X_{l,it})^\top \theta_{j,l}$, where for $\kappa_1 = \kappa^2$, $\theta_{j,l} \in \Re^{\kappa_1}$ and $\Delta\phi^{\kappa_1}(X_{j,it}, X_{l,it}) = [\Delta\phi^{\kappa}(X_{j,it})^\top \otimes \Delta\phi^{\kappa}(X_{l,it})^\top]^\top$ is the tensor product of two basis functions evaluated at different inputs. Thus, $\sum\limits_{1\le j<l\le d}\sum \Delta H_{jl}(X_{j,it}, X_{l,it}) \approx \Delta\phi_H(X_{it})^\top \theta_H$, where

$$\Delta\phi_H(X_{it}) = [\Delta\phi^{\kappa_1}(X_{1,it}, X_{2,it})^\top, \ldots, \Delta\phi^{\kappa_1}(X_{1,it}, X_{d,it})^\top, \ldots, \Delta\phi^{\kappa_1}(X_{d-1,it}, X_{d,it})^\top]^\top$$

is a $(d(d-1)\kappa_1/2) \times 1$ vector and $\theta_H = [\theta_{1,2}^\top, \ldots, \theta_{1,d}^\top, \ldots, \theta_{d-1,d}^\top]^\top$. Finally, to approximate the scaled probability function, we have $p_u^*(W_{it}) \approx \exp(\phi^{\kappa_2}(W_{it})^\top \theta_p)$, where for $\kappa_2 = \kappa^q$, $\theta_p \in \Re^{\kappa_2}$ and $\phi^{\kappa_2}(W_{it}) = [\phi^{\kappa}(W_{1,it})^\top \otimes \phi^{\kappa}(W_{2,it})^\top \otimes \cdots \otimes \phi^{\kappa}(W_{q,it})^\top]^\top$ is the tensor product of all basis functions evaluated at $W_{1,it}, \ldots, W_{q,it}$. Notice that the first element in each $\phi^{\kappa}(W_{s,it})^\top$ is one in order to approximate $\tilde{\mu}_0^u$. The rest elements in each $\phi^{\kappa}(W_{s,it})^\top$ are empirically centered around zero to impose the condition in (A.19). Notice also that, in the first step, there is no need to impose level-restriction on $p_u^*(W_{it})$ in (A.20) because constants will be wiped out by the first-differencing. So, we approximate $\Delta p_u^*(W_{it}) \approx \exp(\phi^{\kappa_2}(W_{it})^\top \theta_p) - \exp(\phi^{\kappa_2}(W_{it-1})^\top \theta_p) = \Delta \exp(\phi^{\kappa_2}(W_{it})^\top \theta_p)$.

For $\Delta\Phi(X_{it}) = [\Delta D_{t,-1}^\top, \Delta\phi_m(X_{it})^\top, \Delta\phi_H(X_{it})^\top]^\top$ and $\theta_f = [\eta_{0,-1}^\top, \theta_m^\top, \theta_H^\top]^\top$, we estimate $\Delta D_{t,-1}^\top \eta_{0,-1} + \sum_{j=1}^d \Delta m_j(X_{j,it}) + \sum\limits_{1\le j<l\le d}\sum \Delta H_{jl}(X_{j,it}, X_{l,it}) \approx \Phi(X_{it})^\top \theta_f$. Hence, the entire differenced model in (A.16) can be approximated as

$$\Delta I_{it} \approx \Delta\Phi(X_{it})^\top \theta_f - \Delta \exp(\phi_p(W_{it})^\top \theta_p) + \Delta\tilde{\epsilon}_{it}.$$

Different from θ_f, coefficients θ_p is not additively separable from basis functions. We propose to estimate $\theta = [\theta_f^\top, \theta_p^\top]$ through its nonlinear least square estimator $\hat{\theta}$ from

$$\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}} \sum_{i=1}^n \sum_{t=2}^T \left(\Delta I_{it} - \Delta\Phi(X_{it})^\top \theta_f + \Delta \exp(\phi_p(W_{it})^\top \theta_p)\right)^2, \tag{A.21}$$

where $\Theta \subset \Re^{S(\kappa)}$ is a compact support, where $S(\kappa) = T - 1 + d\kappa + d(d-1)\kappa_1/2 + \kappa_2$. The estimator in (A.21) is constructed in a similar fashion as in Horowitz and Mammen (2004) for a generalized additive model with known link function and cross-sectional data. The difference is that we use panel data, consider a differenced additive model with interactions, and apply a known (exponential) link function only to the non-negative probability function. Using $\hat{\theta}$, we obtain series estimates for time fixed effects $\hat{\eta}_t$ and functions estimates as

$$\hat{m}_j(x_j) = \phi^{\kappa}(x_j)^\top \hat{\theta}_j, \quad \hat{H}_{jl}(x_j, x_l) = \phi^{\kappa_1}(x_j, x_l)^\top \hat{\theta}_{j,l}, \quad \hat{p}_u^*(w) = \exp(\phi^{\kappa_2}(w)^\top \hat{\theta}_p).$$

In the second step, we propose a one-step kernel backfitting for functions $m_j(\cdot)$, $H_{jl}(\cdot)$, and $p_u^*(\cdot)$, as well as their derivatives $m_j^{(1)}(\cdot)$, $H_{jl}^{(1)}(\cdot)$, and $p_u^{*(1)}(\cdot)$ to improve estimation efficiency.

First, to backfit individual function $m_j(x_j)$, based on (A.16) we consider the following regression model

$$\hat{I}^{(m)}_{it,-j} = m_j(X_{j,it}) + \tilde{\epsilon}_{it},$$

where $\hat{I}^{(m)}_{it,-j} = \Delta I_{it} - \hat{\eta}_t + \hat{m}_j(X_{j,it-1}) - \sum_{l=1,l\neq j}^{d} \Delta\hat{m}_l(X_{l,it}) - \sum_{1\leq j<l\leq d}\sum \Delta\hat{H}_{jl}(X_{j,it}, X_{l,it}) + \Delta\hat{p}^*_u(W_{it})$ is a constructed dependent variable, which collects all estimates from the first step except for $m_j(X_{j,it})$, for $t = 2,\ldots,T$. As in SEM-SP, we obtain the backfitting local-linear estimator for $m_j(x_j)$ as $\tilde{m}_j(x_j) \equiv \tilde{a}_0$, and for $m^{(1)}_j(x_j) \equiv a_1$ as $\tilde{m}^{(1)}_j(x_j) \equiv \tilde{a}_1$, from

$$(\tilde{a}_0, \tilde{a}_1) = \underset{\{a_0,a_1\}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{t=2}^{T} \left[\hat{I}^{(m)}_{it,-j} - a_0 - (X_{j,it} - x_j)a_1\right]^2 K\left(\frac{X_{j,it} - x_j}{h_j}\right). \tag{A.22}$$

Second, to backfit interaction function $H_{jl}(x_j, x_l)$, we consider the following regression model

$$\hat{I}^{(H)}_{it,-jl} = H_{jl}(X_{j,it}, X_{l,it}) + \tilde{\epsilon}_{it},$$

where $\hat{I}^{(H)}_{it,-j} = \Delta I_{it} - \hat{\eta}_t + \hat{H}_{jl}(X_{j,it-1}, X_{l,it-1}) - \sum_{j=1}^{d} \Delta\hat{m}_j(X_{j,it}) - \sum_{\substack{1\leq j'<l'\leq d,\\ (j',l')\neq(j,l)}}\sum \Delta\hat{H}_{j'l'}(X_{j',it}, X_{l',it}) + \Delta\hat{p}^*_u(W_{it})$ collects all estimates from the first step except for $H_{jl}(X_{j,it}, X_{l,it})$. We obtain the backfitting local-linear estimator for $H_{jl}(x_j, x_l)$ as $\tilde{H}_{jl}(x_j, x_l) \equiv \tilde{b}_0$, and for $H^{(1)}_{jl}(x_j, x_l) \equiv b_1^\top$ as $\tilde{m}^{(1)}_j(x_j) \equiv \tilde{b}_1^\top$, from

$$(\tilde{b}_0, \tilde{b}_1) = \underset{\{b_0,b_1\}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{t=2}^{T} \left[\hat{I}^{(H)}_{it,-jl} - b_0 - (X_{jl,it} - x_{jl})^\top b_1\right]^2 K\left(\frac{X_{jl,it} - x_{jl}}{h_j}\right), \tag{A.23}$$

where $(X_{jl,it} - x_{jl}) = [(X_{j,it} - x_j), (X_{l,it} - x_l)]^\top$ and $K\left(\frac{X_{jl,it}-x_{jl}}{h_j}\right) = K\left(\frac{X_{j,it}-x_j}{h_j}\right) K\left(\frac{X_{l,it}-x_l}{h_l}\right)$ is a bivariate product kernel function.

Finally, to backfit probability function $p^*_u(w)$, we consider the following regression model

$$\hat{I}^{(p)}_{it} = -p^*_u(W_{it}) + \tilde{\epsilon}_{it},$$

where $\hat{I}^{(p)}_{it} = \Delta I_{it} - \hat{\eta}_t - \hat{p}^*_u(W_{it-1}) - \sum_{j=1}^{d} \Delta\hat{m}_j(X_{j,it}) - \sum_{1\leq j<l\leq d}\sum \Delta\hat{H}_{jl}(X_{j,it}, X_{l,it})$ collects all estimates from the first step except for $p^*_u(W_{it})$. Inspired by Ziegelmann (2002), we obtain the backfitting exponential local-linear estimator for $g(w)$ as $\tilde{g}(w) \equiv \tilde{c}_0$, and for $g^{(1)}(w)$ as $\tilde{g}^{(1)}(w) \equiv \tilde{c}_1^\top$, from

$$(\tilde{c}_0, \tilde{c}_1) = \underset{\{c_0,c_1\}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{t=2}^{T} \left[\hat{I}^{(p)}_{it} + \exp\left(c_0 + (W_{it} - w)^\top c_1\right)\right]^2 K\left(\frac{W_{it} - w}{h_g}\right), \tag{A.24}$$

where $(W_{it} - w) = [(W_{1,it} - w), \ldots, (W_{q,it} - w)]^\top$ and $K\left(\frac{W_{it}-w}{h_g}\right) = \prod_{s=1}^{q} K\left(\frac{W_{s,it}-w_j}{h_{g,s}}\right)$ is a q-dimensional product kernel function. In practice when one needs to impose the identification condition in (A.20), we propose to empirically center the estimates $\tilde{p}_u^*(w) = \tilde{\mu}_0^* \exp\left(\tilde{g}(w)\right)$ around origin. For instance, if $q = 1$, we construct the centered version as $\tilde{p}_u^*(w) - \tilde{p}_u^*(0)$.

Appendix 1.4.2

We proceed to investigate the finite-sample performance of our proposed two-step estimator for SEM-SP-R in (A.13)-(A.14). We focus on the backfitting kernel estimator $\tilde{m}_j(x_j)$ in (A.22), $\tilde{H}_{jl}(x_j, x_l)$ in (A.23), and scaled probability function $\tilde{p}_u^*(w)$ in (A.24). We retain the previous sample size by setting $T = 15$ and choose $n = (100, 200, 400)$. To reduce computational burden, we set $d = 2$ to generate data according (A.13)-(A.14) as:

$$I_{it} = m_0 + \alpha_i + \eta_t + m_1(X_{1,it}) + m_2(X_{2,it}) + H_{12}(X_{1,it}, X_{2,it}) + \epsilon_{it}, \tag{A.25}$$

where $\epsilon_{it} = v_{it} - u_{it}\mathbb{1}(u_{it} > 0, e_{it} = 0) + e_{it}[1 - \mathbb{1}(u_{it} > 0, e_{it} = 0)]$ follows a two-regime structure such that $\epsilon_{it} = v_{it} - u_{it}$ with probability $p_u(W_{it})$ and $\epsilon_{it} = v_{it} + e_{it}$ with probability $1 - p_u(W_{it})$. We consider a simple setup by choosing a univariate W, so $q = 1$. We specify two DGPs with the same specifications on $m_1(X_1)$ and $m_2(X_2)$ as in (A.10), but consider different form of $H_{12}(X_1, X_2)$ and $p_u^*(W) = \tilde{\mu}_0^u p_u(W)$. In DGP_1, $H_{12}(v_1, v_2) = \sin(1.5v_1)v_2^2$ and $p_u(v) = 3v^2/\pi^3$. In DGP_2, $H_{12}(v) = 0.5v_1v_2$ and $p_u(v) = \sin(v)/2$. For $0 \leq W \leq \pi$, the probability function is designed to increase with W in DGP_1 but decreases with W in DGP_2. Also, $p_u(W)$ satisfies $0 \leq p_u(W) \leq 1$ and $\int_0^\pi p_u(w)dw = 1$.

We generate auxiliary variables $\{X_{j,it}^0\}_{j=1}^2$ from a bivariate normal distribution $\mathcal{N}(\mu_0, \Sigma_0)$, where $\mu_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\Sigma_0 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ is the covariance matrix. We also generate $W_{it}^0 \sim U(0, \pi)$. To introduce dependence across time, we generate $X_{j,it} = X_{j,it}^0 + \zeta_{j,it}^x$ (with $j = 1, 2$), and $W_{it} = W_{it}^0 + \zeta_{it}^w$, where $\zeta_{j,it}^x = 0.25\zeta_{j,it-1}^x + \xi_{it}$ and $\zeta_{it}^w = 0.5\zeta_{it-1}^w + \xi_{it}$ follow an AR(1) process with $\xi_{it} \sim \mathcal{N}(0, 0.25^2)$. Here, we generate $X_{j,it}$ with values falling into $[-2, 2]$ and W_{it} falling into $[0, \pi]$

We construct the firm fixed effect $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} c_0(X_{1,it} + X_{2,it} + W_{it}) + \xi_i^\alpha$, where $\xi_i^\alpha \sim \mathcal{N}(0, 1)$ and $c_0 = 1$. Likewise, we generate time fixed effect as $\eta_t = \frac{1}{n}\sum_{i=1}^{n} c_0(X_{1,it} + X_{2,it} + W_{it}) + \xi_t^\eta$, where $\xi_t^\eta \sim \mathcal{N}(0, 1)$ and $c_0 = 1$. We empirically center (α_i, η_t) to be consistent with our identification condition. Finally, we employ a cubic B-spline series estimator in (A.21) with each knot evenly spaced on the support of X_1 and X_2, and a 2nd-order Gaussian kernel function in (A.22)-(A.24). As discussed in Appendix 1.4.1, we impose the identification condition for $m_1(\cdot)$, $m_2(\cdot)$, $H_{12}(\cdot)$, $g(\cdot)$ and $p_u^*(\cdot)$ by empirically centering their series and kernel estimates around zero. All the turning parameters are chosen following the guidance in Appendix 1.3 by setting $\varrho = 0.1$ and $C = 1$.

We perform $1,000$ repetitions throughout the experiment. We evaluate the performance of func-

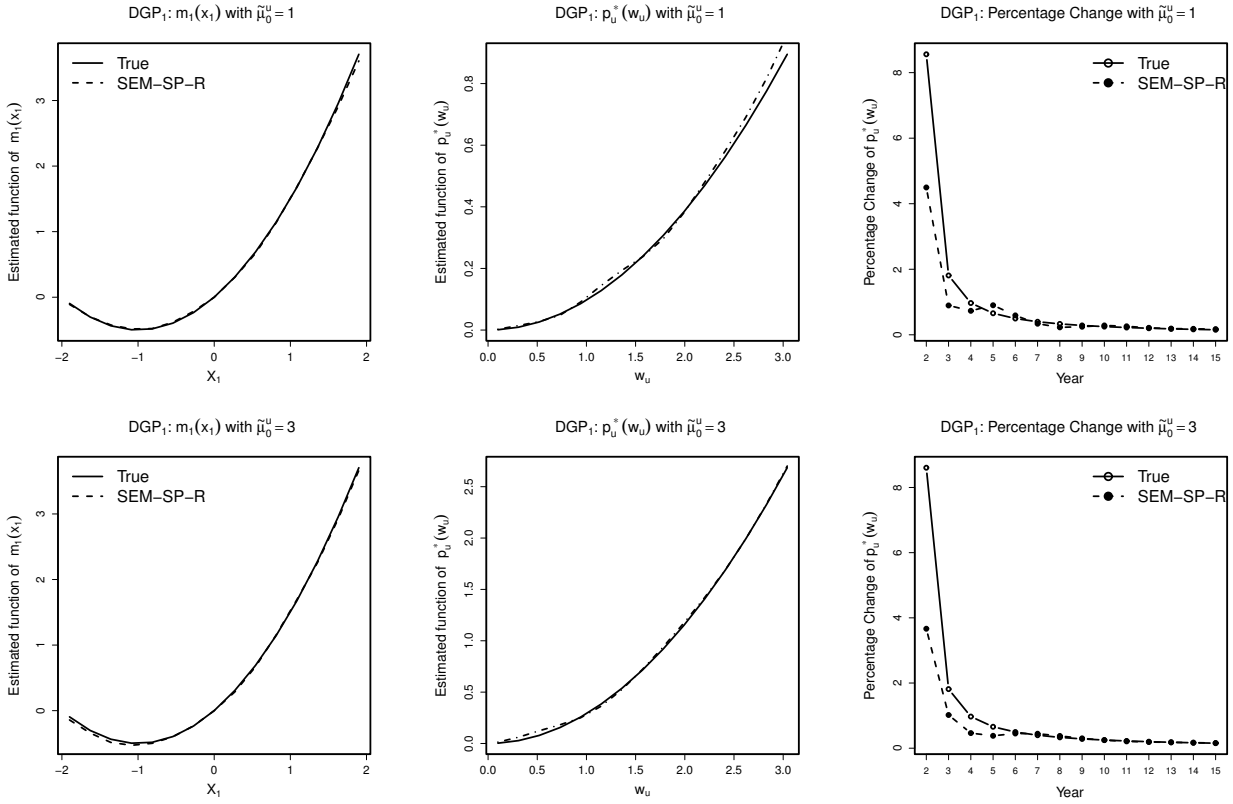
Table A1.4: SEM-SP-R Estimation Results with $T = 15$

			$\tilde{\mu}_0^u = 1$			$\tilde{\mu}_0^u = 3$	
Panel A	DGP_1	$n = 100$	200	400	$n = 100$	200	400
$\tilde{m}_1$	RAMSE	0.0550	0.0420	0.0314	0.0609	0.0472	0.0357
	ABIAS	0.0401	0.0310	0.0237	0.0446	0.0349	0.0267
	ASD	0.0479	0.0377	0.0289	0.0538	0.0428	0.0328
$\tilde{m}_2$	RAMSE	0.1162	0.0910	0.0704	0.1199	0.0940	0.0725
	ABIAS	0.0938	0.0749	0.0583	0.0968	0.0764	0.0592
	ASD	0.0504	0.0379	0.0281	0.0548	0.0435	0.0325
$\tilde{H}_{12}$	RAMSE	0.3592	0.2354	0.1907	0.3600	0.2358	0.1907
	ABIAS	0.2917	0.1868	0.1465	0.2925	0.1871	0.1466
	ASD	0.0763	0.0509	0.0343	0.0829	0.0553	0.0363
$\tilde{p}_u^*$	RAMSE	0.0446	0.0329	0.0250	0.0578	0.0439	0.0319
	ABIAS	0.0314	0.0238	0.0184	0.0407	0.0317	0.0234
	ASD	0.0384	0.0294	0.0225	0.0424	0.0321	0.0242
			$\tilde{\mu}_0^u = 1$			$\tilde{\mu}_0^u = 3$	
Panel B	DGP_2	n=100	200	400	n=100	200	400
$\tilde{m}_1$	RAMSE	0.0719	0.0557	0.0402	0.0820	0.0785	0.0518
	ABIAS	0.0534	0.0419	0.0327	0.0577	0.0461	0.0372
	ASD	0.0465	0.0369	0.0286	0.0603	0.0581	0.0487
$\tilde{m}_2$	RAMSE	0.0650	0.0474	0.0346	0.0786	0.0648	0.0481
	ABIAS	0.0459	0.0339	0.0252	0.0490	0.0391	0.0274
	ASD	0.0484	0.0368	0.0275	0.0628	0.0547	0.0420
$\tilde{H}_{12}$	RAMSE	0.3384	0.2568	0.2040	0.5052	0.4608	0.3204
	ABIAS	0.2556	0.1938	0.1548	0.2844	0.2352	0.1674
	ASD	0.2808	0.2154	0.1716	0.4494	0.4104	0.2808
$\tilde{p}_u^*$	RAMSE	0.2168	0.1332	0.0872	0.2950	0.1682	0.0864
	ABIAS	0.1495	0.0886	0.0515	0.1824	0.1099	0.0665
	ASD	0.1586	0.1025	0.0745	0.1521	0.1257	0.0756

tion estimates $(\tilde{m}_1(\cdot), \tilde{m}_2(\cdot), \tilde{H}_{12}(\cdot), \tilde{p}_u^*(\cdot))$ through the root averaged mean squared error (RAMSE), the averaged absolute BIAS (ABIAS), and averaged standard deviation (ASD) across 50 grid points in $[0, 3.14]$.

Table A1.4 summarizes the estimation performance under DGP_1 (upper panel) and DGP_2 (lower panel). Each panel reports the results for two different scaling parameter, $\tilde{\mu}_0^u = 1$ and $\tilde{\mu}_0^u = 3$, to access their impact on estimation, particularly on $\tilde{p}_u^*(\cdot)$. The results for the two individual functions are qualitatively similar to their early results in Table A1.2. The interaction function estimates $\tilde{H}_{jl}(\cdot)$ exhibit a worsen performance compared to the univariate estimates $\tilde{m}_j(\cdot)$, which aligns with the findings of Wang et al. (2024). This deterioration is expected due to the slower convergence rate of bivariate smoothing. The scaled probability estimates perform nearly as well as $\tilde{m}_1(\cdot)$ under DGP_1, but deteriorate under DGP_2 in terms of larger bias and variance, particular for the smaller sample size ($n = 100$). We observe no significant impact of $\tilde{\mu}_0^u$ on the estimates under DGP_1. For instance, the RAMSE of $\tilde{H}_{12}(\cdot)$ (0.1907) remains the same across different values of $\tilde{\mu}_0^u$

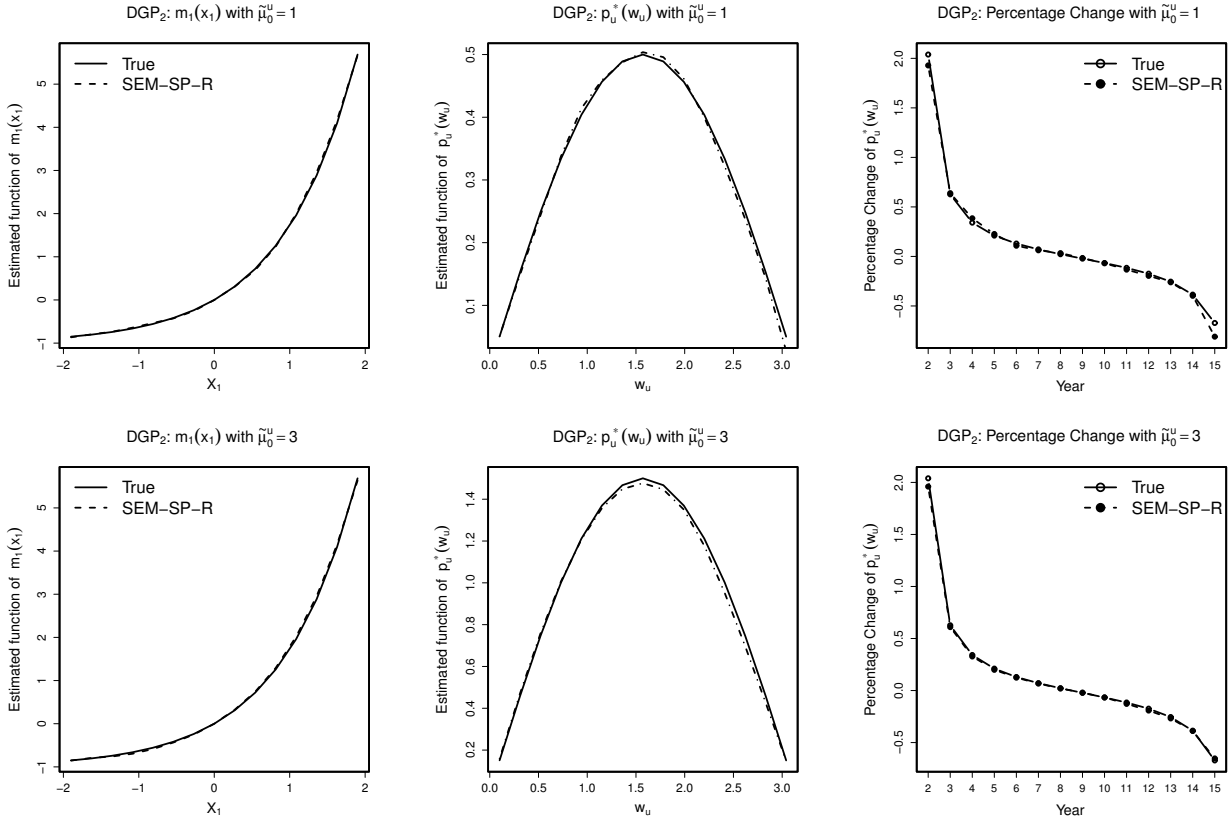
Figure A.3: SEM-SP-R Estimation from DGP_1 for $m_1(x_1)$, $p_u^*(w)$, and Percentage Change



under $n = 400$. However, the impact is more pronounced under DGP_2 by comparing the results for $\tilde{H}_{12}(\cdot)$ and $\tilde{p}_u^*(\cdot)$. This suggests that the periodic function $p_u^*(\cdot)$, when scratched out by $\tilde{\mu}_0^u > 1$, may introduce additional challenges in model estimation. Nonetheless, as n gets large, all estimates improve judged by decreasing RAMSE, ABIAS, and ASD across both DGPs, demonstrating their consistency.

We further provide a vivid picture on the performance of our estimators in Figure A.3 for DGP_1 and Figure A.4 for DGP_2, using the largest sample size of $(n, T) = (400, 15)$. For brevity, each figure presents estimates under $\tilde{\mu}_0^u = 1$ (first row) or $\tilde{\mu}_0^u = 3$ (second row), where the first column plots $m_1(\cdot)$, the second column plots $p_u^*(\cdot)$, and the third column shows the percentage change in $p_u(W_{1t})$, computed as $(p_u(W_{1t}) - p_u(W_{1t-1}))/p_u(W_{1t-1})$ using the first unit $(i = 1)$ for $t = 2, \ldots, 15$. In both DGPs, our function estimates (dashed line) closely capture the nonlinear structure of the true functions $m_1(x_1)$ and $p_u^*(w)$ (solid line), regardless of the value of $\tilde{\mu}_0^u$. Additionally, the estimated percentage change in $\tilde{p}_u^*(W_{1t})$ (dashed line with empty circles) is computed *without* imposing the zero-intercept identification condition (A.20) on $p_u^*(W)$. As expected, irrespective of $\tilde{\mu}_0^u$, the estimated percentage change in $p_u(W_{1t})$ (dashed line with empty circles) closely follows the true percentage change (solid line with solid circles), except at boundary points of w where larger

Figure A.4: SEM-SP Estimation from DGP_2 for $m_1(x_1)$, $p_u^*(w)$, and Percentage Change



deviations occur. This result aligns with our discussion in Appendix 1.4.1, reinforcing that the identification condition for $p_u^*(W)$ is not required when estimating the percentage change in $p_u(W)$ over time.

Overall, our SEM-SP-R demonstrates its promising numerical performance in estimating unknown individual and interaction EIFs, as well as probability scaled by a single parameter. It can be served as a viable alternative in application where the interactive effects and nonparametric probability function are deemed relevant.

Appendix 2: Nonparametric Tests for Function and Composite Error Structures

Section 3.4 of the paper outlines nonparametric tests (I_n and J_n) for correct specifications on parametric efficiency functions under H_{01}, two-regime composite error under H_{02}, and three-regime composite error under H_{03}. Below, we provide detailed implementation procedures of I_n for H_{01} in Appendix 2.1, J_n for H_{02} in Appendix 2.2, and J_n for H_{03} in Appendix 2.3.

Appendix 2.1: Testing for correct parametric efficiency functions

The efficiency functions $\{m_j(\cdot)\}_{j=1}^d$ under SEM-SP in (A.1) are nonparametrically specified to capture possible unknown nonlinear effect of each X_j. Under a special case where X_j enters the model linearly, SEM-SP reduces to SP-OLS (defined in Section 2.2) where each $m_j(\cdot) = X_j\beta_j$ can be parametrically specified for unknown parameter β_j. Thus, a consistent test for correct parametric functional form of $m_j(\cdot)$ is important to improve estimation efficiency in our applied study. We base our first null hypothesis as

$$H_{01} : Pr(m_j(X_j) = X_j\beta_j) = 1, \text{ almost surely,}$$

where the parametric specification on each efficiency function is correct. Inspired by Lin et al. (2014), we test H_{01} by a nonparametric test based on the integrated error of function difference $I_n = \int [m_j(x_j) - x_j\beta_j]^2 \, dx_j$.

To construct a feasible version of I_n for testing the underlying structure of $m_j(\cdot)$, first notice from (A.4) that $I_{it}^{(-j)} = m_0 + \alpha_i + m_j(X_{j,it}) + \tilde{\epsilon}_{it}$, where $I_{it}^{(-j)} = I_{it} - \eta_t - \sum_{l=1,l\neq j} m_l(X_{j,it}) + \pi(W_{it};\omega_0)$ is a psudo-response for X_j. Under H_{01}, we observe that

$$I_{it}^{(-j)} = m_0 + \alpha_i + X_j\beta_j + \tilde{\epsilon}_{it}. \tag{A.26}$$

Using our PNLS estimates $\hat{\omega}$, series estimates $\hat{\eta}$, and kernel backfitting estimates $\tilde{m}_0$ and $\tilde{m}_j(\cdot)$, we estimate $I_{it}^{(-j)}$ by $\tilde{I}_{it}^{(-j)} = I_{it} - \tilde{m}_0 - \hat{\eta}_t - \sum_{l=1,l\neq j} \tilde{m}_l(X_{j,it}) + \pi(W_{it};\hat{\omega})$, and regress $\tilde{I}_{it}^{(-j)}$ on X_j in (A.26) to obtain conventional within estimator $\hat{\beta}_j$. Following the arguments in Lin et al. (2014), we construct the feasible test statistic for I_n as

$$\hat{I}_n = \frac{1}{n^2 h_j} \sum_{i=1}^{n} \sum_{m=1\neq i}^{n} \sum_{t=1}^{T} \sum_{\tau=1\neq t}^{T} K\left(\frac{X_{j,it} - X_{j,m\tau}}{h_j}\right) \hat{\tilde{\epsilon}}_{it}^c \hat{\tilde{\epsilon}}_{m\tau}^c, \tag{A.27}$$

where $\hat{\tilde{\epsilon}}_{it}^c = \hat{\tilde{\epsilon}}_{it} - \frac{1}{T}\sum_{t=1}^{T} \hat{\tilde{\epsilon}}_{it}$ is the within-difference residual from (A.26). We further follow Lin et al. (2014) to employ a centered version of $\hat{I}_n$ as

$$\hat{I}_n^c = \frac{n\sqrt{h_j}\hat{I}_n}{\sqrt{\hat{\sigma}_n^2}}, \tag{A.28}$$

where

$$\hat{\sigma}_n^2 = \frac{2}{n^2 h_j} \sum_{i=1}^{n} \sum_{m=1\neq i}^{n} \sum_{t=1}^{T} \sum_{\tau=1\neq t}^{T} K^2\left(\frac{X_{j,it} - X_{j,m\tau}}{h_j}\right) (\hat{\tilde{\epsilon}}_{it}^c)^2 (\hat{\tilde{\epsilon}}_{m\tau}^c)^2. \tag{A.29}$$

Although $\hat{I}_n^c$ is expected to be asympotically normal as $n \to \infty$ with fixed T, it is known that nonparametric test suffers from finite sample size distortions due to its slower convergence speed. We

Table A2.1: Testing Results of H_{01} for Correct Parametric Efficiency Functions

Function being tested	α	Empirical Size ($\delta = 0$)			Empirical Power ($\delta = 1$)		
		n=100	200	400	n=100	200	400
$H_{01} : m_1(X_1) = X_1\beta_1$	0.01	0.005	0.007	0.013	1	1	1
	0.05	0.041	0.046	0.053	1	1	1
	0.1	0.082	0.093	0.104	1	1	1
$H_{01} : m_2(X_2) = X_2\beta_2$	0.01	0.005	0.008	0.009	1	1	1
	0.05	0.041	0.045	0.052	1	1	1
	0.1	0.084	0.095	0.103	1	1	1
$H_{01} : m_3(X_3) = X_3\beta_3$	0.01	0.003	0.004	0.007	1	1	1
	0.05	0.033	0.040	0.045	1	1	1
	0.1	0.085	0.092	0.098	1	1	1

follow Lin et al. (2014) to implement the following residual-based wild bootstrap testing procedure to better approximate the distribution of $\hat{I}_n^c$ given finite sample.

1) With the original sample $\Theta = \{I_{it}, X_{it}, W_{it}\}_{i=1,t=1}^{n,T}$, compute $\hat{I}_n^c$ in (A.28). Obtain the residual $\hat{\tilde{\epsilon}}_{it}^c = \hat{\tilde{\epsilon}}_{it} - \frac{1}{T}\sum_{t=1}^{T}\hat{\tilde{\epsilon}}_{it}$, where $\hat{\tilde{\epsilon}}_{it} = \tilde{I}_{it}^{(-j)} - X_j\hat{\beta}_j$ defined in (A.26).

2) With $\{\hat{\tilde{\epsilon}}_{it}\}_{i=1,t=1}^{n,T}$, obtain wild-bootstrap residual $\{\epsilon_{it}^*\}_{i=1,t=1}^{n,T}$, where $\epsilon_{it}^* = [(1-\sqrt{5})/2]\hat{\tilde{\epsilon}}_{it}$ with probability $P = (1+\sqrt{5})/2\sqrt{5}$, and $\epsilon_{it}^* = [(1+\sqrt{5})/2]\hat{\tilde{\epsilon}}_{it}$ with probability $1-P$. Construct the bootstrap sample $\Theta^* = \{I_{it}^*, X_{it}, W_{it}\}_{i=1,t=1}^{n,T}$, where $I_{it}^* = X_j\hat{\beta}_j + \epsilon_{it}^*$.

3) Compute $\hat{I}_n^{c*}$ similar as in 1), except replacing sample Θ with Θ^*.

4) Repeat step 2)-3) a large number (B) of times to obtain an empirical distribution from $\{\hat{I}_{n,b}^{c*}\}_{b=1}^{B}$. We reject H_{01} if $p_n^* < \alpha$, where $p_n^* = \frac{1}{B}\sum_{b=1}^{B} 1(\hat{I}_{n,b}^{c*} > \hat{I}_n^c)$ is the empirical p-value and α is the significant level.

We investigate the finite sample performance of $\hat{I}_n^c$ in terms of its empirical size and power. We adopt a similar DGP_1 as in Appendix 1.3, except that $m_1(v) = v + \delta\frac{1}{2}v^2$, $m_2(v) = v^2 - \delta\frac{1}{2}v^3$, and $m_3(v) = v + \delta 1.5v^4$ for a parameter $\delta \geq 0$. Thus, H_{01} is imposed with $\delta = 0$ and any deviation from H_{01} is indicated by $\delta \neq 0$. In our simulation study below, we investigate the size by setting $\delta = 0$ and power by $\delta = 1$. All other specifications under DGP_1 are maintained. We adopt a *rule of thumb* (ROT) bandwidth h_j to implement $\hat{I}_n^c$ and $\hat{I}_n^{c*}$ from $h_j = C_{h_j}\hat{\sigma}_j n^{-\frac{1}{6}}$, where $\hat{\sigma}_j$ is the empirical standard deviation of $\{X_{j,it}\}_{i=1,t=1}^{n,T}$, and $C_{h_j} = 1$ is used in our simulation for all $j = 1, 2, 3$. We fix $T = 15$, choose $n = (100, 200, 400)$, and obtain empirical size or power as the average of rejection frequency from 1000 repetitions, and from each repetition we set $B = 399$ to obtain bootstrap distribution under H_{01}.

Table A2.1 shows the testing results of $\hat{I}_n^c$. The empirical size (left panel) and power (right panel) are obtained for significant levels $\alpha = (0.01, 0.05, 0.1)$, and reported beneath each sample

size n. We observe that the size of $\hat{I}_n^c$ is underestimated when $n = 100$ or 200, but approaches steadily to its corresponding nominal level as n reaches 400. The empirical powers of $\hat{I}_n^c$ are all unities across different sample size, thus indicating that the test is consistent. The results overall reveal reasonable performance of our proposed test for H_{01}, which can be expected to provide valid testing results in our application given our large panel dataset.

Appendix 2.2: Testing for correct two-regime composite error

The composite error from SEM-SP in (A.2) assumes possible existence of under-investment (u) and over-investment (e) under imperfect capital markets. Under SFM in the context of investment efficiency analysis, only under-investment is assumed to exist as a result of financial constraints. If so, SFM is more suitable in our application because agency problem do not constitute firm's inefficient investment. We base our second null hypothesis as

$$H_{02} : Pr(e = 0) = 1, \text{ almost surely,}$$

where over-investment does not exist as assumed under SFM. The alternative is that either under-investment or over-investment may appear with probabilities, as assumed under SEM-SP. We test H_{02} by constructing a F-type test statistic as $J_n = \frac{(SSR_0 - SSR_1)}{SSR_1}$, where SSR_0 (SSR_1) stands for the sum of squared within-differenced residuals under restricted (unrestricted) SEM-SP. The restricted SEM-SP imposes H_{02} by assuming a two-regime composite error $\epsilon_{it} = v_{it} - u_{it}\mathbb{1}(u_{it} > 0, e_{it} = 0)$, from which firms may under-invest with probability $p_u(W_{it}^u; \gamma_0^u)$ or efficiently invest with probability $1 - p_u(W_{it}^u; \gamma_0^u)$. In this case, the inefficiency function is restricted as $\pi^0(W; \omega_0) = \mu_0 p_u(W_{it}^u; \gamma_0^u)$, with $\omega_0 = (\mu_0, \gamma_0^u)$. The unrestricted SEM-SP has the three-regime composite error in (A.2), which properly nests the restrictive SEM-SP under H_{02}.

To construct the feasible version of J_n, we first compute residuals under restricted SEM-SP as $\hat{\epsilon}_{it}^0 = I_{it} - \hat{\eta}_t - \sum_{j=1}^d \tilde{m}_j(X_{j,it}) + \pi^0(W_{it}; \hat{\omega})$, where all estimates above are obtained similarly as discussed in Appendix 1.1, except that μ_0^e and γ_0^e are not estimated. We next compute residuals under unrestricted SEM-SP as $\hat{\tilde{\epsilon}}_{it} = I_{it} - \hat{\eta}_t - \sum_{j=1}^d \tilde{m}_j(X_{j,it}) + \pi(W_{it}; \hat{\omega})$. Define the within-difference residuals $\hat{\tilde{\epsilon}}_{it}^{0,c} = \hat{\tilde{\epsilon}}_{it}^0 - \frac{1}{T}\sum_{t=1}^T \hat{\tilde{\epsilon}}_{it}^0$ under H_{02} and $\hat{\tilde{\epsilon}}_{it}^c = \hat{\tilde{\epsilon}}_{it} - \frac{1}{T}\sum_{t=1}^T \hat{\tilde{\epsilon}}_{it}$ under the alternative. The feasible test statistic of J_n can be obtained as

$$\hat{J}_n = \frac{(\widehat{SSR}_0 - \widehat{SSR}_1)}{\widehat{SSR}_1}, \tag{A.30}$$

where $\widehat{SSR}_0 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \hat{\tilde{\epsilon}}_{it}^{0,c}$ and $\widehat{SSR}_1 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \hat{\tilde{\epsilon}}_{it}^{c}$ are obtained under restricted and unrestricted SEM-SP, respectively.

As in our test $\hat{I}_n^c$ above, we obtain a more accurate finite sample distribution of $\hat{J}_n$ under H_{02} through the following bootstrap procedure.

1) With the original sample $\Theta = \{I_{it}, X_{it}, W_{it}\}_{i=1,t=1}^{n,T}$, compute $\hat{J}_n$ in (A.30). Obtain the restricted residuals $\hat{\epsilon}_{it}^{\hat{0},c} = \hat{\epsilon}_{it}^{\hat{0}} - \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{it}^{\hat{0}}$, and unrestricted residuals $\hat{\hat{\epsilon}}_{it}^{c} = \hat{\hat{\epsilon}}_{it} - \frac{1}{T}\sum_{t=1}^{T}\hat{\hat{\epsilon}}_{it}$, where $\hat{\epsilon}_{it}^{\hat{0}}$ and $\hat{\hat{\epsilon}}_{it}$ are constructed as in (A.30).

2) With $\{\hat{\epsilon}_{it}^{\hat{0}}\}_{i=1,t=1}^{n,T}$, obtain wild-bootstrap residual $\{\epsilon_{it}^{*}\}_{i=1,t=1}^{n,T}$, where $\epsilon_{it}^{*} = [(1-\sqrt{5})/2]\hat{\epsilon}_{it}^{\hat{0}}$ with probability $P = (1+\sqrt{5})/2\sqrt{5}$, and $\epsilon_{it}^{*} = [(1+\sqrt{5})/2]\hat{\epsilon}_{it}^{\hat{0}}$ with probability $1-P$. Construct the bootstrap sample $\Theta^{*} = \{I_{it}^{*}, X_{it}, W_{it}\}_{i=1,t=1}^{n,T}$, where $I_{it}^{*} = \hat{\eta}_t + \sum_{j=1}^{d}\tilde{m}_j(X_{j,it}) - \pi^0(W_{it};\hat{\omega}) + \epsilon_{it}^{*}$ uses all estimates from restricted SEM-SP.

3) Compute $\hat{J}_n^{*}$ similar as in 1), except replacing sample Θ with Θ^{*}.

4) Repeat step 2)-3) a large number (B) of times to obtain an empirical distribution from $\{\hat{J}_{n,b}^{*}\}_{b=1}^{B}$. We reject H_{02} if $p_n^{*} < \alpha$, where $p_n^{*} = \frac{1}{B}\sum_{b=1}^{B}1(\hat{J}_{n,b}^{*} > \hat{J}_n)$ is the empirical p-value and α is the significant level.

We investigate the finite sample size and power of $\hat{J}_n$. We adopt a similar DGP_1 as in Appendix 1.3, except that the composite error is modified as $\epsilon_{it} = v_{it} - u_{it}\mathbb{1}(u_{it} > 0, e_{it} = 0) + \delta e_{it}\mathbb{1}(u_{it} = 0, e_{it} > 0)$. Thus, H_{02} is imposed with $\delta = 0$ and any departure from H_{02} is indicated by $\delta \neq 0$. In our simulation study below, we investigate the size by setting $\delta = 0$ and power by $\delta = 1$. All other specifications under DGP_1 are maintained. We adopt a ROT h_j to implement $\hat{J}_n$ and $\hat{J}_n^{*}$. We fix $T = 15$, choose $n = (100, 200, 400)$, and use $B = 399$ bootstrap repetitions to approximate null distribution under H_{02}. For the sake of computational cost, we perform 500 repetitions to obtain the average of rejection frequency for empirical size or power.

Table A2.2 shows the testing results of $\hat{J}_n$ for H_{02}. The empirical size (left panel) and power (right panel) are obtained for significant levels $\alpha = (0.01, 0.05, 0.1)$, and reported beneath each sample size of n. Different from the qualitative results for H_{01} in Appendix 2.1, $\hat{J}_n$ under H_{02} is typically oversized conditioning on each sample size. Nonetheless, the size steadily reduces toward nominal levels as n doubles. When H_{02} is not true, the empirical power of $\hat{J}_n$ is relatively low with a small sample size of $n = 100$ under $\alpha = 0.01$. As n increases, however, the power converges uniformly to unity under all significance levels, suggesting that $\hat{J}_n$ is consistent. Overall, we observe a promising finite sample performance of our proposed test $\hat{J}_n$ for H_{02}, which is expected to perform well given our large panel dataset in the empirical study.

Appendix 2.3: Testing for correct three-regime composite error

Our three-regime composite error under SEM-SP captures three possible scenarios of under-investment, over-investment, and efficient investment, each of which occurs with certain probability influenced by corresponding financial variables. This is contradictory to the one-regime composite error under SFM-TT by Lian and Chung (2008) or Lin et al. (2017), assuming an infeasible scenario where firms exhibit both under-investment and over-investment at the same time. As discussed in Section

Table A2.2: Testing Results for Correct Two-Regime Composite Error under H_{02}

	Empirical Size ($\delta = 0$)			Empirical Power ($\delta = 1$)		
α	n=100	200	400	n=100	200	400
0.01	0.028	0.022	0.016	0.826	0.928	1
0.05	0.068	0.058	0.052	0.912	1	1
0.1	0.186	0.122	0.106	1	1	1

2.1, the infeasible scenario arises because, under SFM-TT, the composite error $\epsilon = v - u + e$ models each side of investment inefficiency through continuous variables u and e. In this case, the probability of observing $u = 0$ (only over-investment occurs) or observing $e = 0$ (only under-investment occurs) is zero. Due to this reason, it may not be applicable to apply the conventional SFM-TT by Kumbhakar and Parmeter (2009) for investment efficiency analysis. It is therefore of our interest to statistically justify our arguments about the potential problem of SFM-TT through hypothesis testing.

A direct comparison between SEM-SP and SFM-TT, however, is not feasible because the former does not nest the latter as a special case in terms of the composite error structure. The reason is that both under-investment and over-investment appear concurrently with probability zero under SEM-SP but with probability one under SFM-TT. Inspired by this observation, we provide statistical justification on the validity of our three-regime structure by testing whether u and e would jointly appear with non-zero probability. Formally, we base our third null hypothesis as

$$H_{03} : Pr(\mathbb{1}(u > 0, e > 0) = 0) = 1, \text{ almost surely,}$$

under which firms do not under-invest and over-invest simultaneously, as assumed by SEM-SP. We construct our alternative model as a revised SEM-SP with a four-regime composite error

$$\epsilon = v - u1(u > 0, e = 0) + e1(u = 0, e > 0) - (u - e)\mathbb{1}(u > 0, e > 0). \tag{A.31}$$

Different from the three-regime structure in (A.2), the four-regime composite error in (A.31) specifies $\epsilon = v - u$ for under-investment with probability $p_u(W^u; \gamma_0^u)$; $\epsilon = v + e$ for over-investment with probability $p_e(W^e; \gamma_0^e)$; $\epsilon = v - u + e$ for simultaneous occurrence of under and over-investment with probability $p_s(W; \gamma_0^s)$; and $\epsilon = v$ for efficient investment with probability $1 - p_u(W^u; \gamma_0^u) - p_e(W^e; \gamma_0^e) - p_s(W; \gamma_0^s)$. Here, we assume that the probability for both $u > 0$ and $e > 0$ to occur (i.e., $p_s(\cdot)$) is influenced by determinants of both financial constraints (W^u) and agency problem (W^e); thus, the determinants of $p_s(\cdot)$ are $W = (W^u, W^e)$ with associated coefficients $\gamma_0^s \in \Re^{q_u + q_e}$. Therefore, the rejection of H_{03} provides supporting evidence on the validity of SFM-TT, where firms may possibly exhibit under-investment and over-investment at the same time with non-zero probability.

We follow a similar testing procedure as in Appendix 2.2 to test H_{03} by $\hat{J}_n$ in (A.30), except that the restricted and unrestricted SEM-SP are modified as follows. The restricted SEM-SP now refers to (A.1) with a three-regime composite error in (A.2). Thus, the inefficiency mean function $\pi^0(W;\omega_0)$ remains unchanged as in (A.3). The unrestricted SEM-SP contains a four-regime composite error in (A.31), with its corresponding inefficiency mean function defined as

$$\pi(W_{it};\omega_0) \equiv -E(\epsilon_{it}|\alpha_i,\eta_t,X_{it},W_{it}) = \left[\mu_0^u p_u(W_{it}^u;\gamma_0^u) - \mu_0^e p_e(W_{it}^e;\gamma_0^e) - \mu_s^0 p_s(W_{it};\gamma_0^s)\right], \quad \text{(A.32)}$$

where $\mu_0^s = \mu_0^u - \mu_0^e$ is the difference between the mean inefficiency driven by financial constraints or agency problem. For identification of the unrestricted SEM-SP, we assume that $\mu_s \neq 0$. Clearly, our SEM-SP with three-regime under H_{03} is properly nested by SEM-SP with four-regime as a special case. The test statistic $\hat{J}_n$ can be then constructed by computing $\widehat{SSR}_0$ and $\widehat{SSR}_1$ with the within-difference residuals $\hat{\tilde{\epsilon}}_{it}^{0,c}$ from restricted (three-regime) SEM-SP and $\hat{\tilde{\epsilon}}_{it}^{c}$ from unrestricted (four-regime) SEM-SP, respectively. We propose the following bootstrap procedure to refine the finite sample distribution of $\hat{J}_n$ under H_{03}.

1) With the original sample $\Theta = \{I_{it}, X_{it}, W_{it}\}_{i=1,t=1}^{n,T}$, compute $\hat{J}_n$ in (A.30). Obtain the within-difference residuals $\hat{\tilde{\epsilon}}_{it}^{0,c} = \hat{\tilde{\epsilon}}_{it}^{0} - \frac{1}{T}\sum_{t=1}^{T}\hat{\tilde{\epsilon}}_{it}^{0}$ from restricted (three-regime) SEM-SP with $\pi^0(W_{it};\hat{\omega})$ defined in (A.3), and within-difference residuals $\hat{\tilde{\epsilon}}_{it}^{c} = \hat{\tilde{\epsilon}}_{it} - \frac{1}{T}\sum_{t=1}^{T}\hat{\tilde{\epsilon}}_{it}$ from unrestricted (four-regime) SEM-SP with $\pi(W_{it};\hat{\omega})$ defined in (A.32).

2) With $\{\hat{\tilde{\epsilon}}_{it}^{0}\}_{i=1,t=1}^{n,T}$ under restricted (three-regime) SEM-SP in (A.1), obtain wild-bootstrap residual $\{\epsilon_{it}^*\}_{i=1,t=1}^{n,T}$, where $\epsilon_{it}^* = [(1-\sqrt{5})/2]\hat{\tilde{\epsilon}}_{it}^{0}$ with probability $P = (1+\sqrt{5})/2\sqrt{5}$, and $\epsilon_{it}^* = [(1+\sqrt{5})/2]\hat{\tilde{\epsilon}}_{it}^{0}$ with probability $1-P$. Construct the bootstrap sample $\Theta^* = \{I_{it}^*, X_{it}, W_{it}\}_{i=1,t=1}^{n,T}$, where $I_{it}^* = \hat{\eta}_t + \sum_{j=1}^{d}\tilde{m}_j(X_{j,it}) - \pi^0(W_{it};\hat{\omega}) + \epsilon_{it}^*$ uses all estimates from restricted (three-regime) SEM-SP.

3) Compute $\hat{J}_n^*$ similar as in 1), except replacing sample Θ with Θ^*.

4) Repeat step 2)-3) a large number (B) of times to obtain an empirical distribution from $\{\hat{J}_{n,b}^*\}_{b=1}^{B}$. We reject H_{02} if $p_n^* < \alpha$, where $p_n^* = \frac{1}{B}\sum_{b=1}^{B}1(\hat{J}_{n,b}^* > \hat{J}_n)$ is the empirical p-value and α is the significant level.

We proceed by investigating the finite sample size and power of $\hat{J}_n$ under H_{03}. We adopt a similar DGP_1 as in Appendix 1.3, except that the composite error is modified as $\epsilon_{it} = v_{it} - u_{it}\mathbb{1}(u_{it} > 0, e_{it} = 0) + e_{it}\mathbb{1}(u_{it} = 0, e_{it} > 0) - \delta(u_{it} - e_{it})\mathbb{1}(u_{it} > 0, e_{it} > 0)$. Thus, H_{03} is imposed with $\delta = 0$ and any violation of H_{03} is indicated by $\delta \neq 0$. As in our previous cases, we investigate the size by setting $\delta = 0$ and power by $\delta = 1$ in our simulation study below. All other specifications under DGP_1 are maintained. We adopt ROT bandwidth h_j to implement $\hat{J}_n$ and $\hat{J}_n^*$. We fix $T = 15$, choose $n = (100, 200, 400)$, and set $B = 399$ to obtain bootstrap distribution under H_{03}. We again perform 500 repetitions to compute the average of rejection frequency for empirical size or power.

Table A2.3: Testing Results for Correct Three-Regime Composite Error under H_{03}

	Empirical Size ($\delta = 0$)			Empirical Power ($\delta = 1$)		
α	n=100	200	400	n=100	200	400
0.01	0.020	0.012	0.008	0.812	0.986	1
0.05	0.078	0.066	0.056	0.912	1	1
0.1	0.190	0.132	0.118	0.998	1	1

Table A2.3 shows the testing results of $\hat{J}_n$ for H_{03}. We observe a qualitatively similar results compared to those for H_{02} above. The test $\hat{J}_n$ is oversized given small sample sizes of $n = 100$ or 200, but discloses reasonable size as n rises to 400. The test also reveals a non-trivial power upon the violation of H_{03}, albeit with relatively low magnitude when $n = 100$. As n doubles, the power clearly converges to unity under all three significance levels, suggesting its consistency property. Overall, the performance of our test $\hat{J}_n$ for H_{03} is suggested to be fairly reasonable for our applied study.

Appendix 3: Descriptive Data Summary

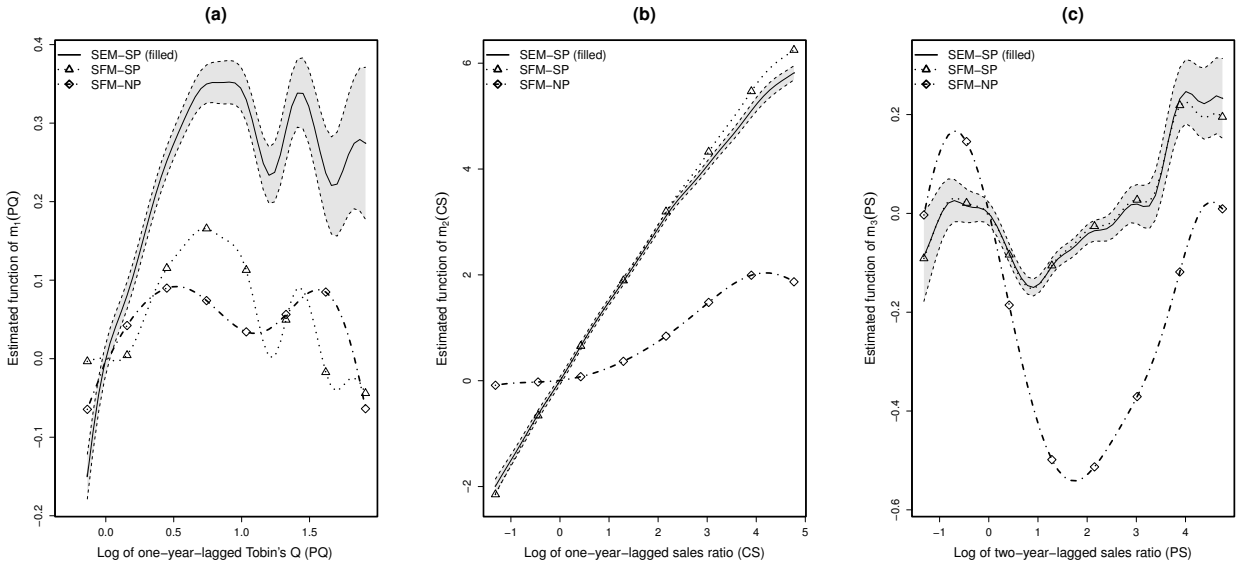
Table A3.1: Data Descriptive Summary Statistics

Classification	Description	Notation	Mean	SD	Min	Max
Dependent-variable	Log of investment over-one-year-lagged fixed assets	I	-2.087	1.415	-12.107	7.763
Variables in efficient-investment function (X)	Log of firm's current sale value over-one-year-lagged fixed assets ($ln(Sale_{it}/K_{it-1})$)	CS	1.222	1.218	-2.657	8.346
	Log of firm's last year sale value over-	PS	1.200	1.209	-2.657	8.346
	Log of one-year-lagged Tobin's q	PQ	0.470	0.455	-0.203	2.872
Variables in-under/over-investment-probability function (W^u/W^e)	Firm's total debt over-firm's total assets	$Debt$	0.484	0.185	0.028	0.937
	Firm's total net income plus-depreciation, divided by one-year-lagged fixed assets	CF	1.111	8.340	-38.525	435.565
Variables in-over-investment-probability function (W^e)	Log of firm's total assets	$Size$	2.336	1.187	-0.850	6.837
	Firm's age, the difference between firm's current year and establishment year	Age	15.515	5.588	2	31
	Board independence, or number of independent directors out of total number of directors	BI	0.367	0.052	0.25	0.6
	Duality, one if the CEO is also the Chairperson of the board, and zero otherwise	$Duality$	0.127	0.333	0	1

Note: The table reports the summary statistics of dependent variables I, efficient investment variables $X = (CS, PS, PQ)$, determinants of under-investment probability $W^u = (Debt, CF)$, and determinants of over-investment probability $W^e = (Debt, CF, Size, Age, BI, Duality)$. The total number of incumbent firms during 2006-2020 is 851. Source: CSMAR dataset 2006-2020.

Appendix 4: Further Empirical Comparison Results from Section 4.3

Figure A.5: Model Comparison of EIFs between SEM-SP and SFM-SP, and SFM-NP



In Section 4.3 of the paper, we compare the empirical results of our proposed SEM-SP model with two benchmark non/semiparametric models: the SFM-SP model by Wang et al. (2024) and the SFM-NP model by Simar et al. (2017), both of which assume under-investment only. This appendix presents the additional empirical results that complement those in Section 4.3.

Figure A.5 presents the estimated efficient investment functions (EIFs) from the three models: SEM-SP (gray-filled solid line), SFM-SP (dotted line with $\triangle$), and SFM-NP (dash-dot line with $\diamond$). Recall that standardized cash flow (CF) is used to ensure the comparability and implementability of all models. In each panel, the SFM-NP estimates are plotted as partial functions, holding all other covariates at their median. For ease of comparison, each estimated function is empirically centered around zero. *Panel A* of Table A4.1 reports the numerical derivative estimates of $\tilde{m}_1^{(1)}(PQ)$, $\tilde{m}_2^{(1)}(CS)$, and $\tilde{m}_3^{(1)}(PS)$ from each model, evaluated at five selected percentiles: 0.10, 0.25, 0.50, 0.75, and 0.90. The final row of *Panel A* summarizes the mean derivative estimates across these percentiles.

Figure A.5(a) is identical to Figure 1(a) in the paper, which reveals a non-monotonic relationship between Tobin's q and investment across all three models, albeit with different degrees of nonlinearity in $\tilde{m}_1(PQ)$. Compared to SEM-SP, both SFM-SP and SFM-NP display lower absolute magnitudes in the estimated $m_1(PQ)$, a steeper decline with respect to PQ, and vary less with PQ. Consequently, the mean marginal effect of PQ (i.e., $\tilde{m}^{(1)}(PQ)$) declines significantly, from 0.1942 under SEM-SP to 0.0933 under SFM-NP, and further to -0.0212 under SFM-SP. Figure A.5(b) presents the estimates of $m_2(CS)$. All models suggest that higher sales levels are associated with

increased investment. Both SEM-SP and SFM-SP indicate a positive, mildly concave relationship, with similar mean partial effects of $\tilde{m}_2^{(1)}(CS)$ at 1.2862 and 1.3781, respectively. In contrast, SFM-NP produces a flatter curve with a substantially smaller mean partial effect of 0.4434. A similar pattern is observed in Figure A.5(c) for $\tilde{m}_3(PS)$. SEM-SP and SFM-SP yield similar shapes, with mean marginal effects of 0.0425 and 0.0374, respectively. In contrast, SFM-NP exhibits greater fluctuations and a negative mean effect of -0.2141 (holding other variables constant). Notably, since SFM-NP requires estimation of a nine-dimensional unknown function, the curse of dimensionality greatly inflates the standard errors, rendering the effects of PQ, CS, and PS statistically insignificant - contradicting well-established investment theories in finance.

Panel B of Table A4.1 reports the estimated coefficients by SEM-SP and SFM-SP in inefficient probability functions $p_u(W;\gamma_0^u)$ and $p_e(W;\gamma_0^e)$. As discussed in Section 4.3 of the paper, SFM-SP is limited by its inability to account for over-investment. As a consequence, the coefficient estimates under SFM-SP are either largely reduced (i.e., CF and Age), changed with sign (i.e., $Debt$, $Size$, and BI), or loss significance (i.e., $Duality$). Notably, the estimated constant $\hat{\mu}^u$ under SFM-SP (6.54), which measures the expectation of under-investment given W (i.e., $\mu^u = E(u|W, u > 0, e = 0)$), is about five times larger than that under our SEM-SP (1.3) which accounts for the presence of over-investment. This indicates that SFM-SP overestimate under-investment. This can be seen clearly from *Panel C* of Table A4.1, which reports that the expected k-fold decrease in investment due to financial constraints (EFD-F) under SFM-SP (1.8173) is more than 1.5 times greater than its counterpart under SEM-SP (0.6582). Similar observation regarding overstated under-investment is made on SFM-NP, which yields a notably larger median of under-investment at 6.1.

Panel B of Table A4.1 reports the estimated coefficients from the inefficient probability functions, $p_u(W;\gamma^u 0)$ and $pe(W;\gamma_0^e)$, under SEM-SP and SFM-SP. As discussed in Section 4.3, SFM-SP is limited by its inability to account for over-investment. As a result, its coefficient estimates are either reduced in absolute value (e.g., CF and Age), exhibit sign change (e.g., $Debt$, $Size$, and BI), or lose statistical significance (e.g., $Duality$). Notably, the estimated constant $\hat{\mu}^u$ under SFM-SP is 6.54, which represents the expected under-investment conditional on W and the absence of over-investment (i.e., $\mu^u = E(u|W, u > 0, e = 0)$). This value is more than five times larger than the corresponding estimate under our SEM-SP (1.3), which explicitly models over-investment. The result suggests that SFM-SP significantly overestimates the degree of under-investment. This overestimation is reflected in Panel C of Table A4.1, which shows that the expected k-fold decrease in investment due to financial constraints (EFD-F) under SFM-SP is 1.8173, more than 1.5 times higher than the corresponding value of 0.6582 under SEM-SP. A similar pattern is also observed for SFM-NP, which yields a notably inflated median level of under-investment at 6.1.

Overall, these results highlight the advantages of our SEM-SP model: it disentangles under-investment from over-investment. The semiparametric EIF specification provides a flexible yet tractable structure, alleviating the curse of dimensionality and is useful in empirical applications.

Appendix 5: Instructions for Data and Code

This section provides instructions for accessing the data and code used in the empirical analysis. The data (labeled Data) are sourced from the China Stock Market and Accounting Research (CSMAR) database. The code (labeled Package Download) is implemented in user-friendly R packages. Both can be downloaded from https://694160821.wixsite.com/taining by selecting **R Code** from the top menu.

The "Package Download" contains two R files: Source code.R and Replication file.R. The Source code.R includes all predefined functions that are needed for estimation. In particular, function [10.1] WWYK.SIF performs the two-step estimation for our proposed SEM-SP model (see Section 3 of the paper). Function [10.2] result.WWYK.SIF collects and summarizes the results from WWYK.SIF into numerical tables and figures. Both functions are designed for ease of use—simply provide the dataset: press <Enter> and the results will be generated. Detailed descriptions for each command are given above the corresponding function. Replication file.R contains the code to reproduce the tables and figures of the main empirical results. To use it, simply upload the source code and dataset (CSMAR-Sample.xlsx) then run each line sequentially. If the users encounter any issues in implementing the code for this or other applications, please contact the first author of this paper.

Please notice that our sample is derived from a subscription-based commercial database CSMAR. The sample is made available here solely for the purposes of replicating the results reported in the paper and demonstrating the use of the accompanying codes. Without a valid subscription or authorized access to the original database, this dataset may not be used for other research projects or redistributed in any form.

Table A4.1: Estimation Results of SEM-SP, SFM-SP, and SFM-NP

Panel A (EIFs)	SEM-SP			SFM-SP			SFM-NP		
	$\tilde{m}_1^{(1)}(PQ)$	$\tilde{m}_2^{(1)}(CS)$	$\tilde{m}_3^{(1)}(PS)$	$\tilde{m}_1^{(1)}(PQ)$	$\tilde{m}_2^{(1)}(CS)$	$\tilde{m}_3^{(1)}(PS)$	$\tilde{m}_1^{(1)}(PQ)$	$\tilde{m}_2^{(1)}(CS)$	$\tilde{m}_3^{(1)}(PS)$
0.1	0.461***	1.5577***	-0.0365	0.0173	1.6391***	-0.0362	0.4205	0.3176	-0.3691
	(0.0803)	(0.0680)	(0.0726)	(0.0804)	(0.0679)	(0.0721)	(39.2319)	(78.1416)	(55.1769)
0.25	0.5954***	1.5665***	-0.2101***	0.4132***	1.5658***	-0.2169***	-0.0718	0.3839	-0.7889
	(0.0811)	(0.0350)	(0.0353)	(0.0816)	(0.0351)	(0.0353)	(35.8246)	(48.1917)	(24.8118)
0.5	0.0045	1.4131***	0.0648**	-0.1090	1.4793***	0.0741***	-0.2352	0.6701	-0.2495
	(0.1225)	(0.0292)	(0.0298)	(0.1227)	(0.0291)	(0.0298)	(42.3493)	(39.0097)	(19.3705)
0.75	0.4545***	1.1385***	0.0040	0.3905***	1.3094***	0.0035	0.5413	0.6379	-0.0326
	(0.1941)	(0.0769)	(0.0778)	(0.1956)	(0.0772)	(0.0782)	(65.2851)	(65.4843)	(73.7143)
0.9	-0.0580	1.0426***	-0.0593	-0.2010	1.156***	-0.0961	0.1259	0.1595	0.1958
	(0.2968)	(0.1482)	(0.1136)	(0.2956)	(0.1507)	(0.1153)	(69.1243)	(74.1624)	(83.9185)
Mean	0.1942	1.2862	0.0425	-0.0212	1.3781	0.0374	0.0933	0.4434	-0.2141
Panel B (Inefficiency)	Under		Over	Under		Over	Under		Over
Constant $\hat{\mu}^u$	1.3083**			6.5355***		-	-		-
	(0.9316)			0.5014					
Constant $\hat{\mu}^e$			10.1414***	-		-	-		-
			(1.9658)	-					
Debt	1.0114***		0.0630***	-0.0753***		-			-
	(0.2871)		(0.0033)	(0.0204)					
CF	-0.8202***		0.0630***	-0.0201***		-			-
	(0.0585)		(0.0033)	(0.0059)					
Age			0.0791***	0.0056***		-			-
			(0.0069)	(0.0015)					
Size			0.1390***	-0.0110***		-			-
			(0.0148)	(0.0031)					
BI			-0.3875***	0.0325*		-			-
			(0.0808)	(0.0177)					
Duality			-0.022***	-0.0034		-			-
			(0.0087)	(0.0026)					
Panel C	Under	Over	Optimal	Under	Over	Optimal	Under	Over	Optimal
Median of probability	0.3816	0.5396	0.0788	1	0	0	1	0	0
	EFD-F	EFI-A	ENF	EFD-F	EFI-A	ENF	EFD-F	EFI-A	ENF
Median of expected fold change	0.6582	3.6768	3.0159	1.8173	0	1.1873	6.0574	0	6.0574

References

Baltagi, B. H., 2013. Econometric Analysis of Panel Data, 2nd Edition. Wiley, New York.

Horowitz, J. L., Mammen, E., 2004. Nonparametric estimation of an additive model with a link function. The Annals of Statistics 32 (6), 2412–2443.

Jensen, M., Meckling, W., 1976. ltheory of the firm: Managerial behavior. Agency Costs and Ownership Structure. mJournal of Financial Economics 3.

Kumbhakar, S. C., Parmeter, C. F., 2009. The effects of match uncertainty and bargaining on labor market outcomes: Evidence from firm and worker specific estimates. Journal of Productivity Analysis 31 (1), 1–14.

Li, Q., 2000. Efficient estimation of additive partially linear models. International Economic Review 41 (4), 1073–1092.

Lian, Y., Chung, C.-F., 2008. Are chinese listed firms over-investing. Working paper.

Lin, Z., Li, Q., Sun, Y., 2014. A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. Journal of Econometrics 178, 167–179.

Lin, Z. J., Liu, S., Sun, F., 2017. The impact of financing constraints and agency costs on corporate r&d investment: Evidence from china. International Review of Finance 17 (1), 3–42.

Simar, L., Van Keilegom, I., Zelenyuk, V., 2017. Nonparametric least squares methods for stochastic frontier models. Journal of Productivity Analysis 47, 189–204.

Wang, T., Tian, J., Yao, F., 2020. Does high leverage ratio influence chinese firm performance? a semiparametric stochastic frontier approach with zero-inefficiency. Empirical Economics 61, 587–636.

Wang, T., Yao, F., Kumbhakar, S. C., 2024. A flexible stochastic production frontier model with panel data. Journal of Applied Econometrics 39 (4), 564–588.

Yao, F., Wang, T., Tian, J., Kumbhakar, S. C., 2018. Estimation of a smooth coefficient zero-inefficiency panel stochastic frontier model: A semiparametric approach. Economics Letters 166, 25–30.

Ziegelmann, F. A., 2002. Nonparametric estimation of volatility functions: the local exponential estimator. Econometric Theory 18 (4), 985–991.